



**HAL**  
open science

# Optimisation des capteurs d'arômes et fusion multisensorielle appliquée à la caractérisation des produits agro-alimentaires

S. Roussel

► **To cite this version:**

S. Roussel. Optimisation des capteurs d'arômes et fusion multisensorielle appliquée à la caractérisation des produits agro-alimentaires. Sciences de l'environnement. Doctorat Sciences pour l'ingénieur, ENSA Montpellier, 1998. Français. NNT: . tel-02578039

**HAL Id: tel-02578039**

**<https://hal.inrae.fr/tel-02578039>**

Submitted on 14 May 2020

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

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

98/0823

MINISTERE DE L'AGRICULTURE

ECOLE NATIONALE SUPERIEURE AGRONOMIQUE DE MONTPELLIER

**THESE**

présentée à l'Ecole Nationale Supérieure Agronomique de Montpellier

pour obtenir le **DIPLOME DE DOCTORAT**

**SPECIALITE** : Sciences pour l'ingénieur

Laboratoire de Génie Instrumental pour la Qualité Agro-alimentaire - CEMAGREF - Montpellier

**OPTIMISATION DES CAPTEURS D'AROMES  
ET FUSION MULTISENSORIELLE  
APPLIQUEE A LA CARACTERISATION  
DES PRODUITS AGRO-ALIMENTAIRES**

par

**Sylvie ROUSSEL**

Soutenue le 14 décembre 1998 devant le Jury composé de :

M. Francis SEVILA, Professeur à l'ENSA.M	Président
M. Jean-Louis BERDAGUE, Directeur de recherche à l'INRA de Theix	Rapporteur
M. Douglas RUTLEDGE, Maître de conférence HDR à l'INA-PG	Rapporteur
Mme Véronique BELLON-MAUREL, Chef de Division au Cemagref	Directeur de thèse
M. Gilles CARAUX, Professeur à l'ENSA.M	Examineur
M. Alain RAZUNGLES, Maître de conférence à l'ENSA.M	Examineur
M. Gilles TRYSTRAM, Professeur à l'ENSIA	Examineur

**MINISTERE DE L'AGRICULTURE**  
**ECOLE NATIONALE SUPERIEURE AGRONOMIQUE DE MONTPELLIER**

**THESE**

présentée à l'Ecole Nationale Supérieure Agronomique de Montpellier

pour obtenir le **DIPLOME DE DOCTORAT**

**SPECIALITE** : Sciences pour l'ingénieur

Laboratoire de Génie Instrumental pour la Qualité Agro-alimentaire - CEMAGREF - Montpellier

**OPTIMISATION DES CAPTEURS D'AROMES**  
**ET FUSION MULTISENSORIELLE**  
**APPLIQUEE A LA CARACTERISATION**  
**DES PRODUITS AGRO-ALIMENTAIRES**

par

**Sylvie ROUSSEL**

Soutenue le 14 décembre 1998 devant le Jury composé de :

M. Francis SEVILA, Professeur à l'ENSA.M	Président
M. Jean-Louis BERDAGUE, Directeur de recherche à l'INRA de Theix	Rapporteur
M. Douglas RUTLEDGE, Maître de conférence HDR à l'INA-PG	Rapporteur
Mme Véronique BELLON-MAUREL, Chef de Division au Cemagref	Directeur de thèse
M. Gilles CARAUX, Professeur à l'ENSA.M	Examineur
M. Alain RAZUNGLES, Maître de conférence à l'ENSA.M	Examineur
M. Gilles TRYSTRAM, Professeur à l'ENSIA	Examineur

## Remerciements

Je tiens à remercier chaleureusement Mme Véronique Bellon-Mauvel pour m'avoir offert l'opportunité d'effectuer cette thèse au Cemagref, dans la division Génie Instrumental pour la Qualité Agro-alimentaire. Mais, je voudrais aussi lui dire un grand **MERCI** pour m'avoir soutenue et encouragée tout au long de ces 3 années, malgré un « nez » souvent enflé, merci pour les slogans du jour (« une idée, une publi » et j'en passe...), merci pour ton portable (private joke ☺)....

Je dois aussi remercier M. Francis Sévila, directeur de l'école doctorale « Sciences pour l'ingénieur » de l'ENSA.M., qui m'a finalement laissé choisir cette thèse.

Un comité de suivi de thèse s'est réuni tous les ans pour des brain-storming visant au recadrage de ma thèse : merci donc à Joseph Hossenlopp, Gilles Trystram et Alain Razungles, Pierre Grenier, Véronique Bellon-Mauvel et Francis Sévila.

Merci à Pierre Grenier sans lequel aucun des projets sur lesquels nous avons travaillé n'aurait vu le jour. Il a aussi été un viticulteur acharné, puisqu'il m'a aidée à recueillir les échantillons de raisins de Pech Rouge et de Nîmes : pour tout ça, merci !

Merci à tous ceux qui m'ont aidé à mener à bien le projet « cépages blancs » : Jean-Luc Favarel, Laure, Eric et François de l'ITV de Gaillac, qui m'ont baladé dans toute la région midi-pyrénées, Denis Caboulet de Pech Rouge et Frédéric Neaudeau de Nîmes, et merci aux courageux dégustateurs de moûts (ITV, ICV et Chambre d'agriculture du Tarn) !

Merci aux stagiaires qui ont bien travaillé et m'ont bien fait travailler : Elizabeth Privat, Gustaf Forsberg, Estelle Dupuit, Laetitia Vidie et Marc Baquelin.

Merci aux organismes qui ont financé et suivi nos projets : la DRAF, l'ACTIA, TRIAL (Michel-Ange Sobas et J. Mihu) et la DGAL.

Merci aux Giqualiens et aux Giqualiennes qui m'ont aidé tous les jours, Serge avec qui j'ai travaillé avant qu'il ne fuit à l'ENSA.M., Jean-Michel qui a pris le relais, Nathalie (reviens vite, on a les mêmes à la maison), Louis et Jean-Luc les rois du plantage imprimante, Michèle et Aline, nos deux secrétaires de choc, et j'en passe...

Merci à mes chers thésards, Rémi et Jérémie, qui m'ont quand même laissé les nuits pour rédiger dans le calme, sans un Karma, sans rien...

Merci aux autres Cémagrefiens, à Jean-Mi pour m'aider à croire que je ne suis pas folle de faire une thèse après l'agro, à Olaf pour ses coups de queue, à Alain pour essayer de me faire croire qu'un jour je serai une star du volley, et j'en passe...

Merci à mes parents, qui sont tellement gentils qu'ils ont même proposé de relire ma thèse...

Merci à Jean-Fran et Mimi pour avoir fait un si beau Zvattito.

Merci à PCM pour m'avoir laissé travailler en paix.

Merci

Merci

Merci

Merci

Merci.....

*A mon* 

# *Sommaire*

*1. Table des matières*

*2. Tableaux*

*3. Figures*

*4. Equations*

# TABLE DES MATIERES

<b>I. INTRODUCTION</b>	<b>I-1</b>
<b>II. ETAT DE L'ART</b>	<b>II-3</b>
<b>II.1. Problématique</b>	<b>II-3</b>
II.1.1. Demande de qualité agro-alimentaire	II-3
II.1.2. Caractériser les arômes	II-4
II.1.2.1. Analyse sensorielle	II-4
II.1.2.2. Chromatographie en Phase Gazeuse	II-4
II.1.2.3. Mesure instrumentale par « nez électronique »	II-5
II.1.3. Mesure de la qualité dans la filière viti-vinicole	II-5
II.1.3.1. Problématique	II-6
II.1.3.2. Arômes variétaux du moût	II-6
<b>II.2. 2. Les capteurs d'arômes</b>	<b>II-8</b>
II.2.1. Définition : « nez électroniques » ou capteurs d'arômes ?	II-8
II.2.2. Eléments sensibles	II-10
II.2.2.1. Oxydes métalliques semi-conducteurs (MOS)	II-10
II.2.2.2. Polymères conducteurs (CP)	II-11
II.2.2.3. Quartz Microbalance (QCM)	II-12
II.2.2.4. Oxydes métalliques semi-conducteurs à effet de champ (MOSFET)	II-12
II.2.2.5. Capteurs à ondes acoustiques de surface (SAW)	II-12
II.2.2.6. Autres types d'éléments sensibles	II-13
II.2.3. Appareillages commerciaux	II-13
II.2.3.1. Oxydes métalliques semi-conducteurs	II-13
II.2.3.2. Polymères conducteurs	II-14
II.2.3.3. Quartz microbalances	II-15
II.2.3.4. Appareillages hybrides	II-15
II.2.4. Caractéristiques métrologiques comparées	II-16
II.2.4.1. Caractéristiques métrologiques	II-16
II.2.4.2. Comparaison d'appareillages	II-18
II.2.4.3. Conclusion	II-19
II.2.5. Différentes applications	II-19
II.2.6. Conclusion	II-21
<b>II.3. Mise en œuvre des capteurs d'arômes</b>	<b>II-22</b>
II.3.1. Conditions expérimentales	II-22
II.3.2. Génération et transfert de l'échantillon	II-24
II.3.2.1. Echantillons non alcoolisés	II-24
II.3.2.2. Echantillons alcoolisés	II-26
II.3.3. Conclusion	II-29

<b>II.4. Traitement des données des systèmes multi-capteurs</b>	<b>II-31</b>
II.4.1. Acquisition et prétraitement des données	II-31
II.4.2. Méthodes de classification	II-35
II.4.2.1. Classifications non supervisées	II-35
II.4.2.2. Classifications supervisées	II-36
II.4.3. Méthodes de sélection des capteurs	II-39
II.4.3.1. Sélection des capteurs d'arômes	II-39
II.4.3.2. Méthodes de sélection de variables	II-40
II.4.4. Méthodes de fusion multi-capteur	II-44
II.4.4.1. Niveaux de fusion	II-45
II.4.4.2. Fusion de niveau haut par théorie de Bayes	II-46
II.4.4.3. Pertinence de la fusion multi-capteur	II-51
II.4.5. Conclusion	II-52
<b>II.5. Conclusions</b>	<b>II-53</b>
<b>III. Méthodologie d'optimisation des mesures par capteurs d'arômes</b>	<b>III-54</b>
<b>III. 1. Matériel</b>	<b>III-55</b>
III. 1.1. Appareillage	III-55
III. 1.2. Echantillons	III-56
III. 1.2.1. Solutions synthétiques	III-57
III. 1.2.2. Moûts de raisin	III-58
<b>III. 2. Sélection des descripteurs</b>	<b>III-60</b>
III. 2.1. Méthodes	III-60
III. 2.1.1. Extraction des descripteurs du signal des capteurs	III-60
III. 2.1.2. Critères de sélection des descripteurs	III-62
III. 2.2. Résultats	III-63
III. 2.2.1. Application aux solutions modèles	III-63
III. 2.2.2. Application aux moûts de raisin	III-67
III. 2.3. Conclusion	III-71
<b>III. 3. Quantification de l'influence des conditions expérimentales</b>	<b>III-71</b>
III. 3.1. Méthodes	III-72
III. 3.1.1. Plans d'expériences	III-72
III. 3.1.2. Sorties analysées	III-75
III. 3.2. Résultats	III-76
III. 3.2.1. Application aux solutions modèles	III-76
III. 3.2.2. Application aux moûts de raisin	III-79
III. 3.3. Conclusion	III-81
<b>III. 4. Sélection des capteurs</b>	<b>III-82</b>
III. 4.1. Méthodes	III-82
III. 4.1.1. Capteurs analysés	III-82
III. 4.1.2. Analyse Factorielle Discriminante Pas à Pas	III-83

III. 4.1.3. Algorithmes génétiques	III-83
III. 4.2. Résultats	III-84
III. 4.2.1. Application aux solutions modèles	III-84
III. 4.2.2. Application aux moûts de raisin	III-86
<b>III. 5. Conclusions</b>	<b>III-88</b>
<b>IV. Système multi-capteur appliqué à la discrimination variétale de moûts</b>	<b>IV-91</b>
<b>IV. 1. Acquisition des données</b>	<b>IV-91</b>
IV. 1.1. Echantillons	IV-91
IV. 1.2. Analyses chimiques	IV-92
IV. 1.3. LCA 1000 modifié	IV-92
IV. 1.4. Spectromètre IR-TF	IV-92
IV. 1.5. Spectromètre UV	IV-94
IV. 1.6. Mesure de référence : l'analyse sensorielle	IV-95
<b>IV. 2. Classification par capteurs individuels</b>	<b>IV-96</b>
IV. 2.1. Méthodes	IV-96
IV. 2.1.1. Classification	IV-96
IV. 2.1.2. Sélection des longueurs d'onde par algorithmes génétiques	IV-98
IV. 2.2. Résultats	IV-99
IV. 2.2.1. LCA 1000 modifié	IV-99
IV. 2.2.2. Spectromètre IR-TF	IV-101
IV. 2.2.3. Spectromètre UV	IV-104
IV. 2.2.4. Discussion et conclusions sur les classifications individuelles	IV-106
<b>IV. 3. Fusion multi-capteur</b>	<b>IV-111</b>
IV. 3.1. Méthode de fusion multi-capteur de niveau bas	IV-111
IV. 3.1.1. Méthodes	IV-111
IV. 3.1.2. Résultats	IV-112
IV. 3.2. Fusion multi-capteur de niveau haut	IV-116
IV. 3.2.1. Méthodes	IV-116
IV. 3.2.2. Résultats	IV-119
IV. 3.3. Pertinence de la fusion	IV-127
IV. 3.3.1. Estimateurs de classification	IV-127
IV. 3.3.2. Redondance des capteurs	IV-129
<b>IV. 4. Conclusion</b>	<b>IV-132</b>
<b>V. Conclusion et perspectives</b>	<b>V-134</b>
<b>VI. Valorisation de la recherche</b>	<b>V-139</b>
<b>VII. Bibliographie</b>	<b>V-140</b>
<b>VIII. Annexes</b>	<b>V-157</b>

# FIGURES

Figure 1 : Système d'olfaction humain (GÖPEL, 1998) et système olfactif artificiel	II-9
Figure 2 : structure des 3 principaux monomères	II-11
Figure 3 : schéma d'un capteur piézo-électrique QCM	II-12
Figure 4 : influence de la température sur la sélectivité d'un capteur MOS (MIELLE, 1998)	II-22
Figure 5 : courbes d'adsorption/désorption avec injection en mode statique et dynamique	II-32
Figure 6 : schéma d'un réseau de neurones multicouche	II-38
Figure 7 : mécanismes de croisement et de mutation	II-43
Figure 8 : montage expérimental	III-56
Figure 9 : courbe d'adsorption et de désorption et ses dérivées première et seconde	III-61
Figure 10 : liaison linéaire entre écart-type et le niveau des descripteurs	III-63
Figure 11 : répétabilité et distance discriminante des 29 descripteurs	III-64
Figure 12 : ratio $q_i$ des 29 descripteurs calculé	III-65
Figure 13 : cercle des corrélations des 12 descripteurs présélectionnés	III-66
Figure 14 : ACP calculée sur le maximum (a) et sur les 5 descripteurs sélectionnés (b)	III-67
Figure 15 : coefficient de variation et distance discriminante des 12 descripteurs présélectionnés	III-68
Figure 16 : comparaison des courbes d'adsorption des solutions modèles et des moûts.	III-69
Figure 17 : ratio $q_i$ des 12 descripteurs présélectionnés	III-69
Figure 18 : cercle des corrélations des 8 descripteurs présélectionnés	III-70
Figure 19 : ACP normée sur les 5 descripteurs sélectionnés	III-77
Figure 20 : discrimination des moûts de raisin dans les 3 conditions expérimentales	III-80
Figure 21 : description des solutions modèles par 2 descripteurs du TGS 815	III-85
Figure 22 : exemple de sélection des capteurs par algorithmes génétiques	III-87
Figure 23 : Spectre électromagnétique	IV-93
Figure 24 : exemple d'un spectre IR-TF de moût de raisin (BUZON et C, 1981)	IV-93
Figure 25 : exemple d'un spectre UV de moût de raisin	IV-94
Figure 26 : prédiction des échantillons par PLS discriminante sur le spectre IR-TF filtré	IV-103
Figure 27 : Sélection des longueurs d'onde du spectre IR-TF	IV-104
Figure 28 : Sélection des longueurs d'onde du spectre UV	IV-106
Figure 29 : taux d'erreur et coefficients de contingentement des différents systèmes de mesure	IV-107
Figure 30 : Analyse en Composantes Principales sur les analyses chimiques	IV-109
Figure 31 : Sélection des longueurs d'onde du spectre commun UV-IR-TF	IV-113
Figure 32 : Taux d'erreur et coefficient de contingentement des différents systèmes de mesure	IV-115
Figure 33 : taux d'erreur et coefficient de contingentement des différentes classifications	IV-125
Figure 34 : taux d'erreur et coefficient de contingentement pour les différents capteurs et de la méthodes de fusion	IV-128
Figure 35 : corrélation entre les analyses chimiques, le LCA1000, les spectres UV et IR-TF	IV-129
Figure 36 : corrélation entre les analyses chimiques, le LCA1000, les spectres UV et IR-TF filtré a priori	IV-130
Figure 37 : corrélation entre les analyses chimiques, le LCA1000, les spectres UV et IR-TF filtré a posteriori	IV-131

# TABLEAUX

Tableau 1 : comparaison des performances des diverses technologies d'éléments sensibles et du système olfactif humain.	II-16
Tableau 2 : applications des capteurs d'arômes	II-19
Tableau 3 : exemple d'une matrice de confusion pour la discrimination des 4 cépages de moûts de raisins blancs	II-37
Tableau 4 : exemples de caractéristiques de capteurs MOS (TGS Figaro de la série 8xx)	II-39
Tableau 5 : matrice de causalité Mi pour le capteur i	II-49
Tableau 6 : description des 32 échantillons appartenant aux 2 classes de solutions modèles	III-58
Tableau 7 : moûts de référence pour la sélection des descripteurs et des conditions expérimentales.	III-59
Tableau 8 : caractéristiques des 29 descripteurs extraits de la courbe et de ses dérivées	III-60
Tableau 9 : 5 groupes de descripteurs redondants pour les solutions modèles	III-66
Tableau 10 : 5 groupes de descripteurs redondants pour les moûts	III-70
Tableau 11 : matrices d'expérience des 3 plans d'expériences.	III-74
Tableau 12 : effets des facteurs et de leurs interactions sur les indices de performances.	III-78
Tableau 13 : indices de performance des états du plan d'expériences	III-79
Tableau 14 : indices de performance pour les 3 états expérimentaux	III-80
Tableau 15 : Liste des capteurs analysés pour les solutions modèles	III-82
Tableau 16 : Liste des capteurs analysés pour les moûts de raisins blancs	III-82
Tableau 17 : résultats de la sélection des capteurs par AFD pas à pas	III-84
Tableau 18 : résultats de la sélection des capteurs par AFD pas à pas	III-86
Tableau 19 : sélection des capteurs par algorithmes génétiques.	III-87
Tableau 20 : tableau comparatif des performances des 4 sélections de capteurs	III-88
Tableau 21 : matrice de confusion des prédictions par AFD	IV-99
Tableau 22 : matrice de confusion des prédictions par réseau de neurones	IV-100
Tableau 23 : matrice de confusion des prédictions par PLS discriminante	IV-100
Tableau 24 : matrice de confusion des prédictions par PLS discriminante	IV-101
Tableau 25 : Résultats d'une PLS discriminante sur les différents spectres IR-TF filtrés	IV-102
Tableau 26 : matrice de confusion des prédictions par PLS discriminante du meilleur spectre IR-TF sélectionné.	IV-102
Tableau 27 : matrice de confusion des prédictions PLS discriminante du spectre UV brut	IV-104
Tableau 28 : Résultats d'une PLS discriminante sur les différents spectres UV filtrés	IV-105
Tableau 29 : matrice de confusion des prédictions par PLS discriminante du spectre UV filtré	IV-105
Tableau 30 : matrice de confusion de classification des moûts par AFD sur les analyses chimiques	IV-109
Tableau 31 : matrice de confusion des prédictions par PLS discriminante sur le spectre commun IR-TF et UV	IV-112
Tableau 32 : matrice de confusion des prédictions par PLS discriminante sur les spectres filtrés a posteriori	IV-113
Tableau 33 : matrice de confusion des prédictions par PLS discriminante	IV-114
Tableau 34 : Matrice de causalité du spectromètre IR-TF	IV-116
Tableau 35 : Matrice de causalité du spectromètre UV	IV-116
Tableau 36 : Matrice de causalité du LCA avec traitement par PLS discriminante	IV-117
Tableau 37 : Matrice de Coût	IV-117
Tableau 38 : matrice d'affectation d'erreur minimale pour les deux capteurs IR-TF et UV	IV-119
Tableau 39 : résultats de la fusion par erreur minimale de Bayes des spectres IR-TF & UV filtrés	IV-120
Tableau 40 : règle d'affectation du risque minimum pour les deux capteurs IR-TF et UV	IV-121
Tableau 41 : résultats de la fusion par la règle du risque minimum de Bayes entre les spectres IR-TF et UV filtrés	IV-121
Tableau 42 : règle d'affectation d'erreur minimale pour les capteurs d'arômes, IR-TF et UV	IV-122
Tableau 43 : résultats de la fusion par erreur minimale de Bayes entre les capteurs d'arômes et les spectromètres	IV-123
Tableau 44 : résultats de la fusion par risque minimum de Bayes entre les capteurs d'arômes et les spectromètres	IV-123

# EQUATIONS

Equation 1 : pouvoir de résolution	II-17
Equation 2 : théorème de Bayes sur les probabilités conditionnelles	II-46
Equation 3 : règle d'affectation d'erreur minimale de Bayes en fonction des probabilités a posteriori	II-47
Equation 4 : règle d'affectation d'erreur minimale de Bayes en fonction des probabilités a priori	II-47
Equation 5 : vraisemblance conditionnelle de X pour le capteur i	II-47
Equation 6 : fusion multi-capteur par l'erreur minimale de Bayes basée sur des probabilités a posteriori	II-48
Equation 7 : fusion multi-capteur par l'erreur minimale de Bayes basée sur des probabilités a priori	II-48
Equation 8 : fusion multi-capteur par l'erreur minimale de Bayes basée sur l'intersection de probabilités a priori	II-48
Equation 9 : formule des probabilités composées pour des événements non indépendants	II-48
Equation 10 : probabilité conjointe d'événements indépendants d'un même espace de probabilité	II-48
Equation 11 : fusion multi-capteur par l'erreur minimale de Bayes basée sur le produit des vraisemblances a priori	II-49
Equation 12 : exemple d'application à la discrimination de 4 classes de cépage	II-50
Equation 13 : coût moyen	II-50
Equation 14 : règle d'affectation du risque minimum de Bayes	II-50
Equation 15 : règle d'affectation de risque minimum de Bayes pour un système multi-capteur	II-51
Equation 16 : fusion multi-capteur par le risque minimum de Bayes basée sur des probabilités a priori	II-51
Equation 17 : coefficient du $\chi^2$	II-52
Equation 18 : coefficient de contingentement	II-52
Equation 19 : coefficient de variation du descripteur i	III-62
Equation 20 : distance discriminante du descripteur i	III-62
Equation 21 : indice de qualité $q_i$ du descripteur i	III-63
Equation 22 : effet moyen du facteur i	III-75
Equation 23 : effet moyen de l'interaction des facteur i et k	III-75
Equation 24 : variance estimée $S^2$ de la mesure	III-75
Equation 25 : Intervalle de confiance	III-75
Equation 26 : coefficient de variation de l'état j du plan d'expériences	III-76
Equation 27 : distance discriminante de l'état j du plan d'expériences	III-76
Equation 28 : indice de qualité de l'état j du plan d'expériences	III-76
Equation 29 : définition de la fonction de coût	III-117
Equation 30 : règle d'affectation de risque minimum de Bayes appliquée à la discrimination de 4 classes de cépage.	IV-118
Equation 31 : règle d'affectation d'erreur minimale de Bayes appliquée à la discrimination de 4 classes de cépage.	IV-119
Equation 32 : Fusion appliquée à la discrimination de 4 classes de cépage par les spectromètres IR-TF et UV	IV-120
Equation 33 : règle d'affectation d'erreur minimale de Bayes sous forme matricielle	IV-126
Equation 34 : fusion par la règle d'affectation d'erreur minimale de Bayes pour des classifications booléennes	IV-126
Equation 35 : fusion pour les classifications en degrés d'appartenance	IV-126
Equation 36 : fusion pour les classifications en degrés d'appartenance et le choix d'un opérateur d'agrégation	IV-127
Equation 37 : fusion pour les classifications en degrés d'appartenance et le choix d'un opérateur d'agrégation	IV-127

# *Introduction*

# I. INTRODUCTION

Pour les produits agro-alimentaires, l'exigence de qualité n'est pas seulement une contrainte commerciale, mais devient au travers de la future loi d'orientation agricole une priorité affichée du ministère de l'Agriculture et de la Pêche. Cette loi met l'accent sur les productions sous des signes de qualité, tels que les Certifications de Conformité, les Appellations d'Origine Contrôlée, etc.

Cette stratégie ne peut devenir une réalité que dans la mesure où des systèmes de contrôle fiables sont disponibles pour quantifier les paramètres de qualité organoleptique ou pour garantir l'origine de produits (authentification). La réussite de ces systèmes est fondée sur la multiplicité des capteurs et le traitement approprié de l'information. Parmi les capteurs spécifiques, les « nez électroniques » sont les plus récents et les moins bien maîtrisés.

Cette thèse a pour objectif d'aborder la problématique de la fusion multisensorielle, à partir de données multivariées de qualité très variée et partiellement redondantes, en vue de la caractérisation de la qualité organoleptique ou variétale de produits agro-alimentaires. Une démarche d'optimisation de ces données - principalement issues des capteurs d'arômes - constitue l'axe expérimental de cette thèse. Les produits d'application sont issus de la filière viti-vinicole.

Une *revue bibliographique* (Partie II) décrit tout d'abord le *contexte* du domaine d'application choisi : la caractérisation organoleptique ou variétale de produits oenologiques. Elle expose ensuite les connaissances actuelles sur les « nez électroniques » et leur mise en œuvre ainsi que les divers types de traitements de données dédiés à la classification multivariée. Cet état de l'art nous conduit à déterminer les problématiques scientifiques inhérentes à la mesure des arômes par « nez électronique ».

Ces problématiques sont abordées au travers d'une *démarche méthodologique*, proposée en troisième partie ; celle-ci a pour objectif d'optimiser différentes étapes décisives de la mesure aromatique et pourra être généralisée à toute mesure par capteurs d'arômes. Elle consiste à :

- ⊗ extraire l'information la plus discriminante des courbes du système multi-capteur d'arômes, reposant sur des critères de performance que nous établissons ;
- ⊗ quantifier l'influence de divers paramètres extérieurs sur la performance de la mesure afin de déterminer les conditions expérimentales optimales, grâce à la mise en œuvre de plans d'expériences ;
- ⊗ sélectionner les capteurs les plus adaptés à chaque discrimination, en s'appuyant sur diverses méthodes de sélection de variables.

Dans un premier temps, cette approche méthodologique est appliquée à l'étude de solutions modèles connues ; celle-ci est ensuite validée sur des produits agro-alimentaires naturels complexes : *des moûts de raisins blancs de cépages différents*.

Enfin, la quatrième partie est dédiée à la classification supervisée des cépages de moûts de raisins blancs, grâce à la fusion multisensorielle ; celle-ci intègre la mesure aromatique ainsi que les mesures issues des spectromètres *moyen infrarouge et ultraviolet*, qui analysent les composés en phase liquide.

Différentes méthodes de fusion multi-capteur sont étudiées :

- ⊗ la fusion de *niveau bas*, grâce à la mise en commun des données brutes ;
- ⊗ la fusion de *niveau haut*, prenant en compte les résultats des classifications individuelles, en s'appuyant sur la théorie Bayésienne.

Les démarches d'optimisation des capteurs d'arômes et des mesures spectrométriques ainsi que les techniques de fusion proposées pourront être généralisées à toute problématique de discrimination faisant intervenir des capteurs complexes.

# *Etat de l'art*

*1. Problématique*

*2. Capteurs d'arômes*

*3. Mise en œuvre*

*4. Traitement des données des systèmes*

*multi-capteurs*

## II. ETAT DE L'ART

Après avoir situé le contexte d'application de nos travaux de recherche, cet état de l'art présente les connaissances actuelles sur les « nez électroniques » et leur mise en œuvre, celles-ci ayant évolué très rapidement durant les 3 années des travaux de thèse.

La synthèse bibliographique se décompose en 3 parties :

- ⊗ la première partie rappelle quelques concepts liés à la qualité aromatique ;
- ⊗ la seconde partie est consacrée aux capteurs d'arômes eux-mêmes, à leur principe de fonctionnement et à la comparaison de leurs caractéristiques ainsi qu'aux différentes applications actuelles (pII-8) ;
- ⊗ la troisième partie présente les diverses conditions expérimentales inhérentes à la mise en œuvre des capteurs d'arômes, ainsi que leur influence sur la mesure aromatique (pII-22);
- ⊗ enfin, la quatrième partie expose de façon plus générale divers traitements multivariés, applicables aux mesures des capteurs d'arômes ; en outre, un chapitre est dédié à la fusion multi-capteur, en détaillant particulièrement les méthodes employées ultérieurement (pII-31).

### II.1. PROBLEMATIQUE

Notre étude se situe dans une démarche de caractérisation instrumentale de la qualité de produits viticoles.

#### II.1.1. Demande de qualité agro-alimentaire

Devant l'exigence croissante des consommateurs en matière de qualité des produits agro-alimentaires, l'industrie a développé des mesures instrumentales permettant de contrôler la qualité de leur production ; la qualité comprend classiquement 4 concepts (LAURET, 1982) : la *santé* (propriétés nutritionnelles), la *sécurité* (absence de composés nocifs), le *service* (adéquation du produit et de son utilisation) et la *qualité organoleptique* (goût, odeur, consistance, couleur, aspect, etc.).

Nous nous sommes spécifiquement intéressés à la mesure instrumentale de la *qualité organoleptique*. Celle-ci fait appel aux 5 sens (DULL, 1978). *L'ouïe* est un sens rarement simulé (à part le craquant de pains suédois, par exemple (BRESSOLETTE-HENRY, 1995)). Le *toucher* est généralement réduit à

des analyses de fermeté destructives ou non destructives (*DELWICHE, 1987 & FRANCOIS, 1996 & MOLTO et al., 1996*).

La *vision* est le sens le plus couramment simulé, grâce à des caméras CCD, permettant de caractériser les produits alimentaires en fonction de leur couleur, taille, forme ou défauts d'aspects (le tri de biscuits (*PERROT et al., 1996*), de fruits (*GUIZARD et al., 1992*), etc.).

La *gustation*, qui réunit dans un continuum sensoriel *principalement* les sensations de sucré, salé, acide et amer (*MACLEOD et SAUVAGEOT, 1986*), peut être grossièrement approchée par des mesures en spectrométrie infrarouge et ultraviolet, permettant par exemple de quantifier le taux de sucre dans les fruits (*BELLON et al., 1993*), etc.

L'*olfaction* est le sens le plus important mis en œuvre lors de la dégustation, comprenant une phase d'olfaction, puis de rétro-olfaction après la mise en bouche. Ce sens extrêmement complexe est difficilement simulé de façon instrumentale.

## II.1.2. Caractériser les arômes

L'évaluation des arômes d'un produit agro-alimentaire est généralement confiée à un *jury d'analyse sensorielle*. La mesure de référence instrumentale est la *Chromatographie en Phase Gazeuse*, capable de quantifier les arômes. Récemment, une méthode instrumentale alternative séduisante est apparue sur le marché : les « nez électroniques ».

### II.1.2.1. Analyse sensorielle

Dans l'industrie agro-alimentaire, la qualité organoleptique de produits est généralement évaluée par *analyse sensorielle* : c'est une démarche de description sensorielle objective d'un produit agro-alimentaire lourde à mettre en place (*MACLEOD et SAUVAGEOT, 1986*) ; en effet, elle nécessite la disponibilité à long terme et l'entraînement de nombreux dégustateurs, afin de déterminer et quantifier les descripteurs sensoriels décrivant les mêmes sensations organoleptiques pour tout le jury (*ETIEVANT et ISSANCHOU, 1987*). De plus, l'expert se doit de détecter et de caractériser les goûts ou arômes anormaux, en s'affranchissant de toute subjectivité liée à la méthodologie ou à la psychologie (stress, accoutumance, rhumes, etc.) (*LANGLOIS, 1998*).

### II.1.2.2. Chromatographie en Phase Gazeuse

La *Chromatographie en Phase Gazeuse* (CPG) est une technique séparative, capable de quantifier les molécules composant un mélange et de les identifier par couplage à un spectromètre de masse (*TRANCHANT, 1996*). Cette méthode, précise et reproductible, est considérée comme la *méthode de référence*. Cependant, elle est lourde et coûteuse, nécessitant généralement le traitement préalable des échantillons ainsi que l'emploi de personnel qualifié.

En outre, la disparité des seuils de perception entre les différentes molécules volatiles explique la très faible corrélation entre la perception humaine de l'intensité des odeurs et les pics chromatographiques. De même, la CPG ne permet pas de caractériser complètement un arôme, constitué par l'association de nombreuses molécules odorantes, en raison des phénomènes de masquage et de renforcement entre les molécules volatiles. Alors, le couplage entre olfaction humaine et mesure chimique peut être effectué par des techniques de *CPG-olfactométrie* ou « *sniffing* », afin de mettre en évidence les zones du chromatogramme qui contiennent des composés odorants (MOLIMARD, 1997).

Ces 2 méthodes ne peuvent pas être employées comme technique de routine en contrôle qualité de production dans une industrie agro-alimentaire.

### II.1.2.3. Mesure instrumentale par « nez électronique »

Nous nous sommes donc intéressés à de nouveaux instruments de mesure des arômes, généralement appelés « nez électroniques », réalisant une mesure plus proche de l'analyse sensorielle que la CPG grâce à la quantification globale des arômes. De nombreux articles de vulgarisation montrent l'engouement et l'espoir considérables suscités par cette nouvelle technologie dans la filière agro-alimentaire et cosmétique, au début des années 90 (MOY *et al.*, 1991a, 1991b, 1993 & TJOMB, 1993).

Cependant, quelques années après, ces systèmes multi-capteurs d'arômes sont encore au stade du développement, aucun appareillage n'étant mis en place pour un contrôle qualité systématique dans une industrie agro-alimentaire. Après une phase d'enthousiasme aveugle puis de déception et de rejet (ZINI, 1997), le milieu des années 90 a marqué le retour à l'étude réaliste de ces systèmes.

Ce type de mesure instrumentale paraît prometteur, mais les capacités métrologiques des diverses technologies de capteurs restent encore très imprécises. Il est donc nécessaire de caractériser les performances et les limites de ces appareillages, dans la mesure où leurs données doivent être intégrées à un système multi-capteur plus complexe.

### II.1.3. Mesure de la qualité dans la filière viti-vinicole

Dans la filière viti-vinicole, la qualité d'un vin est principalement évaluée par les oenologues. En parallèle, de multiples mesures analytiques sont réalisées afin de quantifier les nombreux composés contenus dans le raisin, le moût et le vin et de les corrélés aux pratiques culturelles et oenologiques ainsi qu'aux cépages et aux terroirs. L'apparition des « nez électroniques » a suscité de nouvelles ambitions dans la filière viti-vinicole, grâce à leur mesure rapide et globale des arômes.

### II.1.3.1. Problématique

Ces systèmes peuvent être appliqués à différentes problématiques liées à l'analyse aromatique :

⊗ *la détection rapide des défauts organoleptiques dans les vins*. Ce dépistage a pour objectif d'aider les oenologues dans de nombreuses dégustations triviales et fastidieuses de présélection des vins, afin de les préserver les sens des dégustateurs pour des analyses sensorielles plus subtiles.

Cependant, comme nous le verrons dans l'état de l'art (pII-26), cette voie se heurte rapidement au **verrou technologique constitué par la forte présence d'alcool dans les vins**, qui masque le signal des autres molécules volatiles. Des travaux de recherches doivent être menés afin de concevoir des méthodes de traitement préalable des échantillons de boissons alcoolisées.

Devant le délai imposé par ces travaux, nos recherches se sont orientées vers l'étude des moûts de raisin avant fermentation.

⊗ Les vins de cépages étant de plus en plus appréciés et commercialisés, l'identification et la révélation de la *typicité du cépage* sont des caractéristiques très recherchées. De plus, devant l'augmentation des achats de vendanges sur pieds ou après récolte, se pose le problème de l'authentification des cépages et de l'adultération des moûts.

⊗ L'objectif final de la caractérisation de la typicité des cépages blancs est l'estimation du *potentiel aromatique du raisin* ; celui-ci permettrait d'évaluer la qualité aromatique du vin à venir et de déterminer les méthodes de vinification appropriées. Ce diagnostic très rapide de la qualité du vin, évitant le recours à une (micro-)vinification, pourrait permettre d'étendre le diagnostic de la vendange à d'autres mesures que le degré alcoolique potentiel par réfractomètre ou taux de pourriture grise (estimation du *Botrytis cinerea* par la mesure de l'activité de l'enzyme laccase grâce au système Raisytis (SERES S.A., F)) (BARCELO, 1997).

La discrimination des défauts organoleptiques du vin par capteurs d'arômes ne sera pas exposée dans ce manuscrit ; nous ne détaillerons donc pas les composés volatils responsables des défauts olfactifs. Seule la classification des cépages de moûts de raisins blancs sera présentée ; la partie suivante est donc destinée à la description du *potentiel aromatique des moûts* et en particulier des arômes variétaux et de ceux développés au cours de la vinification.

### II.1.3.2. Arômes variétaux du moût

Le *potentiel aromatique variétal du raisin* peut se décomposer en deux parties (BAYONOVE, 1998) :

- ① *L'arôme variétal libre*, constitué par les substances volatiles typiques du cépage ;
- ② Les *précurseurs d'origine variétale*, essentiellement non volatils (glycosides, acides phénols, acides gras, etc.).

❶ Les substances volatiles typiques du cépage se décomposent en deux groupes : les *pyrazines*, caractéristiques de l'*arôme herbacé*, en particulier « poivron vert » dans les Cabernet Sauvignon et dans le Sauvignon (DUBOURDIEU, 1995), dont le seuil olfactif se situe aux alentours du  $\text{ng.l}^{-1}$  et les *terpénols* (linalol, nérol, géraniol, etc.), caractéristiques de l'odeur *fleurie* que l'on retrouve fortement dans le Muscat. Leur seuil individuel de détection avoisine les  $0.1\text{-}5\text{mg.l}^{-1}$ , mais ce seuil s'abaisse par phénomène de renforcement. Une Analyse Factorielle Discriminante basée sur la mesure des alcools terpéniques est la meilleure façon de discriminer les cépages des vins (RAPP, 1994), plutôt que de prendre en considération la fraction volatile fermentaire qui caractérise principalement l'*arôme vineux*.

❷ Les précurseurs d'arômes sont essentiellement non volatils et situés dans la pellicule et la pulpe (monoterpénols, polyols terpéniques, acides gras, caroténoïdes (RAZUNGLES, 1985), composés phénoliques, etc.). Cependant, les acides gras insaturés (acides linoléiques, linoléique, etc.), généralement estérifiés, sont transformés au cours des premières minutes de broyage à l'air des baies : cette phase préfermentaire de macération pelliculaire forme des aldéhydes et des alcools en C6, ayant l'odeur de feuille froissée et de verdure (RAPP, 1976). Mais, la proportion de ces composés semble peu varier d'un cépage à l'autre.

De nombreux précurseurs d'arômes non volatils se transforment au cours de la fermentation et/ou du vieillissement, et sont caractéristiques de l'arôme du cépage dans le vin. Une mesure de la composition en phase liquide est donc indispensable afin compléter la caractérisation variétale des moûts.

Les acides-phénols (*acides cinnamique, benzoïque*) sont un facteur de différenciation de cépages (ROMEYER, 1984), en se transformant en phénols volatils (*éthylphénol, gaiacol*) au cours de la fermentation ou restant sous forme glycosilée et ne se libérant que dans les vins vieillis (*arômes de brûlé, vanillé*).

Les arômes du vin de Sauvignon ont été étudiés par (DUBOURDIEU, 1995) ; différentes molécules volatiles spécifiques des arômes de Sauvignon ont été identifiées, telles que les *méthoxy-pyrazines* caractéristiques de l'arôme de poivron vert, mais aussi des molécules volatiles soufrées (TOMINAGA *et al.*, 1998), telles que la *4-méthyl-4-mercapto-pentan-2-one* (4MMP), caractéristique de l'odeur de « buis ». Ces composés soufrés, très peu concentrés dans le vin (concentration inférieure à  $15\text{ng.l}^{-1}$ ), se forment au cours de la fermentation à partir de précurseurs d'arômes non volatils, présents dans le moût (DUBOURDIEU *et* DARRIET, 1993).

La méthode de préparation des échantillons de moûts de raisin sera donc élaborée en tenant compte des connaissances précédentes, c'est-à-dire grâce à un foulage et une macération pelliculaire.

De plus, les capteurs à mettre en œuvre pour la discrimination variétale des moûts doivent être capables de mesurer la fraction volatile ainsi que les composés liés ; les capteurs d'arômes et les analyses spectrométriques IR-TF et UV semblent correspondre à ces spécifications.

## II.2. 2. LES CAPTEURS D'AROMES

Dans un premier temps, les « nez électroniques » et leur principe de fonctionnement sont définis, menant au choix de la dénomination de « capteurs d'arômes » plutôt que « nez électroniques » (pII-8). Ces systèmes multi-capteurs d'arômes sont constitués de transducteurs ou *éléments sensibles*, relevant de différentes technologies, qui sont détaillées dans la seconde partie (pII-10).

Les divers appareillages commerciaux sont présentés dans le troisième paragraphe (pII-13) et leurs caractéristiques sont comparées dans une quatrième partie (pII-16), afin de guider l'utilisateur dans le choix des technologies d'éléments sensibles ou d'appareillages.

Enfin, les différents types d'applications actuelles des capteurs d'arômes sont exposées (pII-19), celles-ci étant menées en laboratoire et non en contrôle en ligne.

### II.2.1. Définition : « nez électroniques » ou capteurs d'arômes ?

Une définition générale des « nez électroniques » est fournie par (GARDNER et BARTLETT, 1994a), de l'université de Warwick (GB) :

*« Un nez électronique est un instrument, qui comprend un groupe de capteurs chimiques électroniques avec une spécificité partielle et un système de reconnaissance des formes approprié, capable de reconnaître des odeurs simples et complexes. »*

Un « nez électronique » est donc construit sur un schéma similaire à celui d'un nez de mammifères (PERSAUD et DODD, 1982). Ce parallèle est illustré par la Figure 1.

Un « nez électronique » est composé de  $n$  capteurs, sensibles à une large gamme de molécules volatiles et comportant chacun une légère spécificité ; cependant,  $n$  est égal à environ 50 millions dans le neuro-épithélium olfactif humain alors que  $n = 5$  à 32 dans un « nez électronique ». Ces capteurs sont peu spécifiques, comme la centaine de types de chimiorécepteurs humains (JOURDAN, 1997).

Ces capteurs sont le lieu d'une adsorption réversible des composés volatils qui provoque un signal électrique, homologue de l'influx nerveux parcourant les nerfs olfactifs. Celui-ci est transmis au bulbe olfactif situé dans le cerveau, après une amplification du signal  $10^9$  fois, une compression de l'information et un accroissement du contraste spatial, permettant au nez humain d'atteindre des seuils de détection de l'ordre du  $\text{ng.l}^{-1}$  (JOURDAN, 1997). Cette phase d'amplification et d'amélioration du signal est absente des systèmes « nez électroniques », faiblement imité par les prétraitements appliqués aux courbes, mais ne permettant pas aux capteurs d'atteindre des seuils aussi bas.

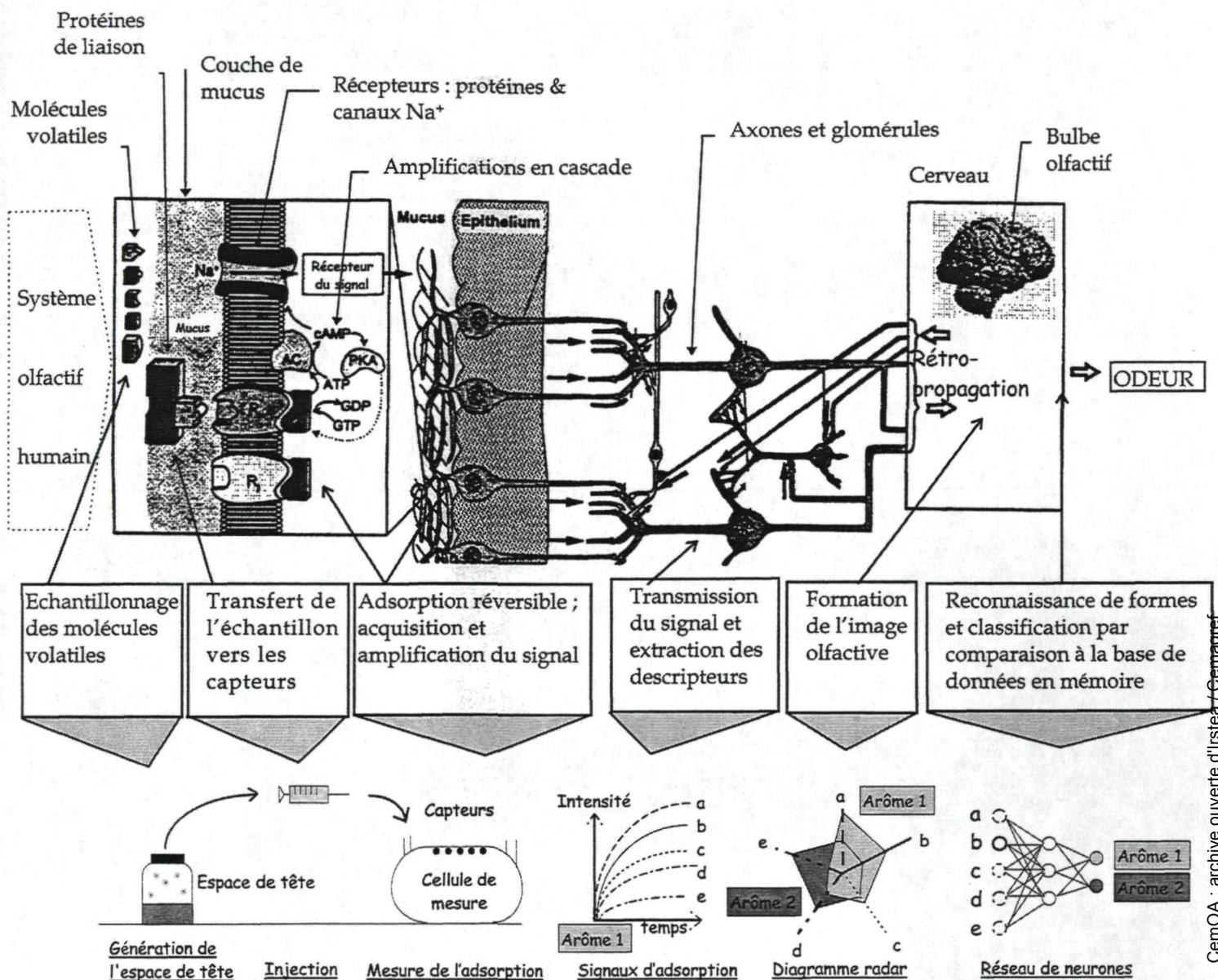


Figure 1 : Système d'olfaction humaine (GÖPEL, 1998) et système olfactif artificiel

L'*image olfactive* ou *signature olfactive* qui se crée dans le bulbe olfactif est l'homologue des *diagrammes radar*, représentant en étoile l'intensité de la réponse de chaque capteur (C.f. Figure 1). Cette image est alors reconnue par le cerveau grâce à sa comparaison avec la *base de données* des arômes en *mémoire* (DAVAL *et al.*, 1980). Ce processus correspond à la classification d'un arôme par un *système de reconnaissance des formes*, à partir de l'information extraite des courbes des capteurs. Le système de reconnaissance des formes peut être constitué par des réseaux de neurones artificiels, afin de se « rapprocher » de la structure du cerveau, ou par d'autres systèmes de traitement multivarié. La « mémoire » est constituée par les échantillons de calibration du modèle et les échantillons aromatiques nouveaux sont testés, classés et vont enrichir la base de données.

Etant données les différences notables entre les « nez électroniques » et les nez humains, comme le nombre de capteurs et leur sensibilité, la faiblesse du traitement du signal et des données, la pauvreté de la base de données ainsi que des problèmes que nous détaillerons ultérieurement, tels que le manque de reproductibilité, le masquage des signaux, l'influence des conditions expérimentales, etc., les « nez électroniques » seront plus modestement appelés *capteurs d'arômes* ou *systèmes multi-capteurs d'arômes*.

## II.2.2. Eléments sensibles

Les éléments sensibles composant les systèmes multi-capteurs d'arômes sont des capteurs chimiques très divers : les capteurs en oxydes métalliques semi-conducteurs (MOS), les capteurs en polymères conducteurs (CP) et les capteurs microbalances à quartz (QCM), les transistors à effet de champ en oxyde métallique semi-conducteur (MOSFET), les capteurs à onde acoustique de surface (SAW). La plupart des appareillages commercialisés comportent 3 premiers types de capteurs. De nombreux articles de synthèse présentent de façon détaillée les principes de fonctionnement de ces capteurs, tels que (SARIG, 1998) ou (MIELLE, 1998).

### II.2.2.1. Oxydes métalliques semi-conducteurs (MOS)

Les capteurs en oxydes métalliques semi-conducteurs employés pour la mesure aromatique sont formés d'un substrat en céramique sur lequel est déposé une couche d'oxyde métallique (en général  $\text{SnO}_2$ ), dopé par des matériaux, tels que le platine ou le palladium ; ces capteurs ont été largement étudiés et comparés (PAPADOPOULOS *et al.*, 1996 & SUZUKI *et TAKADA*, 1995 & VLACHOS *et al.*, 1996a). Ils sont chauffés à une température comprise entre 200 et 500°C, grâce à une résistance insérée dans la céramique, ce qui induit une forte consommation d'énergie (CORCORAN, 1993).

Ces oxydes métalliques sont sensibles aux composés réducteurs (MOS de *type n*), qui s'adsorbent de façon réversible à leur surface : les réactions d'oxydoréduction (CHANG, 1980) qui ont lieu entre l'oxygène préalablement adsorbé ( $\text{O}^-$  ou  $\text{O}^{2-}$ ) et les molécules volatiles provoquent une diminution de résistance du capteur, grâce au captage d'électrons (C.f. Annexe1). Ces réactions d'oxydoréduction en surface du semi-conducteur sont étudiées et peuvent être modélisées afin d'obtenir la sensibilité du capteur en fonction de la concentration du gaz, de la température de fonctionnement et des matériaux employés (BAO *et al.*, 1996 & BREUIL, 1989).

Les capteurs présents sur le marché sont commercialisés par des industries japonaises (FIGARO ENGINEERING, 1990) et (FIS INTELLIGENT SENSORS, 1994) ; ils ont été mis au point initialement pour la détection de gaz toxiques dans l'environnement (CO, NOx, alcanes, Fréon, etc.).

Ces capteurs sont sensibles à une très large gamme de molécules volatiles, avec une légère spécificité liée au type et à l'épaisseur de l'oxyde de métal employé, au type et à la concentration du « dopant »,

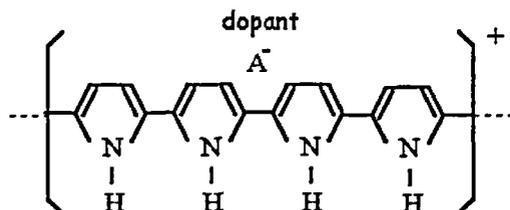
ainsi qu'à la température de surface du capteur. Néanmoins, leur spécificité est faible, ce qui peut gêner la distinction des arômes, malgré une sensibilité plus forte que les autres types d'éléments sensibles.

La méthode d'acquisition de la résistance des capteurs MOS est détaillée en [Annexe 2](#); celle-ci est une fonction non linéaire de la concentration en molécules volatiles (logarithmique ou  $C^n$  avec  $n < 1$ ).

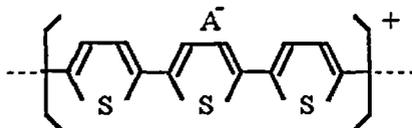
### II.2.2.2. Polymères conducteurs (CP)

Les capteurs en polymères conducteurs sont constitués par la polymérisation de monomères (*pyrrole, aniline, thiophène, etc.*) par voie chimique ou électrochimique.

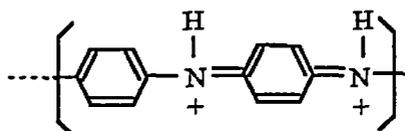
Les polymères employés pour la mesure des arômes sont des *polymères conducteurs intrinsèques*, i.e. leur conductivité provient de l'introduction par voie chimique d'impuretés non conductrices (dopant), qui produisent un excédent de charges positives ou négatives au cours de la polymérisation (*BARTLETT et al.*, 1989) (C.f. Figure 2).



**Polypyrrole conducteur**



**Polythiophène conducteur**



**polyaniline conductrice**

Les orbitales  $\pi$  se « conjuguent » pour créer une orbitale supramoléculaire délocalisée. Sans dopage, cette orbitale est « pleine » d'électrons, ce qui ne provoque aucun phénomène de conduction.

Le dopage par un constituant donneur d'électrons (comme le lithium) ou accepteur d'électrons (comme l'iode) crée un transfert d'électrons qui augmente la conductivité des polymères jusqu'à  $10^2 \Omega^{-1}\text{cm}^{-1}$  pour le polypyrrole, ou  $10 \Omega^{-1}\text{cm}^{-1}$  pour le polythiophène (*GARNIER*, 1987).

**Figure 2: structure des 3 principaux monomères**

L'adsorption réversible des composés volatils modifie la conductivité des polymères selon des mécanismes peu connus. Leur réponse est indépendante de la taille du capteur, ce qui autorise leur miniaturisation en un réseau sur une puce.

De plus, ils sont conducteurs à température ambiante, contrairement aux oxydes métalliques (*BARTLETT et GARDNER*, 1992). Leur variété peut être beaucoup plus large que celle des MOS, grâce à l'emploi de divers monomères ; cependant, ils sont beaucoup moins sensibles aux composés volatils, et sont particulièrement sensibles à l'eau (*NICOLAS et al.*, 1998b).

### II.2.2.3. Quartz Microbalance (QCM)

Les cristaux piézo-électriques, généralement appelés quartz (*Quartz Cristal Microbalance*), sont recouverts d'une fine couche (*coating*) sensible aux molécules volatiles (C.f. Figure 3). La fréquence de résonance du quartz est modifiée du fait de l'accroissement du poids dû à l'adsorption réversible des composés.

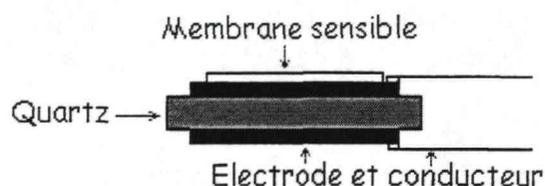


Figure 3 : schéma d'un capteur piézo-électrique QCM

La sensibilité de ces capteurs est fonction de la membrane sensible, mais aussi de la fréquence de vibration ainsi que de certaines conditions expérimentales, telles que la température (STOLL *et MITROVICS*, 1998). (DI NATALE *et al.*, 1997b) ont étudié différentes membranes, montrant la grande sélectivité de ces capteurs pour divers composés volatils lourds. Ils sont moins sensibles que les capteurs MOS, mais peuvent être extrêmement spécifiques, en particulier vis à vis de molécules volatiles lourdes, allant jusqu'à discriminer des isomères (IDE *et al.*, 1995) ou des énantiomères de monoterpènes, grâce à des QCM recouverts de films sensibles à base de  $\beta$ -cyclodextrines (MAY *et al.*, 1997).

### II.2.2.4. Oxydes métalliques semi-conducteurs à effet de champ (MOSFET)

Les capteurs en oxydes métalliques semi-conducteurs transistors à effet de champ (*Metal Oxide Semiconductor Field Effect Transistor*) sont composés d'une grille en métal catalytique (Pt, Pd, etc.) et d'une couche d'oxyde de Silicium entre lesquelles est interposé un matériau isolant. Les molécules qui s'adsorbent créent une polarisation électrique, qui provoque un changement de tension (LUNDSTRÖM *et PETERSSON*, 1996).

Leur température de fonctionnement est de l'ordre de 50 à 200°C, et celle-ci induit un changement de sélectivité, comme pour les capteurs MOS (WIDE *et al.*, 1997).

### II.2.2.5. Capteurs à ondes acoustiques de surface (SAW)

Les capteurs à ondes acoustiques de surface (*Surface Acoustic Waves*) sont similaires aux QCM, mais leurs vibrations se propagent en surface et non au sein du cristal. Aucun appareillage commercial ne contient ce type d'éléments sensibles, même s'ils ont été largement étudiés, par exemple pour la détection spécifique de NO<sub>x</sub> (PENZA *et VASANELLI*, 1997).

### II.2.2.6. Autres types d'éléments sensibles

Face à l'impossibilité de miniaturisation et à la forte consommation des capteurs MOS, des *capteurs semi-conducteurs fins* ont été développés. (ALTHAINZ *et al.*, 1996) ont fabriqué une barrette de moins de 1cm<sup>2</sup>, comportant 39 micro-capteurs en oxyde d'étain semi-conducteur ; un gradient de température est établi d'une extrémité à l'autre de la barrette (250-300°C). Leur seuil de détection est situé en dessous du mg.l<sup>-1</sup>.

Une toute autre démarche a été développée grâce à la collaboration du département de chimie et de neurosciences de l'université de Tufts (MA) et l'université de Pennstate (DIKINSON *et al.*, 1996 & JOHNSON *et al.*, 1997 & WHITE *et al.*, 1996) : la conception de *capteurs optiques* sensibles aux molécules volatiles. Ils sont constitués d'un faisceau de 19 fibres optiques, dans lequel est envoyé une radiation à 535 nm ; à son extrémité, des polymères de caractéristiques très différentes sont fluorescents à 610 nm lors de l'adsorption des composés volatils. Une caméra CCD acquiert l'*image olfactive* de l'arôme en fonction du temps, via le faisceau de fibres optiques et un microscope.

Ces capteurs ont l'avantage d'avoir une taille miniature, de pouvoir effectuer une mesure déportée et d'être insensibles aux interférences électriques. La mesure est rapide (quelques secondes) ainsi que le temps de récupération (1 à 2 minutes).

## II.2.3. Appareillages commerciaux

Les principales sociétés commercialisant des capteurs d'arômes sont recensées sur un serveur internet, géré par E. Vanneste (université d'Antwerpen, B), à l'adresse suivante :

<http://sch-www.uia.ac.be/struct/review/research@sales.html>

### II.2.3.1. Oxydes métalliques semi-conducteurs

2 appareillages commercialisés comportent uniquement des capteurs en oxyde métallique semi-conducteur ; il sont vendus par les sociétés MIDIVALEUR (Toulouse, F) et ALPHAMOS (Toulouse, F).

⊗ Le LCA 1000 est un prototype, mis au point par l'ENSC.T (TALOU, 1992) et commercialisé par la société MIDIVALEUR ; il comporte 5 éléments sensibles, de type FIGARO TGS série 8xx, dans une cellule de mesure en verre de 500ml. L'injection de l'échantillon est manuelle, grâce à une seringue à gaz de 50 ml. L'acquisition est effectuée par un logiciel, permettant de moduler la consigne en tension appliquée aux résistances chauffantes. Ce prototype ne comporte aucun contrôle des conditions expérimentales ; certains travaux de thèse ont cherché à améliorer ce prototype (BOURROUNET-DUBREUIL, 1998) ; mais, la société MIDIVALEUR a cessé son activité en 1998.

⊗ La société **ALPHAMOS** (<http://www.alpha-mos.com/>) commercialise différents types d'appareillage, contenant diverses technologies d'éléments sensibles ; les FOX 2000, 3000 et 4000 contiennent respectivement 6, 12 et 18 capteurs en oxydes métalliques semi-conducteurs. Les capteurs sont regroupés par 6, dans une cellule de mesure individuelle en plastique ; les capteurs MOS commerciaux à inclure dans l'appareillage sont sélectionnés par la société ALPHAMOS, qui compare leurs réponses à un standard grâce à un banc de dilution. L'injection de l'échantillon peut être manuelle ou automatique, grâce au couplage à un échantillonneur. Le logiciel d'acquisition permet aussi de réaliser les traitements statistiques multivariés classiques : ACP, AFD et réseaux de neurones. La température et l'hygrométrie sont mesurées ; la régulation en température des cellules de mesure est en cours de mise en œuvre dans certains appareillages, afin d'améliorer la reproductibilité des mesures.

### II.2.3.2. Polymères conducteurs

Trois appareillages sont uniquement constitués de capteurs en polymères conducteurs. Il s'agit de :

⊗ l'Aromascanner A32S de la société **AROMASCAN** (Manchester, GB) (*AROMASCAN*, 1995) (<http://www.aromascan.com/>), réalisé en collaboration avec l'université de MANCHESTER : **UMIST** (*NEAVES et HATFIELD*, 1995). Il contient une barrette de 32 polymères conducteurs. L'injection de l'échantillon est automatique, grâce à une poche souple ou une boucle d'échantillonnage. Malgré le contrôle de l'hygrométrie, la forte présence d'eau dans les espaces de tête masque fortement le signal des autres molécules volatiles ; la société AROMASCAN a donc proposé récemment de nouvelles améliorations afin de pouvoir discerner aussi les produits riches en eau (*VISSER et TAYLOR*, 1998).

⊗ l'appareillage e-NOSE<sup>®</sup>4000 (*Neotronics Olfactory Sensing Equipment*) de **NEOTRONICS** (Stansted Mountfitchet, GB) est composé de 12 capteurs polymères (<http://www.neotronics.com/>). L'espace de tête est injecté automatiquement et le logiciel d'acquisition permet d'effectuer quelques traitements statistiques (ACP, AFD, réseaux de neurones). Cet appareillage est moins répandu que l'Aromascanner.

⊗ La société **BLOOD HOUND SENSOR** (Leeds, GB) fabrique un capteur d'arômes, le BH114, contenant 14 polymères, développés en étroite collaboration avec l'université de **LEEDS** (<http://www.leeds.ac.uk/ulis/Bloodhound/Homepage001.htm>). L'échantillonnage est automatique, grâce à une poche souple ou par barbotage. Leurs travaux de recherche ont pour objectif d'élaborer des polymères présentant une moindre sensibilité à l'eau (*HULBERT*, 1998).

### II.2.3.3. Quartz microbalances

Deux appareillages comportent exclusivement des capteurs QCM, mais sont peu répandus :

⊗ **HKR SensorSystem** (Munich, D) a mis au point un dispositif qui contient 6 QCM, en collaboration avec l'université technique de **MUNICH** (<http://home.t-online.de/home/0897460207-0001/welcom~1.htm>).

⊗ **RST Rostock** (DK) (<http://www.rst-rostock.de/>), filiale de Daimler-Benz Aerospace commercialise un appareillage, appelé SamSelect, qui comprend 6 QCM (RAKOV, 1998). Une version simplifiée de cet instrument est embarqué à bord de la station MIR, afin de rendre compte de la qualité de l'air dans la cabine !

### II.2.3.4. Appareillages hybrides

Les appareillages hybrides, i.e. comportant plusieurs types de technologies, se sont développés plus récemment :

⊗ La société **ALPHAMOS** commercialise un système contenant 6 QCM, 6 CP et 6 à 12 MOS : le FOX 5000. Le protocole d'injection est modifié par rapport aux appareillages ne contenant que des capteurs MOS, car les QCM et CP sont moins sensibles : une boucle d'échantillonnage permet d'augmenter la concentration des molécules volatiles au contact des QCM, puis l'échantillon est chassé jusqu'aux autres cellules de mesure par le gaz vecteur.

⊗ La société **LENNARTZ** (Tübingen, D) commercialise le MOSES II (*Modular Sensor System*) ([http://www.lennartz-electronic.de/Lennartz\\_Homepage.html](http://www.lennartz-electronic.de/Lennartz_Homepage.html)). Il a été mis au point en collaboration avec l'université de **TÜBINGEN** (GÖPEL, 1998) et la société MOTECH. Il est modulaire et peut contenir 6 MOS et 6 QCM (STOLL et MITROVICS, 1998). L'injection est automatique, grâce à un module d'entrée ou au branchement d'un échantillonneur automatique.

⊗ La seule société commercialisant des capteurs MOSFET est la compagnie **NORDIC SENSORS** (Linköping, Se) sous la dénomination de NST (<http://www.nordicsensor.se/>). Ces capteurs ont été élaborés en collaboration avec l'université de **LINKÖPING**. Les appareillages NST contiennent 10 MOSFET, 5 MOS et un capteur infrarouge optionnel pour la mesure du CO<sub>2</sub> (KRAMER, 1998).

## II.2.4. Caractéristiques métrologiques comparées

Il est extrêmement difficile de comparer les performances des différentes technologies d'éléments sensibles, en l'absence de bancs d'essais normalisés, car les caractéristiques métrologiques dépendent des conditions de mesure ainsi que des molécules volatiles analysées (MARQUIS, 1997). Néanmoins, une revue bibliographique permet d'évaluer de façon très approximative - et parfois contradictoire - les performances des différents capteurs. Certains auteurs ont alors proposé des critères de performance adaptés aux systèmes multi-capteurs ainsi que des conditions expérimentales standard.

La seconde partie fait la synthèse de différentes comparaisons d'appareillages, rendant compte des performances des éléments sensibles ainsi que des montages expérimentaux.

### II.2.4.1. Caractéristiques métrologiques

Différentes caractéristiques métrologiques permettent d'estimer les performances des capteurs, basées sur les spécifications de la norme AFNOR (AFNOR, 1993). Les diverses performances recensées dans la bibliographie sont regroupées dans le Tableau 1.

Tableau 1 : comparaison des performances des différentes technologies d'éléments sensibles et du système olfactif humain.

①(BARTLETT et GARDNER, 1992); ② (CORCORAN, 1993); ③ (GARDNER et BARTLETT, 1995); ④(GARDNER et BARTLETT, 1994a); ⑤ (GARDNER et al., 1994) ; ⑥(MAY, et al., 1997) ; ⑦(DI NATALE, et al., 1997b) ; ⑧(DI NATALE et al., 1995a) ; ⑨(TAN et al., 1997) ; ⑩(MIELLE, 1996a).

CAPTEURS	SYSTEME BIOLOGIQUE	MOS EPAIS	MOS FINS	PC	QCM
Caractéristiques					
Seuil de détection & Sensibilité	1ppt-1ppm <sup>1</sup> ①	10-1000ppm ① ; 10ppm ② ; 0.1-100ppm pour les combustibles ③	1-100ppm ① 10ppm ②	vapeurs saturantes ② ; >20ppm ① ; ppb-ppm ③ ; 0.1-100ppm ④	forte pour les composés lourds ⑦
Sélectivité	faible	forte pour certains composés (triméthylamine) ④		faible ③	forte pour certains composés ⑥ ⑦ ⑩
Stabilité / dérive	Régénération ①	faible ① ④ <2%/h ②	faible ① <0.1%/h ②	modérée ① ; <0.1%/h ② ; Dérive induite par CO <sub>2</sub> ⑤	dérive aléatoire à long terme ③
Production Transférabilité		automatisée, mais problèmes de transférabilité ⑨ ⑩	non automatisée	non automatisée, transférabilité faible ⑩	non automatisée ⑩
Interférences	Infection ①	Cl <sub>2</sub> , H <sub>2</sub> O ①	Cl <sub>2</sub> , H <sub>2</sub> O ①	H <sub>2</sub> O ③ ⑤	
Choix du matériel	Vaste ①	Limité ①	Limité ①	Vaste ①	Vaste ①
Temps de réponse	~1 ms ①	~20 ms ① ; 20 s ② ; 0.5-5s ⑩	~20 ms ① 12 s ②	~60 ms ① ; 3-4min ② ; 20-40s ⑩	20-40s ⑩
Température	37°C ①	250-600°C ①	250-600°C ①	10-110°C ①	10-100°C
Consommation	nW ①	800 mW ① ⑩	80mW ①	<10mW ①	<10mW
Taille	Sub-micron ①	1 cm <sup>2</sup> ①	<1mm <sup>2</sup> ①	<1mm <sup>2</sup> ①	<1mm <sup>2</sup>

<sup>1</sup> 1ppm=1mg.l<sup>-1</sup> ; 1ppb=1µg.l<sup>-1</sup> ; 1ppt=1ng.l<sup>-1</sup>

Malgré l'imprécision de ces informations, il ressort que les capteurs MOS sont plus sensibles que les autres capteurs, mais que la sélectivité des QCM est bien supérieure.

La dérive à court terme est non négligeable, quelle que soit la technologie et la durée de vie des capteurs est limitée à quelques mois ou années. Malgré l'automatisation de certaines productions, la fabrication des capteurs semble peu reproductible, induisant une faible transférabilité.

Les possibilités de fabrication de divers CP et de QCM sont infiniment plus grandes que celles des MOS ; cependant, il semble difficile de trouver des monomères adaptés à la fabrication de membranes sensibles. La sensibilité à l'eau des CP est un véritable obstacle à la discrimination de la plupart des produits agro-alimentaires, alors que celle des MOS à l'alcool interdit toute discrimination de boissons alcoolisées (basée sur leurs arômes).

Les capteurs MOS sont beaucoup plus volumineux que les CP ou QCM, qui sont intégrés en une barrette ; des MOS fins ont été conçus afin d'améliorer la miniaturisation et la consommation, mais ne sont insérés dans aucun appareillage commercial.

Peu d'informations sont disponibles concernant les MOSFET et les SAW : la sensibilité des SAW avoisinerait les 0.1-100ppm (GARDNER et BARTLETT, 1994a) et leur sélectivité peut être extrême avec les composés lourds, tout comme les QCM (MIELLE, 1996a).

Devant la difficulté associée à la comparaison des diverses technologies de capteurs, (DAVIDE, et al., 1993) ainsi que (GARDNER et BARTLETT, 1996) ont défini des critères de performance pour les systèmes multi-capteurs d'arômes, pouvant servir de base à la mise en place d'un standard européen :

(DAVIDE, et al., 1993) généralisent 2 caractéristiques métrologiques aux systèmes multi-capteurs gaz : la sensibilité et la résolution.

(GARDNER et BARTLETT, 1996) reprennent le découpage de l'espace olfactif, proposé par (AMOORE, 1970), en 7 grandes familles d'arômes « primaires » (floral, éthéré, camphré, musqué, aigre/épicé, mentholé, putride). Les deux critères de performance sont :

⊗ la « gamme » du système multi-capteur, qui est sa capacité à discriminer des arômes issus de familles aromatiques différentes ;

⊗ le « pouvoir de résolution » qui est sa capacité à discriminer deux arômes similaires, i.e. issus de la même famille d'odeurs. Ce paramètre est considéré comme le plus important, car il est le plus représentatif des discriminations usuelles par capteurs d'arômes.

Le « pouvoir de résolution » (RP), déjà proposé par (MÜLLER, 1991), est défini comme :

$$RP = \frac{|S_{AB}|}{\sqrt{\sigma_A^2 + \sigma_B^2}} \quad \text{où : } |S_{AB}| \text{ est la distance entre les réponses des capteurs pour les arômes A et B ;}$$

$$\sigma_i \text{ est l'écart-type des répétitions de mesure de } i \text{ calculé le long de l'axe (AB).}$$

**Equation 1 : pouvoir de résolution**

Aucun banc d'essai n'a encore été mis en place. Les performances des éléments sensibles recensées dans le Tableau 1, sont très approximatives et n'ont pas vocation à aider l'utilisateur dans le choix d'un appareillage dont les paramètres de fonctionnement sont très différents.

Récemment, certains travaux ont été menés sur la comparaison des performances d'appareillages, en essayant d'optimiser le fonctionnement de chacun.

### II.2.4.2. Comparaison d'appareillages

(GARDNER et BARTLETT, 1994b) ont comparé 12 capteurs MOS et 12 polymères conducteurs fabriqués à Warwick-Southampton pour la discrimination de cafés et de bières ; ces deux technologies sont en cours de développement commercial, la première par la société Alphamos, la seconde par Neotronics. Les temps de réponse, sensibilité et stabilité dans le temps se sont avérés meilleurs pour les MOS alors que la discrimination était meilleure grâce aux polymères.

Plus récemment, des appareillages ont été comparés sur différents produits agro-alimentaires : les huiles alimentaires étudiées par l'institut Leatherhead Food RA (UK) (ADECHY, 1998 & SHIERS et ADECHY, 1998) ou le lait par TetraPak (Se) (DE LA PINSONNAIS, 1998).

⊗ La rancidité des huiles a été étudiée grâce à 4 appareillages, dont les conditions expérimentales ont été optimisées : l'A32S (Aromascan), contenant 32 PC, les FOX 3000 et 4000 (Alphamos) contenant respectivement 12 et 18 MOS et le NST 3220 (Nordic Sensors) (SHIERS et ADECHY, 1998). D'autres mesures ont été effectuées avec les appareillages BH 114 (Blood Hound), e-Nose 5000 (Neotronics) et MOSES II (Lennartz) : les échantillons codés ont été envoyés aux constructeurs, qui ont réalisé les mesures sans optimisation préalable ni répétitions dans le temps. Préalablement, les taux de peroxyde et d'anisidine ont été mesurés afin d'estimer la rancidité des huiles qui ont ensuite été congelées.

Cette étude a permis d'établir un classement dans les résultats des appareillages : l'A32S n'a pas été capable de discrimination, malgré l'absence d'eau dans les échantillons, l'espace de tête injecté n'étant pas suffisant ; l'appareillage NST est capable de discriminer les huiles rances des fraîches, mais peut être amélioré en réduisant le volume des flacons contenant l'espace de tête, afin de le concentrer (SHIERS et ADECHY, 1998). Les FOX ont pu distinguer les huiles fraîches des huiles rances (teneur en peroxyde > 10 mEq/kg) ; ils ont aussi été capables de discriminer plusieurs classes de rancidité à court-terme, mais ils manquent de reproductibilité à 1 semaine ou 1 mois d'intervalle.

Aucun des systèmes n'a été capable de prédire la durée de conservation de l'huile (ADECHY, 1998).

Sans phase d'optimisation, le BH 114 a pu discriminer les huiles rances des fraîches et le MOSES II a pu distinguer plusieurs classes de rancidité, mais aucune répétition de mesure n'a été faite à court ou long terme (ADECHY, 1998).

⊗ L'étude des arômes de « lait bouilli » a permis de comparer les appareillages A32S (Aromascan) et NST (Nordic Sensors) (DE LA PINSONNAIS, 1998). Les échantillons sont composés d'un lait de référence contenant 0 à 100% du même lait bouilli, dont le taux de matière grasse est constant (0.5%); le seuil de différence du panel sensoriel est de 30% de lait bouilli. L'A32S n'a détecté aucune différence entre le lait bouilli (à 100%) et le lait de référence, probablement à cause du fort taux d'humidité de l'espace de tête ; les capteurs d'arômes NST ont été capables de distinguer du lait référence les échantillons contenant 10% de lait bouilli. Mais, la quantification du taux de lait bouilli a été impossible à cause de la dérive des capteurs, qui ne permet pas d'établir une gradation entre les pourcentages successifs de lait bouilli.

### II.2.4.3. Conclusion

Il apparaît dans ces deux études que les polymères conducteurs (Aromascan & Bloodhound) semblent être moins performants que les autres technologies, d'une part à cause de leur sensibilité à l'humidité (la plupart des produits agro-alimentaires étant composés majoritairement d'eau, comme dans l'application sur le lait bouilli) et d'autre part à cause de leur manque de sensibilité (malgré l'absence d'eau dans les huiles, volume d'espace de tête injecté est trop faible). Ceci est confirmé grâce au FOX 5000 (Alphamos), contenant 12 MOS, 6 CP et 6 QCM : dans la majorité des applications, les signaux provenant des capteurs polymères ne sont pas discriminants, alors que la classification est possible grâce aux MOS (ALPHAMOS, 1997).

Les MOS semblent être plus sensibles aux arômes et moins sensibles à l'eau, cependant, leur problème majeur réside dans la forte dérive à court et long terme (ADECHY, 1998 & LUCAS, 1997).

La combinaison des technologies telles que les MOS et QCM (FOX5000 ou MOSES II) ou les MOS et MOSFET (NST) semble être la solution d'avenir, exploitant les caractéristiques complémentaires des 2 technologies, bénéficiant par exemple de la sensibilité des MOS et de la sélectivité des QCM (STOLL et MITROVICS, 1998).

### II.2.5. Différentes applications

Les applications des capteurs d'arômes concernent des mesures en laboratoire ainsi que quelques applications en contrôle des procédés.

Tableau 2 : applications des capteurs d'arômes

TYPE DE DISCRIMINATION	PRODUITS	REFERENCES
DETECTION DE DEFAUTS	Moisissure des céréales	(BORJESSON et al., 1996)
	Moisissure des bouchons de liège	(ROCHA et al., 1998)
	Arômes présents dans le sucre	(ROBILLARD, 1998)
	Rancidité des huiles	(ADECHY, 1998)
	Défauts dans les bières	(GARDNER, et al., 1994)

DETECTION DE DEFAUTS	Mauvaises odeurs dans les boîtes de médicaments	(SCHIFFMAN <i>et al.</i> , 1996)
	Produits carnés	(BERDAGUE <i>et TALOU</i> , 1993) (BOURROUNET <i>et al.</i> , 1995) (VERNAT-ROSSI, 1998)
	Défauts des oranges	(NAKAMOTO <i>et al.</i> , 1993)
CLASSIFICATION	Types d'arbres	(ZÜRCHER <i>et KUCERA</i> , 1997)
	Millésimes de vins	(DI NATALE <i>et al.</i> , 1995b)
	Terroirs de vins	(DI NATALE <i>et al.</i> , 1996)
	Variétés de carottes	(BRODA <i>et al.</i> , 1998)
	Qualité du papier	(HOLMBERG <i>et al.</i> , 1995)
PROCEDES	Contrôle de la maturité et tri des fruits	(DI NATALE <i>et al.</i> , 1998) (BENADY <i>et al.</i> , 1995) (SIMON <i>et al.</i> , 1996)
	Fraîcheur du poisson	(ETIENNE <i>et al.</i> , 1997) (OLAFSDOTTIR <i>et al.</i> , 1997) (SMELT <i>et al.</i> , 1997)
	Qualité du fromage	(MUIR <i>et al.</i> , 1997) (JOU <i>et HARPER</i> , 1998)
	Qualité du jus de fruit	(KONDOH <i>et SHIOKAWA</i> , 1994)
	Appétence de chats pour des nourritures animales	(NICOLAS <i>et al.</i> , 1997)
	Contrôle de bio-procédés	(NAMDEV <i>et al.</i> , 1998)
	Lait brûlé	(DI NATALE, <i>et al.</i> , 1998)
	Lait bouilli	(DE LA PINSONNAIS, 1998)
	Fermentation des saucisses	(ELKÖV <i>et al.</i> , 1998)
	Détection et identification de souches bactériennes	(GARDNER <i>et al.</i> , 1998) (GIBSON <i>et al.</i> , 1997) (HOLMBERG <i>et al.</i> , 1998)
CONTROLE ENVIRONNEMENTAL	Détection de pollen	(KALMAN <i>et al.</i> , 1997)
	Analyse des odeurs de fumier	(MISSELBROOK <i>et al.</i> , 1997) (PERSAUD <i>et al.</i> , 1996)
	Acétone, isopropanol, etc.	(KELLER <i>et al.</i> , 1994)
CONTROLE MEDICAL	Etat de santé des troupeaux de vaches	(ELLIOTMARTIN <i>et al.</i> , 1997)
	Détection du diabète	(PING <i>et al.</i> , 1997)

## II.2.6. Conclusion

Lors de la revue bibliographique des diverses technologies d'éléments sensibles et d'appareillages disponibles, différentes problématiques sont apparues :

❶ Quel(s) appareillage(s) et quelle(s) technologie(s) d'éléments sensibles sont les plus performants ?

Les indications issues des différentes caractéristiques métrologiques ainsi que de la comparaison d'appareillages mettent en avant les capteurs MOS, QMB et MOSFET. Le choix *a priori* des capteurs semble impossible, car leurs caractéristiques sont très floues. Néanmoins, les connaissances actuelles permettent d'affirmer que : les capteurs MOS sont extrêmement sensibles à l'alcool, donc ne sont pas appropriés à la mesure des boissons alcoolisées ; les capteurs polymères sont très sensibles à l'eau, donc ils discriminent difficilement des produits contenant majoritairement de l'eau ; les QCM sont particulièrement sensibles aux molécules volatiles lourdes.

Nous avons donc choisi d'employer un appareillage commercial simple, ouvert et flexible, doté de capteurs MOS : le LCA1000, prototype commercialisé par la société MIDIVALEUR. Il sera fortement modifié afin d'améliorer le contrôle des conditions expérimentales, l'acquisition et le traitement des données (C.f. chapitre III).

❷ Comment choisir les capteurs (et/ou l'appareillage) en fonction de l'application ? Parmi les capteurs MOS, de nombreux transducteurs sont disponibles. Des méthodes de sélection des éléments sensibles seront proposées et appliquées dans la suite de nos travaux (C.f. chapitre III).

La revue bibliographique a montré les limitations des éléments sensibles, du point de vue de leur seuil de détection, sensibilité, sélectivité et dérive. Les capteurs d'arômes sont loin d'atteindre les performances du nez humain ; il est donc essentiel d'améliorer les caractéristiques des capteurs, afin de pouvoir faire progresser les mesures aromatiques, notamment au niveau des seuils de détection et de la dérive à court terme.

La dérive à court et long terme semble aléatoire et particulièrement difficile à corriger (HOLMBERG *et al.*, 1996, 1997 & BAZZO *et al.*, 1998). Durant ces travaux de recherche, nous ne traiterons pas ce point particulier, mais nous nous donnerons les moyens de l'étudier ultérieurement grâce à la mesure quotidienne de solutions de référence durant nos analyses.

## II.3. MISE EN ŒUVRE DES CAPTEURS D'AROMES

Ce chapitre propose une revue bibliographique des différentes méthodes et conditions de mesures inhérentes aux capteurs d'arômes.

Les conditions expérimentales influent fortement sur la réussite de la discrimination ; elles sont détaillées dans la première partie. En outre, la formation de l'espace de tête ainsi que sa mise au contact des capteurs sont déterminantes dans la discrimination d'échantillons faiblement concentrés ou peu différents (pII-24) ; une partie spécifique (pII-26) est consacrée à la mesure des échantillons alcoolisés, qui nécessitent des précautions supplémentaires.

### II.3.1. Conditions expérimentales

Les conditions expérimentales influencent considérablement la qualité de la mesure des capteurs d'arômes, et donc leur capacité à discriminer les divers produits agro-alimentaires. La température et l'hygrométrie sont les facteurs les plus influents sur la mesure des capteurs d'arômes.

La température des capteurs MOS et de leur environnement (cellule de mesure) doit être précisément contrôlée. En effet, leur conductivité évolue de façon non linéaire en fonction de leur température de surface (C.f. Figure 4) (*FIGARO ENGINEERING*, 1990 & *FIS INTELLIGENT SENSORS*, 1994 & *MIELLE*, 1998).

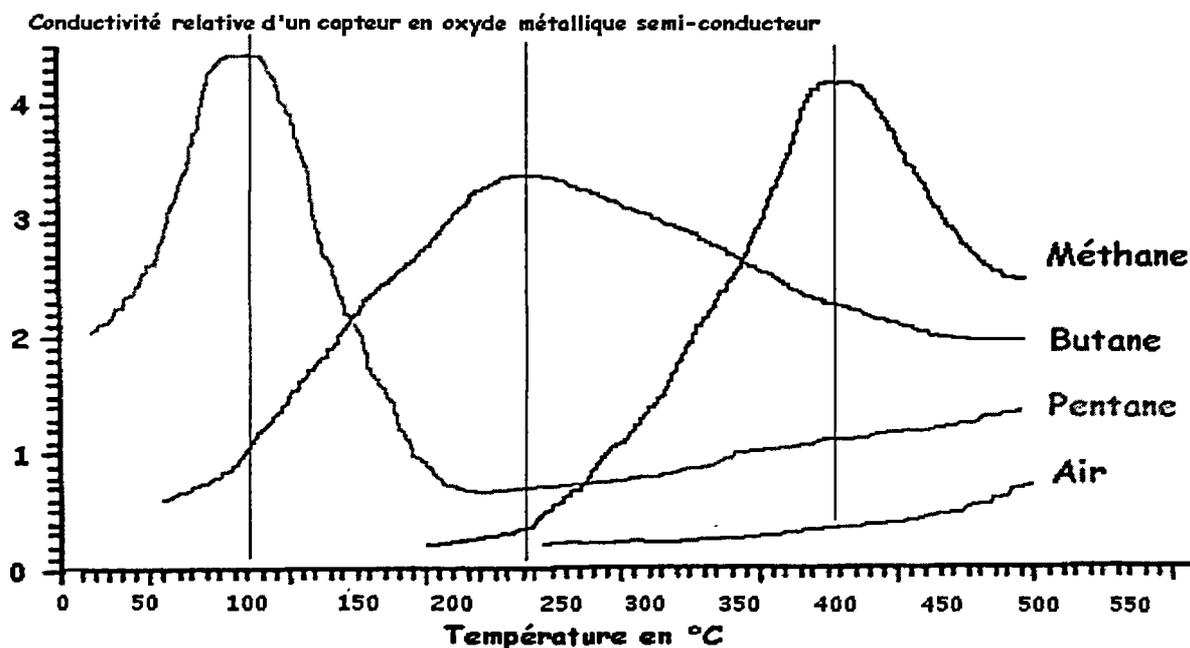


Figure 4 : influence de la température sur la sélectivité d'un capteur MOS (*MIELLE*, 1998)

De plus, diverses caractéristiques métrologiques, telles que la sensibilité et la résolution, varient fortement en fonction de la température (*DI NATALE et al.*, 1995c). En particulier, leur sensibilité est influencée de manière différente selon le composé volatil analysé : par exemple, la sensibilité de capteurs MOSFET à l'éthanol est beaucoup plus affectée par un changement de température que celle à l'hydrogène (*ELKÖV et al.*, 1997b).

La température de surface des capteurs est très sensible aux conditions régnant dans la cellule de mesure, telles que la température ou le débit du gaz vecteur : un accroissement du débit de 300ml.min<sup>-1</sup> induit une baisse de la température d'un capteur MOS de 20°C (*MIELLE*, 1996). C'est pourquoi il est nécessaire de réguler la température de surface des capteurs MOS ; ceci peut être effectué en maintenant toutes les conditions expérimentales stables dans l'environnement des capteurs, ou en régulant la température intrinsèque des capteurs, grâce au contrôle de leur résistance chauffante (*JONDA et al.*, 1996 & *MIELLE*, 1996).

L'hygrométrie influence aussi considérablement la réponse des capteurs d'arômes MOS. Plus l'humidité relative dans l'environnement du capteur est élevée, plus sa conductivité augmente. Mais, l'impact d'une variation d'hygrométrie dépend des composés volatils ainsi que du niveau d'humidité relative (*VLACHOS et al.*, 1995) : la sensibilité des MOS à certains composés, comme les alcanes, diminue avec l'hygrométrie, alors que la sensibilité à l'éthanol ou au CO reste inchangée. L'hygrométrie conditionne totalement l'image aromatique d'un composé volatil perçue par les capteurs : (*ROMAIN et al.*, 1997) ont montré qu'un réseau de neurones est incapable de reconnaître un arôme, s'il a été entraîné avec les mesures de cet arôme effectuées à un taux d'humidité différent.

La forte présence d'eau dans les échantillons peut être un obstacle à la discrimination, si les capteurs sont extrêmement sensibles à l'eau, tels que les polymères conducteurs. (*HOBBS et al.*, 1995) ont constaté que les déjections brutes (humides) de porcs et de poulets sont moins bien discernées que les mêmes échantillons purgés à l'air sec. Du fait de leurs différentes teneurs en eau, les produits agro-alimentaires conduisent à des espaces de tête d'une hygrométrie très hétérogène, induisant une forte variabilité intraclasse qui perturbe fortement la discrimination (*NICOLAS*, 1997, 1998).

D'autres paramètres expérimentaux causent des dysfonctionnements des capteurs d'arômes, tels que les volumes morts dans les tuyauteries et la cellule de mesure (*MIELLE*, 1998), la condensation des arômes dans des parties froides (*SMELT*, 1998) ou la durée insuffisante entre 2 injections induisant un fort « effet mémoire » (*ROBILLARD*, 1998). (*ELKÖV, et al.*, 1997b) ont conçu une cellule de mesure originale contenant des capteurs MOSFET : les capteurs sont placés en ligne et l'action catalytique à la surface de chacun d'eux change la composition du gaz, ce qui permet d'accroître leur distinction.

Mais, la qualité des mesures peut également être altérée par l'échantillon aromatique, dont la génération et le transfert vers la cellule de mesure sont extrêmement délicats.

## II.3.2. Génération et transfert de l'échantillon

Dans un premier temps, le cas général des produits agro-alimentaires sera traité, à l'exclusion des boissons alcoolisées. Celles-ci seront analysées séparément (C.f. pII-26).

### II.3.2.1. Echantillons non alcoolisés

Lors de la génération de l'espace de tête, il est extrêmement important de contrôler rigoureusement les paramètres expérimentaux que sont la température, l'hygrométrie, la durée d'équilibre de l'échantillon, la pression, etc. Il en est de même pour le transfert de l'échantillon dont la température, le volume, la vitesse et la pression doivent être constants.

Les températures de génération de l'espace de tête sont très diverses, variant de la température ambiante jusqu'à 80-85°C (ETIENNE, *et al.*, 1997 & NICOLAS, *et al.*, 1997) afin d'exacerber les arômes sans les altérer. La température de l'espace de tête doit être inférieure à celle de l'appareillage (tuyauteries et cellule de mesure) afin d'éviter la condensation des molécules volatiles dans l'eau contenue dans le gaz (SMELT, 1998).

Les durées d'équilibre sont aussi très différentes (de quelques minutes à quelques heures ; par exemple, 6h pour estimer la qualité olfactive du papier (HOLMBERG, *et al.*, 1995)), dépendant de l'inertie thermique de l'échantillon, de la volatilité des arômes et du volume du flacon contenant l'échantillon ; par exemple, il peut être nécessaire d'attendre 12 heures pour atteindre l'équilibre thermique au sein d'un melon (BENADY, *et al.*, 1995).

(MARQUIS, 1997) s'est intéressée aux différents modes de génération et de transfert de l'échantillon gazeux disponibles dans les divers appareillages commerciaux.

L'espace de tête peut être généré de façon *statique*, c'est-à-dire que l'échantillon liquide ou solide est enfermé dans un flacon hermétique dans lequel les composés volatils se dégagent, en tendant vers leur concentration d'équilibre liquide/gaz ou solide/gaz. La concentration d'une molécule volatile dans l'espace de tête dépend de sa nature (volatilité, coefficient de partage...) et de la matrice agro-alimentaire, mais aussi des conditions expérimentales, telles que la température, la durée d'équilibre, l'agitation ou la pression.

Cet espace de tête statique peut être généré dans la cellule de mesure contenant les capteurs (BENADY, *et al.*, 1995) ou dans un flacon indépendant, une partie de l'espace de tête étant ensuite prélevée et injectée dans la cellule de mesure.

L'avantage de la première méthode (génération de l'espace de tête dans la cellule de mesure) est de pouvoir suivre la cinétique de formation de l'espace de tête, mais elle provoque une pollution importante des capteurs, nécessitant des cycles de nettoyage très longs ; de plus, lors de la mise en place de l'échantillon liquide ou solide dans la cellule de mesure, il est extrêmement difficile de contrôler la qualité de la phase gazeuse emprisonnée.

La seconde méthode (génération de l'espace de tête dans un flacon indépendant) est plus couramment utilisée, quel que soit le type de d'élément sensible employé : capteurs MOS (PICQUE *et al.*, 1997), MOSFET (BORJESSON, *et al.*, 1996) ou QCM (DI NATALE, *et al.*, 1997b). Une fraction d'espace de tête peut être injectée grâce à une seringue dans la cellule de mesure, en présence ou non d'un flux de gaz vecteur (30 à 300 ml.min<sup>-1</sup>). Cependant, les proportions entre le volume de l'échantillon et celui de la cellule de mesure ainsi que le volume de gaz vecteur ajouté peut entraîner un taux de dilution élevée ; en outre, ce type d'injection est parfois manuel, ce qui peut induire un manque de répétabilité.

Le transfert de l'espace de tête peut aussi être effectué via une *boucle d'échantillonnage* (« *sample loop* »), souvent employée pour les capteurs les moins sensibles, tels que les PC et les QCM dans les Fox 5000 (ALPHAMOS), Moses (LENNARTZ) et AromaScanner (AROMASCAN). Pour réduire la dilution, l'espace de tête peut être aspiré à partir d'une poche souple (AROMASCAN). Cependant, l'utilisation de poches à usage unique est très onéreuse et le volume de la poche ou de la boucle d'échantillonnage doit être suffisant par rapport à celui de la cellule de mesure ; de plus, la boucle d'échantillonnage est souvent couplée à une méthode de pressurisation du flacon contenant l'échantillon, ce qui ne favorise pas le passage en phase gazeuse des composés volatils.

L'espace de tête peut être généré de façon *dynamique*, c'est-à-dire par barbotage du gaz vecteur dans les échantillons liquides ou par balayage en surface d'échantillons visqueux ou solides, l'échantillon gazeux passant directement dans la cellule de mesure (EMA *et al.*, 1989). L'équilibre des phases n'étant pas atteint, la concentration instantanée des molécules volatiles dans le gaz vecteur est plus faible que dans un espace de tête statique (NAKAMOTO *et al.*, 1991), mais cette méthode évite la dilution subie par un espace de tête statique injecté en présence d'un flux de gaz vecteur. Ce transfert d'échantillon peut être précédé par une phase de génération statique, afin d'augmenter la concentration initiale des arômes (HOLMBERG, *et al.*, 1995).

### II.3.2.2. Échantillons alcoolisés

La discrimination de boissons alcoolisées est particulière en raison de la présence d'un composé majoritaire, l'éthanol, qui masque le signal des autres molécules volatiles (leur concentration allant du mg.l<sup>-1</sup> au ng.l<sup>-1</sup>). L'éthanol est une molécule très concentrée dans les boissons alcoolisées, d'une centaine de g.l<sup>-1</sup> dans un vin jusqu'à 300-400g.l<sup>-1</sup> dans un whisky ; elle est en outre extrêmement volatile et oxydable, ce qui la rend très réactive, quelle que soit la technologie de capteurs d'arômes employée.

L'éthanol peut induire des différences prépondérantes dans la composition d'un espace de tête qui ne sont pas liées à la qualité aromatique des échantillons analysés. Ce type d'artefact, dû à la présence d'un composé extrêmement majoritaire, peut être occasionné de façon similaire par la présence d'eau ou de lipides : la forte hétérogénéité intraclasse introduite par différentes activités de l'eau altère fortement la discrimination basée sur des polymères conducteurs (NICOLAS, 1998 & NICOLAS, *et al.*, 1997) ; de même, le contrôle du taux de lipides (avec lesquels les arômes se combinent) est important pour certains produits agro-alimentaires, tels que le lait (DE LA PINSONNAIS, 1998).

L'objectif des mesures par capteurs d'arômes est généralement la discrimination d'échantillons alcoolisés, sans connaître les molécules volatiles qui provoquent le signal global des éléments sensibles. Ainsi, la distinction entre boissons alcoolisées peut simplement être due à d'infimes écarts de la concentration en éthanol, si celle-ci n'est pas mesurée *précisément* ; en effet, une variation de 10 millièmes de la concentration en éthanol (10<sup>-5</sup>), correspond à un écart de 1mg.l<sup>-1</sup>, ce qui équivaut à la concentration des molécules aromatiques les plus concentrées ! (ROBILLARD, 1998) a montré que la discrimination de moûts de raisin en début de fermentation, contenant en moyenne 0.5mg.l<sup>-1</sup> d'éthanol, est essentiellement basée sur la teneur en alcool.

Certains auteurs ont même utilisé ces capteurs pour la mesure de l'éthanol, dans des gammes de 100-1000mg.l<sup>-1</sup> (AUSTIN *et* RUSSELL, 1996 & MAEKAWA *et al.*, 1992 & ZIMMERMANN *et* LECLERCQ, 1995).

Nous avons abordé le problème en faisant une bibliographie des travaux concernant la mesure de boissons alcoolisées par capteurs d'arômes ainsi que les diverses techniques permettant de surmonter ce verrou technologique (PRIVAT *et al.*, 1998).

☒ Les premières études de discrimination de boissons alcoolisées par capteurs d'arômes ne se sont pas préoccupés de la présence d'éthanol.

De nombreux travaux ont été effectués par les équipes de recherches de Warwick (GB) sur la discrimination de bières (CORCORAN, 1993 & PEARCE *et al.*, 1993 & SHURMER *et* GARDNER, 1992), sans mesurer les teneurs en éthanol. Les derniers travaux montrent une prise en

compte progressive de l'importance de l'alcool, puisque (*GARDNER, et al.*, 1994) ont développé un système d'injection automatique pour la détection de traces de diacétyle (de l'ordre d'un  $\text{mg.l}^{-1}$ ) dans des solutions éthanolées modèles à 4% v:v.

(*ALEXANDER et al.*, 1998) propose même un système portable pour le contrôle de procédé dans les brasseries, basé sur 6 MOS et un réseau de neurones, sans jamais aborder le problème de l'alcool.

De même, l'équipe de l'université de Roma Tor Vergata (I) s'est intéressée à la classification de vins de différents terroirs (*DI NATALE et al.*, 1996) ou millésimes (*DI NATALE et al.*, 1995b). Les résultats sont critiquables car certains vins ont vieilli en fûts de chêne ayant contenu au préalable du cognac et sont alors correctement distingués des autres millésimes (*DI NATALE, et al.*, 1995b). De plus, les vapeurs analysées sont des dilutions de 2 et 6% v:v de l'espace de tête initial ; ainsi, une molécule volatile concentrée à  $1\text{mg.l}^{-1}$  est alors présente aux concentrations de 20 à  $60\mu\text{g.l}^{-1}$  au contact des capteurs d'arômes, ce qui est largement en dessous du seuil de perception des capteurs QCM !

De la même manière, (*NANTO, et al.*, 1993) discriminent correctement un vin rosé, un vin blanc, un vin rouge, un brandy et un whisky à l'aide de capteurs QCM combinés à un réseau de neurones, puis avec un seul capteur QCM (*NANTO et al.*, 1995). En employant un seul capteur, il est clair que la distinction entre les échantillons est effectuée sur « l'intensité aromatique », principalement due au composé majoritaire présent dans des concentrations très différentes : l'éthanol.

⊗ Néanmoins, certains auteurs ont cherché à intégrer la teneur en alcool dans la mesure des arômes, ou à diminuer la teneur en alcool des échantillons.

(*NAKAMOTO, et al.*, 1991) distinguent 5 whiskies à l'aide de 8 QCM ; afin d'accroître les infimes différences dues aux molécules volatiles, le signal d'un des échantillons alcoolisés est pris comme référence et soustrait aux autres. Mais, cette méthode revient à soustraire un biais constant qui n'est pas en relation avec la teneur en alcool de chaque échantillon.

(*EMA, et al.*, 1989) utilisent 6 capteurs QCM et un réseau de neurones pour identifier 11 types de liqueurs. Travaillant avec des whiskies, un brandy, une liqueur et un vin, ils montrent justement que le système peut être utilisé pour mesurer des teneurs en alcool. Pour identifier la typicité, et non la teneur en alcool, à chaque mesure est soustraite la valeur obtenue sur un témoin (solution pure de même concentration en éthanol). Le taux de reconnaissance des 11 produits est alors accru de 67 à 81%. De façon similaire, (*SLATER et al.*, 1993) discriminent 28 échantillons de deux brandy, un gin et une liqueur à partir d'un capteur en polymère conducteur grâce à une ACP. Ils observent que la meilleure discrimination est obtenue lorsque l'on soustrait du signal celui de la vapeur d'éthanol de même concentration.

Cette méthode peut être appliquée à condition de connaître *la teneur exacte* en alcool des produits à analyser : (ROBILLARD, 1997) montre que la standardisation à un degré alcoolique commun (par exemple 12.1% v:v) génère une dispersion dans les échantillons, de l'ordre de  $\pm 0.1\%$  v:v (soit  $1000\text{g.l}^{-1}$ ), qui perturbe la discrimination aromatique par capteurs d'arômes ; ainsi, les différences mesurées ne sont pas celles perçues par le panel sensoriel.

De plus, il est nécessaire que les capteurs ne soient pas saturés par de telles concentrations en éthanol ; ceci est plus facile à réaliser avec les capteurs QCM ou CP qui sont beaucoup moins sensibles que les capteurs  $\text{SnO}_2$ .

⊗ D'autres auteurs ont cherché à diminuer la concentration en éthanol.

(AISHIMA, 1991) propose d'éliminer l'effet de l'alcool en incluant une étape de concentration aromatique et d'élimination de l'éthanol par des pièges de composés hydrophobes en polymères poreux. Avec 6 capteurs MOS et une classification hiérarchique ascendante, il discrimine 11 liqueurs. Toutefois, il ne vérifie pas le taux d'éthanol restant en sortie du concentrateur d'espace de tête.

Une autre méthode permet d'abaisser le taux en éthanol de l'espace de tête : quelques gouttes du liquide alcoolisé sont déposées sur une languette en papier, de laquelle une partie de l'éthanol s'évapore, avant l'introduction de la languette dans un flacon. Une bonne discrimination de menthol dissous dans 20% v:v d'éthanol a été obtenue, bien qu'une grande partie des molécules volatiles s'évaporent simultanément à l'éthanol (ROBIE, 1997). Cette méthode ne s'applique que si la concentration en composés aromatiques est élevée par rapport à celle de l'éthanol, comme par exemple dans les parfums. De plus, il est préférable que la volatilité des arômes étudiés soit inférieure à celle de l'éthanol.

⊗ Pour les systèmes actuels, l'influence de l'alcool est encore réhibitoire.

En collaboration avec la société ALPHAMOS, nous avons mené une étude préliminaire pour caractériser l'influence de l'alcool sur la discrimination des arômes, grâce à l'étude de solutions hydro-alcooliques modèles et de vins, contenant des ajouts dosés de molécules responsables du goût de piqué du vin (acide acétique à  $1\text{ g.l}^{-1}$  ou acétate d'éthyle à  $300\text{ mg.l}^{-1}$ ). Les mesures ont été réalisées grâce au FOX 4000, contenant 18 capteurs MOS et au FOX 5000, composé de 12 capteurs MOS, 6 PC et 6 QCM (LUCAS, 1997).

Le premier problème identifié est le manque de reproductibilité des mesures au cours du temps : quel que soit l'appareillage employé, la dérive au cours d'une journée induit plus de variabilité que la différence intergroupe des répétitions de mesures de solutions modèles contenant 10, 11 et 12% v:v d'éthanol.

La discrimination d'une solution de référence à 12% v:v d'éthanol de la même solution contenant de l'acide acétique à  $1\text{ g.l}^{-1}$  ou acétate d'éthyle à  $300\text{ mg.l}^{-1}$  est impossible grâce à l'ensemble des capteurs du FOX 5000 ; néanmoins, après sélection de trois capteurs appartenant à chacune des technologies,

la discrimination entre la solution de référence et celle contaminée par l'acétate d'éthyle est possible, grâce à une ACP. En effet, cette molécule possède une volatilité supérieure à l'acide acétique (environ mille fois) et également une plus grande réactivité vis à vis des capteurs.

L'étude de la discrimination d'un vin de référence et du même vin contenant les contaminants fournit des résultats variables selon les technologies employées : la discrimination basée sur le FOX 5000 est impossible ; cependant, le FOX 4000 distingue le vin contaminé par l'acétate d'éthyle du vin de référence.

Ce résultat rejoint les études menées sur les vins par le laboratoire de recherche de Moët & Chandon : (**ROBILLARD**, 1997) a montré qu'il est impossible de discriminer un vin de référence et ce même vin contenant des ajouts dosés de 100mg.l<sup>-1</sup> de lactate d'éthyle ou d'acide acétique avec un FOX 3000 contenant 12 MOS. Afin d'exacerber les arômes en phase volatile, des sels de sodium ont été introduits dans les vins, mais la baisse de la répétabilité induite par ce procédé annihile tout gain de concentration.

**La présence d'alcool constitue un verrou technologique dans la discrimination aromatique des boissons alcoolisées. Les solutions peuvent être issues de l'amélioration des capteurs ou de traitements préalables de l'échantillon.**

La première approche, encore peu explorée aujourd'hui, consiste à développer de nouveaux capteurs (C.f. pII-13) plus sensibles et sélectifs vis à vis de molécules volatiles composant les espaces de tête de produits agro-alimentaires. (**VLACHOS et al.**, 1996b) proposent d'employer l'effet de courants de « Seebeck » entre 2 capteurs MOS, proportionnels à l'adsorption différentielle des molécules volatiles à la surface de 2 capteurs catalytiques afin d'améliorer leur sélectivité. De même, (**PIJOLAT et al.**, 1997) ont cherché à diminuer la sensibilité des capteurs MOS à l'alcool.

La seconde approche, relevant du domaine du génie des procédés, consiste à modifier l'échantillon mis au contact des capteurs. Cette modification peut aussi bien avoir lieu au niveau de l'échantillon liquide ou gazeux. Celle-ci a été ébauchée par (**AISHIMA**, 1991a).

### **II.3.3. Conclusion**

---

---

Les différentes possibilités de mise en œuvre des capteurs d'arômes ont été explorées, faisant apparaître divers points délicats de la mesure aromatique :

- ❶ De nombreuses conditions expérimentales influent largement sur la mesure des éléments sensibles, quelle que soit leur technologie. Comment choisir les conditions expérimentales optimales, de la génération de l'espace de tête à la mesure ?

Dans le chapitre III, nous mettrons en œuvre un plan d'expériences, permettant de quantifier l'influence de 3 facteurs expérimentaux sur la qualité des mesures. La méthode exposée pourra être appliquée à tout appareillage et toute discrimination aromatique.

- ② La distinction des boissons alcoolisées semble être essentiellement basée sur la teneur alcoolique. Peut-on réellement discriminer les arômes de boissons alcoolisées, et non pas l'intensité alcoolique ?

Il semble que la discrimination aromatique d'échantillons alcoolisés nécessite un traitement préalable, visant à diminuer le taux d'éthanol et à concentrer les autres molécules volatiles. La classification des boissons alcoolisées par capteurs d'arômes ne sera donc pas traitée dans nos exemple d'application. L'étude de la typicité des vins sera alors réalisée sur les moûts de raisins exempts d'alcool.

En parallèle aux travaux de thèse, les conclusions de cette synthèse bibliographique ont servi de base au montage d'un projet multidisciplinaire visant à éliminer l'alcool dans l'espace de tête.

Ces travaux de recherches, financés par la DRAF<sup>2</sup>, l'ACTIA, TRIAL et la DGAL, sont menés en collaboration avec des industriels (les Chais Beaucairois S.A., ALPHAMOS), des centres techniques (Institut Technique de la Vigne et du Vin (ITV) de Gaillac, Pech Rouge et Nîmes) et des centres de recherches (CEMAGREF (GIQUAL), INRA-IPV<sup>3</sup>, ENSC.M (LMPM<sup>4</sup>), UM II (LGBSA<sup>5</sup>)).

Les différentes techniques de désalcoolisation ont été recensées par (PRIVAT, *et al.*, 1998) et sont actuellement mises en œuvre et optimisées par des équipes de l'université de Montpellier et de l'IPV. Les procédés thermiques ont été éliminés, car la plupart des molécules volatiles disparaissent en même temps que l'alcool (ESCUDIER, 1990). Les méthodes d'extraction liquide / liquide ont été testées avec succès (PRIVAT, 1997), parvenant à diviser par 4 la teneur en éthanol d'un vin modèle à 12% v.v. Les techniques membranaires de concentration par piégeage/relargage (*purge & trap*) ainsi que de pervaporation (BALDET, 1999) sont actuellement étudiées et leurs résultats seront publiés fin 1999.

---

<sup>2</sup> DRAF : Direction Régionale de l'Agriculture et de la Forêt - Filière viti-vinicole ; ACTIA : Association de Coordination Technique pour l'Industrie Agro-alimentaire ; TRIAL : Transfert Innovation Agro-alimentaire ; DGAL : Direction Générale de l'Alimentation.

<sup>3</sup> Institut de Production de la Vigne de Montpellier et Pech Rouge (11)

<sup>4</sup> Laboratoire des Matériaux et Procédés Membranaires (Montpellier)

<sup>5</sup> Université Montpellier II - Laboratoire de Génie Biologique et Sciences des Aliments

## II.4. TRAITEMENT DES DONNEES DES SYSTEMES MULTI-CAPTEURS

Ce chapitre est consacré à divers aspects du traitement des données de systèmes multi-capteurs, tout en analysant spécifiquement les problématiques inhérentes aux capteurs d'arômes.

Dans un premier temps, les diverses *méthodes d'acquisition et de prétraitement des données* issues des capteurs d'arômes sont détaillées. Dans un second paragraphe (pII-35), les *méthodes de reconnaissance des formes*, employées pour les classifications à partir des mesures de capteurs d'arômes sont exposées. La troisième partie présente les rares travaux portant sur la *sélection des capteurs* d'arômes (pII-39) et détaille les diverses techniques de sélection de variables applicables à la sélection de capteurs (pII-40). Enfin, la quatrième partie introduit le concept de *fusion multi-capteur*, en développant plus particulièrement 2 méthodes de fusion, l'une destinée à la fusion de courbes (pII-45), l'autre à la fusion de classifications grâce à la théorie Bayésienne (pII-46).

### II.4.1. Acquisition et prétraitement des données

Le signal de chaque élément sensible est une courbe constituée par l'acquisition de son changement de résistance (ou de fréquence s'il s'agit de QCM) lors de l'adsorption réversible des molécules volatiles à sa surface. Cette courbe comporte une *partie transitoire* lors de l'adsorption des composés, pendant laquelle la résistance chute, donc la tension acquise augmente (C.f. Annexe 2).

Elle est généralement suivie d'une *partie statique*, d'équilibre entre l'adsorption et la désorption ; le temps de réponse dépend du mode d'injection de l'échantillon (C.f. Figure 5). Enfin, lors de l'injection du gaz vecteur qui nettoie la cellule de mesure et la surface des capteurs, la *phase de désorption* montre le retour de la tension vers l'état initial, appelée *ligne de base*.

Dans la plupart des travaux, seul le maximum d'adsorption est extrait du signal, correspondant à l'intensité aromatique perçue par l'élément sensible durant la phase stationnaire. Divers *prétraitements* peuvent être effectués sur le maximum d'adsorption, prenant en compte la ligne de base, avant d'introduire ce paramètre dans un système de classification (*GARDNER et BARTLETT, 1994a*). L'un des prétraitements les plus efficaces est la « différence fractionnaire » (*fractional difference*) :  $Y = (V_{\text{max.}} - V_{\text{ligne de base}}) / V_{\text{ligne de base}}$ .

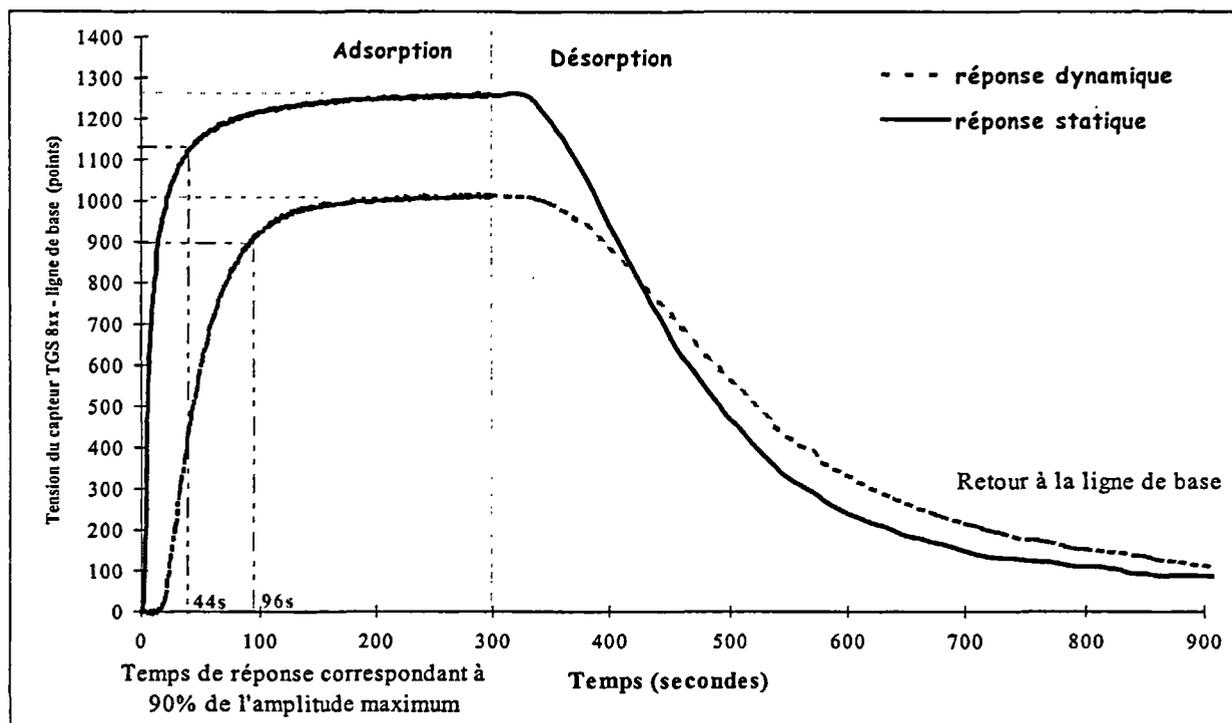


Figure 5 : courbes d'adsorption/désorption avec injection en mode statique et dynamique

Néanmoins, de rares articles exploitent non seulement la réponse stationnaire (maximum de la courbe), mais aussi la réponse transitoire pour la discrimination des arômes. Les trois stratégies développées dans la littérature sont les suivantes :

❶ Une première approche consiste à échantillonner le signal.

(*YEA*, 1994) a introduit dans un réseau de neurones la résistance de capteurs MOS acquise toutes les secondes. De la même manière, (*NANTO, et al.*, 1995) ont extrait la fréquence de résonance de capteurs QCM toutes les minutes, divisée par la fréquence de vibration maximale.

Cette méthode présente l'avantage de prendre en considération l'intégralité de la courbe ; cependant, en raison du grand nombre de variables extraites (points échantillonnés), le nombre d'échantillons mesurés doit être extrêmement conséquent, afin de pouvoir bâtir des modèles statistiquement valides. De plus, ce type d'acquisition ne prend pas réellement en compte la forme de la courbe (pente, point d'inflexion, etc.) et les acquisitions successives sont extrêmement corrélées.

Cette méthode est améliorée par (*NICOLAS et al.*, 1998a), qui sélectionnent les parties discriminantes de la courbe de réponse de polymères conducteurs, grâce à une analyse de variance en chaque point de la courbe. Les zones les plus intéressantes se situent en fin d'adsorption. Cette démarche, applicable à tout signal, permet de réduire le nombre d'échantillons mesurés, mais ne prend pas en compte la redondance des informations prélevées.

**② Une seconde méthode consiste à modéliser la courbe d'adsorption et à introduire les coefficients du modèle dans le système de classification.**

Les fonctions utilisées sont non linéaires, telles que la sigmoïde de Gompertz<sup>6</sup> employée par (VERNAT-ROSSI *et al.*, 1996), la fonction de Weibull<sup>7</sup> utilisée par (PICQUE, *et al.*, 1997), un modèle exponentiel plus simple<sup>8</sup> appliqué par (ELKÖV *et al.*, 1997a), ou encore un modèle polynomial de degré 3 (ELKÖV, *et al.*, 1997a).

Les résultats de classification employant les coefficients du modèle sont meilleurs que ceux utilisant uniquement la réponse stationnaire ; en effet, les coefficients modélisant la partie transitoire du signal sont beaucoup moins influencés par les conditions expérimentales que le maximum du signal (VERNAT-ROSSI, *et al.*, 1996).

Cette méthode est très intéressante car elle permet d'extraire un nombre réduit de paramètres qui caractérisent totalement la courbe d'adsorption.

Cependant, elle comporte des inconvénients : le choix du modèle n'est pas évident a priori puisque chacun des travaux cités ci-dessus a appliqué une équation de modélisation différente. (ELKÖV, *et al.*, 1997a) ont montré que les modèles polynomiaux étaient plus adaptés à la modélisation de courbes de MOSFET que les modèles exponentiels, alors que ces modèles ont été choisis par (PICQUE, *et al.*, 1997 & VERNAT-ROSSI, *et al.*, 1996) pour la modélisation de courbes de capteurs MOS.

En outre, la précision du point de départ de l'acquisition par rapport à l'injection est cruciale, introduisant du bruit dans le modèle, particulièrement en début d'adsorption (ELKÖV, *et al.*, 1997a & VERNAT-ROSSI, *et al.*, 1996).

**③ Une troisième approche consiste à prendre en considération la forme des courbes au travers de divers paramètres extraits.**

Cette démarche a été abordée par (NANTO, *et al.*, 1993), qui ont extrait 4 paramètres de la courbe de fréquence de vibration de QMB : la fréquence maximale et le temps correspondant, la fréquence après 14 minutes et l'aire sous la courbe après 20 minutes. De même, (LLOBET *et al.*, 1997) emploient la réponse stationnaire de 4 capteurs MOS, caractérisée par la conductance maximale, ainsi que la *constante de temps* des capteurs (« *rise time* ») ; ce dernier s'est révélé être plus répétable que la réponse stationnaire et indépendant de la concentration de la molécule analysée. L'extraction de ce paramètre fait suite aux travaux de (VILANOVA *et al.*, 1996) dans lesquels une constante  $\tau'$ ,

<sup>6</sup> Sigmoïde de Gompertz :  $y = a \cdot \exp\left(\frac{-b \cdot \exp(-c \cdot t)}{c}\right)$  ; (a) correspond à l'amplitude et (b) et (c) modélisent la phase transitoire du signal.

<sup>7</sup> Fonction de Weibull :  $y = a \cdot \left(1 - \exp\left(-b \cdot t^c\right)\right)$

<sup>8</sup> Modèle exponentiel simple :  $y = a \cdot (1 - \exp(-x/b))$

caractéristique de chaque gaz et indépendante de la concentration, avait été mise en évidence grâce aux équations de diffusion dans l'élément sensible MOS TGS 822.

Dans la même optique, (MIELLE, 1998) s'est intéressé à l'extraction de paramètres non influencés par la dérive des capteurs : il a identifié 2 caractéristiques qui sont la cinétique d'équilibre des arômes, variant en fonction de leur volatilité, et le temps de réponse des capteurs, indépendant de la dérive sauf en cas d'empoisonnement.

(ELKÖV, *et al.*, 1997a) ont étudié divers paramètres extraits de la courbe d'un MOSFET (niveaux, pente, aires, temps, etc.), d'un point de vue de leur rapport signal/bruit, leur corrélation et l'erreur de prédiction par réseaux de neurones. Les problèmes de répétabilité se sont avérés liés à la précision de l'injection et du début de l'acquisition. Cette méthode d'extraction d'information a été comparée à la modélisation par des modèles polynomiaux et exponentiels, et a fourni des performances similaires aux modèles polynomiaux, les modèles exponentiels étant moins satisfaisants.

(WILSON *et DEWEERTH*, 1997) ont eu une approche originale, pour accélérer la durée de mesure des arômes : ils ont conçu un système de micro-capteurs MOS, dont la température est contrôlable. Les réponses transitoires des capteurs sont seuillées toutes les secondes pendant 8 secondes après l'injection de l'espace de tête et ceci à 8 températures différentes. La sortie de chaque capteur est binaire : 1 si l'intensité du signal est supérieure à un seuil, 0 sinon. Cette technique de « *wavefront* » procure une carte de 64 points 0/1 par capteur, constituée des 8 durées de seuillage à 8 températures différentes : c'est une version simplifiée de l'échantillonnage de la courbe toutes les secondes, rendant compte du temps de réponse des capteurs (démarche similaire à celle de (LLOBET, *et al.*, 1997) & (MIELLE, 1998)). Cette carte est caractéristique de chaque molécule, permettant la discrimination de composés tels que le CO<sub>2</sub>, l'éthanol, l'isopropanol et l'ammoniaque après quelques secondes.

## Conclusion

Comme le montre la bibliographie, les travaux sur l'extraction de paramètres à partir de la courbe des capteurs sont très récents. La démarche que nous avons adoptée correspond à la troisième approche. Celle-ci, mise en œuvre avant même la publication des travaux de Elköv, est détaillée dans la partie III. Cependant, nous proposerons également une autre méthode, encore jamais appliquée aux capteurs d'arômes et permettant de traiter l'intégralité du signal sans modélisation ou échantillonnage : la PLS discriminante (C.f. Partie IV).

## II.4.2. Méthodes de classification

Les diverses méthodes de *reconnaissance des formes* ou de *classification* employées pour le traitement des données issues de capteurs d'arômes sont exposées. Elles ont été largement présentées, en particulier dans des synthèses bibliographiques sur les traitements linéaires (GARDNER, 1991), non linéaires (GARDNER *et al.*, 1992) et une compilation de toutes les méthodes (GARDNER *et HINES*, 1997).

La première partie de ce chapitre est consacrée aux classifications non supervisées, linéaires ou non. La seconde présente les traitements multivariés supervisés appliqués couramment aux variables extraites des signaux des capteurs d'arômes ; en outre, l'application à la classification d'une technique statistique multivariée généralement appliquée aux spectres est exposée : la PLS discriminante. Cette technique sera employée dans la partie IV pour le traitement des spectres IR-TF et UV ainsi que pour celui des courbes entières des capteurs d'arômes.

### II.4.2.1. Classifications non supervisées

Les classifications non supervisées ont pour objectif d'attribuer les échantillons à des classes, *sans connaissance a priori* de l'affectation des échantillons.

#### (1) Linéaires

Une des méthodes de classification non supervisée linéaire est la Classification Ascendante Hiérarchique (CAH) (*clustering*), méthode de coalescence hiérarchique. Nous ne détaillerons pas d'autre technique non supervisée linéaire, telle que la méthode des centres mobiles ou des nuées dynamiques, qui ne seront pas employées dans ces travaux. La CAH est généralement précédée d'une Analyse en Composantes Principales (ACP), *méthode descriptive* visant à réduire le nombre de variables.

L'objectif de l'ACP est de réduire la dimension de l'espace de représentation des échantillons, pour ne garder que les informations les plus importantes. Elle consiste à construire les axes orthogonaux, ou *composantes principales*, qui maximisent la variance (ou *l'inertie*) du nuage de points formé par les échantillons. Ainsi, les composantes principales, constituées par la combinaison linéaire de toutes les variables caractérisant les échantillons, permettent de projeter les échantillons (et les variables) dans le nouvel espace factoriel, pour visualiser les différences intrinsèques des échantillons. L'inertie associée à chaque composante principale permet de sélectionner les premiers axes, afin de réduire le nombre de variables définissant chaque échantillon. Cette méthode est couramment employée pour *l'analyse descriptive* des échantillons aromatiques ainsi que pour l'étude de la contribution des variables à la répartition des échantillons et de leurs corrélations.

La CAH peut être réalisée dans l'espace des variables ou des composantes principales. Celle-ci consiste à associer 2 à 2 les échantillons (ou les groupes d'échantillons formés) qui minimisent

l'augmentation de la variance intraclasse : les groupes (*clusters*) sont constitués par les échantillons les plus « proches » en fonction de la métrique définie (la distance euclidienne, par exemple). Un seuil de variance fixé arbitrairement permet de scinder les classes issus du *dendrogramme* résultant des associations successives (C.f. Annexe 5).

## (2) Non linéaires

Les réseaux de Kohonen ou SOM (*Self Organizing Map*) sont des réseaux de neurones non supervisés (KOHONEN, 1990). L'espace des échantillons est quadrillé par des neurones interconnectés, qui s'organisent progressivement en fonction de la proximité des variables associées à ses entrées. La carte de Kohonen résultante, semblable à une carte topographique, rend compte de la répartition des échantillons, mais n'est pas vraiment adaptée à la classification (KOHONEN, 1990).

Ce type d'*analyse descriptive non linéaire* a été employée par DI NATALE (1995, 1997, 1998), pour le contrôle de la fraîcheur du poisson grâce à 4 capteurs quartz. La contribution de chaque capteur peut être analysée localement, contrairement à l'ACP, ce qui traduit plus finement la participation des capteurs dans la construction de chaque classe.

### II.4.2.2. Classifications supervisées

Les classifications supervisées nécessitent 2 phases de construction ; la base de données est scindée en 2 lots d'échantillons : le premier est employé pour l'*apprentissage* du système de classification (ou *calibration*) grâce à la connaissance *a priori* de l'affectation des échantillons et le second lot sert de *prédiction* (ou *test*) afin de tester la performance du modèle (DUBUISSON, 1990).

Le choix de l'estimateur de l'erreur d'un modèle est toujours délicat (PALAGOS, 1997). En discrimination, généralement l'*erreur de prédiction* est calculée ; celle-ci est définie comme la probabilité d'affectation d'un nouvel échantillon à une classe incorrecte, et est calculée comme la moyenne du taux d'erreur sur le lot de prédiction. Cependant, si les échantillons sont peu nombreux (ce qui est notre cas), le taux d'erreur de prédiction peut être calculé grâce à une *validation croisée par blocs* ou « *leave-one-out* ».

La *validation croisée par blocs* consiste à scinder l'ensemble des échantillons en  $p$  blocs ;  $p-1$  blocs sont employés pour la calibration tandis que la prédiction est réalisée sur le bloc restant. Ce processus est réitéré  $p$  fois, afin de prédire tous les échantillons. Le taux d'erreur de prédiction est alors calculé comme la moyenne des erreurs sur les  $p$  blocs prédits. Cette technique peut être poussée à l'extrême, en scindant la base de données en  $n$  « blocs » contenant chacun 1 échantillon ; alors,  $n$  modèles sont créés à base de  $(n-1)$  échantillons et l'échantillon exclu est prédit ; le taux d'erreur de cette *validation croisée* « *leave-one-out* » est la moyenne des erreurs des  $n$  modèles :

$$\text{Taux d'erreur} = \frac{\sum_{i=1}^n \text{erreur}_{\text{modèle } i}}{n}, \text{ où } n \text{ est le nombre total d'échantillons ; } \text{erreur}_{\text{modèle } i} \in \{0; 1\}.$$

Les résultats de prédiction sont décrits sous la forme d'une *matrice de confusion*, exposant le nombre d'échantillons affectés à chaque classe ainsi que les taux d'erreur de 1<sup>er</sup> et 2<sup>nd</sup> ordre (POIRIER, 1989) (C.f. Tableau 3).

**Tableau 3 : exemple d'une matrice de confusion pour la discrimination des 4 cépages de moûts de raisins blancs (Sauvignon, Mauzac, Colombard et Autres cépages)**

$n_{ij}$  est le nombre d'échantillons attribués à la classe  $i$  alors qu'ils appartiennent à la classe  $j$ .  
 $n_{i+}$  est la somme marginale en ligne et  $n_{+j}$  en colonne.

	CLASSES REELLES				TOTAL
Nombre d'échantillons	$N_{sa}=n_{+1}$	$N_{Ma}=n_{+2}$	$N_{Co}=n_{+3}$	$N_{Au}=n_{+4}$	$N=n_{++}$
CLASSES PREDITES $\hat{p}$	Sauvignon	Mauzac	Colombard	Autres cépages	% erreur (2 <sup>nd</sup> ordre) $\hat{p}$
Sauvignon	$n_{11}$	$n_{12}$	$n_{13}$	$n_{14}$	$(n_{12}+n_{13}+n_{14})/n_{1+}$
Mauzac	$n_{21}$	$n_{22}$	$n_{23}$	$n_{24}$	$(n_{21}+n_{23}+n_{24})/n_{2+}$
Colombard	$n_{31}$	$n_{32}$	$n_{33}$	$n_{34}$	$(n_{31}+n_{32}+n_{34})/n_{3+}$
Autres cépages	$n_{41}$	$n_{42}$	$n_{43}$	$n_{44}$	$(n_{41}+n_{42}+n_{43})/n_{4+}$
% bons classements	$n_{11}/N_{sa}$	$n_{22}/N_{Ma}$	$n_{33}/N_{Co}$	$n_{44}/N_{Au}$	$BC=(n_{11}+n_{22}+n_{33}+n_{44})/N$
Taux d'erreur (1 <sup>er</sup> ordre) $\hat{e}$	$1-(n_{11}/N_{sa})$	$1-(n_{22}/N_{Ma})$	$1-(n_{33}/N_{Co})$	$1-(n_{44}/N_{Au})$	Taux d'erreur global : 1-BC

### (1) Classifications linéaires

L'Analyse Factorielle Discriminante (AFD) est une classification linéaire paramétrique supervisée. Elle consiste à trouver les axes de représentation des échantillons discriminant au mieux les  $C$  classes, i.e. qui maximisent la variance interclasse ( $B$  : *between-class*) et minimisent la variance intraclasse ( $W$  : *within-class*) (ROMEDER, 1973). Ceci revient à maximiser la matrice  $(T^{-1}.B)$ , sachant que  $B+W=T$  ( $T$  : *matrice de variance Totale*). Les échantillons tests sont alors projetés sur les  $(C-1)$  axes factoriels construits à partir des échantillons de calibration, et sont attribués à la classe dont le centre de gravité est le plus proche.

Ce type de modélisation supervisée linéaire est couramment employé pour la classification multivariée, en particulier dans le cas des capteurs d'arômes.

### (2) Classifications non linéaires

Nous n'exposerons ici qu'un seul type de système supervisé non linéaire : les *réseaux de neurones multicouches* (ou *perceptron multicouche*), systèmes les plus répandus, que nous appliquerons dans la partie IV. Ces réseaux neuro-mimétiques sont des *classifieurs non paramétriques adaptatifs*, car ils ne sous entendent aucune hypothèse de distribution statistique des classes et les paramètres du modèle sont ajustés itérativement au cours de l'apprentissage.

Les *neurones* sont les constituants élémentaires du réseau, interconnectés par couche. Le perceptron multicouche est formé généralement de 3 couches (C.f. Figure 6) : la couche d'entrée composée de  $n_1$  neurones, correspondant aux  $n_1$  variables caractérisant les échantillons, une seconde couche, dite

couche cachée, formée d'un nombre variable de nœuds  $n_2$ , et la couche de sortie dont le nombre de nœuds  $n_3$ , est égal au nombre de sorties à prédire ; dans le cas d'une classification, le nombre de sorties à prédire correspond au nombre de classes, codées en disjonctif complet.

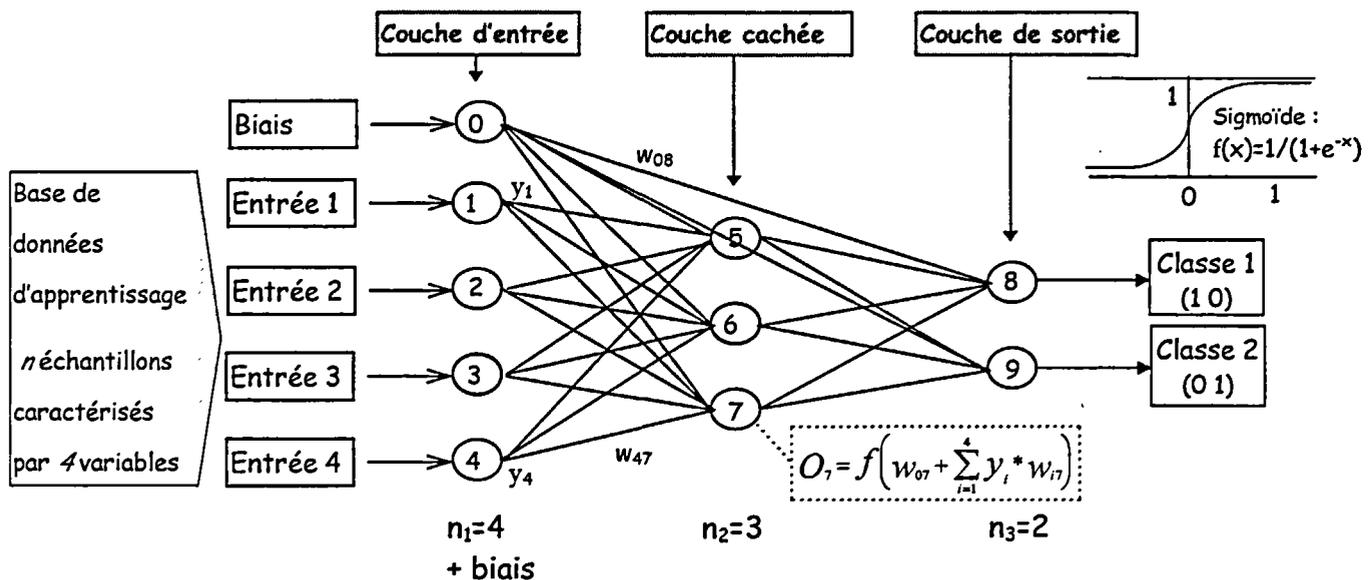


Figure 6 : schéma d'un réseau de neurones multicouche

La sortie estimée  $O_k$  d'un neurone  $k$  est calculée grâce à la combinaison linéaire pondérée (poids :  $w_{ik}$ ) des sorties  $y_i$  des  $n$  neurones de la couche précédente et de l'ajout d'un biais ( $w_{0k}$ ). Cette sortie est alors transformée par le passage dans une fonction d'activation  $f$  à seuil binaire (généralement sigmoïde) : (COMON, 1991).

L'apprentissage le plus courant est réalisé par *rétro-propagation du gradient de l'erreur* calculée entre la sortie vraie (0 ou 1 en classification) et la sortie estimée (RUMELHART *et al.*, 1986). Ce processus est répété jusqu'à convergence de l'erreur. Il est préférable de normaliser les données ou de réduire leur gamme de variation à l'intervalle  $[-1 ; 1]$ , car la fonction d'activation sature rapidement à l'extérieur de cet intervalle (C.f. Figure 6) (KVAAL *et McEWAN*, 1996).

Le nombre d'itérations (*epochs*) peut être délibérément restreint, afin d'éviter le *sur-apprentissage*, par la méthode du « *stopped-learning* » (GALLINARI *et GASCUEL*, 1996) ; celle-ci détermine le nombre d'itérations optimal grâce à une validation croisée, arrêtant l'apprentissage dès que *l'erreur de prédiction* s'accroît. Ainsi, le pouvoir de généralisation du réseau de neurones est accru, en évitant l'apprentissage par cœur.

(GRANDVALET, 1995) a montré que l'introduction de bruit dans les variables d'entrée peut favoriser la généralisation du réseau de neurones lors de la prédiction, en particulier quand le nombre d'échantillons est faible. Un bruit aléatoire peut être introduit sous forme d'une distribution Gaussienne de moyenne nulle et d'un écart-type proportionnel à celui de chaque entrée  $\sigma_{\text{bruit}} : p * \sigma_{\text{entrée}}$ .

### II.4.3. Méthodes de sélection des capteurs

L'objectif de la sélection des capteurs est d'introduire dans les systèmes multi-capteurs d'arômes les éléments sensibles les plus discriminants pour chaque application.

Les travaux concernant le choix des transducteurs sont présentés dans un premier paragraphe. Face à leur insuffisance, des méthodes de sélection de variables sont exposées, afin de choisir indirectement les capteurs. Les techniques employées ultérieurement seront particulièrement développées : il s'agit des méthodes de sélection de variables pas à pas et des algorithmes génétiques.

#### II.4.3.1. Sélection des capteurs d'arômes

Même si certains capteurs d'arômes permettent d'inclure jusqu'à 24 capteurs MOS (ALPHAMOS, Toulouse, F) ou 32 capteurs polymères (AROMASCAN, Warwick, GB), il est impossible de rassembler tous les capteurs commercialisés dans un seul et même appareillage.

En outre, quelle que soit la technologie employée, les éléments sensibles composant les capteurs d'arômes répondent à une large gamme de molécules volatiles ; leurs sensibilités se recouvrent en grande partie, chacun présentant une légère spécificité ( $r^2$  généralement de l'ordre de 0.8-0.9 entre les maxima des courbes de capteurs MOS !).

Il est donc extrêmement important de choisir les capteurs les plus appropriés pour chaque application. Or, les caractéristiques de sensibilité et de sélectivité des capteurs, fournies par les constructeurs, sont peu précises et ne correspondent pas aux arômes des produits agro-alimentaires. Le Tableau 4 présente diverses sélectivités de capteurs MOS FIGARO TGS de la série 8xx (*FIGARO ENGINEERING*, 1990), fournies par les constructeurs : il est très difficile de se baser sur ces indications pour choisir les capteurs les plus adaptés à chaque discrimination.

Tableau 4 : exemples de caractéristiques de capteurs MOS (TGS Figaro de la série 8xx)

TYPE DE CAPTEUR	CARACTERISTIQUES
TGS 800	CO, H <sub>2</sub> , fumée de cigarettes, (1 ~ 10ppm)
TGS 812	gaz combustibles et toxiques
TGS 813	gaz naturels (CH <sub>4</sub> , C <sub>2</sub> H <sub>6</sub> , C <sub>3</sub> H <sub>8</sub> )
TGS 815	CH <sub>4</sub> , combustibles
TGS 821	H <sub>2</sub>
TGS 822, 823	alcools, et solvants organiques (50 ~5000ppm)
TGS 824	NH <sub>3</sub>
TGS 825	H <sub>2</sub> S (5 ~ 100ppm) et gaz toxiques
TGS 830	CFC, HCFC (100 ~ 3000ppm)
TGS 831	Fréon, CHCl <sub>2</sub> F, CHClF <sub>2</sub>
TGS 842	gaz domestiques, CH <sub>4</sub>
TGS 880, 881	vapeur d'eau, contrôle de cuisson
TGS 882	alcool (50 ~ 5000ppm), contrôle de cuisson
TGS 883T, 884	vapeurs d'eau et gaz de cuisson

De rares travaux ont été menés sur la sélection des éléments sensibles. (WALMSLEY *et al.*, 1991) ont analysé la corrélation de 3 capteurs TGS (813, 820 et 816), en employant une classification hiérarchique ascendante ; ils ont aussi établi une corrélation entre les groupes fonctionnels de composés volatils (alcool, alcane, alcène...) et le type de capteur répondant à ces gaz.

Une analyse plus fine des corrélations entre capteurs peut être réalisée grâce aux SOM<sup>9</sup>, en calculant les contributions de chaque élément sensible dans tout l'espace couvert par le réseau de Kohonen. Ainsi, (DINATALE *et al.*, 1997a) ont pu déterminer quels sont les capteurs QCM responsables de la discrimination de la qualité du poisson, pour chacune des classes de fraîcheur.

Si l'appareillage est capable de contenir de nombreux capteurs, les meilleurs peuvent être sélectionnés a posteriori. C'est le cas de (ROCHA, *et al.*, 1998) qui ont sélectionné 17 capteurs polymères parmi les 32 capteurs contenus dans un appareillage AROMASCAN. Ce choix a été effectué grâce à la méthode de sélection de variables procustéenne (KRZANOWSKI, 1987), en conservant 94% de la variance des échantillons.

La sélection des capteurs peut aussi être conduite individuellement sur chaque molécule volatile composant un arôme, en combinaison avec des méthodes séparatives : il suffit de remplacer le détecteur FID<sup>10</sup> d'une chromatographie à phase gazeuse par la cellule de mesure contenant les capteurs d'arômes. Ainsi, (HOFMANN *et al.*, 1997) ont pu identifier les MOS et MOSFET sensibles au diacétyle, ou à l'acide butanoïque contenu dans du beurre rance. De la même manière, (HIVERT *et al.*, 1995) sélectionnent des capteurs MOS et emploient cette méthode comme technique de calibration des capteurs, en comparant leur signaux aux chromatogrammes.

### II.4.3.2. Méthodes de sélection de variables

Paradoxalement, face à la pauvreté des travaux sur la sélection des capteurs d'arômes, les méthodes de sélection de variables ont été largement explorées. Cette démarche résulte de l'application du principe de parcimonie qui stipule que : « *si deux systèmes modélisent correctement un ensemble de données, alors celui qui est bâti sur le plus petit nombre de variables sera le plus apte à prédire de nouvelles données* » (SEASHOLTZ *et KOWALSKI*, 1993).

Pour sélectionner un sous-ensemble de variables (ou de capteurs au travers des variables qui les caractérisent), il est généralement impossible d'étudier de manière exhaustive toutes leurs combinaisons<sup>11</sup>.

Les méthodes de type « *Branch & Bound* » ne sont pas exhaustives, mais garantissent l'optimalité de la sélection des variables sous l'hypothèse de monotonie de l'erreur (NARENDRA *et*

<sup>9</sup> « SOM » : « Self Organizing Map » ou réseau de Kohonen (C.f. pII-36)

<sup>10</sup> FID : Détecteur à Ionisation de Flamme

<sup>11</sup> Le nombre de combinaisons à étudier pour retenir p variables parmi n est :  $C_n^p = \frac{n!}{(n-p)! \cdot p!}$

FUGUNAGA, 1977) ; l'évaluation implicite de tous les sous-ensembles grâce à la possibilité d'un retour en arrière évite l'énumération complète des variables, mais si le nombre initial de variables est important (supérieur à une dizaine), la charge de calcul devient considérable (SIEDLECKI et SKLANSKY, 1988).

Alors, diverses méthodes sous-optimales de sélection de variables sont employées, basées sur des méthodes statistiques pas à pas (« *Stepwise Multivariate Analysis* » ou « *Sequential Forward and Backward Selection* ») (DUBUISSON, 1990). Elles seront décrites dans le paragraphe suivant et employées lors de la sélection des éléments sensibles à inclure dans les capteurs d'arômes (C.f. Chapitre III).

Un autre type de méthode, le parcours stochastique de l'espace des solutions, est offert par le *recuit simulé* qui consiste à simuler les processus de cuisson en métallurgie ou les *algorithmes génétiques* qui miment le processus de sélection naturelle biologique où seuls les individus les plus adaptés survivent. Ces algorithmes peuvent être employés aussi bien pour des problématiques d'optimisation (HÖRCHNER et KALIVAS, 1995) ou d'apprentissage (VENTURI, 1996) que de sélection de variables (CHTIOUI et al., 1998 & YANG et HONAVAR, 1998).

La méthode des algorithmes génétiques a déjà été citée pour la sélection de longueurs d'onde dans des spectres infrarouges (FEARN et DAVIES, 1997) ou pour réduire l'information d'une image spectroscopique (VAN DEN BROEK et al., 1997), où la sélection de longueurs d'onde améliore largement les prédictions par régression multiple et perfectionne légèrement celles par PLS (LEARDI et al., 1994).

Toutes ces méthodes consistent à comparer les performances des différents sous-ensembles de variables sélectionnées, grâce à un *critère de sélection* ou *d'adéquation*. Celui-ci est bâti à partir de diverses techniques de classification, telles que l'Analyse Factorielle Discriminante (ROMEDER, 1973) ou la méthode des *k-Plus Proches Voisins* (DEVROYE et al., 1996).

Classiquement, ce critère est le *pourcentage d'échantillons bien classés*, calculé à partir de l'affectation des échantillons aux classes grâce à diverses métriques ; le processus de sélection s'arrête quand le pourcentage de bons classements ne croît plus. Un autre critère couramment employé est la *trace de la matrice* ( $T^1.B$ ) (variance interclasse ( $B$ ) divisée par la variance totale ( $T$ ) ; C.f. définition de l'AFD, pII-37) qui croît continuellement à l'ajout d'une variable, même si la discrimination ne s'améliore pas : il n'y a donc pas de critère d'arrêt naturel.

De façon similaire, le *Lambda de Wilks* ( $\Lambda = |W|/|T|$  : déterminant de la variance intraclasse ( $W$ ) divisée par celui de la variance totale ( $T$ )) peut être calculé et minimisé.

Sous les hypothèses de multinormalité et d'équicovariance des classes, des tests peuvent être effectués sur les moyennes conditionnelles des groupes, tels que le *test de Fisher*.

Ces 2 derniers *critères de sélection* seront développés dans la Partie III, pour la sélection des éléments sensibles au sein des capteurs d'arômes grâce aux méthodes pas à pas.

Les méthodes de sélection détaillées ci-dessous seront employées lors de la sélection de capteurs d'arômes (C.f. Partie III) ainsi que pour la sélection de longueurs d'onde des spectres IR-TF et UV (C.f. Partie IV).

### (1) Méthodes Pas à Pas

Les méthodes de sélection de variables pas à pas consistent à choisir itérativement chaque variable optimisant un *critère de sélection* choisi, afin de les introduire dans le système de classification. Elles peuvent être ascendantes (« *Sequential Forward Selection* »), i.e. incluant les variables une à une, ou descendantes (« *Sequential Backward Selection* »), i.e. enlevant les variables une à une à partir de la totalité. A chaque itération, les 2 processus d'inclusion/exclusion peuvent être exécutés par groupe de variables et non une par une (« *Generalised Sequential Selection* ») (DUBUISSON, 1990).

### (2) Algorithmes génétiques

Le principe des algorithmes génétiques est de trouver les individus qui maximisent un critère donné, à partir d'une population initiale générée au hasard. Chaque individu, appelé « *chromosome* », est constitué de  $n$  « *gènes* » qui codent la prise en compte des variables à sélectionner. Si le gène vaut 1, alors la variable est choisie, s'il vaut 0, alors la variable ne participe pas au calcul du *critère d'adéquation*. (CHTIOUI, 1997) a montré que le taux de gènes actifs évolue peu au cours des générations : si l'on veut sélectionner  $p$  variables parmi  $n$ , chaque chromosome initial doit contenir  $p$  gènes à 1.

La population initiale, contenant  $m$  individus, évolue naturellement grâce à la *reproduction*, aux *croisements* et aux *mutations*. A chaque génération, la population se reproduit. Elle subit un croisement, qui consiste à croiser 2 chromosomes, en 1 ou deux points, afin de créer 2 descendants possédant une partie de chaque parent (C.f. Figure 7) ; celui-ci est effectué à chaque reproduction sur tout ou partie de la population. La mutation, dont la fréquence est réglable, consiste à inverser un gène au hasard (C.f. Figure 7), faisant apparaître ou disparaître une variable du chromosome.

La nouvelle population est alors évaluée afin de conserver les meilleurs individus, c'est-à-dire ceux ayant le meilleur *critère d'adéquation*. Au fur et à mesure, la population converge vers les individus présentant la meilleure adéquation, générant de nombreux chromosomes semblables. La multiplication cesse en respectant un *critère d'arrêt*, qui peut être un nombre maximum de générations ou de chromosomes identiques.

**1. Sélection des parents (5 variables présentes sur 14) :**

Parent 1 : 10010101001000

Parent 2 : 00100110001010

**2. Cross-over :**

Parent 1 : 10010 101001000

Parent 2 : 00100 110001010

2 enfants : 10010110001010

00100101001000

**3. Mutation :**

Enfant 1 avant : 10010110001010

Enfant 1 après : 10000110001010

**Figure 7 : mécanismes de croisement et de mutation**

Cette méthode permet la sélection de très nombreuses variables, mais présente certains inconvénients : elle nécessite le réglage de nombreux paramètres afin de contrôler l'évolution de la population (*SEBAG et SCHOENAUER*, 1996), tels que le nombre initial de chromosomes, le taux initial de sélection des variables, le critère d'arrêt de la multiplication de la population, la probabilité de croisement et de mutation et le critère d'adéquation.

Ce réglage est un compromis entre la durée d'exécution des algorithmes génétiques et l'exploration du plus large espace des solutions ainsi que la précision du calcul du critère d'adéquation.

De plus, elle est non déterministe, donc nécessite plusieurs exécutions afin d'obtenir des résultats exploitables. En outre, le résultat final est une population optimale comprenant un certain nombre de chromosomes ordonnés en fonction du critère d'adéquation croissant. Il est alors nécessaire de choisir parmi toutes ces solutions possibles ; nous proposerons et analyserons donc différentes stratégies de choix ultime des variables, afin d'établir une démarche de sélection pertinente (C.f. partie IV).

## II.4.4. Méthodes de fusion multi-capteur

Une définition des plus générales de la fusion multi-capteur est énoncée par (AYARI et HATON, 1995) :

« la fusion multi-capteur est le procédé permettant d'intégrer l'information de sources multiples afin de produire les données les plus spécifiques et les plus pertinentes sur une entité, une activité ou un événement ».

La fusion multi-capteur est principalement utilisée pour la prédiction de position d'objets, en particulier dans le domaine militaire (LUO et KAY, 1989) ; les techniques de fusion employées ne peuvent pas s'appliquer à notre problématique.

Pour (FILDES, 1992), les 4 principaux avantages de la fusion sont :

- ⊗ La *redondance* : plusieurs mesures de la même caractéristique sont réalisées par différents capteurs. La fiabilité de la classification est donc améliorée ;
- ⊗ La *complémentarité* : différents types de mesure sont effectués sur un même échantillon ;
- ⊗ Le *gain de temps* : le travail en parallèle augmente la rapidité de l'évaluation ;
- ⊗ Le *moindre coût* : le système ne nécessite pas des capteurs extrêmement performants, du fait de la complémentarité et de la redondance.

L'application de la fusion multi-capteur à la caractérisation de produits agro-alimentaires peut être intitulée *fusion multisensorielle* (STEINMETZ, 1997). Celle-ci est relativement rare :

(EDAN et al., 1994) et (OZER et al., 1995) ont appliqué la fusion par réseaux de neurones, respectivement à la classification de tomates et melons, grâce aux mesures organoleptiques (sucré, acide, arômes), d'apparence (couleur, taille, forme et poids) et de texture (fermeté, sensations en bouche). (STEINMETZ et al., 1995) ont effectué des mesures de fermeté, de sucre, de couleur et d'arômes pour l'estimation de la qualité des melons, fusionnées par réseaux de neurones.

Ce dernier (1997) a organisé la fusion multi-capteur en 3 niveaux, en fonction du type de données fusionnées. Elles sont détaillées ci-après.

### II.4.4.1. Niveaux de fusion

#### (1) Niveau bas

La *fusion de niveau bas* consiste à concaténer les signaux bruts de chaque capteur afin de les traiter comme un seul et même signal (STEINMETZ, 1997). Ce type de fusion n'est possible qu'entre capteurs présentant des signaux commensurables, tels que les spectres issus de divers spectromètres ; il est impossible de combiner des valeurs uniques, comme un taux de sucre ou une valeur de pH.

Les techniques de traitement multivarié sont alors appliquées à ce signal commun, telles qu'une PCR (Régression sur les Composantes Principales) ou PLS (*Partial Least Squares*), car elle sont particulièrement adaptées à l'analyse de variables très corrélées. Nous adapterons cette méthode à la classification, dans la Partie IV, basée sur la concaténation des signaux issus des capteurs d'arômes et des spectromètres IR-TF et UV.

#### (2) Niveau intermédiaire

Ce type de fusion consiste à extraire séparément du signal de chaque capteur des variables et de les introduire dans un système de classification commun afin d'obtenir une classification conjointe. Cette fusion, généralement effectuée par réseaux de neurones, a été étudiée au cours de différents travaux (STEINMETZ, 1997 & STEINMETZ et BELLON, 1995). Cependant, si les capteurs employés produisent un grand nombre de variables, alors l'application de cette méthode exige de mesurer un très grand nombre d'échantillons.

#### (3) Niveau haut

La *fusion de niveau haut* fournit une classification commune basée sur les résultats de discrimination individuelle des divers capteurs à fusionner.

Parmi toutes les techniques possibles (vote, etc.), la méthode bayésienne a été largement développée, car elle est optimale<sup>12</sup> pour la gestion de l'incertain dans des données (CARAUX et LECHEVALLIER, 1996). Cette démarche, qui sera adoptée dans la suite des travaux sur la fusion de niveau haut (C.f. Chapitre IV), est détaillée ci-dessous.

La mise en œuvre de cette technique de fusion nécessite la définition de *concepts probabilistes* (pII-46). La fusion peut être effectuée en employant la *règle d'affectation d'erreur minimale de Bayes* (pII-46) ; celle-ci a été appliquée à la fusion d'images à rayon X pour la détection de fissures (DROMIGNY et ZHU, 1997), ou au diagnostic d'alarmes (CHATILA, 1996). Nous proposerons et appliquerons une extension de cette méthode, en employant la *règle d'affectation de risque minimum de Bayes* (pII-50).

---

<sup>12</sup> C.f. pII-47 : Règle de décision d'erreur minimale de Bayes

## II.4.4.2. Fusion de niveau haut par théorie de Bayes

### (1) Notations et concepts probabilistes

#### (a) Notations

Tout objet à classer  $x$  appartient à une population théorique  $\mathcal{X}$ . Cette population est munie d'une partition  $(\Pi_1, \dots, \Pi_p, \dots, \Pi_g)$  en  $g$  classes ( $G = \{1, \dots, g\}$ ) définies a priori grâce à la fonction d'affectation  $Y$ . L'objet  $x$  est défini par  $p$  descripteurs élémentaires :  $x' = (x_1, \dots, x_p, \dots, x_p)$  de  $D_x$ , l'espace de description. Le classement d'un objet  $x$  est noté  $y$  et sa classe d'appartenance est  $\Pi_y$ .

$\hat{Y}$  est la règle d'affectation ou de décision, induisant sur  $D_x$  une partition en classes d'équivalence appelées régions d'affectations  $(R_1, \dots, R_p, \dots, R_g)$ . Dans le cas idéal :

$$Y = \hat{Y}, \text{ i.e. } (\Pi_1, \dots, \Pi_g) = (\hat{\Pi}_1, \dots, \hat{\Pi}_g)$$

#### (b) Concepts probabilistes

Soit  $p_k$ , la probabilité a priori de la classe  $k$  :

$$p_k = \Pr(\Pi_k) = \Pr(Y = k), \text{ avec } \sum_{k=1}^g p_k = 1 \quad ^{13}$$

Soit  $L(x)$  la *vraisemblance* de  $X$  :  $L(x) = \Pr(X=x)$  ; un objet  $x$  est d'autant plus vraisemblable qu'il a de fortes chances d'apparaître.

Soit  $L_k(x)$  la *vraisemblance conditionnelle* de  $x$  pour la classe  $k$  :

$$L_k(x) = \Pr(X = x | \Pi_k) = \Pr(x|k)$$

Le théorème de Bayes sur les probabilités conditionnelles, exprime la probabilité *a posteriori* de la classe  $k$  sachant  $x$  ( $\Pr(k|x)$ ), en fonction des probabilités *a priori* de  $x$  sachant  $k$  ( $\Pr(x|k)$ ) et de la classe  $k$  ( $p_k$ ) :

$$\Pr(k|x) = \frac{\Pr(x|k) * \Pr(\Pi_k)}{\Pr(x)} = \frac{L_k(x) * p_k}{L(x)} = \frac{L_k(x) * p_k}{\sum_{k=1}^g L_k(x) * p_k}$$

**Equation 2 : théorème de Bayes sur les probabilités conditionnelles**

### (2) Fusion par la règle d'erreur minimale de Bayes

Dans un premier temps, la règle d'affectation d'erreur minimale de Bayes sera exposée (CARAUX et LECHEVALLIER, 1996) ; puis, elle sera généralisée à la fusion multi-capteur (CHATILA, 1996) (DROMIGNY et ZHU, 1997).

<sup>13</sup> Dans tout le manuscrit, la probabilité est notée Pr.

## (a) Règle de décision d'erreur minimale de Bayes

La règle d'affectation d'erreur minimale  $\hat{Y}$  attribue un objet  $x$  à la classe la plus probable, compte tenu de sa description  $\Pr(k|x)$ , probabilité a posteriori de  $k$  sachant  $x$  :

$$\forall x \hat{Y}(x) = \arg \max_{k \in G} \{\Pr(k|x)\}^{14}$$

**Equation 3 : règle d'affectation d'erreur minimale de Bayes en fonction des probabilités a posteriori**

En utilisant le théorème de Bayes (Equation 2), cette règle peut s'exprimer grâce aux probabilités conditionnelles *a priori*, afin de pouvoir la mettre en pratique :

$$\forall x \hat{Y}(x) = \arg \max_{k \in G} \left\{ \frac{L_k(x) * P_k}{L(x)} \right\} = \arg \max_{k \in G} \{L_k(x) * P_k\}, \quad \text{car } \forall k \in G \ L(x) \text{ est constante}$$

**Equation 4 : règle d'affectation d'erreur minimale de Bayes en fonction des probabilités a priori**

Cette règle d'affectation est la probabilité d'erreur minimum en tout point du domaine de description  $D_x$ . L'intégration sur  $D_x$  de cette équation montre que la performance globale de cette règle est optimale, i.e. aucune règle d'affectation ne donne un taux d'erreur inférieur.

## (b) Fusion multi-capteur par l'erreur minimale de Bayes

La règle de décision d'erreur minimale de Bayes peut être généralisée à  $N$  capteurs :  $i = \{1, \dots, N\}$  (CHATILA, 1996) (DROMIGNY et ZHU, 1997). Un objet  $X$  est caractérisé par  $N$  capteurs :  $X^i = \{X^1, \dots, X^N\}$ .

Soit  $L_k^i(x)$  la vraisemblance conditionnelle, pour le capteur  $i$ , que la loi de probabilité de  $X$  prenne la valeur  $x$  dans la classe  $I_k$  :

$$L_k^i(x) = \Pr(X^i = x | \prod_k) = \Pr(X^i | k)$$

**Equation 5 : vraisemblance conditionnelle de  $X$  pour le capteur  $i$**

<sup>14</sup> La notation  $\arg \max_{k \in G} (f(k))$  correspond à la valeur de  $k$  qui rend  $f(k)$  maximale.

Alors, la règle d'affectation d'erreur minimale de Bayes d'un objet  $X$  (C.f. Equation 3) s'écrit :

$$\forall X \hat{Y}(x) = \arg \max_{k \in G} \{ \Pr(k | (X^1, \dots, X^N)) \}$$

**Equation 6 : règle d'affectation d'erreur minimale de Bayes pour un système multi-capteur en fonction des probabilités a posteriori**

En appliquant le théorème de Bayes (Equation 2),  $\hat{Y}$  s'exprime :

$$\forall X \hat{Y}(x) = \arg \max_{k \in G} \left\{ \frac{\Pr((X^1, \dots, X^N) | k) * p_k}{\Pr(X^1, \dots, X^N)} \right\} = \arg \max_{k \in G} \{ \Pr((X^1, \dots, X^N) | k) * p_k \}$$

car  $\forall k \in G \Pr(X^1, \dots, X^N) = Cste$

**Equation 7 : règle d'affectation d'erreur minimale de Bayes pour un système multi-capteur en fonction des probabilités a priori**

La probabilité conjointe d'événements peut s'exprimer grâce à divers opérateurs (ET, OU, MAX, MIN, etc.). L'opérateur d'agrégation le plus strict est le ET, pour lequel la probabilité conjointe s'exprime par l'intersection des probabilités individuelles. Ce type d'agrégation, employé par (CHATILA, 1996) et (DROMIGNY et ZHU, 1997), est adopté dans la suite de notre étude.

$$\forall X \hat{Y}(x) = \arg \max_{k \in G} \{ \Pr((X^1, \dots, X^N) | k) * p_k \} = \arg \max_{k \in G} \left\{ \Pr \left( \left( \bigcap_{i=1}^N X^i \right) | k \right) * p_k \right\}$$

**Equation 8 : règle d'affectation d'erreur minimale de Bayes pour un système multi-capteur par l'intersection des probabilités a priori**

La formule des probabilités composées exprime la probabilité conjointe grâce aux probabilités conditionnelles des événements (C.f. Equation 9).

$$\Pr \left( \bigcap_{i=1}^N X^i \right) = \Pr(X^1) * \Pr(X^2 | X^1) * \Pr(X^3 | X^1 \cap X^2) * \dots * \Pr(X^N | X^1 \cap \dots X^{N-1})$$

**Equation 9 : formule des probabilités composées pour des événements non indépendants**

En admettant l'hypothèse d'indépendance entre les capteurs, alors cette probabilité conjointe s'écrit comme le produit des probabilités individuelles (C.f. Equation 10).

$$\Pr \left( \bigcap_{i=1}^N X^i \right) = \prod_{i=1}^N \Pr(X^i)$$

**Equation 10 : probabilité conjointe d'événements indépendants d'un même espace de probabilité**

(DROMIGNY et ZHU, 1997) affirment que cette hypothèse exprime simplement l'absence d'effet de la mesure d'un capteur sur celle des autres. Cependant, si l'on considère que 2 capteurs sont indépendants si leurs mesures ne sont pas corrélées, alors cette hypothèse est très lourde. Afin de simplifier le modèle, nous admettrons dans un premier temps cette hypothèse d'indépendance, qui permet d'appliquer la règle d'affectation d'erreur minimale de Bayes. La validité de cette hypothèse sera ensuite analysée grâce à l'étude des corrélations des mesures et le modèle pourra éventuellement être rectifié (C.f. Partie IV).

Alors, en appliquant l'Equation 10 qui suppose l'indépendance des capteurs, l'Equation 8 s'écrit :

$$\forall X \hat{Y}(x) = \arg \max_{k \in G} \left\{ p_k * \prod_{i=1}^N \Pr(X^i | k) \right\} = \arg \max_{k \in G} \left\{ p_k * \prod_{i=1}^N L_k^i(x) \right\}$$

Equation 11 : règle d'affectation d'erreur minimale de Bayes pour un système multi-capteur par le produit des vraisemblances a priori

Toutes les vraisemblances associées au capteur  $i$  sont compilées dans la matrice de causalité  $M^i$  :

$$M^i = \{ \Pr(X^i | k) = L_k^i(x) \}$$

Les probabilités a priori ne sont pas bâties à partir de loi de distribution probabiliste, comme c'est le cas lors de classifications Bayésiennes. Elle sont estimées à partir des classifications individuelles des capteurs, regroupées dans la matrice de confusion obtenue après prédiction de tous les échantillons (C.f. Tableau 3, pII-37). La matrice de causalité, contenant toutes les probabilités a priori, est bâtie directement à l'aide de la matrice de confusion : ce calcul est détaillé dans le Tableau 5.

Dans le cas de la discrimination de cépages de raisins blancs, les  $g=4$  classes sont les cépages « Sauvignon », « Mauzac », « Colombard » et tous les « autres cépages » blancs, notées « Sa », « Ma », « Co » et « Au », et numérotées de 1 à 4, donc  $X^i = [\hat{S}a', \hat{M}a', \hat{C}o', \hat{A}u']$ .

Tableau 5 : matrice de causalité  $M^i$  pour le capteur  $i$

$n_{ij}$  est un élément de la matrice de confusion pour un capteur donné.

CLASSES PREDITES 	CLASSES REELLES (N ECHANTILLONS AU TOTAL)			
	Sauvignon (Sa)	Mauzac (Ma)	Colombard (Co)	Autres cépages (Au)
Nombre d'échantillons	$N_{Sa}=n_{+1}$	$N_{Ma}=n_{+2}$	$N_{Co}=n_{+3}$	$N_{Au}=n_{+4}$
$p_k$	$p_{Sa}=n_{+1}/N$	$p_{Ma}=n_{+2}/N$	$p_{Co}=n_{+3}/N$	$p_{Au}=n_{+4}/N$
$\hat{S}a$	$\Pr(\hat{S}a'   Sa) = n_{11}/N_{Sa}$	$\Pr(\hat{S}a'   Ma) = n_{12}/N_{Ma}$	$\Pr(\hat{S}a'   Co) = n_{13}/N_{Co}$	$\Pr(\hat{S}a'   Au) = n_{14}/N_{Au}$
$\hat{M}a$	$\Pr(\hat{M}a'   Sa) = n_{21}/N_{Sa}$	$\Pr(\hat{M}a'   Ma) = n_{22}/N_{Ma}$	$\Pr(\hat{M}a'   Co) = n_{23}/N_{Co}$	$\Pr(\hat{M}a'   Au) = n_{24}/N_{Au}$
$\hat{C}o$	$\Pr(\hat{C}o'   Sa) = n_{31}/N_{Sa}$	$\Pr(\hat{C}o'   Ma) = n_{32}/N_{Ma}$	$\Pr(\hat{C}o'   Co) = n_{33}/N_{Co}$	$\Pr(\hat{C}o'   Au) = n_{34}/N_{Au}$
$\hat{A}u$	$\Pr(\hat{A}u'   Sa) = n_{41}/N_{Sa}$	$\Pr(\hat{A}u'   Ma) = n_{42}/N_{Ma}$	$\Pr(\hat{A}u'   Co) = n_{43}/N_{Co}$	$\Pr(\hat{A}u'   Au) = n_{44}/N_{Au}$

Un exemple d'application de l'Equation 11 à la discrimination des cépages est détaillé ci-après. Dans le cas de deux capteurs, si le capteur 1 affecte l'objet  $X$  à la classe Sauvignon (Sa) et le capteur 2 à la classe « Autres cépages » (Au), alors la règle d'affectation est la suivante :

$$\hat{Y}(x) = \arg \max \left\{ \begin{array}{l} P_{Sa} * (\Pr(\hat{S}a^1|Sa) * \Pr(\hat{A}u^2|Sa)) \\ P_{Ma} * (\Pr(\hat{S}a^1|Ma) * \Pr(\hat{A}u^2|Ma)) \\ P_{Co} * (\Pr(\hat{S}a^1|Co) * \Pr(\hat{A}u^2|Co)) \\ P_{Au} * (\Pr(\hat{S}a^1|Au) * \Pr(\hat{A}u^2|Au)) \end{array} \right\}$$

Equation 12 : exemple d'application de la règle d'affectation d'erreur minimale de Bayes à la discrimination de 4 classes de cépage.

La règle d'affectation d'erreur minimale de Bayes peut être améliorée en introduisant une expertise dans le système de discrimination sous la forme d'un coût de classification.

### (3) Fusion par la règle de risque minimum de Bayes

Dans un premier temps, la règle d'affectation de risque minimum de Bayes est exposée (CARAUX et LECHEVALLIER, 1996). Dans un second temps, nous proposerons sa généralisation à la fusion multi-capteur, de façon similaire à la fusion par la règle d'affectation d'erreur minimale de Bayes.

#### (a) Règle de décision de risque minimum de Bayes

La fonction de coût  $C$  associée à tout couple  $(k, h)$  le coût  $C(k, h)$  de classement d'un objet de  $\Pi_k$  comme un objet de  $\Pi_h$ <sup>15</sup> :

$$C : G \times G \rightarrow \mathfrak{R}^+$$

$$(k, h) \mapsto \begin{cases} C(k|k) = 0, \forall k \in G \\ C(h|k) \geq 0, \forall h \neq k \in G \end{cases}$$

Le coût moyen  $\bar{C}(k|x)$  de l'affectation d'un objet  $x$  à la classe  $k$  est défini comme la somme sur toutes les classes du produit du coût d'affectation et de la probabilité a posteriori de la classe (Equation 13):

$$\bar{C}(k|x) = \sum_{h=1}^g C(k|h) * \Pr(h|x)$$

Equation 13 : coût moyen

La règle d'affectation du risque minimum de Bayes revient à attribuer l'échantillon à la classe dont le coût moyen est minimal. En appliquant le théorème de Bayes, la règle d'affectation s'exprime alors grâce aux probabilités a priori (Equation 14) :

$$\forall x \hat{Y}(x) = \arg \min_{k \in G} (\bar{C}(k|x)) = \arg \min_{k \in G} \left( \sum_{h=1}^g C(k|h) * \Pr(h|x) \right) = \arg \min_{k \in G} \left( \sum_{h=1}^g C(k|h) * p_h * \Pr(x|h) \right)$$

Equation 14 : règle d'affectation du risque minimum de Bayes

<sup>15</sup> le coût  $C(k, h)$  de classement d'un objet de  $\Pi_k$  comme un objet de  $\Pi_h$  est noté  $C(h|k)$ ,  $h$  étant la classe prédite et  $k$  la classe réelle.

## (b) Fusion multi-capteur par le risque minimum de Bayes

De la même façon que pour la règle de l'erreur minimale de Bayes, la règle d'affectation du risque minimum peut être généralisée à  $N$  capteurs, en utilisant l'intersection comme opérateur d'agrégation:

$$\forall X \hat{Y} = \arg \min_{k \in G} \left\{ \bar{C} \left( k \left( X^1, \dots, X^N \right) \right) \right\} = \arg \min_{k \in G} \left\{ \bar{C} \left( k \left| \bigcap_{i=1}^N X^i \right. \right) \right\} = \arg \min_{k \in G} \sum_{h=1}^g \left( C(k|h) * p_h * \Pr \left( \bigcap_{i=1}^N X^i | h \right) \right)$$

**Equation 15 : règle d'affectation de risque minimum de Bayes pour un système multi-capteur**

Le coût  $C(k, h)$  du classement d'un objet de  $\mathcal{I}_k$  comme un objet de  $\mathcal{I}_h$  est identique quel que soit le capteur, car il est établi à partir de l'expertise liée à la classification et non aux capteurs.

De plus, sous l'hypothèse d'indépendance des capteurs, l'Equation 10 exprimant la probabilité conjointe s'applique, et on obtient :

$$\forall X \hat{Y} = \arg \min_{k \in G} \left\{ \sum_{h=1}^g \left( C(k|h) * p_h * \prod_{i=1}^N \Pr(X^i|h) \right) \right\} = \arg \min_{k \in G} \left\{ \sum_{h=1}^g \left( C(k|h) * p_h * \prod_{i=1}^N L_h^i(x) \right) \right\}$$

**Equation 16 : règle d'affectation de risque minimum de Bayes pour un système multi-capteur par le produit des probabilités a priori**

### II.4.4.3. Pertinence de la fusion multi-capteur

Cette partie expose les outils permettant d'évaluer et d'analyser la pertinence de la fusion multi-capteur grâce à la redondance des capteurs et à la comparaison de leurs performances.

#### (1) Corrélations

L'analyse descriptive des données par une Analyse en Composantes Principales (FENELON, 1981) permet d'estimer la corrélation entre les capteurs, c'est-à-dire leur redondance ou leur complémentarité. Les coefficients de corrélation des variables sont calculés et la position de ces données peut être visualisée dans les plans principaux, formant les *cercles de corrélation*, de rayon 1.

#### (2) Evaluation des modèles

Il est extrêmement important de vérifier que la fusion multi-capteur fournit de meilleurs résultats que chaque classification individuelle. Pour ce faire, il existe de nombreux indices qui permettent d'évaluer la qualité d'un modèle (PALAGOS, 1997).

Dans le cas de classifications et non de prédictions quantitatives, le *taux d'erreur* est couramment employé ; il est calculé sur des échantillons tests, n'ayant pas participé à la construction du modèle. Cependant, si la base de données est constituée d'un nombre réduit d'échantillons, alors des *validations croisées par blocs* ou *leave-one-out* peuvent être réalisées (C.f. pII-36).

Le taux d'erreur prend en considération les éléments mal classés dans leur globalité ; ce critère peut être élargi à la prise en compte de la répartition des échantillons dans toutes les classes grâce au

*coefficient de contingement*. Celui-ci rend compte de l'indépendance de 2 classifications, en l'occurrence l'indépendance entre la classification réelle et celle du capteur analysé. Il est calculé à partir du coefficient du  $\chi^2$  (C.f. Equation 17).

Si l'accord est total, c'est-à-dire que le taux d'erreur est nul, alors seuls les  $n_{ii} \neq 0$ , et  $\chi^2 = n(g-1)$ .

$$\chi^2 = \sum_{i=1}^g \sum_{j=1}^g \left[ \frac{\left( n_{ij} - \frac{n_{i+} \cdot n_{+j}}{n} \right)^2}{\frac{n_{i+} \cdot n_{+j}}{n}} \right]$$

où  $g$  est le nombre de classes et  $n$  le nombre total d'échantillons ;

$n_{ij}$  est le nombre d'objets de  $P_j$  affectés à  $P_i$  ;

$n_{i+}$  est la somme marginale sur la ligne  $i$  ;

$n_{+j}$  est la somme marginale sur la colonne  $j$ .

**Equation 17 : coefficient du  $\chi^2$**

Le coefficient de contingement est calculé à partir du  $\chi^2$ , afin de le normaliser :  $C \in [0 ; 1]$  (MAUNG, 1941). Si  $C$  vaut 1, alors la classification est parfaite.

$$C = \sqrt{\frac{\chi^2}{n(g-1)}}$$

**Equation 18 : coefficient de contingement**

## II.4.5. Conclusion

Ce chapitre bibliographique consacré au traitement des données multi-capteur a permis de soulever différentes problématiques inhérentes aux capteurs d'arômes ainsi qu'à la fusion multi-capteur :

- ❶ Quelle est la méthode la plus performante pour *extraire l'information discriminante* des courbes des capteurs d'arômes ? L'extraction du seul maximum d'adsorption semble extrêmement restrictive et beaucoup moins performante que les 3 types de méthodes exposées (C.f. pII-31).
- ❷ Quelle est la *méthode de traitement multivariée* la plus adaptée aux données des capteurs d'arômes (C.f. pII-35) ? Les méthodes les plus couramment employées sont l'Analyse en Composantes Principales suivie d'une Classification Ascendante Hiérarchique, l'Analyse Factorielle Discriminante et les Réseaux de Neurones. Nous proposerons l'adaptation d'un traitement multivarié habituellement appliqué à la prédiction quantitative basée sur des spectres : la PLS discriminante.
- ❸ La *sélection des éléments sensibles* à inclure dans les capteurs d'arômes est une condition nécessaire au succès de la discrimination aromatique (C.f. pII-39). Quelles sont les techniques de sélection de variables les plus efficaces pour le choix des capteurs d'arômes ?
- ❹ Différentes *méthodes de fusion multi-capteur*, appliquées à l'estimation de propriétés de produits agro-alimentaires, ont été exposées (C.f. pII-44). La *fusion de haut niveau* semble particulièrement adaptée à notre problématique. Une extension de la règle d'affectation de risque minimum de Bayes a été proposée, afin de pouvoir introduire des contraintes liées à l'expertise au cours de la fusion.

## II.5. CONCLUSIONS

Au terme de cette revue bibliographique, nous avons choisi de nous intéresser à la discrimination des cépages de moûts de raisins blancs. En outre, devant la difficulté des mesures de boissons alcoolisées, nous avons initié des travaux de recherche visant à lever le verrou technologique constitué par la forte présence d'éthanol dans les vins.

L'état de l'art sur les capteurs d'arômes met en évidence de nombreux points métrologiques et méthodologiques qui ne sont pas maîtrisés. Ainsi, nous avons orienté les travaux de thèse autour d'un axe méthodologique expérimental, visant à comprendre et optimiser différentes étapes de la mesure aromatique, en particulier pour l'amélioration des conditions expérimentales.

La discrimination variétale des cépages nécessite le concours de plusieurs systèmes capables d'analyser la fraction volatile ainsi que la phase liquide. La combinaison de capteurs complexes en vue de la classification supervisée des moûts constitue donc l'axe théorique de la thèse. Celui-ci comprend une phase de sélection de variables issues des signaux bruités par des algorithmes déterministes itératifs (méthodes pas à pas) ou stochastiques (algorithmes génétiques), une phase de classification grâce à l'adaptation à la reconnaissance des formes d'une méthode de régression multiple (PLS) et une ultime phase de fusion multi-capteur à divers niveaux.

# *Méthodologie d'optimisation des mesures par capteurs d'arômes*

*1. Matériel*

*2. Sélection des descripteurs*

*3. Quantification de l'influence des conditions  
expérimentales*

*4. Sélection des capteurs*

### III. METHODOLOGIE D'OPTIMISATION DES MESURES PAR CAPTEURS D'AROMES

Lors de l'étude bibliographique, nous avons soulevé diverses problématiques inhérentes à discrimination aromatique. Ce chapitre est donc destiné à proposer des approches méthodologiques visant à optimiser les mesures des capteurs d'arômes. Celles-ci sont appliquées à des solutions synthétiques, qui représentent de façon extrêmement simplifiée 2 classes de vin à discriminer : les vins « témoins » satisfaisants et les vins comportant un défaut olfactif ; elles sont ensuite mises en œuvre et validées sur la discrimination variétale d'échantillons naturels de moûts de raisin.

L'appareillage employé et les échantillons mesurés sont présentés dans le premier chapitre. Dans les chapitres suivants, nous développerons plus particulièrement les méthodologies permettant de répondre aux questions suivantes :

❶ **Quelle information faut-il extraire des courbes de mesure des capteurs d'arômes ?** La démarche choisie est l'extraction de descripteurs des courbes d'adsorption et de désorption ; ceux-ci sont calculés et sélectionnés à partir d'un très petit nombre d'échantillons (pIII-60) (*ROUSSEL et al.*, 1998b).

❷ **Dans quelles conditions expérimentales est-il préférable d'effectuer les mesures ?** La méthodologie des plans d'expériences est appliquée au choix des conditions expérimentales (pIII-71) (*ROUSSEL et al.*, 1998a).

❸ **Comment choisir les éléments sensibles à inclure dans le système multi-capteur d'arômes ?** Trois méthodes différentes sont proposées afin de sélectionner les capteurs : 2 méthodes pas à pas, basées sur l'Analyse Factorielle Discriminante pas à pas - méthode supervisée paramétrique - (*ROUSSEL et al.*, 1998c), ainsi qu'une méthode stochastique non déterministe (Algorithmes Génétiques) (pIII-82).

## III. 1. MATERIEL

### III. 1.1. Appareillage

L'appareillage employé est dérivé du LCA 1000 (MIDIVALEUR, F), décrit pII-13. Il peut contenir 5 capteurs semi-conducteurs en oxyde métallique de type FIGARO (TGS de la série 8xx).

La chaîne d'acquisition a été modifiée afin d'augmenter la précision du signal : le signal est obtenu en mesurant la tension dans un pont diviseur (C.f. Annexe 2), par un convertisseur A/D 12 bits (et non plus 8 bits) situé dans un micro-ordinateur (PC). Les courbes d'adsorption correspondent donc à des tensions exprimées sur 4096 points (C.f. Figure 5, pII-32).

La tension d'alimentation de la résistance chauffante insérée au sein de chaque élément sensible peut être modifiée via le PC ; mais, la stabilisation de la température des capteurs étant extrêmement délicate (MIELLE, 1996b), la tension des résistances sera maintenue à 5V tout au long des expérimentations. La température des capteurs se situe alors entre 280 et 430°C.

La flexibilité du montage expérimental permet d'effectuer des expérimentations dans des conditions diverses ; un capteur mesurant la température et l'hygrométrie a été inséré dans la cellule de mesure. En particulier, les températures de la cellule de mesure et du gaz vecteur sont modulables, ainsi que celle des échantillons (C.f. Figure 8). De même, l'hygrométrie est régulée en dessiccant l'air vecteur, mais elle peut aussi être fixée à divers niveaux en faisant barboter l'air dans des solutions salines saturées.

L'échantillon gazeux est généré comme un espace de tête statique, c'est-à-dire par équilibre des phases liquide/gaz d'un échantillon liquide de 10ml contenu dans un flacon hermétique de 100ml, durant 16 minutes, soumis à une agitation magnétique. Ces conditions se sont avérées les plus répétables et constituent le meilleur compromis entre la durée de génération de l'espace de tête et la concentration les molécules volatiles (FORSBERG, 1997).

L'injection de l'échantillon est manuelle, soit en mode « statique », c'est-à-dire qu'une fraction de 50ml de l'espace de tête est injecté à l'aide d'une seringue dans la cellule de mesure de 500ml hermétiquement fermée ; soit en mode « dynamique » où l'espace de tête généré est injecté dans la cellule de mesure grâce au passage du gaz vecteur de 500ml.min<sup>-1</sup> de débit dans le flacon contenant l'échantillon (C.f. Figure 8).

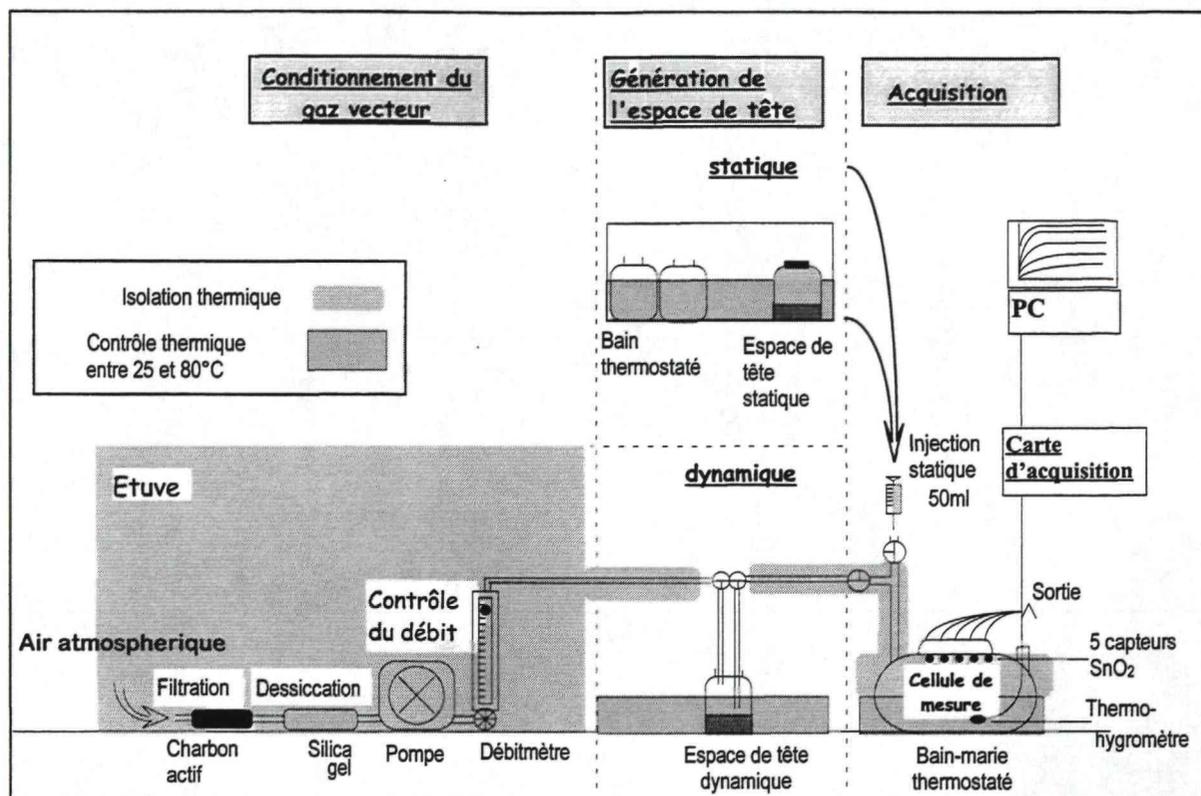


Figure 8 : montage expérimental

Les capteurs employés dans les phases d'optimisation des conditions expérimentales et de la sélection des descripteurs ont été choisis a priori, pour leur efficacité lors d'expérimentations précédentes. Il s'agit des *TGS 800, 822, 824, 825 et 842* (FIGARO) pour l'application concernant les solutions modèles et des *TGS 813, 815, 822, 830, 882* pour la discrimination des goûts de raisin.

Dans toute cette étude, nous ne nous intéresserons qu'aux capteurs semi-conducteurs en oxyde métallique FIGARO TGS de la série 800 ; nous posons comme hypothèse que les conditions qui se révéleront optimales pour ces 5 capteurs dans l'appareillage employé sont aussi les conditions idéales quels que soient les capteurs semi-conducteurs en oxyde métallique employés.

### III. 1.2. Echantillons

Nous avons cherché à minimiser le nombre d'expérimentations ; le nombre d'échantillons mesurés est un compromis entre la durée des expérimentations et la possibilité d'appliquer des outils statistiques permettant d'améliorer significativement les discriminations aromatiques.

Les 2 premières méthodes d'optimisation (la sélection des descripteurs (§III.2) et des conditions expérimentales (§III.3)) sont basées sur 5 répétitions de la mesure d'un seul échantillon représentatif de chaque « classe ».

Cependant, les techniques statistiques multivariées nécessitent un plus grand nombre d'échantillons par classe ; la méthode de choix des capteurs (§III.4) est donc basée sur la mesure d'une dizaine d'échantillons par classe (ce qui reste encore très réduit).

### III. 1.2.1. Solutions synthétiques

Les solutions synthétiques visent à simuler de façon extrêmement simplifiée des vins « satisfaisants » et des vins contenant un défaut olfactif. Le défaut de *piqué* a été choisi, car il est très répandu et les composés incriminés sont connus et relativement concentrés (quelques  $\text{mg.l}^{-1}$ ) par rapport aux autres défauts olfactifs.

Cependant, l'alcool, fortement concentré dans les vins (environ 12% v :v, soit  $94.7\text{g.l}^{-1}$ )<sup>16</sup>, masque le signal des autres composés volatils dont la teneur varie du  $\text{mg.l}^{-1}$  au  $\mu\text{g.l}^{-1}$ , voire même au  $\text{ng.l}^{-1}$  (C.f. pII-26). Il est donc envisagé, en parallèle de la thèse, de désalcooliser les échantillons de vins.

En conséquence, des solutions synthétiques représentant les vins sont générées sans alcool ; comme les molécules volatiles sont généralement hydrophobes, un solvant remplace l'alcool : la triacétine a été choisie, essentiellement pour sa faible volatilité.

Pour les 2 premières méthodes, uniquement 2 solutions mères sont préparées : le témoin est composé de triacétine seule, et l'autre solution contenant 3 composés aromatiques responsables du goût de « piqué », aux concentrations correspondant aux seuils de détection du défaut, dissous dans la triacétine : l'acide acétique à  $700\text{mg.l}^{-1}$ , l'acétate d'éthyle à  $200\text{mg.l}^{-1}$  et Acétaldéhyde à  $300\text{mg.l}^{-1}$ .

Les méthodes de sélection des capteurs nécessitent un plus grand nombre d'échantillons pour chaque classe. Les 2 classes sont définies chacune par 16 échantillons, couvrant l'ensemble de la dispersion des concentrations en acétate d'éthyle et acide acétique (C.f. Tableau 6). L'acétaldéhyde n'a pas été inclus dans ces échantillons, car il est extrêmement volatil et donc très difficile à quantifier précisément.

<sup>16</sup> Densité de l'éthanol =  $0.7896 \text{ kg.l}^{-1} + / - 0.003$

**Tableau 6 : description des 32 échantillons appartenant aux 2 classes de solutions modèles**

« VINS » TEMOINS (numéro des échantillons)					« VINS PIQUES » (numéro des échantillons)				
Acétate d'éthyle (mg/l) <sup>φ</sup>					Acétate d'éthyle (mg/l) <sup>φ</sup>				
Acide acétique (mg/l) <sup>φ</sup>	50	66.6	83.2	100	Acide acétique (mg/l) <sup>φ</sup>	200	233.3	266.6	300
200	1	2	3	4	700	17	18	19	20
300	5	6	7	8	800	21	22	23	24
400	9	10	11	12	900	25	26	27	28
500	13	14	15	16	1000	29	30	31	32

### III. 1.2.2. Moûts de raisin

Les moûts de raisin analysés sont répartis en 4 groupes à discriminer, qui correspondent à 4 types de cépages : le Sauvignon (Sa), le Mauzac (Ma), le Colombard (Co) et tous les autres cépages blancs (Au). Au total, 138 échantillons sont générés ; leur description est détaillée en [Annexe 3](#).

Chaque moût est préparé à partir de 200 grains de raisins, cueillis grain par grain selon la méthode ITV, afin d'être représentatifs des grappes de raisin présentes sur la parcelle de vigne. On cherche à récolter des échantillons de maturité homogène, celle-ci étant estimée par la mesure du taux de sucre, de l'acidité totale et du pH (C.f. [Annexe 3](#)).

Ces grains sont alors transformés en moût par une méthode visant à enrichir l'échantillon en composés volatils grâce à l'évolution des précurseurs d'arômes et des arômes liés (C.f. pII-6).

Les grains de raisin sont soigneusement séchés et pesés. Ils sont ensuite foulés à la main afin d'établir un contact étroit entre la pellicule, la pulpe et le jus. Une macération pelliculaire est alors effectuée, durant 2 heures à température ambiante (environ 25°C) afin d'exacerber les arômes. Puis, le moût est pressuré dans un presseur hydraulique de laboratoire Parapresse (Dekra, D) à 2 bars, jusqu'à obtenir un ratio de 60% entre le poids des raisins et le volume de jus. Ce jus est filtré au travers d'un chinois, réparti dans des flacons et congelé.

Pour les 2 premières méthodes, un échantillon de référence est sélectionné par classe de cépage. Etant donné que la typicité des moûts est difficile à déterminer a priori, les échantillons de référence sont choisis parmi ceux dont le volume est le plus important (C.f. [Tableau 7](#)).

**Tableau 7 : moûts de référence pour la sélection des descripteurs et des conditions expérimentales.**

CEPAGE	REFERENCE	N°
Sauvignon	SA-PER-9	SA44
Mauzac	738R2 SX MA	MA8
Colombard	EFTEM1 GE CO	CO6
Autres cépages	CHA-PER-T	AU24

REFERENCE : nom de l'échantillon correspondant  
au cépage et au terroir.

N° : numérotation des échantillons pour les  
mesures chimiques (C.f. Annexe 3)

La méthode de sélection des capteurs nécessite la mesure d'échantillons plus nombreux. Ces échantillons sont choisis afin de couvrir au mieux la dispersion de chaque classe, du point de vue des analyses chimiques (Sucres, Acidité totale, pH) (C.f. figure 29). Chaque classe est représentée par 10 échantillons, détaillés en Annexe 4.

## III. 2. SELECTION DES DESCRIPTEURS

Parmi les 3 méthodes d'exploitation des courbes des capteurs d'arômes décrites en bibliographie (C.f. pII-31), nous avons approfondi la démarche d'extraction de descripteurs de forme de la courbe (méthode ③). En effet, l'échantillonnage de la courbe (méthode ①) fournit un trop grand nombre de variables et la modélisation (méthode ②) est extrêmement délicate quand l'injection de l'échantillon est manuelle.

### III. 2.1. Méthodes

Notre démarche consiste à calculer divers paramètres, appelés descripteurs, et à les sélectionner en fonction de leur *répétabilité*, leur *pouvoir de discrimination* et leur *redondance*.

Afin de généraliser cette approche et de sélectionner les descripteurs les moins influencés par la fluctuation de paramètres extérieurs, les acquisitions sont effectuées dans diverses conditions expérimentales : la température de la cellule de mesure et du gaz vecteur varient de 35°C à 60°C ainsi que celle de l'espace de tête et les échantillons sont injectés en mode statique ou dynamique (C.f. montage expérimental, pIII-56).

Ces différentes expérimentations s'organisent au sein d'un plan d'expériences comprenant 9 « états » (i.e. 9 combinaisons de conditions expérimentales) détaillés dans le Tableau 11 (pIII-74) ; celui-ci sera analysé dans le chapitre III-3 (pIII-71). Dans chaque état, les mesures de chaque échantillon représentatif d'une classe sont répétées 5 fois, donc chaque échantillon est mesuré 45 fois au total.

#### III. 2.1.1. Extraction des descripteurs du signal des capteurs

La courbe acquise comporte 5 minutes d'adsorption jusqu'à atteindre la phase stationnaire et 10 minutes de désorption ; elle est lissée par moyenne glissante et ses dérivées première et seconde sont calculées par la méthode Savitsky-Golay (SAVITZKY *et* GOLAY, 1964) de la PLS\_TOOLBOX (EigenVector, USA) (WISE *et* GALLAGHER, 1997). A partir de ces 3 courbes, 29 descripteurs sont calculés et détaillés dans le Tableau 8 et la Figure 9.

**Tableau 8 : caractéristiques des 29 descripteurs extraits de la courbe et de ses dérivées**

Le « temps » d'un descripteur représente la durée entre le début de l'acquisition et l'apparition du descripteur (C.f. Figure 9).

NUMERO DU DESCRIPTEUR	NIVEAUX MOYENS SUR DES INTERVALLES DE TEMPS	NUMERO DU DESCRIPTEUR	PENTES MOYENNES SUR DES INTERVALLES DE TEMPS
n° 1.	{3 - 4} secondes	n° 8.	n° 2 - n° 1, i.e. {3 - 6} secondes
n° 2.	{5 - 8}	n° 9.	n° 3 - n° 2, i.e. {6 - 12}
n° 3.	{9 - 16}	n° 10.	n° 4 - n° 3, i.e. {12 - 24}
n° 4.	{17 - 32}	n° 11.	n° 5 - n° 4, i.e. {24 - 48}
n° 5.	{33 - 64}	n° 12.	n° 6 - n° 5, i.e. {48 - 96}
n° 6.	{65 - 128}	n° 13.	n° 7 - n° 6, i.e. {96 - 192}
n° 7.	{129 - 256} (correspondant au maximum)		

NUMERO DU DESCRIPTEUR	DERIVEE 1 <sup>ERE</sup>		NUMERO DU DESCRIPTEUR	DERIVEE 2 <sup>NDE</sup>	
n° 14.	Maximum	Adsorption	n° 22.	Maximum	Adsorption
n° 15.	Temps du maximum	Adsorption	n° 23.	Temps du maximum	Adsorption
n° 16.	Minimum	Adsorption	n° 24.	Minimum	Adsorption
n° 17.	Temps du minimum	Adsorption	n° 25.	Temps du minimum	Adsorption
n° 18.	Maximum	Désorption	n° 26.	Maximum	Désorption
n° 19.	Temps du maximum	Désorption	n° 27.	Temps du maximum	Désorption
n° 20.	Minimum	Désorption	n° 28.	Minimum	Désorption
n° 21.	Temps du minimum	Désorption	n° 29.	Temps du minimum	Désorption

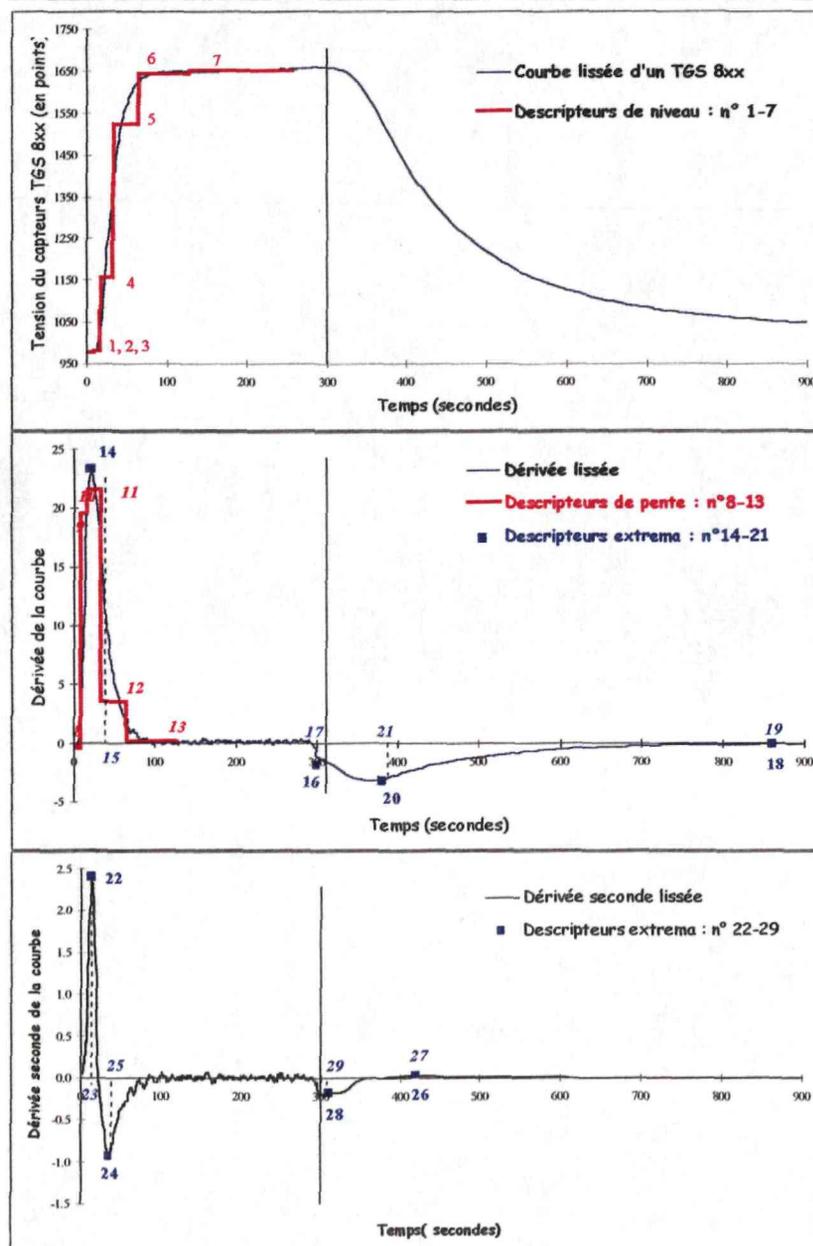


Figure 9 : courbe d'adsorption et de désorption lissée d'un capteur TGS 8xx et ses dérivées première et seconde ainsi que les descripteurs extraits.

### III. 2.1.2. Critères de sélection des descripteurs

La classification de ces 29 descripteurs est établie sur les critères de *répétabilité*, de *distance discriminante* et de *corrélation*.

#### (1) Répétabilité

La répétabilité est souvent estimée par le calcul de l'écart-type des répétitions de mesure ; mais, si cet écart-type s'accroît de façon linéaire en fonction du niveau du paramètre étudié, il est préférable d'estimer les répétabilités à l'aide d'un coefficient de variation afin de les rendre comparables. Nous verrons que c'est le cas des descripteurs extraits des courbes des capteurs d'arômes (C.f. Figure 10, pIII-63). Ainsi, la répétabilité est estimée par le calcul d'un coefficient de variation  $C_i$  (C.f. Equation 19). Celui-ci est calculé pour chaque descripteur  $i$  ( $i \in [1 ; F]$  ;  $F=29$ ) dans chaque état du plan d'expériences  $j$  ( $j \in [1 ; E]$  ;  $E=9$ ), pour les classes d'échantillon  $k$  ( $k \in [1 ; C]$ ) et les capteurs  $s$  ( $s \in [1 ; N]$  ;  $N=5$ ).  ${}^k_s\sigma_i^j$  est l'écart-type du descripteur  $i$  calculé sur les 5 répétitions de mesure par le capteur  $s$  effectuées dans l'état  $j$  pour la classe  $k$  et  ${}^k_s m_i^j$  est sa moyenne.

$$C_i = \sqrt{\frac{1}{N} \sum_{s=1}^N \left( \frac{1}{E} \sum_{j=1}^E \left( \frac{1}{C} \sum_{k=1}^C \left( \frac{{}^k_s \sigma_i^j}{{}^k_s m_i^j} \right)^2 \right) \right)}$$

**Equation 19 : coefficient de variation du descripteur  $i$**

De plus, les descripteurs recherchés doivent aussi être les plus stables par vis à vis des fluctuations des conditions expérimentales ; afin de les distinguer, des répétabilités seront également calculées en prenant en considération les mesures effectuées à différentes températures de la cellule de mesure ou de l'espace de tête.

#### (2) Distance discriminante

La distance discriminante  $D_i$  est calculée pour chaque descripteur  $i$  comme la somme des rapports entre la distance des centres de gravité des classes et les écart-types intraclasses (THOMPSON, 1990) (C.f. Equation 20).  ${}^{k,l}_s d_i^j$  est la distance euclidienne entre les classes  $k$  et  $l$  du descripteur  $i$  pour l'état expérimental  $j$  calculée sur les mesures du capteur  $s$ . Pour  $C$  classes,  $\sum_{m=1}^{C-1} m$  distances euclidiennes sont calculées (i.e. pour 2 classes, 1 distance ; pour 4 classes, 6 distances calculées).

$$D_i = \frac{1}{N} \sum_{s=1}^N \left( \frac{1}{E} \sum_{j=1}^E \left( \frac{1}{\sum_{m=1}^{C-1} m} \sum_{k \neq l} \frac{{}^{k,l}_s d_i^j}{\sqrt{({}^k_s \sigma_i^j)^2 + ({}^l_s \sigma_i^j)^2}} \right) \right)$$

**Equation 20 : distance discriminante du descripteur  $i$**

Les descripteurs les plus performants sont ceux qui présentent le coefficient de variation le plus bas et la plus forte distance discriminante ; afin de les distinguer, le rapport  $q_i$  est calculé (Equation 21) : plus il est haut, meilleur est le descripteur.

$$q_i = \frac{D_i}{C_i}$$

Equation 21 : indice de qualité  $q_i$  du descripteur  $i$

### (3) Corrélation

La corrélation entre les descripteurs est calculée par capteur et pour le système multi-capteur, en employant les mesures effectuées dans toutes les conditions expérimentales. Les descripteurs sont alors assemblés en groupes, en fonction de leur redondance. Le meilleur descripteur de chaque groupe, i.e. celui qui a le meilleur  $q_i$ , est sélectionné. La redondance des capteurs est visualisée sur le cercle des corrélations d'une ACP normée.

## III. 2.2. Résultats

### III. 2.2.1. Application aux solutions modèles

Les 29 descripteurs issus des capteurs TGS 800, 822, 824, 825 et 842 sont évalués en fonction de leur *répétabilité, distance discriminante et redondance*.

#### (1) Répétabilité

La répétabilité est estimée par un coefficient de variation, sous réserve de la croissance linéaire de l'écart-type en fonction du niveau du descripteur. La Figure 10 montre la liaison linéaire entre ces 2 caractéristiques, ce qui justifie pleinement l'emploi du coefficient de variation.

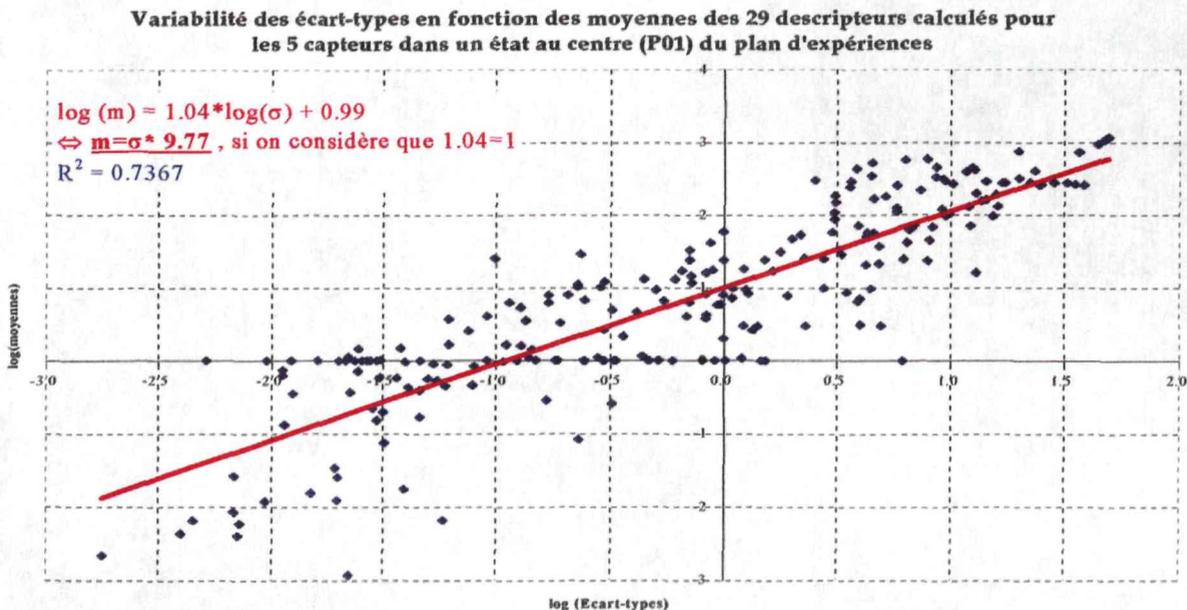


Figure 10 : liaison linéaire entre écart-type et le niveau des descripteurs

La répétabilité des 29 descripteurs est présentée dans la Figure 11.

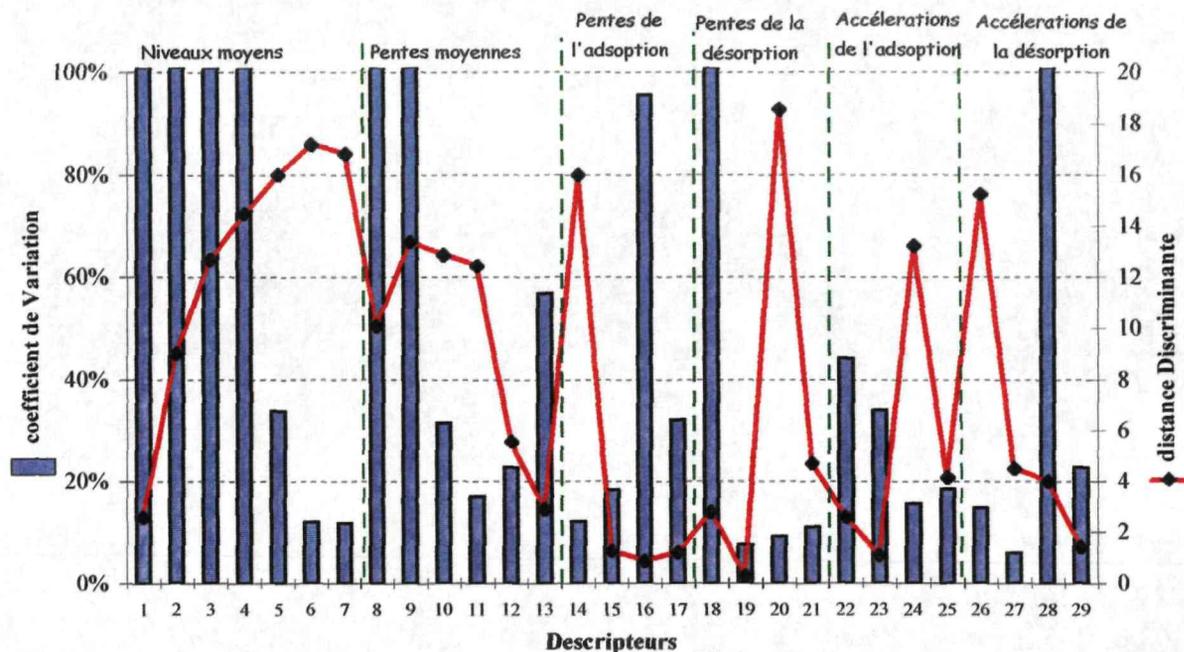


Figure 11 : répétabilité et distance discriminante des 29 descripteurs

Cette figure montre que plus les descripteurs de type « niveaux moyens » sont éloignés du point de départ de l'acquisition, plus ils sont répétables (descripteurs n°5, 6 et 7). Ceci confirme que la précision du début de l'injection et de l'acquisition du signal sont primordiales (ELKÖV *et al.*, 1997 & VERNAT-ROSSI *et al.*, 1996), quelle que soit la démarche d'extraction de l'information des courbes.

La meilleure répétabilité des descripteurs de niveau est fourni par le maximum (descripteur n° 7), ce qui justifie le choix classique de ce descripteur. Mais, certains descripteurs extraits de la partie dynamique du signal présente une répétabilité similaire, tels que la pente maximale et la décélération de l'adsorption (n°14 et 24), ou la pente minimale de la désorption (n°20). D'autres sont très répétables car ils sont constants quel que soit l'échantillon, tels que le temps d'apparition de la décélération maximale (n°29), marquant le début de la désorption ou la pente maximale de la désorption (n°19), à la fin de l'acquisition. Il faut cependant être vigilant vis à vis de ces critères « calculés » (dérivées numériques), car leur manque de stabilité peut provoquer une insuffisance de robustesse du modèle.

Cette répétabilité est une moyenne calculée sur tous les capteurs et individuellement dans toutes les conditions expérimentales. Néanmoins, nous avons déterminé que, quel que soit le capteur, les répétabilités des divers descripteurs sont ordonnés de la même façon, malgré des performances très différentes d'un capteur à l'autre.

De plus, les répétabilités des descripteurs ont été calculées lors de changements de température, afin d'estimer leur stabilité. Celles-ci s'ordonnent de la même manière que lors du calcul des répétabilités précédentes : donc, les descripteurs les plus répétables dans un état expérimental donné sont aussi ceux qui sont le moins influencés par la fluctuation des conditions expérimentales.

CemOA : archive ouverte d'Irstea / Cemagref

## (2) distance discriminante

La Figure 11 montre la distance discriminante calculée à partir des 29 descripteurs. Sachant que la principale différence entre les solutions modèles est l'intensité olfactive, on remarque que plus les descripteurs de *niveaux* sont proches du maximum de la courbe, plus ils sont discriminants. Il en est de même pour les *pentés moyennes* dont les niveaux proches du maximum (n°14) sont les plus discriminants (n°9, 10, 11). Parmi les dérivées première et seconde, les descripteurs n°20 (*pente minimale de la désorption*), 24 (*décélération maximale de l'adsorption*) et 26 (*accélération maximale de la désorption*, marquant le début de la fin de la désorption) sont aussi discriminants que le maximum de la courbe ; par contre, certains descripteurs sont inutiles à la discrimination, car ils sont constants, tels que la *pente minimale de l'adsorption* (n°16) et le *temps correspondant* (n°17), la *pente maximale de la désorption* et son *temps* (n°18-19) ou encore la *décélération minimale de la désorption* (n°28) se produisant à chaque début de désorption (n°29).

Afin de sélectionner les descripteurs présentant les meilleures distances discriminantes et répétabilités, leur ratio  $q_i$  est calculé ; il est représenté dans la Figure 12. Un seuil est fixé arbitrairement à 20, pour présélectionner les descripteurs avant de les grouper en famille grâce à leurs corrélations : il s'agit des n°5, 6, 7 (*niveaux*), 10, 11, 12 (*pentés moyennes*), 14 (*pente maximale*), 20, 21 (*pente minimale de la désorption et son temps*), 24 (*décélération maximale*), 26 et 27 (*dérivée 2<sup>nde</sup> maximale de la désorption*).

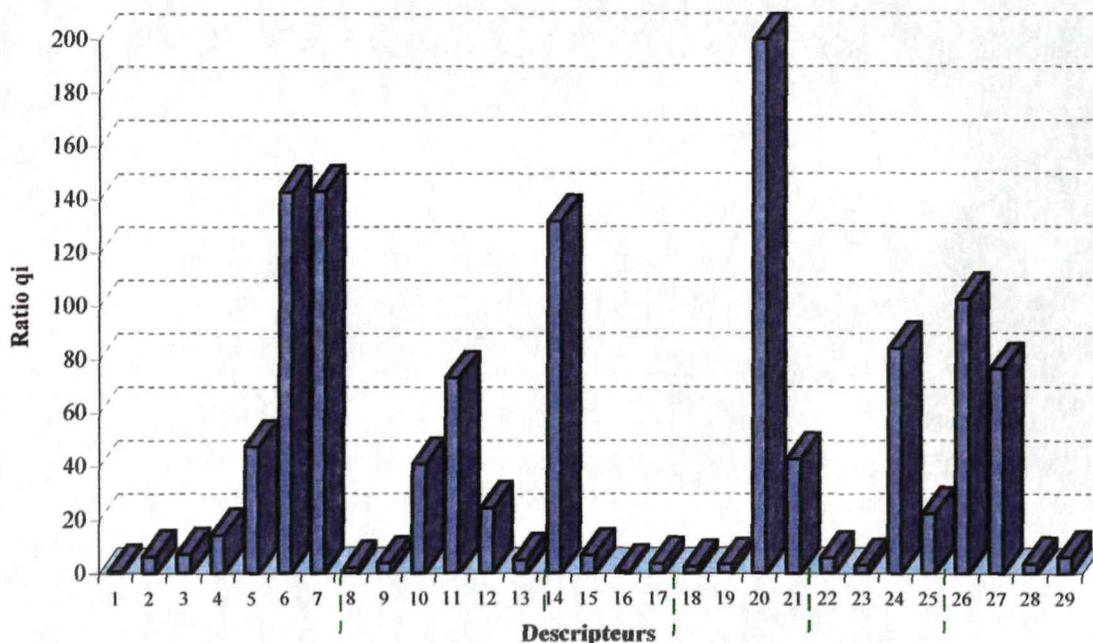


Figure 12 : ratio  $q_i$  des 29 descripteurs calculé

## (3) Corrélation

A partir des 12 descripteurs présélectionnés, 5 familles contenant des éléments très corrélés sont construites ( $r^2 > 0.80$ ). Celles-ci sont détaillées dans le Tableau 9 et les coordonnées des descripteurs

sont exprimées dans le premier plan principal d'inertie après une ACP normée, afin de visualiser leur proximité (C.f. Figure 13).

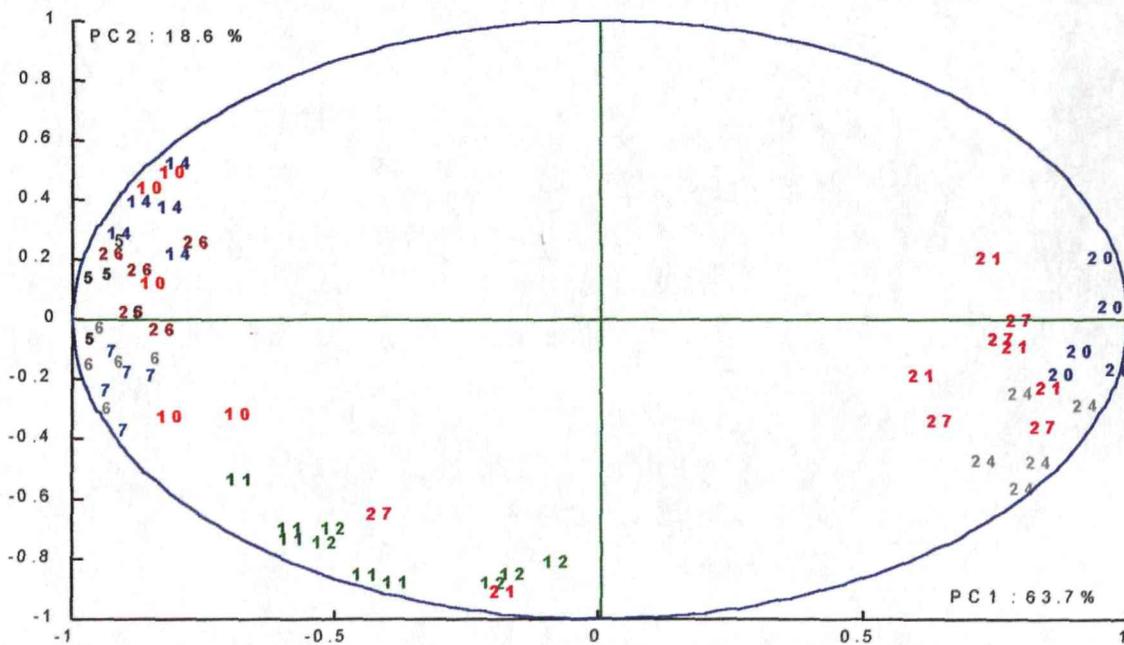
Quel que soit le capteur, les mêmes descripteurs se trouvent dans les mêmes zones de la Figure 13, révélant que les capteurs sont extrêmement corrélés par type de descripteur ; il est donc extrêmement important d'extraire des descripteurs exprimant différentes caractéristiques de la courbe de réponse des capteurs, afin d'augmenter la quantité d'information extraite du système multi-capteur.

Cependant, 2 descripteurs (n°21 et 27) du capteur TGS 825 se trouvent éloignés des autres descripteurs du même type, mais restent corrélés entre eux (C.f. Figure 13) ; en effet, ce capteur revient à la ligne de base beaucoup plus rapidement que les autres capteurs, donc ses temps de désorption sont atypiques par rapport aux autres.

**Tableau 9 : 5 groupes de descripteurs redondants pour les solutions modèles**

GROUPES	NUMERO DU DESCRIPTEUR	CARACTERISTIQUES
Groupe I	5, 6 et 7	Niveaux successifs, proches du maximum
Groupe II	11 et 12	Pentes moyennes successives
Groupe III	10, 20 et 26	Une pente d'adsorption et 2 extrema de désorption (dérivée 1 <sup>ère</sup> et 2 <sup>nde</sup> )
Groupe IV	14 et 24	Extrema d'adsorption (dérivée 1 <sup>ère</sup> et 2 <sup>nde</sup> )
Groupe V	21 et 27	Temps de descripteurs de désorption (dérivée 1 <sup>ère</sup> et 2 <sup>nde</sup> )

De chacun de ces 5 groupes, le descripteur présentant le meilleur  $q$ , est extrait ; il s'agit du *maximum d'adsorption* (n°7), la *pente moyenne calculée entre 24 et 48 secondes* (n°11), la *pente maximale* (n°14), la *pente minimale de la désorption* (n°20) et le *temps correspondant à l'apparition de la décélération maximum de la désorption* (n°27).



**Figure 13 : cercle des corrélations des 12 descripteurs présélectionnés**

#### (4) Conclusion

Afin de comparer les performances de cette méthode d'extraction et de sélection des descripteurs par rapport à la méthode classique d'extraction du maximum d'adsorption, la Figure 14 montre la dispersion des échantillons, calculée avec le maximum seul (a) ou avec les descripteurs sélectionnés (b). Lorsque l'ACP est réalisée sur les 5 descripteurs, la dispersion entre les répétitions (*variance intraclasse*) est réduite et la distance séparant les groupes (*variance interclasse*) est accrue.

Néanmoins, l'écart entre ces deux performances est atténué par le fait que la différence principale des 2 solutions modèles est l'intensité aromatique ; en effet, si la différence était surtout qualitative, l'importance des descripteurs qui ne sont pas liés à l'intensité aromatique serait amplifiée.

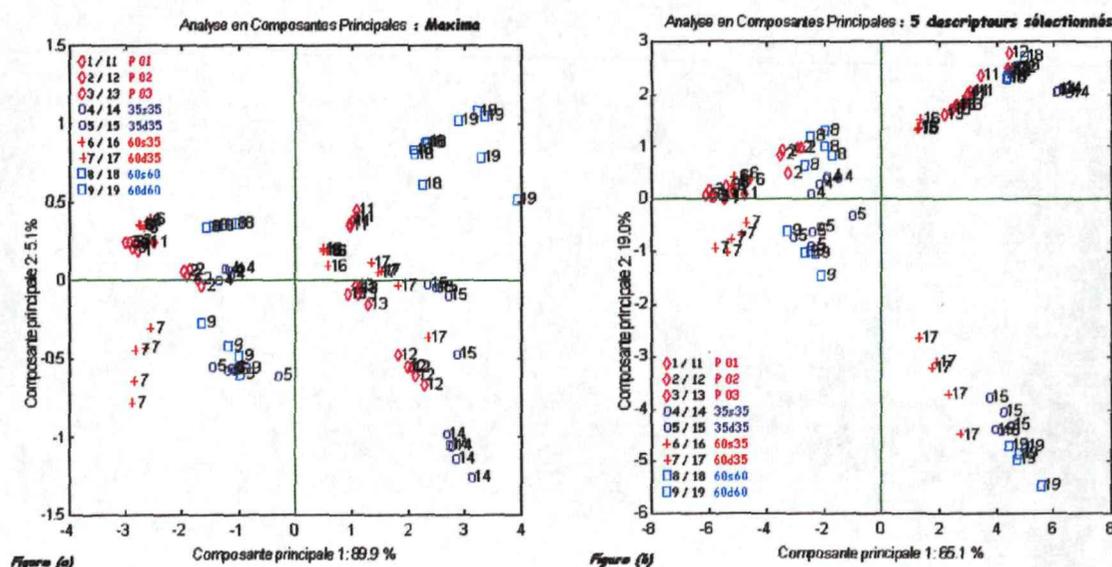


Figure 14 : ACP calculée sur le maximum (a) et sur les 5 descripteurs sélectionnés (b)

#### III. 2.2.2. Application aux moûts de raisin

L'application de la méthodologie de choix des descripteurs aux moûts de raisins blancs s'appuie sur les conclusions établies précédemment. Celles-ci nous permettent d'éviter le calcul des 29 descripteurs initiaux ; seuls les 12 descripteurs présélectionnés, en fonction de leur répétabilité et distance discriminante (C.f. Tableau 9, pIII-66), sont pris en considération. Ils sont calculés à partir des répétitions de mesures des 4 échantillons de cépages différents (C.f. Tableau 7, pIII-59) effectuées par les capteurs TGS 813, 815, 822, 830, 882.

**(1) indices de performance**

La répétabilité et distance discriminante des 12 descripteurs présélectionnés sont détaillées dans la Figure 15.

Les répétabilités moyennes sont sensiblement meilleures que celles calculées sur les solutions modèles ; en effet, toutes les injections des échantillons de moûts de raisin ont été effectuées en mode statique, grâce auquel la précision des mesures est meilleure. Ce changement de mode opératoire est consécutif à l'analyse des conditions expérimentales grâce à un plan d'expériences (C.f. §III.3), qui a déterminé que le transfert de l'échantillon en mode statique est préférable.

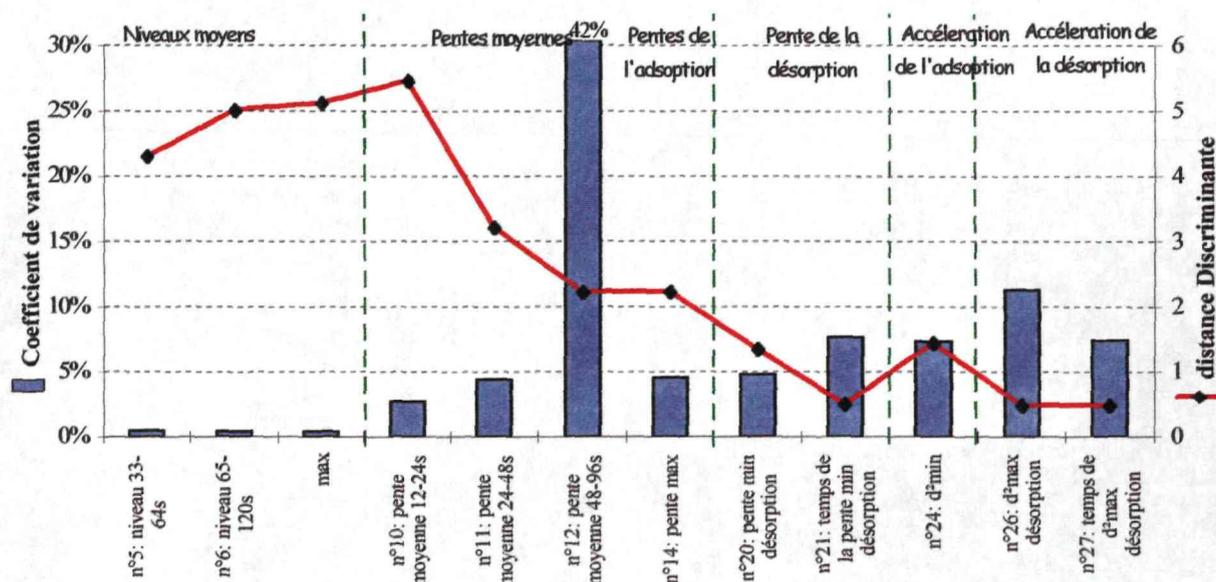


Figure 15 : coefficient de variation et distance discriminante des 12 descripteurs présélectionnés

Cependant, contrairement aux solutions modèles, les pentes moyennes calculées en début d'adsorption sont plus répétables et discriminantes que les suivantes ; lors de ces mesures, l'espace de tête est injecté beaucoup plus rapidement après le début de l'acquisition, donc les pentes n°10 et 11 sont calculées dans la zone de la pente maximale alors que la pente n°12 correspond au début du plateau de la réponse stationnaire (C.f. Figure 16).

La vitesse d'adsorption maximale est beaucoup moins discriminante que dans l'application précédente, car le début d'adsorption est très abrupt quel que soit l'échantillon (il correspond à la pente maximale) ; néanmoins, une pente moyenne calculée en début d'adsorption entre 12 et 24 secondes (n°10) est plus discriminante que le maximum, tout en étant relativement répétable.

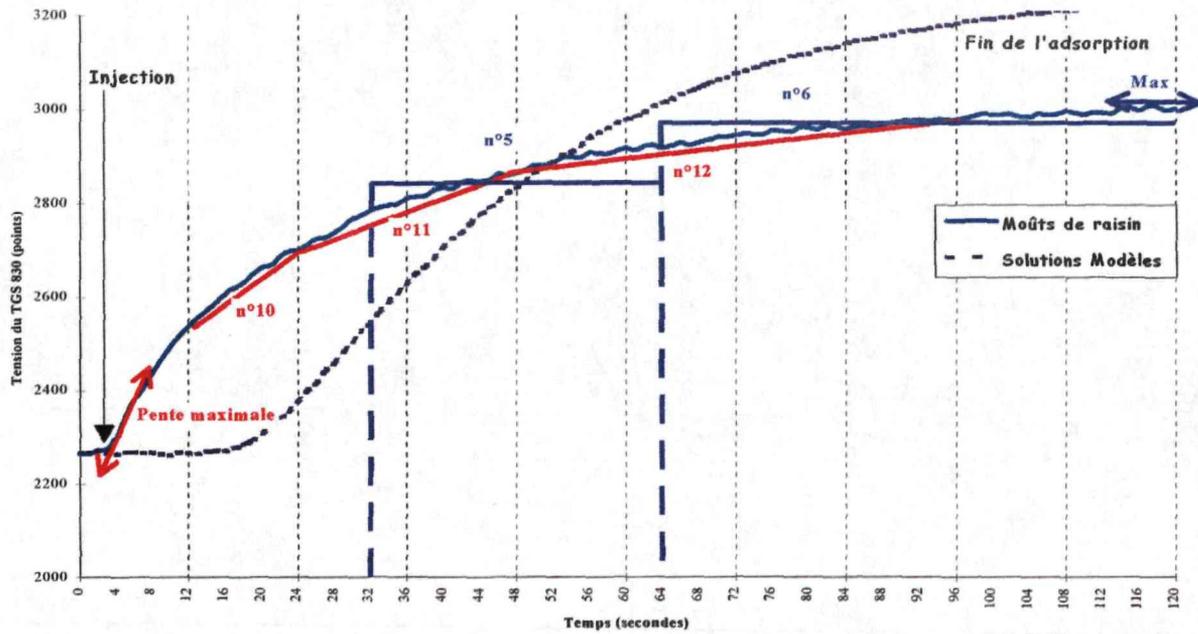


Figure 16 : comparaison des courbes d'adsorption des solutions modèles et des moûts. Extraction des descripteurs de la courbe de mesure d'un moût.

La Figure 17 montre que les descripteurs issus de la désorption sont beaucoup moins performants que lors des mesures précédentes, indiquant que les différences de désorption pourraient être principalement dues à des différences de quantités adsorbées, i.e. aux concentrations. La différence entre les cépages n'étant pas effectuée sur l'intensité aromatique, ces descripteurs sont peu efficaces.

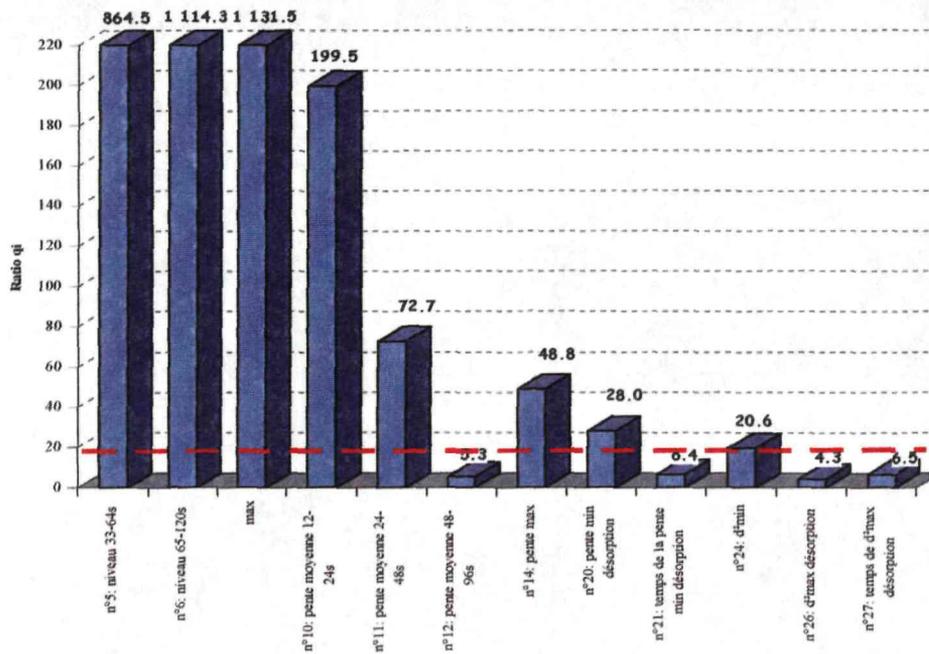


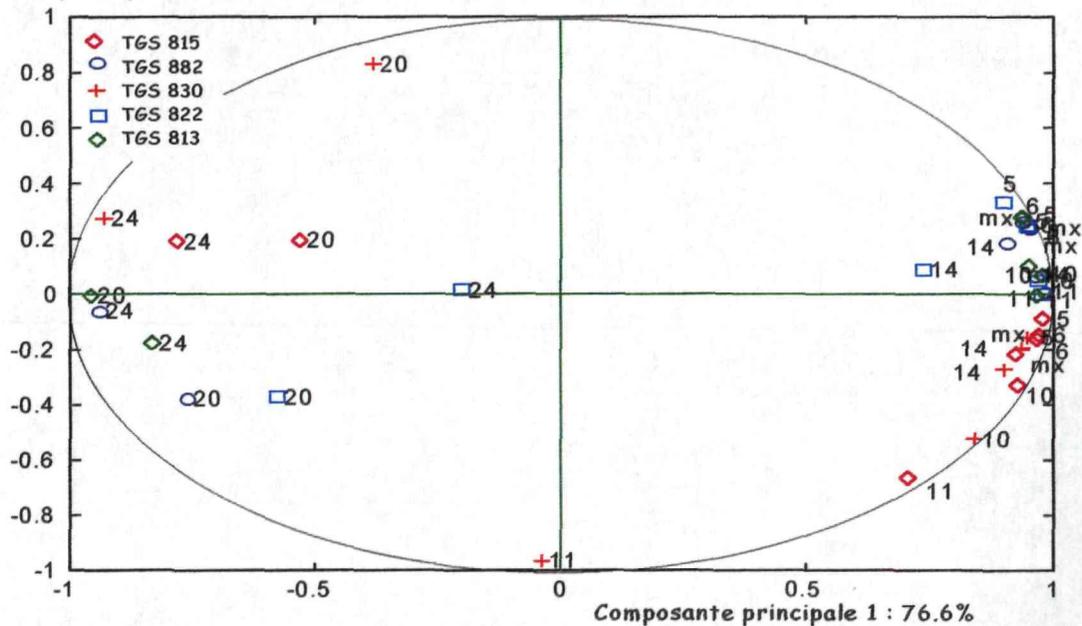
Figure 17 : ratio  $q_i$  des 12 descripteurs présélectionnés

Donc, lors de la présélection, quatre descripteurs sur les 12 sont éliminés car leur performance est insuffisante, i.e. leur  $q_i$  est inférieur à 20 ; il s'agit des n°12, 21, 26 et 27.

**(2) corrélations**

Les corrélations des 8 descripteurs présélectionnés sont visualisées sur le cercle des corrélations, issus d'une ACP normée (C.f. Figure 18); elles induisent la construction de familles de descripteurs redondants, présentées dans le Tableau 10.

Composante principale 2 : 9.7%



**Figure 18 : cercle des corrélations des 8 descripteurs présélectionnés**

Les descripteurs étant extrêmement redondants, les familles sont formées par des descripteurs corrélés à plus de 90% ( $r^2 > 0.90$ ).

**Tableau 10 : 5 groupes de descripteurs redondants pour les moûts**

GROUPES	NUMERO DU DESCRIPTEUR	CARACTERISTIQUES
Groupe I	5, 6, maxima	Niveaux successifs, proches du maximum
Groupe II	10 et 11	Pentes moyennes
Groupe II	14	pente maximale
Groupe III	20	Pente minimale de désorption
Groupe IV	24	Décélération maximale d'adsorption

Le descripteur présentant la meilleure performance ( $q_i$ ) est sélectionné dans chacune des familles. Il s'agit du *maximum d'adsorption*, de la *pente moyenne calculée entre 12 et 24 secondes* (n°10), de la *pente maximale de l'adsorption* (n°14), de la *pente minimale de la désorption* (n°20) et de la *décélération maximale de l'adsorption* (n°24).

**(3) Conclusion**

Grâce à la sélection de ces 5 descripteurs, la discrimination des répétitions des 4 échantillons de cépages différents est légèrement améliorée, comme le montrent les diverses classifications hiérarchiques ascendantes de l'Annexe 5.

### III. 2.3. Conclusion

Les deux applications précédentes ont montré l'intérêt de l'extraction de plusieurs descripteurs des signaux des capteurs d'arômes. En effet, les maxima des divers signaux étant extrêmement corrélés (C.f. Figure 13 et Figure 18), il est nécessaire de calculer des descripteurs, rendant compte de différentes caractéristiques de forme de la courbe, afin d'accroître l'information issue des capteurs d'arômes. De la même façon, les descripteurs d'un même type sont très corrélés.

Néanmoins, le maximum de l'adsorption est un des descripteurs les plus performants, du point de vue de la répétabilité et de la distance discriminante. Les descripteurs sélectionnés pour chaque application sont similaires (à l'exception d'un seul), et proviennent de l'adsorption ainsi que de la désorption ; il s'agit du *maximum* d'adsorption, d'une *pente moyenne* calculée sur un intervalle de l'adsorption, de la *pente maximale* de l'adsorption, de la *décélération maximale* de l'adsorption, de la *pente minimale de la désorption* et du *temps* correspondant à l'apparition de la décélération maximale de la désorption.

Les descripteurs employés lors de l'optimisation des conditions expérimentales ainsi que pour le choix des capteurs seront choisis parmi ceux-ci ; en effet, les descripteurs sont sélectionnés de manière globale pour les 5 capteurs analysés, il peut être judicieux ultérieurement de choisir individuellement le(s) descripteur(s) à calculer pour chacun des capteurs, afin de minimiser la redondance et le bruit (C.f. p III-87).

### III. 3. QUANTIFICATION DE L'INFLUENCE DES CONDITIONS EXPERIMENTALES

La bibliographie a montré que les conditions expérimentales influent considérablement sur les mesures des capteurs d'arômes (C.f. pII-22). Nous avons donc choisi d'élaborer une *méthodologie capable de quantifier l'effet des paramètres expérimentaux sur la performance des capteurs d'arômes et de déterminer les conditions expérimentales optimales*. La démarche que nous proposons consiste à appliquer la théorie des *plans d'expériences* à des critères de performance calculés à partir de mesures des capteurs d'arômes.

#### III. 3.1. Méthodes

La démarche des plans d'expériences est appliquée à 3 paramètres extérieurs, qui sont la *température de la cellule* de mesure et du gaz vecteur qui lui est injecté, celle de *l'espace de tête* et son *mode d'injection*. La démarche proposée est basée sur la méthodologie des plans d'expériences.

##### III. 3.1.1. Plans d'expériences

###### (1) Définition

La technique des plans d'expériences est une méthode chimiométrique couramment employée dans de nombreux domaines, tels que la chimie ou l'agro-alimentaire (DANZART, 1997). Son objectif est de quantifier l'influence de divers paramètres (appelés *facteurs*) et de leurs interactions sur un phénomène (mesure ou procédé) en utilisant aussi peu d'expérimentations que possible (SADO et SADO, 1991).

Divers plans d'expériences peuvent être appliqués en fonction de l'objectif des expérimentations : l'évaluation de l'effet de facteurs et de leurs interactions sur une mesure est effectué par des « plans d'expériences factoriels à  $n$  niveaux » (PLACKETT et BURMAN, 1946), sous l'hypothèse de *monotonie des effets* sur les sorties analysées. Le nombre d'expérimentations peut être réduit grâce aux « plan d'expériences fractionnaires », dans lesquels les interactions supérieures à l'ordre 2<sup>17</sup> sont négligées.

Si l'on cherche à modéliser un phénomène non linéaire afin de déterminer son point de fonctionnement optimal, il est préférable d'employer par exemple des « plans composites à surface de réponse » (BOX et HUNTER, 1957).

---

<sup>17</sup> L'interaction d'ordre 2 correspond à celle de 3 facteurs

## (2) Identification des paramètres influents

Seuls les paramètres expérimentaux modifiant de façon monotone la réponse des capteurs seront étudiés, afin de pouvoir appliquer les plans d'expériences factoriels à 2 niveaux ; par exemple, l'effet d'un changement de température intrinsèque des capteurs ne sera pas étudié, car il induit des changements non monotones sur le maximum d'adsorption (C.f. Figure 4, pII-22), donc nous pouvons supposer que c'est aussi le cas sur des critères de performance.

La première phase de construction d'un plan d'expériences est l'identification exhaustive des paramètres contrôlables pouvant influencer la mesure du phénomène analysé et la délimitation du domaine d'intérêt des facteurs étudiés. Lors de la revue bibliographique (C.f. pII-22) nous avons mis en évidence de nombreux facteurs ayant un effet sur la mesure des capteurs d'arômes ; en particulier, l'hygrométrie, la température des capteurs, de la cellule de mesure, de l'espace de tête, la pression, la durée de génération de l'espace de tête, la durée entre 2 mesures successives, le mode d'injection de l'échantillon (durée, quantité, vitesse, présence d'un flux de gaz vecteur,...), etc.

Nous nous sommes intéressés à 3 conditions expérimentales fondamentales sur lesquelles nous appliquons la méthodologie des plans d'expériences, tout en maintenant les autres paramètres constants : la température de la cellule de mesure et du gaz vecteur, celle de l'espace de tête et son mode d'injection.

La température de l'espace de tête est étudiée entre 35°C (température de dégustation et de rétro-olfaction) et 60°C (température visant à augmenter la concentration des arômes sans les dénaturer). La température de la cellule de mesure, ainsi que celle du gaz vecteur qui lui est injecté, varie également entre 35°C et 60°C. Enfin, le mode d'injection, qui est un critère qualitatif, est « statique » ou « dynamique », ces conditions expérimentales étant définies dans le chapitre III.1.1 (pIII-55).

## (3) Planification des expérimentations

Dans un plan d'expériences, les paramètres expérimentaux étudiés varient simultanément, pendant que les autres conditions expérimentales sont fixées. Le domaine d'intérêt de chaque facteur ayant été déterminé, les expérimentations ne seront effectuées qu'à ses bornes, en considérant que l'effet du facteur évolue de façon monotone au sein du domaine.

Les bornes sont appelées *niveaux*,  $-1$  pour la borne inférieure (niveau bas) et  $+1$  pour la borne supérieure (niveau haut) : il s'agit d'un plan d'expériences factoriel à 2 niveaux. Un plan analysant 3 facteurs comporte donc  $2^3 = 8$  états expérimentaux. Dans notre étude, la température de la cellule de mesure ne peut pas être inférieure à celle de l'espace de tête, afin d'éviter toute condensation : un état expérimental est donc impossible (température de la cellule de mesure à 35°C et celle de l'espace de tête à 60°C), ce qui nous a conduit à décomposer ce plan en 3 plans d'expériences à 2 facteurs.

Ceux-ci sont détaillés dans le Tableau 11 : le *plan d'expériences I* où la température de l'espace de tête est fixée au niveau bas (35°C) ; le *plan d'expériences II* où la température de la cellule de mesure est fixée au niveau haut (60°C) et le *plan d'expériences III* dans lequel les deux températures sont égales.

Une expérimentation supplémentaire est répétée 3 fois au centre du domaine de définition de chaque facteur, afin d'estimer la variance des mesures et de vérifier la monotonie des effets ainsi que l'absence de dérive ; le mode d'injection statique est choisi pour le niveau « 0 » du facteur qualitatif « mode d'injection ». Sous l'hypothèse d'homoscédasticité<sup>18</sup>, nous considérerons que l'intervalle de confiance calculé au centre est valable dans tout le domaine d'intérêt.

**Tableau 11 : matrices d'expérience des 3 plans d'expériences.**

Le nom de l'état correspond au codage suivant : température de la cellule de mesure, mode d'injection et température de l'espace de tête.

(a) Plan d'expériences I : la température de l'espace de tête est fixée au niveau bas (35°C).

NUMERO DE L'ETAT (j)	NOM DE L'ETAT	TEMPERATURE DE LA CELLULE DE MESURE	MODE D'INJECTION	TEMPERATURE DE L'ESPACE DE TETE
2	35s35	-1	-1	-1
3	35d35	-1	+1	-1
6	60d35	+1	+1	-1
7	60s35	+1	-1	-1
Niveau +		60°C	Dynamique	60°C
Niveau -		35°C	Statique	35°C
Niveau 0		47.5°C	Statique	47.5°C

(b) Plan d'expériences II : la température de la cellule de mesure est fixée au niveau haut (60°C).

NUMERO DE L'ETAT	NOM DE L'ETAT	TEMPERATURE DE LA CELLULE DE MESURE	MODE D'INJECTION	TEMPERATURE DE L'ESPACE DE TETE
4	60s60	+1	-1	+1
6	60d35	+1	+1	-1
7	60s35	+1	-1	-1
8	60d60	+1	+1	+1

(c) Plan d'expériences III : les deux températures sont égales

NUMERO DE L'ETAT	NOM DE L'ETAT	TEMPERATURE DE LA CELLULE DE MESURE	MODE D'INJECTION	TEMPERATURE DE L'ESPACE DE TETE
1	P01	0	0	0
2	35s35	-1	-1	-1
4	60s60	+1	-1	+1
5	P02	0	0	0
9	P03	0	0	0

<sup>18</sup> Homoscédasticité : variance identique en tout point du domaine d'étude

#### (4) Quantification des effets

Afin de quantifier l'influence de  $p$  facteurs sur les sorties analysées, leurs effets sont calculés grâce à l'Equation 22. L'effet moyen du facteur  $i$  est calculé grâce aux sorties analysées  $y_j$  et aux  $N$  ( $N=2^p$ ) niveaux  $n_j^i \in \{-1 ; 0 ; 1\}$  des états  $j$  (SADO et SADO, 1991).

$$\text{Effet}_{\text{facteur } i} = \frac{1}{N} \sum_{\forall j} n_j^i * y_j$$

Equation 22 : effet moyen du facteur  $i$

L'effet de l'interaction des facteurs  $i$  et  $k$  se calcule de façon similaire (C.f. Equation 23) :

$$\text{Effet}_{\text{interaction } i,k} = \frac{1}{N} \sum_{\forall j} (n_j^i * n_j^k) * y_j$$

Equation 23 : effet moyen de l'interaction des facteur  $i$  et  $k$

L'effet d'un facteur est significativement différent de zéro s'il est à l'extérieur de l'intervalle de confiance (C.f. Equation 25) ; celui-ci est calculé grâce à de la variance estimée  $S^2(y_0)$  à partir des  $n$  répétitions au centre  $\{y_{p01}, y_{p02}, y_{p03}\}$  (C.f. Equation 24) et au coefficient  $t_{\nu, 1-\alpha/2}$  de Student ( $\nu=n-1$  est le degré de liberté et  $\alpha$  le risque de 1<sup>ère</sup> espèce<sup>19</sup>) (C.f. Equation 25).

$$S^2(y_0) = \frac{\sum_{j=1}^N (y_{p0j} - \bar{y}_{p0})^2}{n-1}$$

Equation 24 : variance estimée  $S^2$  de la mesure

$$\text{Intervalle de confiance} = \left[ \pm t_{\nu, 1-\alpha/2} * \left( \frac{S(y_0)}{\sqrt{N}} \right) \right]$$

Equation 25 : Intervalle de confiance

Si le pourcentage de risque est  $\alpha=10\%$ , pour  $n=3$  répétitions, le coefficient de Student est  $t_{2,0.95} = 2.9$  (pour 5% de risque,  $t_{2,0.975} = 4.3$ ). Alors, si l'effet est supérieur à l'intervalle défini, l'hypothèse d'effet négligeable  $H_0$  est rejetée avec un risque de 10%.

#### III. 3.1.2. Sorties analysées

Les phénomènes sur lesquels l'influence des facteurs est analysée sont généralement directement issus de la mesure, comme par exemple la mesure d'un appareillage, ou le rendement d'une réaction (DANZART, 1997).

Ici, l'étude ne porte pas sur l'influence directe des facteurs sur la courbe des capteurs d'arômes, mais sur l'effet des conditions expérimentales sur des *indices de qualité*. Ceux-ci sont similaires aux indices d'évaluation des performances des descripteurs décrits dans les Equation 19 & Equation 20 (pIII-62). Il s'agit des critères de *répétabilité* et de *distance discriminante*.

<sup>19</sup>  $\alpha$ , le risque de 1<sup>ère</sup> espèce, est la probabilité (rejet  $H_0$  /  $H_0$  vraie) ; ici  $H_0$  est l'hypothèse d'un effet négligeable.

### (1) Répétabilité

La répétabilité  $C^j$  est calculée pour chaque état  $j$  du plan d'expériences. Comme précédemment,  ${}^k_s\sigma_i^j$  est l'écart-type du descripteur  $i$  calculé sur les 5 répétitions de mesure du capteur  $s$ , effectuées dans l'état  $j$  pour la classe  $k$  et  ${}^k_s m_i^j$  est sa moyenne. Cette étude sera menée sur les  $F$  descripteurs sélectionnés lors de la première phase d'optimisation.

$$C^j = \sqrt{\frac{1}{N} \sum_{s=1}^N \left( \frac{1}{F} \sum_{i=1}^F \left( \frac{1}{C} \sum_{k=1}^C \left( \frac{{}^k_s\sigma_i^j}{{}^k_s m_i^j} \right)^2 \right) \right)}$$

Equation 26 : coefficient de variation de l'état  $j$  du plan d'expériences

### (2) Distance discriminante

De même, la distance discriminante  $D^j$  est calculée pour chaque état  $j$ .

$$D^j = \frac{1}{N} \sum_{s=1}^N \left( \frac{1}{F} \sum_{i=1}^F \left( \frac{1}{\sum_{m=1}^{C-1} m} \sum_{\forall k \neq l} \frac{{}^{k;l}_s d_i^j}{\sqrt{({}^k_s\sigma_i^j)^2 + ({}^l_s\sigma_i^j)^2}} \right) \right)$$

Equation 27 : distance discriminante de l'état  $j$  du plan d'expériences

Le rapport entre la distance discriminante et le coefficient de variation est le ratio  $q^j$  : l'état du plan d'expériences le plus performant sera caractérisé par le plus haut  $q^j$ .

$$q^j = \frac{D^j}{C^j}$$

Equation 28 : indice de qualité de l'état  $j$  du plan d'expériences

## III. 3.2. Résultats

Les 3 plans d'expériences sont appliqués dans un premier temps aux solutions modèles présentées dans la partie III.1. Les résultats généralisables, issus de cette première étude, sont exploités pour l'élaboration des différentes conditions expérimentales testées lors de la mesure des moûts de raisin.

### III. 3.2.1. Application aux solutions modèles

L'influence de 3 paramètres expérimentaux est analysée grâce au calcul des effets sur les indices de performances (répétabilité et distance discriminante) en utilisant les descripteurs sélectionnés précédemment (C.f. pIII-66). Les 3 paramètres analysés grâce aux plans d'expériences sont : la température de la cellule de mesure et du gaz vecteur, celle de l'espace de tête et son mode d'injection.

### (1) Discrimination

Quelles que soient les conditions expérimentales, les échantillons de triacétine pure (n°1 à 9) sont correctement séparés des échantillons contenant les composés volatils (n°11 à 19) sur le 1<sup>er</sup> plan factoriel d'une ACP normée (C.f. Figure 19).

La première composante principale correspond à la distinction des 2 types de solutions modèles et le deuxième axe est surtout lié à la méthode d'injection. L'influence des 2 autres facteurs est beaucoup moins visible.

Les répétitions au centre (p01, p02 et p03) montrent que la répétabilité des mesures est bien meilleure que sa reproductibilité, malgré le retour précis aux mêmes conditions expérimentales ; néanmoins, aucune dérive linéaire n'est perceptible.

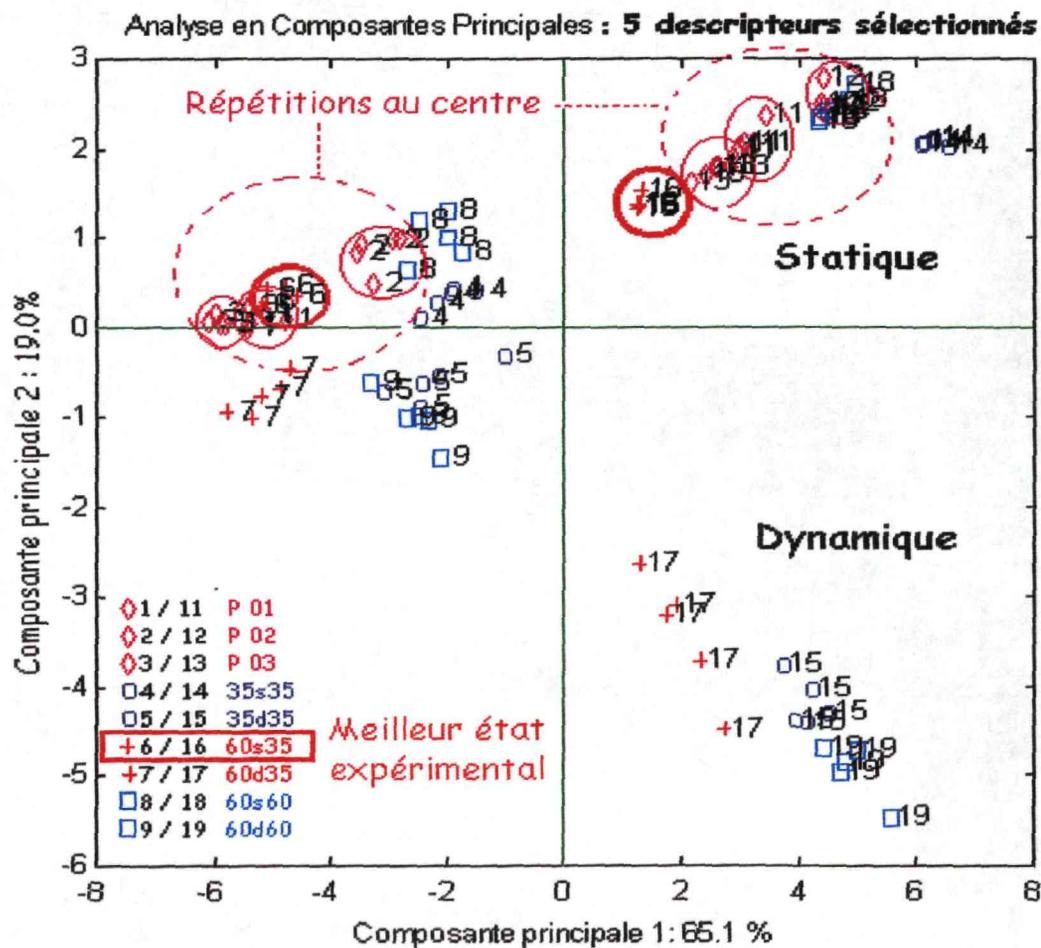


Figure 19 : ACP normée sur les 5 descripteurs sélectionnés

Les échantillons de solvant pur sont moins éparpillés que ceux contenant les molécules volatiles ; en effet, leur signal étant moins important, il est moins influencé par les fluctuations des conditions expérimentales.

## (2) Analyse du plan d'expériences

Les effets des 3 facteurs et de leurs interactions sont calculés conformément aux Equation 22 & Equation 23 (pIII-75). Ils sont détaillés dans le Tableau 12. L'influence des facteurs se manifeste principalement sur la répétabilité des mesures du système multi-capteur ; les 3 facteurs ont une influence significative (avec un risque de 10%).

Quelles que soient les conditions de température, le *mode d'injection* est beaucoup plus répétable en statique<sup>20</sup>. De même, la distance discriminante est significativement plus faible en mode d'injection dynamique, si l'espace de tête est généré à 35°C.

Les températures de la cellule de mesure et de l'espace de tête ont des effets opposés sur la répétabilité des mesures : celle-ci est améliorée si la température de la cellule de mesure est haute (60°C) (Tableau 12(a)) et si celle de l'espace de tête est basse (35°C) (Tableau 12(b)). Ceci explique que l'effet de la température globale ne soit pas significatif (Tableau 12(c)).

L'augmentation de la température de l'espace de tête devrait accroître la distance entre les 2 échantillons, grâce à amplification de l'intensité aromatique de l'échantillon contenant les molécules volatiles de défaut. Cependant, celle-ci est compensée par la détérioration de l'écart-type des répétitions à 60°C, qui participe au calcul de la distance discriminante (C.f. Equation 23, pIII-75).

**Tableau 12 : effets des facteurs et de leurs interactions sur les indices de performances.**

Les sous-tableaux (a), (b) et (c) analysent les 3 plans d'expériences à 2<sup>2</sup> expérimentations. L'intervalle de confiance est de 0.7% pour le coefficient de variation et de 7.4 pour la distance discriminante, avec un risque de 10% (1.0% et 11.0, respectivement pour un risque de 5%). Les cases noires correspondent aux effets significatifs.

(a) PLAN D'EXPERIENCES I (TEMPERATURE DE L'ESPACE DE TETE =35°C)	REPETABILITE	DISTANCE DISCRIMINANTE
Effet de la température de la cellule de mesure	-1.9 %	-1.1
Effet du mode d'injection	3.4 %	-11.1
Effet de leur interaction	0.6 %	0.7

(b) PLAN D'EXPERIENCES II (TEMPERATURE DE LA CELLULE DE MESURE =60°C)	REPETABILITE	DISTANCE DISCRIMINANTE
Effet de la température de l'espace de tête	2.3 %	-3.0
Effet du mode d'injection	2.1 %	-4.5
Effet de leur interaction	-1.9 %	5.9

(c) PLAN D'EXPERIENCES III (MEMES TEMPERATURES)	REPETABILITE	DISTANCE DISCRIMINANTE
Effet de la température globale	0.3 %	-4.1
Effet du mode d'injection	1.5 %	-5.2
Effet de leur interaction	-1.3 %	6.6

<sup>20</sup> L'effet du mode d'injection sur la répétabilité est positif : cela signifie que le passage du niveau bas (statique) au niveau haut (dynamique) du mode d'injection entraîne l'augmentation du coefficient de variation, donc altère la répétabilité.

### (3) Conditions expérimentales optimales

Les différents effets analysés précédemment conduisent à déterminer les conditions expérimentales optimales : l'injection en mode statique de l'espace de tête généré à 35°C dans la cellule de mesure à 60°C. Le Tableau 13 montre les indices de performance en fonction des états du plan d'expériences.

**Tableau 13 : indices de performance des états du plan d'expériences**

INDICES DE PERFORMANCE	Moyenne des points au centre	35S35	35D35	60S35	60D35	60S60	60D60
D <sub>i</sub>	41.5	45.2	21.7	41.5	20.8	23.8	26.6
C <sub>i</sub>	5%	8%	14%	3%	11%	12%	12%
q <sub>i</sub>	801.5	532.1	152.9	1 216.1	181.0	201.7	219.6

Les performances du mode d'injection sont indépendantes de la nature de l'échantillon ; cette propriété peut donc être généralisée à tout échantillon analysé à l'aide de cet appareillage. Ainsi, lors des expérimentations suivantes, nous n'emploierons que le mode d'injection statique à l'aide de la seringue à gaz, car celui-ci présente la meilleure répétabilité.

Quant aux conditions de température, elles semblent liées à l'échantillon, tout au moins en ce qui concerne la température de génération de l'espace de tête. En effet, même si la répétabilité est meilleure à 35°C, ce résultat peut être corrélé à la volatilité des molécules volatiles présentes dans cet échantillon ; en outre, il peut être nécessaire d'augmenter la température afin d'exacerber les arômes lors de discriminations délicates. Alors, la température de la cellule de mesure doit être fixée en conséquence, afin d'éviter toute condensation.

#### III. 3.2.2. Application aux moûts de raisin

Ainsi, seuls 3 états expérimentaux seront analysés lors de la mesure des moûts de raisin, car le mode d'injection est désormais fixé au *mode statique* ; les états sont caractérisés par les températures de la cellule de mesure et de l'espace de tête variant entre 35 et 60°C, en évitant la génération d'un échantillon gazeux plus chaud (60°C) que la cellule de mesure (35°C).

La discrimination des 4 groupes de répétitions n'est pas aussi aisée que pour les solutions modèles ; la Figure 20 montre que la différence induite par le changement de conditions expérimentales est supérieure à celle des 4 groupes d'échantillons. Il est donc extrêmement important de contrôler précisément les températures de l'espace de tête et de la cellule de mesure.

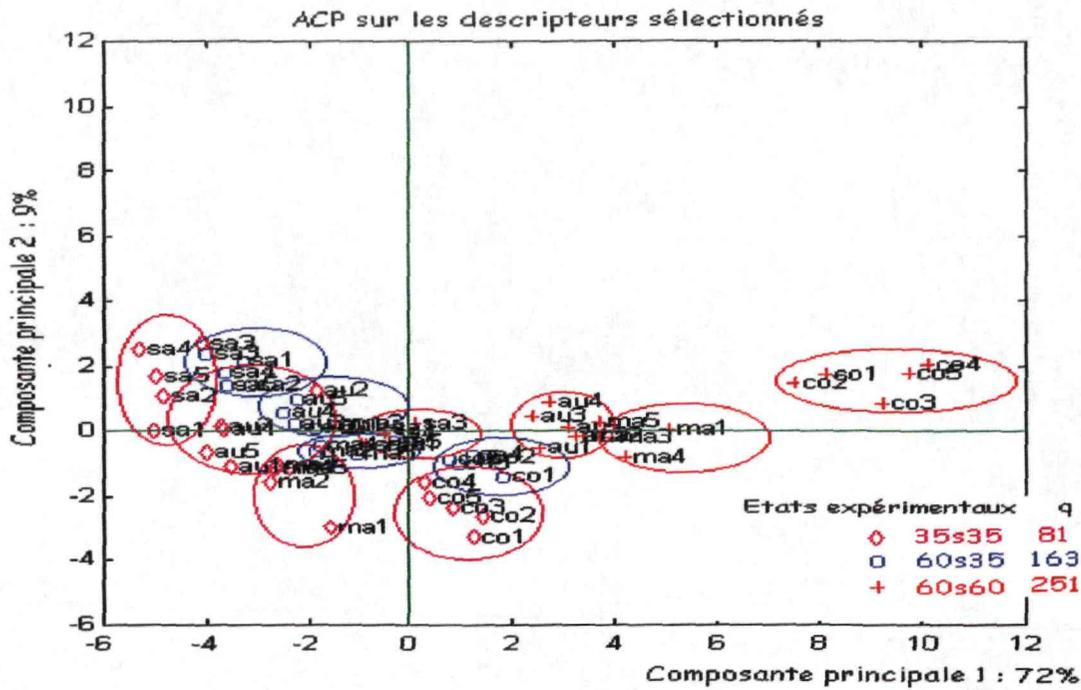


Figure 20 : discrimination des moûts de raisin dans les 3 conditions expérimentales

Les indices de performance du Tableau 14 conduisent à déterminer les conditions expérimentales optimales : l'injection en mode statique de l'espace de tête généré à 60°C dans la cellule de mesure à 60°C.

Tableau 14 : indices de performance pour les 3 états expérimentaux

NOM DES ETATS : cellule de mesure / mode d'injection / espace de tête			
INDICES DE PERFORMANCE	35S35	60S35	60S60
D <sub>i</sub>	2.3	3.8	3.9
C <sub>i</sub>	2.8%	2.4%	1.6%
q <sub>i</sub>	81	163	251

Les performances de la discrimination sont nettement moins bonnes quand la température de la cellule de mesure est basse (35S35), comme c'était le cas pour les solutions modèles. Par contre, quand la température de la cellule de mesure est de 60°C, la distance discriminante ne varie pas significativement en fonction la température de l'espace de tête.

Cependant, la répétabilité des mesures est meilleure quand l'espace de tête est généré à 60°C, contrairement aux solutions modèles. Les arômes présents dans les moûts de raisin étant moins volatils que l'acétate d'éthyle, une température élevée (60°C) permet aux échantillons d'être plus proches de l'équilibre liquide/gaz après 16 minutes, donc plus répétables.

Les mesures suivantes seront donc effectuées dans les conditions expérimentales optimales, c'est-à-dire que l'espace de tête des échantillons sera généré à 60°C et injecté en mode statique dans la cellule de mesure à 60°C.

### III. 3.3. Conclusion

La démarche des *plans d'expériences* a permis de quantifier l'influence de paramètres expérimentaux sur la performance d'un système multi-capteur d'arômes, caractérisée par sa *répétabilité* et *distance discriminante*. Celle-ci a été appliquée à 3 facteurs extérieurs : la *température de la cellule de mesure* et du gaz vecteur, la *température* et le *mode d'injection de l'espace de tête*.

L'analyse des plans d'expériences a montré l'influence significative de ces 3 facteurs, plus particulièrement sur la répétabilité.

❶ *Certains paramètres ne sont pas liés au type d'échantillon, mais uniquement au montage expérimental* ; leur influence peut donc être évaluée lors d'une seule série de mesures et généralisée à l'ensemble des discriminations ultérieures avec le même appareillage. C'est le cas du *mode d'injection*, qui s'est avéré plus répétable en statique, quelles que soient les conditions de température.

Dans les 2 applications étudiées précédemment, la *température de la cellule de mesure* s'est révélée plus efficace à 60°C. Donc, celle-ci ne semble pas directement liée à la qualité des composés volatils analysés ; elle agit sur la dispersion des molécules volatiles au sein de la cellule de mesure, en particulier au contact des capteurs. Néanmoins, cette température est en partie conditionnée par la celle de l'espace de tête, car elle doit lui être supérieure.

Les paramètres extérieurs dont l'effet est généralisable à toute application sont le mode d'injection, le débit de gaz vecteur, la température du gaz vecteur et de la cellule de mesure, la vitesse d'injection de l'échantillon, etc.

❷ *Certains paramètres sont directement conditionnés par l'échantillon analysé*. La qualité de l'espace de tête en fonction de sa température est directement corrélée à l'échantillon ; en effet, la volatilité des composés ainsi que la nature de la matrice de l'échantillon conditionnent la cinétique d'équilibre de l'espace de tête. Ainsi, la répétabilité de l'échantillon gazeux issu des solutions modèles était meilleure à basse température (35°C), alors que celle des moûts est meilleure à haute température (60°C). Ce facteur doit donc être adapté à chaque discrimination.

En conclusion, les paramètres extérieurs qu'il faut optimiser à chaque application sont la durée de génération de l'échantillon, sa température, le volume d'espace de tête injecté, la durée d'un cycle de mesure, l'hygrométrie du gaz vecteur, la température des capteurs.

### III. 4. SELECTION DES CAPTEURS

La revue bibliographique révèle la pauvreté des travaux concernant la sélection des éléments sensibles à inclure dans le système multi-capteur d'arômes (C.f. pII-39) ; paradoxalement, les méthodes de sélection des variables sont extrêmement développées (C.f. pII-40), et peuvent être appliquées indirectement à la sélection des capteurs.

Parmi les méthodes usuelles de sélection de variables, nous avons choisi de développer deux techniques, faisant appel à des stratégies différentes : une méthode paramétrique supervisée pas à pas (*Analyse Factorielle Discriminante pas à pas*) et une méthode évolutionniste (*algorithmes génétiques*).

#### III. 4.1. Méthodes

##### III. 4.1.1. Capteurs analysés

La cellule de mesure du LCA 1000 modifié ne pouvant contenir que 5 capteurs, nous appliquerons ces méthodologies à la sélection de 5 capteurs parmi 13 capteurs MOS de la série TGS 8xx. Ainsi, les mesures seront effectuées en 3 séries consécutives, en conservant un capteur commun aux 3 séries, afin de servir de référence en cas de dérive. Cette démarche est étudiée lors de deux applications différentes, pour lesquelles les capteurs testés sont décrits dans les Tableau 15 & Tableau 16.

Tableau 15 : Liste des capteurs analysés pour les solutions modèles

N° DE SERIE	CAPTEUR 1	CAPTEUR 2	CAPTEUR 3	CAPTEUR 4	CAPTEUR 5
1	TGS 825	TGS 800	TGS 824	TGS 842	TGS 822
2	TGS 812	TGS 815	TGS 882	TGS 880	TGS 822
3	TGS 830	TGS 831	TGS 813	TGS 4C23-3	TGS 822

Tableau 16 : Liste des capteurs analysés pour les moûts de raisins blancs

N° DE SERIE	CAPTEUR 1	CAPTEUR 2	CAPTEUR 3	CAPTEUR 4	CAPTEUR 5
1	TGS 815	TGS 882	TGS 830	TGS 822	TGS 813
2	TGS 800	TGS 812	TGS 830	TGS 821	TGS 824
3	TGS 825	TGS 831	TGS 830	TGS 842	TGS 882

Les conditions expérimentales des mesures et les descripteurs extraits des courbes sont ceux déterminés lors des 2 premières phases d'optimisation ; le choix des capteurs est donc effectuée indirectement, grâce à la sélection des descripteurs qui les caractérisent.

### III. 4.1.2. Analyse Factorielle Discriminante Pas à Pas

L'Analyse Factorielle Discriminante est décrite pII-37 et le principe des méthodes de sélection pas à pas d'un sous-ensemble de  $q$  variables parmi  $P$  est présenté pII-42. Nous emploierons les processus d'inclusion/exclusion d'une variable à chaque itération de l'algorithme.

Deux critères de sélection sont bâtis à partir de l'AFD (ROMEDER, 1973) :

❶ Le Lambda de Wilks à minimiser :  $\Lambda = \frac{|W|}{|T|} = \frac{\text{déterminant (variance intraclasse)}}{\text{déterminant (variance totale)}}$  ; il ne fournit pas de

critère d'arrêt naturel, donc ordonne toutes les variables en fonction de leur aptitude à discriminer.

❷ Le test de Fisher estimant l'homogénéité des moyennes conditionnelles des différentes classes, constituées à l'aide les  $q$  variables sélectionnées et les  $(P-q)$  variables restantes. Il faut donc maximiser la différence entre les moyennes conditionnelles des groupes ; le meilleur sous-ensemble de variables est atteint quand aucune variable ne peut améliorer significativement la discrimination.

### III. 4.1.3. Algorithmes génétiques

La méthode de sélection de variables par algorithmes génétiques est détaillée pII-42. Cet algorithme, programmé sous MATLAB™ version 5.0 (Mathworks, Inc., USA), est l'adaptation d'un logiciel issu de la PLS\_TOOLBOX. Il est appliqué dans ce chapitre à la sélection de capteurs d'arômes et sera également employé lors de la sélection de longueurs d'onde de spectres infrarouge et ultraviolet dans la partie IV.

Les paramètres de fonctionnement des algorithmes génétiques sont les suivants :

(1) la population initiale est constituée par 96 chromosomes, contenant chacun les  $p=5$  descripteurs caractéristiques des 15 capteurs à sélectionner, soient  $p*15$  gènes ;

(2) 10% des gènes sont présents dans un chromosome, c'est-à-dire que 10% des descripteurs participent à l'AFD ; cependant, étant donné que la population initiale est générée au hasard, le pourcentage de capteurs présents peut atteindre  $p$  descripteurs/capteur \* 10%, soit 50%. Néanmoins, au cours des générations, la sélection se restreint aux capteurs les plus discriminants, notre objectif étant de retenir 5 capteurs parmi les 15.

(3) le critère d'arrêt est l'obtention de plus de 40% de chromosomes identiques ou l'atteinte de 100 générations ;

(4) le croisement (*cross-over*) est double (i.e. 2 croisements par couple de chromosomes), et est appliqué à la moitié de la population à chaque génération, afin d'explorer le plus largement l'espace des individus ; ce croisement est réduit à un croisement simple après une trentaine de générations afin de converger plus facilement.

(5) le taux de mutation est abaissé de 1% à 0.1% après une trentaine de générations de la même manière que le croisement ; ainsi, il contribue à la diversité au départ puis facilite la convergence de la population à la fin.

(6) La fonction d'adéquation employée est l'AFD. Le critère d'adéquation à minimiser est le pourcentage d'erreur de prédiction en validation croisée sur 5 blocs. Les échantillons sont sélectionnés par la méthode des « *venetian blinds* », c'est-à-dire par le prélèvement d'1 échantillon tous les 5 successivement à une réorganisation aléatoire.

Après chaque exécution de l'algorithme, une population contenant les meilleurs chromosomes environ est déterminée ; les capteurs sélectionnés sont ceux pour lesquels les descripteurs sont présents majoritairement.

Cet algorithme, non déterministe, est exécuté une dizaine de fois. Le choix final des capteurs est obtenu grâce à l'élection des 5 capteurs sélectionnés le plus souvent au cours de toutes les exécutions de l'algorithme.

### III. 4.2. Résultats

Seules les méthodes pas à pas seront appliquées à la sélection des capteurs pour la discrimination des solutions modèles, alors que toutes les méthodes exposées précédemment seront exploitées lors du choix des capteurs pour la mesure des goûts de raisin.

#### III. 4.2.1. Application aux solutions modèles

la sélection est réalisée sur les descripteurs suivants : le maximum d'adsorption (n°7), la pente moyenne calculée entre 24 et 48 secondes (n°12), la pente maximale de l'adsorption (n°14), la pente minimale de la désorption (n°20) et le temps correspondant à l'apparition de la décélération maximum de la désorption (n°27).

Le Tableau 17 montre le résultat de la sélection des variables par le critère du Lambda de Wilks et du test de Fisher.

Tableau 17 : résultats de la sélection des capteurs par AFD pas à pas

METHODES	Λ WILKS		FISHER	
	Capteurs	Descripteurs	Capteurs	Descripteurs
Capteur 1	TGS 815	n°20	TGS 815	n°20
Capteur 2	TGS 825	n°27	TGS 825	n°27
Capteur 3	TGS 822	n°12, 14		
Capteur 4	TGS 830	n°14		
Capteur 5	TGS 800	n°7, 27		

La méthode basée sur le  $\Lambda$  de Wilks n'est pas munie d'un critère d'arrêt, les 5 premiers capteurs sont donc sélectionnés. Au contraire, la méthode du test de Fisher sélectionne seulement 2 capteurs, qui suffisent à obtenir la meilleure discrimination.

Cette discrimination est très facile : la Figure 21 montre la dispersion des 32 échantillons en fonction de 2 descripteurs extraits de la courbe du TGS 815. Néanmoins, lors de l'application des méthodes de sélection des capteurs à la classification variétale des moûts de raisin, nous verrons que des discriminations plus complexes nécessitent de plus nombreux capteurs.

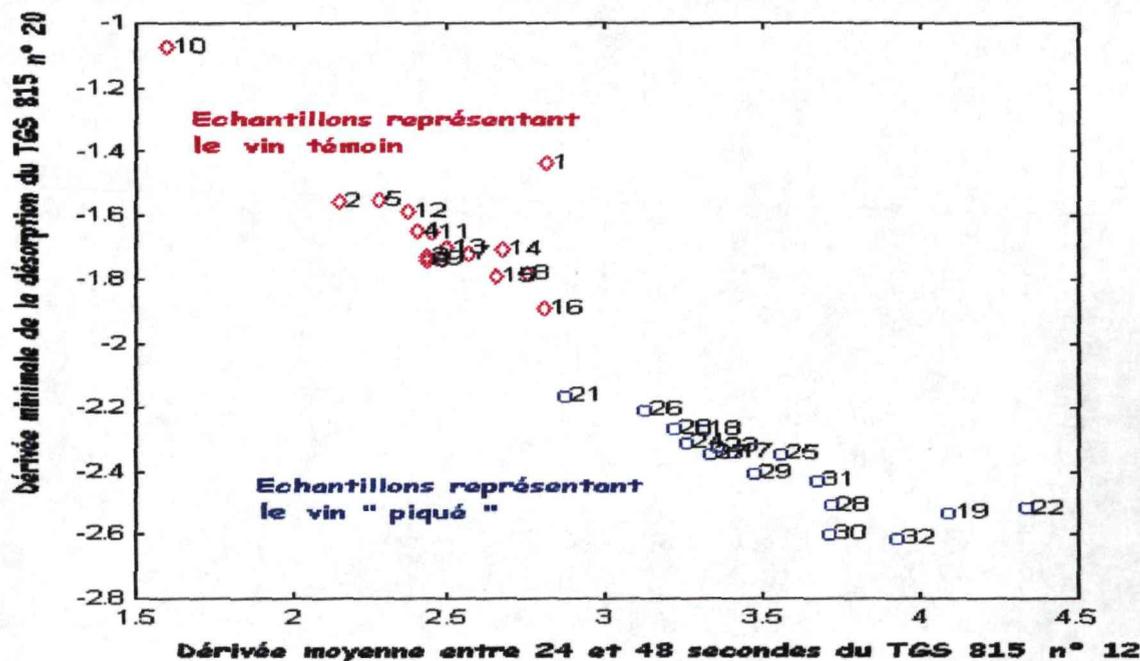


Figure 21 : description des solutions modèles par 2 descripteurs du TGS 815

Le capteur TGS 822 a participé aux 3 séries de mesure afin d'estimer une éventuelle dérive due aux fluctuations des conditions expérimentales, qui pourraient désavantager une série de capteurs par rapport à une autre lors de la sélection. Or, les 2 capteurs sélectionnés par le test de Fisher appartiennent à 2 séries différentes et le  $\Lambda$  de Wilks sélectionne des éléments sensibles de toutes séries. En outre, si les descripteurs issus de la mesure du TGS 822 de la série sélectionnée sont exclus, alors, celui-ci est sélectionné grâce à une autre série. Tout ceci tend à démontrer qu'aucune série de capteurs n'est désavantagée et que la dérive est suffisamment faible pour être négligée.

Chaque sous-ensemble de capteurs sélectionné fournit 100% de bons classements. Compte tenu du principe de parcimonie (SEASHOLTZ et KOWALSKI, 1993), il serait préférable de sélectionner le plus petit nombre de capteurs possible ; néanmoins, afin d'envisager des applications plus complexes, nous pouvons conserver les 5 premiers capteurs sélectionnés, afin d'être capable de couvrir l'intégralité de la variabilité de chaque classe.

### III. 4.2.2. Application aux moûts de raisin

Les descripteurs sur lesquels se basent cette sélection ont été sélectionnés pIII-67 et les 40 échantillons mesurés sont détaillés en Annexe 4. Il s'agit du *maximum d'adsorption* (n°7), de la *pen­te moyenne calculée entre 12 et 24 secondes* (n°10), de la *pen­te maximale de l'adsorption* (n°14), de la *pen­te minimale de la désorption* (n°20) et de la *décélération maximale de l'adsorption* (n°24).

#### (1) Analyse Discriminante Pas à Pas

Le Tableau 18 montre la sélection des capteurs par les critères du  $\Lambda$  de Wilks et du test de Fisher.

**Tableau 18 : résultats de la sélection des capteurs par AFD pas à pas**

METHODES	$\Lambda$ WILKS	FISHER	
Capteurs	Descripteurs	Capteurs	Descripteurs
TGS 882	n°14	TGS 882	n°14
TGS 821	n°24	TGS 815	n°24
TGS 815	n°20, 24		
TGS 822	n°20		
TGS 800	n°14		

De la même façon que précédemment, le test de Fisher détermine que 2 capteurs seulement suffisent à fournir la meilleure discrimination (C.f. Tableau 18) ; le petit nombre de capteurs sélectionné se justifie par la forte redondance des éléments sensibles TGS 8xx, dont l'information est amplifiée grâce au calcul de descripteurs faiblement corrélés. Si le maximum d'adsorption est le seul paramètre extrait des courbes, alors seul le TGS 882 est sélectionné par le test de Fisher.

#### (2) Algorithmes génétiques

La méthode des algorithmes génétiques est appliquée à la sélection de capteurs, en employant les paramètres définis dans le chapitre (pIII-83). Pour chaque exécution d'un algorithme génétique, les capteurs choisis sont ceux dont un des descripteurs est sélectionné majoritairement.

La Figure 22 montre l'exemple de la première exécution de l'algorithme génétique (AG1 du Tableau 19), grâce auquel les *capteurs TGS 812, 824, 825, 831, 842 et 882* ont été retenus.

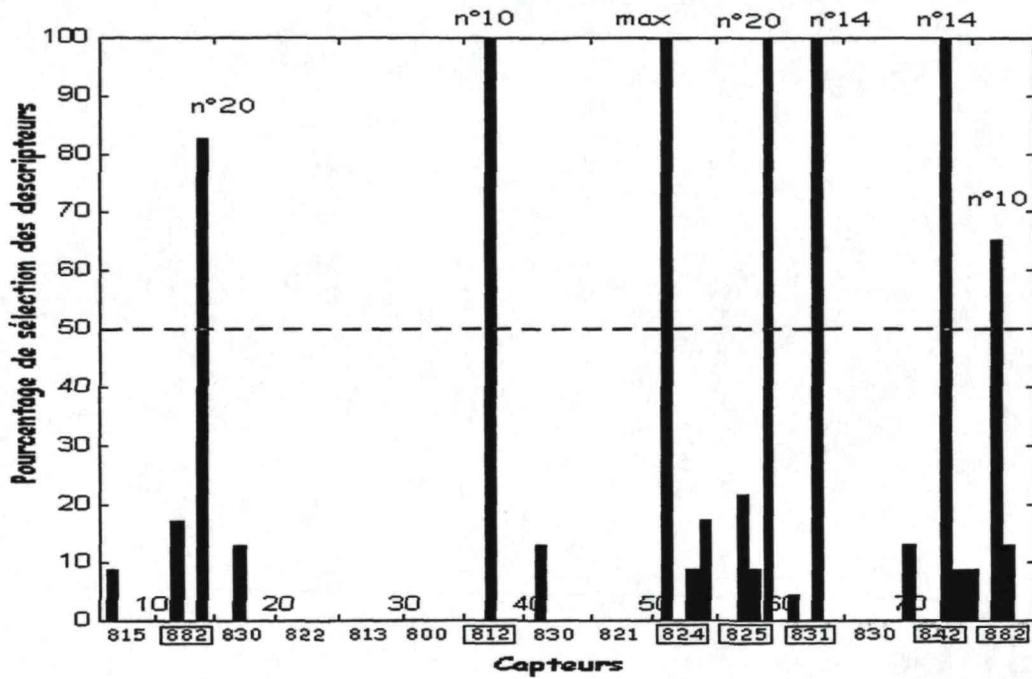


Figure 22 : exemple de sélection des capteurs par algorithmes génétiques

Sept itérations ont permis de sélectionner les 5 capteurs majoritairement choisis ; il s'agit des TGS 812, 824, 825, 830 et 882 (C.f. Tableau 19).

Tableau 19 : sélection des capteurs par algorithmes génétiques.

CAPTEURS	AG1	AG2	AG3	AG4	AG5	AG6	AG7	TOTAL
800				n°10			n°10	2
812	n°10	n°10	n°10		n°14		n°14	5
813				n°10	n°14	n°10		3
815								0
821							n°24	1
822			n°14	n°10		n°10		3
824	max.	n°10	max., n°10	max.	max., n°20	max.	max.	9
825	n°10	n°10,20	n°20	max., n°10,20	n°20	n°10,20	n°20	11
830			n°24	n°10,24	n°24	max., 24	n°24	7
831	n°14							1
842	n°14	n°14		max.			n°14	4
882	n°10,20	n°14	n°24	n°24	n°24		n°14	7
<b>Nombre de capteurs sélectionnés</b>	<b>6</b>	<b>5</b>	<b>6</b>	<b>8</b>	<b>6</b>	<b>5</b>	<b>8</b>	
<b>Meilleur critère d'adéquation</b>	<b>35%</b>	<b>27.5%</b>	<b>30%</b>	<b>25%</b>	<b>27.5%</b>	<b>25%</b>	<b>22.5%</b>	

**(3) Conclusion**

Dans le Tableau 20 sont comparées les performances des sous-ensembles de capteurs, sélectionnés grâce aux divers algorithmes de sélection de variables ; les capteurs choisis sont différents, à l'exception du TGS 882 qui est constamment sélectionné.

La sélection par la méthode des algorithmes génétiques fournit les meilleurs résultats en validation croisée par AFD. La classification est meilleure si seuls les descripteurs sélectionnés participent au modèle ; outre la sélection des capteurs, le choix des descripteurs est donc extrêmement important.

**Tableau 20 : tableau comparatif des performances des 4 sélections de capteurs**

(C) signifie que la classification est basée sur l'ensemble des 25 descripteurs extraits des 5 Capteurs et (D) signifie qu'elle est effectuée sur les seuls *Descripteurs* sélectionnés des 5 capteurs choisis.

METHODES DE SELECTION	$\Lambda$ DE WILKS	TEST DE FISHER	ALGORITHMES GENETIQUES
Capteur 1	TGS 800	TGS 815	TGS 812
Capteur 2	TGS 815	TGS 882	TGS 824
Capteur 3	TGS 821		TGS 825
Capteur 4	TGS 822		TGS 830
Capteur 5	TGS 882		TGS 882
erreur de prédiction AFD (C)	60%	60%	62.5%
erreur de prédiction AFD (D)	50%	60%	37.5%

La meilleure méthode de sélection de capteurs est celle des *algorithmes génétiques*, avec 15 échantillons mal classés sur 40 avec les 5 capteurs. Il s'agit des capteurs *TGS 812, 824, 825, 830 et 882*.

Les descripteurs choisis influencent la qualité de la discrimination. Néanmoins, dans les expérimentations suivantes réalisées sur tous les échantillons de moûts de raisins blancs, nous emploierons les 5 descripteurs issus des 5 capteurs sélectionnés, afin de pouvoir caractériser au mieux la variabilité des classes.

### III. 5. CONCLUSIONS

Au cours de ce chapitre, nous avons proposé et appliqué 3 méthodologies visant à améliorer les performances des capteurs d'arômes. Ces 3 méthodes successives font partie de la phase d'optimisation préalable à la mesure de tout échantillon, quel que soit l'appareillage ou les éléments sensibles utilisés.

❶ Une démarche de *calcul et de sélection de descripteurs de forme de la courbe d'adsorption et de désorption* a été conçue, s'appuyant sur des critères de *répétabilité*, de *discrimination* et de *redondance*. Ces *indices de performance* sont estimés grâce aux répétitions de la mesure d'un échantillon représentatif par classe. La mise en œuvre de cette méthode, pour la discrimination de solutions modèles très simplifiées de vins, a permis d'améliorer les résultats par rapport à une

classification uniquement basée sur le maximum d'adsorption ; de plus, celle-ci a été appliquée et validée sur un cas réel : la discrimination variétale de moûts de raisins blancs.

Il est important de souligner que l'analyse de la pertinence des descripteurs est réalisée de manière globale pour le système multi-capteur. Lors des mesures suivantes, il pourra être préférable de choisir les descripteurs pour chaque capteur parmi ce sous-ensemble de descripteurs sélectionnés ; ceci peut être accompli par les méthodes de sélection de variables, développées dans la troisième partie de ce chapitre.

☒ Cette méthodologie peut être généralisée à tout système multi-capteur d'arômes, dont le signal est composé de diverses courbes. Elle ne nécessite pas *la connaissance a priori de la forme des courbes* et peut donc s'appliquer à tout type d'élément sensible et tout échantillon ; la seule différence résidera dans les descripteurs choisis. De plus, le nombre d'échantillons nécessaires à la mise en œuvre de cette démarche est relativement restreint : 5 répétitions de la mesure d'un échantillon représentatif de chaque classe. *Le nombre de descripteurs sélectionné est réduit*, ce qui permet l'emploi ultérieur de traitements statistiques multivariés, basés sur un nombre raisonnable de mesures.

② Afin d'améliorer les conditions expérimentales de la mesure par capteurs d'arômes, la *méthodologie des plans d'expériences* a été appliquée à 3 paramètres expérimentaux : la *température de la cellule de mesure* et du gaz vecteur, la *température* et le *mode d'injection de l'espace de tête*. L'influence de ces facteurs sur des critères complexes a été analysée : la *répétabilité* et la *distance discriminante* du système multi-capteur d'arômes.

L'influence de ces facteurs s'est avérée significative, en particulier sur la répétabilité, quelle que soit l'application. L'analyse de l'effet des facteurs a mené à la détermination d'un *l'état expérimental optimal* différent pour les 2 discriminations : le *mode d'injection statique* est indéniablement le meilleur, par transfert manuel de l'espace de tête grâce à une seringue à gaz ; de même, la *température de la cellule de mesure* est plus performante à 60°C. Par contre, les échantillons étant différents, ils induisent des conditions optimales d'équilibre différentes : les solutions modèles sont plus répétables à 35°C alors que les moûts de raisin le sont plus à 60°C.

☒ Ces conclusions conduisent à la séparation des paramètres expérimentaux en 2 catégories : les *conditions extérieures exclusivement liées au montage expérimental* et au protocole de mesure et les *conditions inhérentes aux échantillons analysés*, et donc caractéristiques de l'application considérée. Les premiers peuvent être optimisés lors de la première série de mesures et généralisés à toute discrimination ultérieure. Les seconds doivent être adaptés à chaque application et donc remis en cause au sein d'un nouveau plan d'expériences.

☒ Ce type de démarche peut être adapté à tout appareillage contenant des capteurs d'arômes, grâce à une analyse exhaustive préalable de tous les facteurs contrôlables pouvant influencer la mesure. Il convient ensuite de les répartir dans les 2 catégories précitées et de déterminer les facteurs les plus

influent sur la discrimination. Enfin, un plan d'expériences complet ou fractionnaire à 2 niveaux doit être mis en œuvre, afin de quantifier les effets des conditions expérimentales sur les critères de performance. Si l'effet de certains facteurs est non monotone, un plan d'expériences à surface de réponse doit être réalisé, afin de déterminer les conditions expérimentales optimales.

③ Des *méthodes de sélection de variables* ont été proposées, afin de choisir indirectement les éléments sensibles à inclure dans le système multi-capteur d'arômes. Ces démarches, qui n'ont jamais été employées pour la sélection de capteurs d'arômes, se sont avérées très efficaces. Une méthode multivariée pas à pas paramétriques (AFD) a été appliquée à des solutions modèles et au cas réel des moûts de raisin ; la méthode stochastique des algorithmes génétiques a été mise en œuvre seulement lors de la classification des moûts.

La discrimination des solutions modèles étant extrêmement facile ; un seul capteur suffit à obtenir 100% de bons classements, ce qui rend superflu l'emploi de ces méthodologies. Mais, l'application de ces méthodes à la discrimination de produits naturels complexes, tels que des moûts de raisins, est extrêmement bénéfique ; en adoptant le sous-ensemble des  $n$  capteurs les plus performants et les moins redondants, la capacité à discriminer du système multi-capteur d'arômes est amplifiée. La démarche de sélection des capteurs basée sur les algorithmes génétiques s'est avérée la plus efficace.

⊗ La sélection des éléments sensibles est extrêmement importante, quel que soit l'appareillage et la technologie d'éléments sensibles employés. Celle-ci est très rarement effectuée, que ce soit dans une phase d'optimisation préalable à la mesure de tous les échantillons, ou après la mesure effectuée par un système comprenant de nombreux capteurs. Les méthodes développées précédemment peuvent s'appliquer à tout appareillage, consécutivement à l'extraction de descripteurs. Malgré le non déterminisme des algorithmes génétiques, cette technique est très performante et peut être adaptée à des problématiques de quantification (et non de classification), grâce à un simple changement de la fonction d'adéquation.

Après l'optimisation du système d'un point de vue du traitement du signal, du montage expérimental et des capteurs choisis, les mesures de tous les échantillons peuvent être réalisées grâce au LCA 1000 modifié, en vue de la discrimination variétale aromatique des moûts de raisins blancs.

# *Système multi-capteur appliqué à la discrimination variétale des moûts*

- 1. Acquisition des données*
- 2. Classification par capteurs individuels*
- 3. Fusion multi-capteur*

## IV. SYSTEME MULTI-CAPTEUR APPLIQUE A LA DISCRIMINATION VARIETALE DE MOUTS

L'objectif de ce chapitre est de traiter une problématique de classification et de fusion multisensorielle à partir de variables très corrélées extraites de courbes ou de spectres. Le support de cette étude est la classification variétale des moûts de raisins blancs. Afin d'établir la discrimination sur les composés aromatiques volatils et sur les précurseurs d'arômes, nous avons combiné plusieurs types de mesures.

La mesure aromatique est effectuée grâce au LCA1000 modifié (pIV-92), dans des conditions optimales définies dans le chapitre précédent. Celle-ci est complétée par des mesures spectrométriques effectuées en phase liquide dans le moyen infrarouge (IR-TF) (pIV-92) et l'ultraviolet (UV) (pIV-94) ; elles seront comparées aux performances d'un jury sensoriel (pIV-95) ainsi qu'à des analyses chimiques (pIV-94). L'acquisition de ces données est présentée dans une première partie.

Dans un second temps, les classifications individuelles de chaque capteur sont comparées (pIV-96) ; en particulier, une méthode de régression multivariée (*PLS*) est appliquée à la discrimination. Diverses méthodes de sélection des longueurs d'onde par algorithmes génétiques sont proposées (pIV-98) et appliquées aux spectres IR-TF (pIV-101) et UV (pIV-104).

Dans une troisième partie, une méthode de fusion de niveau bas est appliquée afin de combiner les signaux bruts ou filtrés issus de tous les capteurs (pIV-111).

Enfin, les classifications individuelles sont fusionnées (fusion de niveau haut : pIV-116) afin d'obtenir un résultat commun plus performant, grâce à des techniques de fusion bayésienne, présentées pII-46.

Finalement, la pertinence des différents types de fusion est comparée et analysée (pIV-127).

### IV. 1. ACQUISITION DES DONNEES

#### IV. 1.1. Echantillons

Les échantillons analysés sont des moûts de raisins blancs. Ils se répartissent en 4 classes de cépage : le Sauvignon, le Mauzac, le Colombard et d'autres cépages blancs (Aranel, Chardonnay, Loin de l'œil, Muscadelle, Perdea, Pinot blanc, Riesling, Sémillon, Vermentino, Vionier) qui forment la quatrième classe (C.f. Annexe 3). Elles seront notées : « Sa », « Ma », « Co » et « Au ».

Au total, 138 lots de 200 grains de raisins ont été cueillis à maturité dans différentes parcelles du Tarn (Gaillac - Rabastens), Tarn et Garonne (S' Sardos), du Gers, du Lot (Cahors), de l'Aude (Narbonne) et du Gard (Nîmes), afin d'obtenir des échantillons créés dans des conditions homogènes et variées.

Les moûts sont générés de façon à exacerber les arômes, selon le protocole décrit dans pIII-58. Certains échantillons ne sont pas mesurés par tous les capteurs en raison de la faible quantité de moût (C.f. Annexe 3). Mais, en moyenne, les classes sont constituées d'une cinquantaine de Sauvignon, une vingtaine de Mauzac et de Colombard et une quarantaine d'autres cépages blancs.

#### IV. 1.2. Analyses chimiques

La maturité est estimée par la mesure *du taux de sucre, de l'acidité totale et du pH* des raisins pressés. Le taux de sucre est mesuré grâce à un réfractomètre manuel, l'acidité totale par un dosage acide/base et le pH par un pH-mètre électronique. Ceux-ci sont détaillés pour les 138 échantillons en Annexe 3.

#### IV. 1.3. LCA 1000 modifié

Les mesures aromatiques sont obtenues à l'aide de l'appareillage LCA modifié (C.f. pIII-55) ; elles sont effectuées dans les conditions optimales de discrimination décrites ci-dessous, déterminées lors des phases d'optimisation (C.f. pIII-80).

La température de la cellule de mesure et de l'air vecteur est fixée à 60°C (+/-0.5°C). L'espace de tête est généré à 35°C, de façon statique durant 16 minutes, en présence d'agitation magnétique de 10ml de moût dans un flacon de 100ml. 50ml de cet espace de tête est injecté à l'aide d'une seringue à gaz à l'intérieur de la cellule de mesure de 500ml.

L'acquisition de l'adsorption dure 2 minutes, la cellule de mesure étant scellée hermétiquement ; ensuite, l'air vecteur filtré et asséché nettoie la cellule de mesure à un débit de 500ml.min<sup>-1</sup> durant 10 min. d'acquisition supplémentaire. La période d'acquisition étant de 1 seconde, la courbe obtenue pour chaque capteur contient 721 points (C.f. Figure 5, pII-32).

Les échantillons sont analysés sur 4 jours en alternant les classes de cépage afin d'éviter tout effet bloc qui coïnciderait avec les classes.

#### IV. 1.4. Spectromètre IR-TF

La zone du moyen infrarouge est définie par des longueurs d'onde ( $\lambda$ ) comprises entre 2.5 et 25 $\mu$ m, nombres d'onde ( $\nu=1/\lambda$ ) compris entre 400 et 4000 cm<sup>-1</sup> (C.f. Figure 23).

Le spectromètre *infrarouge à Transformée de Fourier* (IR-TF) est couramment employé pour des produits agro-alimentaires, par exemple pour des mesures des sucres (BELLON, 1993 & CADET et OFFMANN, 1997 & DEFENEZ et WILSON, 1995 & KEMSLEY et al., 1993), le suivi de fermentation (DAVENEL et al., 1991), l'adultération des huiles (DUPUY et al., 1996 & LAI et al., 1994 & VAN DE VOORT et al., 1995), etc.

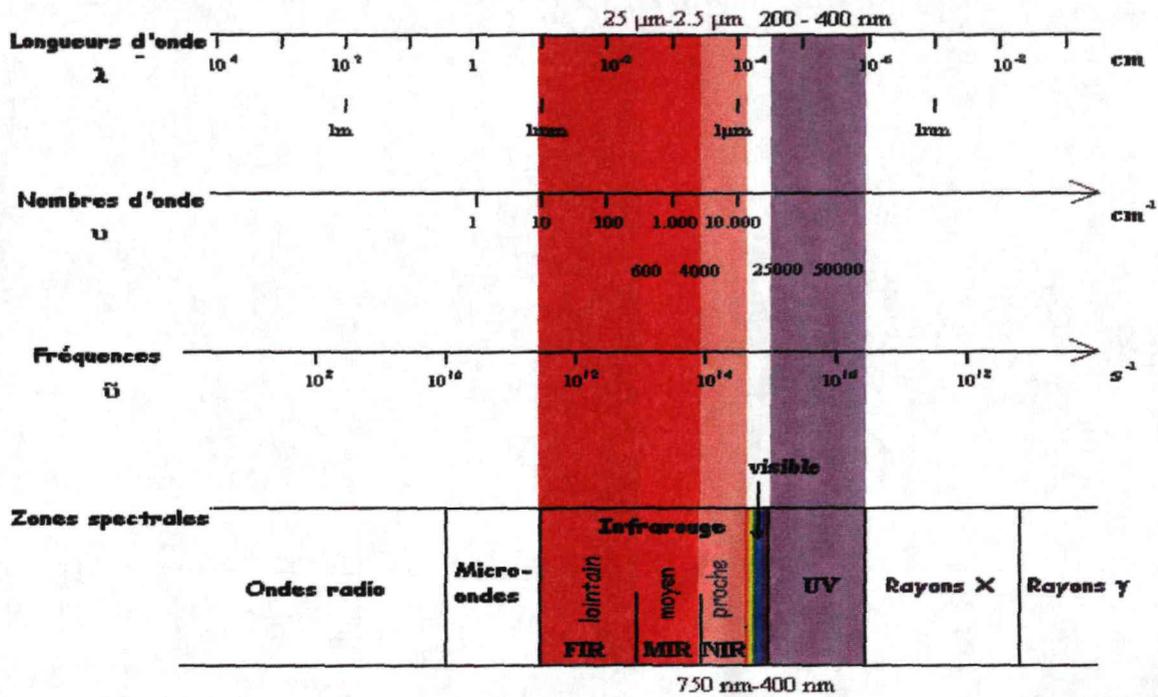


Figure 23 : Spectre électromagnétique

(tiré de : Fourier Transform Infrared Spectroscopy at Aberystwyth (the Univ. of Wales) : THE ELECTROMAGNETIC SPECTRUM : <http://gepasi.dbs.aber.ac.uk/roy/ftir/emspec.htm>)

Le spectromètre moyen IR-TF employé (Bruker IFS 25) contient une source de type Globar, dont le domaine spectral couvre de 6000 à 100  $\text{cm}^{-1}$  et un détecteur MCT. La bande spectrale utilisée dans notre application se limite à 4000-800  $\text{cm}^{-1}$ ; le pas d'acquisition est de 4  $\text{cm}^{-1}$ , donc le spectre obtenu contient 800 points (C.f. Figure 24).

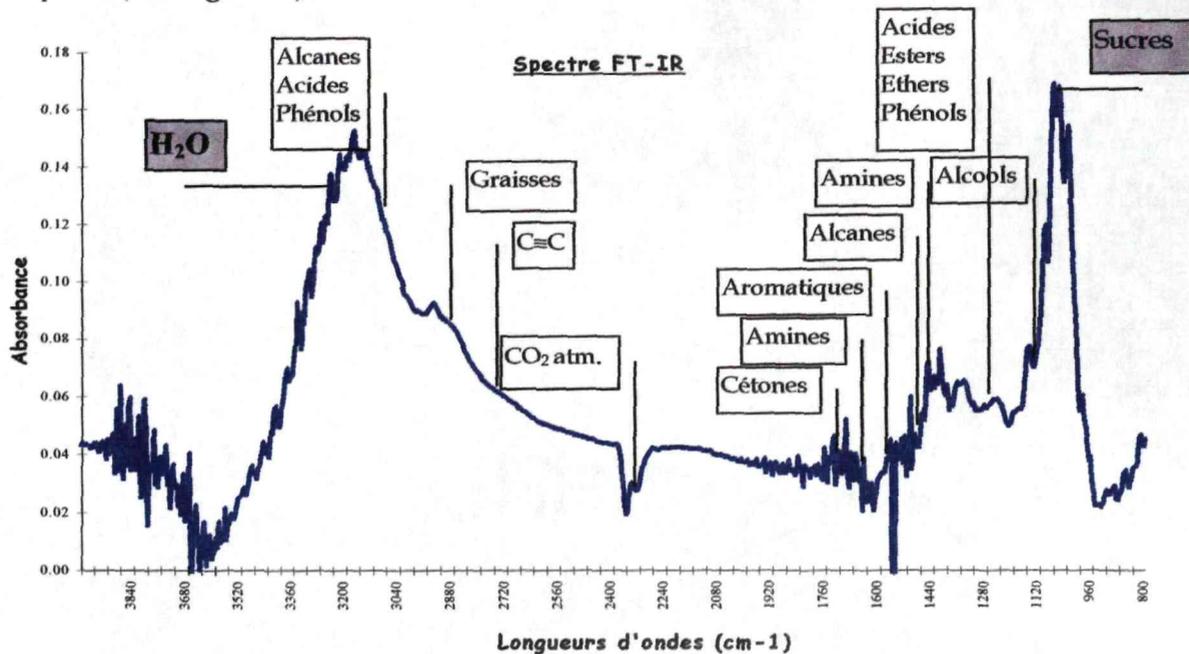


Figure 24 : exemple d'un spectre IR-TF de moût de raisin (BUZON et C, 1981)

Après une étude portant sur la qualité des vins (*DUPUIT, 1997*), la *fonction d'apodisation* choisie est triangulaire. Les moûts sont placés dans une cellule ATR (Réflexion Totale Atténuée en ZnSe, à 10 réflexions), permettant l'analyse d'échantillons fortement absorbants, notamment à cause de la présence d'eau (*VAN DE VOORT, 1992*). Le rayon incident est réfléchi par l'interface cristal/liquide et les spectres obtenus sont indépendants de l'épaisseur de l'échantillon.

L'Absorbance  $A$  se calcule par la formule suivante :  $A(\lambda) = -\log\left(\frac{I(\lambda)}{I_0(\lambda)}\right)$ , où  $I$  est l'intensité du produit analysé et  $I_0$  celle du produit de référence, en l'occurrence de l'eau distillée.

Un spectre d'absorbance de moût de raisin est présenté sur la Figure 24. Les échantillons mesurés dans un ordre aléatoire par le spectromètre IR-TF sont détaillées en Annexe 3.

#### IV. 1.5. Spectromètre UV

Les zones de longueurs d'onde explorées sont la région de l'ultraviolet entre 200 et 400 nm (C.f. Figure 23) et le visible jusqu'à 500 nm (*MALINGREY, 1987*).

La spectrométrie ultraviolet est généralement utilisée pour l'analyse de composés contenant des doubles ou des triples liaisons avec des atomes de C, H, O ou N ; pour être visibles dans l'UV, les doubles liaisons C=C doivent être conjuguées (-C=C-C=C-) (*SEKI et al., 1997*) ou cycliques (*VARDIN et VIDAL, 1991*) comme par exemple dans les polyphénols.

Le spectromètre UV utilisé est le Secomam S1000 ; la lumière est générée grâce à un lampe à décharge au deutérium sous moyenne pression pour l'UV et une lampe halogène pour la partie visible (*BAGUELIN et VIDIE, 1998*).

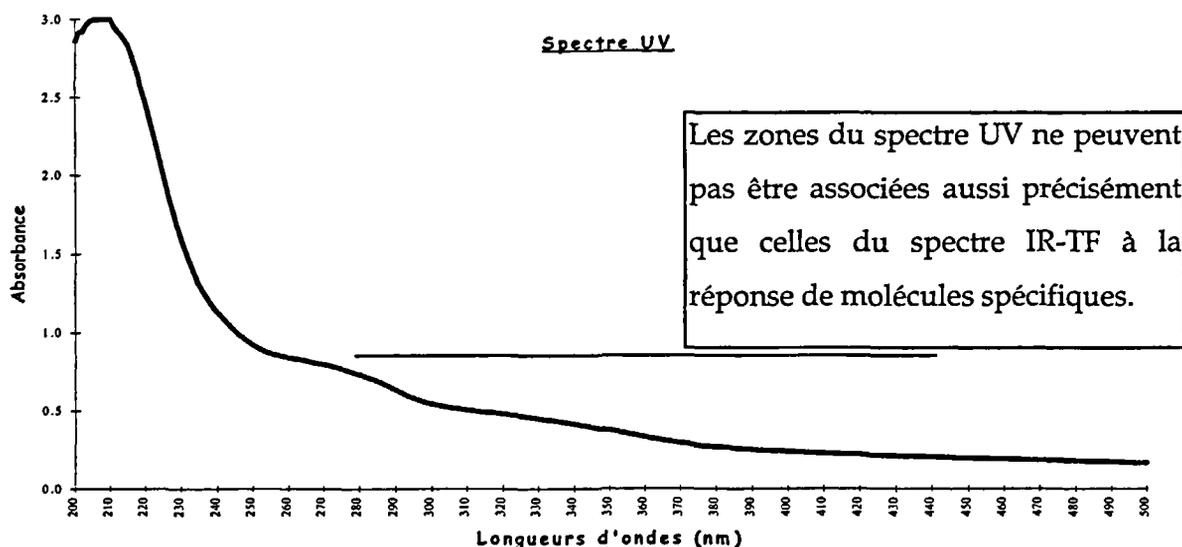


Figure 25 : exemple d'un spectre UV de moût de raisin

Un monochromateur permet d'effectuer une acquisition tous les nanomètres ; le spectre acquis contient donc 300 points, de 200 à 500 nm (C.f. Figure 25).

Les mesures des échantillons de moût sont réalisées dans des cuves en quartz, d'une épaisseur de 2mm et de 1 cm<sup>2</sup> de section.

109 échantillons de moût, détaillés en Annexe 3, sont mesurés dans un ordre aléatoire.

#### IV. 1.6. Mesure de référence : l'analyse sensorielle

Afin de comparer les performances instrumentales aux performances humaines, un jury sensoriel a été réuni, composé de 8 oenologues de l'ITV de Gaillac et de la Chambre d'Agriculture du Tarn. Ceux-ci sont entraînés à la dégustation des vins tout au long de l'année, et lors des vendanges, à la dégustation plus délicate des raisins ou des moûts.

Tous les échantillons n'ont pas pu être dégustés par le panel sensoriel, car cette analyse est trop longue et fastidieuse.

Les 2 objectifs de la dégustation étaient d'estimer la capacité de l'olfaction humaine à *distinguer 2 moûts* de cépages différents et l'aptitude à *identifier le cépage d'un moût*.

Nous avons mis en place 16 tests de différence sur l'olfaction des moûts en verre noir, afin de ne pas être influencé par la couleur des jus. Ce test est complété par l'attribution du cépage ; 4 échantillons de référence (1 par classe de cépage) étaient à la disposition des dégustateurs en permanence.

Les analyses de données des capteurs d'arômes ayant été effectuées antérieurement, les échantillons correctement discriminés analytiquement étaient connus. Donc, 8 tests de différence ont porté sur des échantillons correctement classés, et 8 sur des échantillons ambigus ; les échantillons soumis à dégustation sont détaillés en Annexe 6.

La loi binomiale, de paramètres  $p=1/3$  et 8 répétitions (i.e. 8 dégustateurs), permet de calculer le nombre significatif de différences : si 6 dégustateurs sur 8 ont identifié l'échantillon unique, alors la différence entre les 2 échantillons est significative avec un risque de  $\alpha < 5\%$  ; le risque est de  $\alpha < 1\%$  pour 7/8 différences,  $\alpha < 0.1\%$  pour 8/8 différences.

## IV. 2. CLASSIFICATION PAR CAPTEURS INDIVIDUELS

Dans le présent chapitre, l'étude de la discrimination variétale des moûts de raisins blancs porte sur la comparaison des classifications individuelles fournies par chaque capteur. Dans un premier temps, les 3 méthodes de classification sont exposées : l'AFD, les réseaux de neurones et la PLS discriminante, nouvellement proposée. Elles sont appliquées dans un second temps aux données issues des 3 systèmes de mesure (pIV-99). De même, la méthode de sélection des longueurs d'onde des spectres est détaillée (pIV-98) et appliquée aux spectres IR-TF (pIV-101) et UV (pIV-104). Enfin, les performances des diverses classifications sont comparées entre elles (pIV-106) ainsi qu'aux classifications issues du jury sensoriel et d'analyses chimiques (pIV-108).

### IV. 2.1. Méthodes

Le résultat d'une acquisition par capteurs d'arômes est une courbe d'adsorption - désorption. Le traitement des données peut donc être effectué en extrayant des descripteurs des courbes ; celui-ci permet d'appliquer des méthodes classiques de classification, telles que l'AFD (pII-37) ou les réseaux de neurones (pII-37). Il peut être intéressant de considérer d'autres techniques, encore jamais rencontrées dans le domaine des capteurs d'arômes, et qui consistent à analyser la courbe entière sans échantillonnage ni modélisation. Ces techniques de régression généralement sont généralement dédiées aux spectres, la plus performante étant la PLS (*Partial Least Squares*) ; cette méthode sera adaptée à la *reconnaissance des formes*, devenant « une PLS discriminante ». Les mesures spectrométriques seront uniquement traitées par PLS discriminante après un éventuel prétraitement (sélection des longueurs d'onde par algorithmes génétiques (pII-42)).

#### IV. 2.1.1. Classification

##### (1) Analyse Factorielle Discriminante

25 descripteurs sont extraits des courbes d'acquisition du capteur d'arômes LCA1000 modifié, à raison de 5 descripteurs par capteur. Ces descripteurs ont été préalablement sélectionnés au cours de la phase d'optimisation des capteurs d'arômes (C.f. Partie III, pIII-70).

Une AFD est effectuée sur ces 25 descripteurs, bâtissant un modèle linéaire, contenant 3 axes discriminants. Ce modèle doit être testé par des échantillons tests indépendants de ceux ayant participé à la modélisation ; Etant donné le faible nombre d'échantillons ( $n=138$ ), la prédiction sera effectuée par *validation croisée* « *leave-one-out* » (C.f. pII-36).

## (2) Réseaux de neurones

La modélisation par réseaux de neurones est effectuée sur les 25 descripteurs **normés** afin de prédire les 4 classes de cépage. Donc, le réseau contient 25 neurones d'entrée, 4 en sortie contenant les classes codées en disjonctif complet et un nombre variable de neurones dans la couche cachée. L'apprentissage est effectué par rétro-propagation du gradient de l'erreur (C.f. pII-37).

Plusieurs apprentissages sont nécessaires à l'optimisation de divers paramètres du réseaux de neurones :

- ❶ le nombre optimal de neurones dans la couche cachée ;
- ❷ le nombre d'itérations jusqu'à convergence de l'apprentissage grâce à une validation croisée par blocs (C.f. p II-38);
- ❸ le niveau de bruit introduit dans les données d'entrée (C.f. pII-38).

## (3) PLS discriminante

La PLS (*Partial Least Squares*) est une méthode de régression multivariée, qui bâtit un modèle de régression sur les composantes principales des variables d'entrée, comme la Régression sur les Composantes Principales (*PCR : Principal Component Regression*), tout en tenant compte de la variance interclasse des sorties à prédire (*MARTENS et NAES, 1984*). L'intérêt de cette technique est de pouvoir construire un modèle de régression sur des variables très corrélées, ce qui est très difficile avec une régression multiple (*MLR : Multi-Linear Regression*) ; ainsi, la PLS est couramment utilisée pour la prédiction de critères quantitatifs grâce à des mesures spectrométriques.

Toutefois, nous avons adapté la PLS à la *reconnaissance des formes*, grâce à la prédiction de classes codées en disjonctif complet. Pour chaque échantillon, la *PLS discriminante* fournit alors des prédictions pour chaque classe (qui s'apparentent à des degrés d'appartenance, même s'ils ne sont pas compris entre 0 et 1, mais plutôt entre -0.5 et +1.5). L'échantillon est alors attribué à la classe pour laquelle sa prédiction est la plus forte.

Afin de prendre en compte l'ensemble des courbes, les signaux des 5 MOS du capteur d'arômes LCA1000 modifié sont concaténés, après avoir enlevé la ligne de base de chacun. La courbe résultante, contenant 3605 points (721points \* 5capteurs), est traitée comme un signal unique, grâce à une PLS discriminante.

De la même façon, cette technique sera appliquée aux spectres.

## IV. 2.1.2. Sélection des longueurs d'onde par algorithmes génétiques

### (1) Paramètres des algorithmes génétiques

La méthode de sélection de variables par algorithmes génétiques est détaillée pII-42. Elle est appliquée à la sélection de longueurs d'onde de spectres IR-TF et UV.

Afin de minimiser les temps de calcul, le nombre de variables à sélectionner est réduit : chaque variable est la moyenne du spectre sur une fenêtre de longueurs d'onde déterminée. En IR-TF, les variables à sélectionner sont composées 20 fenêtres comportant chacune 40 points, soit une bande de  $160 \text{ cm}^{-1}$ . En UV, les 30 fenêtres contiennent 10 points chacune, soit 10nm.

La fonction d'adéquation utilisée est la PLS discriminante. Le critère d'adéquation à minimiser est le pourcentage d'erreur de prédiction en validation croisée sur 5 blocs. Les échantillons sont sélectionnés de façon aléatoire par la méthode des *venetian blinds*.

Les paramètres de fonctionnement des algorithmes génétiques sont les suivants :

- ❶ la population initiale est constituée par 48 chromosomes ;
- ❷ 30% des gènes sont présents dans un chromosome, c'est-à-dire que 30% des longueurs d'onde sont sélectionnées dans chaque spectre initial.
- ❸ le critère d'arrêt est l'obtention d'au moins 40% de chromosomes identiques ou l'atteinte de 100 générations ;
- ❹ le critère à optimiser est le taux d'erreur en validation croisée par bloc (sur 5 blocs) d'une PLS discriminante, avec au maximum 15 composantes principales intermédiaires (*latent variables*) ;
- ❺ un croisement simple à chaque génération sur la moitié de la population la moins performante;
- ❻ un taux de mutation de 0.5%.

### (2) Choix final des longueurs d'onde

La méthode des algorithmes génétiques étant non déterministe, le choix final des longueurs d'onde est différent à chaque exécution, en fonction de la population initiale générée au hasard et des phénomènes de croisement et mutation aléatoires. Il est donc nécessaire d'exécuter cet algorithme plusieurs fois. De plus, le résultat de chaque exécution de l'algorithme est une population de chromosomes ordonnés en fonction du critère d'adéquation.

Le passage de cette population de chromosomes à un lot de longueurs d'onde sélectionnées peut être effectué par différentes méthodes décrites ci-après.

☒ Pour chaque population, i.e. pour le résultat d'un algorithme génétique, le spectre pourra être convolué par la courbe :

- (1) des longueurs d'onde sélectionnées par le meilleur chromosome ;
- (2) du taux de sélection des longueurs d'onde dans la population finale ;
- (3) des longueurs d'onde sélectionnées à plus de 80% dans la population finale ;
- (4) des longueurs d'onde constamment sélectionnées (100%) dans la population finale.

☒ Pour l'ensemble des chromosomes créés, i.e. les résultats de tous les algorithmes génétiques, le spectre pourra être convolué par la courbe :

- (5) du taux moyen de sélection dans tous les résultats d'algorithmes génétiques et les convoluer avec le spectre (généralisation de la méthode ②);
- (6) des longueurs d'onde sélectionnées à plus de 80% dans toutes les populations finales (généralisation de la méthode ③);
- (7) des longueurs d'onde constamment sélectionnées (100%) dans toutes les populations finales (généralisation de la méthode ④).

Chacune de ces méthodes sera employée afin de déterminer quel est le choix final des longueurs d'onde le plus pertinent.

## IV. 2.2. Résultats

### IV. 2.2.1. LCA 1000 modifié

#### (1) Analyse Factorielle Discriminante

La classification par AFD appliquée aux 25 descripteurs extraits des courbes des 5 capteurs (C.f. p IV-96) fournit 37.6% d'erreur en calibration et 63 % d'erreur en prédiction en *validation croisée leave-one-out* (C.f. Tableau 21). Ce résultat très insuffisant peut être comparé au résultat d'un tirage aléatoire de 4 classes qui générerait 75% d'erreur. La discrimination linéaire des 4 classes de cépage est donc impossible avec le LCA 1000 modifié et traitement par AFD.

Tableau 21 : matrice de confusion des prédictions par AFD

Nombre d'échantillons	CLASSES REELLES				TOTAL
	Sauvignon	Mauzac	Colombard	Autres cépages	138
CLASSES PREDITES					% erreur
Sauvignon	17	6	9	8	57.5%
Mauzac	11	2	3	6	90.9%
Colombard	11	5	15	8	61.5%
Autres cépages	13	4	3	17	54.1%
% bons classements	32.7%	11.8%	50%	43.6%	37.0%
Taux d'erreur	67.3%	88.2%	50%	56.4%	63.0%

## (2) Réseaux de neurones

Les différents apprentissages par réseaux de neurones (C.f. p IV-96) nous ont conduits à choisir 5 neurones sur la couche cachée. Le nombre d'époques est optimisé par « stopped-learning » à 80 itérations, ce qui est faible mais s'explique par le petit nombre d'échantillons (138) et le peu de différence entre les classes. Un bruit aléatoire est introduit dans les entrées, avec un écart-type égal à  $0.2 \sigma$ .

L'erreur de prédiction est alors de 55.8% (C.f. Tableau 22), améliorant sensiblement les résultats de la discrimination linéaire. Cependant, les Mauzac restent complètement confondus avec les Sauvignon et plus de la moitié des échantillons sont attribués à la classe Sauvignon, quel que soit leur cépage réel, car cette classe est majoritaire et très éparpillée.

**Tableau 22 : matrice de confusion des prédictions par réseau de neurones**

	CLASSES REELLES				TOTAL
Nombre d'échantillons	52	17	30	39	138
CLASSES PREDITES	Sauvignon	Mauzac	Colombard	Autres cépages	% erreur
Sauvignon	32	16	13	17	59.0%
Mauzac	0	0	0	0	0%
Colombard	9	1	12	5	55.6%
Autres cépages	11	0	5	17	48.5%
% bons classements	61.5%	0%	40%	43.6%	44.2%
Taux d'erreur	38.5%	100%	60%	56.4%	55.8%

## (3) PLS discriminante

La classification par PLS discriminante prend en compte les courbes de tous les capteurs (C.f. p IV-97). Le meilleur taux d'erreur est obtenu par le calcul du modèle sur 6 composantes principales intermédiaires, fournissant une erreur de calibration de 41.3% et une erreur de prédiction en validation croisée de 47.1% (C.f. Tableau 23).

**Tableau 23 : matrice de confusion des prédictions par PLS discriminante**

	CLASSES REELLES				TOTAL
Nombre d'échantillons	52	17	30	39	138
CLASSES PREDITES	Sauvignon	Mauzac	Colombard	Autres cépages	% erreur
Sauvignon	35	16	12	15	55.1%
Mauzac	0	1	0	0	0%
Colombard	8	0	17	4	41.4%
Autres cépages	9	0	1	20	33.3%
% bons classements	67.3%	5.9%	56.7%	51.3%	52.9%
Taux d'erreur	32.7%	94.1%	43.3%	48.7%	47.1%

Ces résultats de classification linéaire sont à comparer avec ceux de l'AFD sur les descripteurs : il semblerait que les descripteurs extraits des courbes n'expriment pas de manière exhaustive l'information qu'elles contiennent.

Malgré les bons résultats de la PLS discriminante, il serait nécessaire de tester plus avant cette méthode.

En effet, c'est la première fois qu'elle est appliquée à des courbes et non à des spectres, et de manière discriminante. En outre, étant donné le grand nombre de variables prises en compte (3605 points), la calibration peut être proche de l'apprentissage par cœur et être très sensible au bruit. De plus, la valeur prédite par la PLS pour la classe réelle dépasse rarement 0.5, alors que la classe apprise lors de la calibration est 1 (et 0 pour les autres). L'affectation à la classe se fait donc plutôt par élimination que par une réelle appartenance.

#### IV. 2.2.2. Spectromètre IR-TF

##### (1) Classification sur les spectres bruts

La classification des cépages en appliquant la PLS discriminante (C.f. p IV-97) sur les spectres moyen infrarouge bruts fournit 30.7% d'erreur en prédiction (C.f. Tableau 24). Comme dans le cas des capteurs d'arômes, le cépage Mauzac est très mal classé, alors que le cépage Colombard est correctement distingué des autres classes.

Tableau 24 : matrice de confusion des prédictions par PLS discriminante

Nombre d'échantillons	CLASSES REELLES				TOTAL
	Sauvignon	Mauzac	Colombard	Autres cépages	% erreur
Sauvignon	34	4	4	8	32%
Mauzac	0	5	0	2	28.6%
Colombard	4	1	16	1	27.3%
Autres cépages	7	4	0	24	31.4%
% bons classements	75.6%	28.6%	80.0%	68.6%	69.3%
Taux d'erreur	24.4%	71.4%	20.0%	31.4%	30.7%

##### (2) Sélection des longueurs d'onde

Cette première classification peut être améliorée grâce à la sélection des longueurs d'onde par algorithmes génétiques (C.f. p IV-98). Les algorithmes génétiques ont été exécutés 12 fois et les sept méthodes de sélection des longueurs d'onde dans les populations finales de chromosomes (C.f. p IV-98) sont testées et détaillées dans le Tableau 25.

**Tableau 25 : Résultats d'une PLS discriminante sur les différents spectres IR-TF filtrés**

Ep : Erreur en prédiction (en %) sur 10 lots en validation croisée ; Ec : erreur en calibration ; Cp : nombre de composantes principales intermédiaires.

METHODES 	①			②			③			④		
	Ep	Ec	Cp									
Algorithmes Génétiques 												
Ag1	15%	7%	13	17%	8%	12	16%	9%	13	14%	14%	10
Ag2	11%	10%	10	13%	9%	10	11%	11%	11	26%	31%	7
Ag3	15%	7%	12	16%	11%	11	13%	11%	9	14%	11%	10
Ag4	13%	11%	11	14%	11%	11	18%	10%	11	21%	10%	11
Ag5	11%	11%	9	11%	11%	9	16%	10%	11	15%	10%	11
Ag6	14%	8%	9	15%	9%	9	16%	6%	10	15%	11%	9
Ag7	14%	11%	10	13%	11%	10	13%	10%	9	19%	14%	10
Ag8	14%	11%	10	14%	11%	10	14%	11%	10	14%	11%	10
Ag9	13%	11%	10	13%	16%	10	13%	11%	10	12%	12%	11
Ag10	14%	11%	10	14%	9%	12	15%	11%	10	16%	11%	10
Ag11	17%	11%	10	16%	9%	11	16%	11%	10	17%	9%	12
Ag12	11%	10%	12	10%	9%	12	12%	15%	9	16%	12%	12
METHODES 				⑤			⑥			⑦		
				Ep	Ec	Cp	Ep	Ec	Cp	Ep	Ec	Cp
∪ Ag				13%	11%	8	16%	12%	12	30%	30%	5

Les meilleurs résultats sont obtenus avec méthodes de sélection des longueurs d'onde ① et ②, qui correspondent à la sélection du meilleur chromosome et la convolution des spectre avec le taux de sélection des variables dans la population finale.

Parmi les 12 exécutions des algorithmes génétiques, seuls 3 résultats fournissent un taux d'erreur minimal similaire, ce qui semble indiquer qu'il est nécessaire de se livrer à plusieurs exécutions pour obtenir les meilleurs résultats.

Le meilleur taux de prédiction en validation croisée est de 9.6% (C.f. Tableau 26), ce qui améliore d'un facteur 3 les résultats obtenus grâce aux spectres bruts (C.f. Tableau 24) !

**Tableau 26 : matrice de confusion des prédictions par PLS discriminante du meilleur spectre IR-TF sélectionné.**

Nombre d'échantillons	CLASSES REELLES				TOTAL
	Sauvignon	Mauzac	Colombard	Autres cépages	% erreur
CLASSES PREDITES					
Sauvignon	44	0	0	3	6.4%
Mauzac	1	13	0	6	35%
Colombard	0	1	20	0	4.8%
Autres cépages	0	0	0	26	0%
% bons classements	97.8%	92.9%	100%	74.3%	90.3%
Taux d'erreur	2.2%	7.1%	0%	25.7%	9.6%

La Figure 26 montre la prédiction des échantillons de moûts par PLS discriminante basée sur le meilleur spectre (Ag12 avec la méthode de sélection ②).

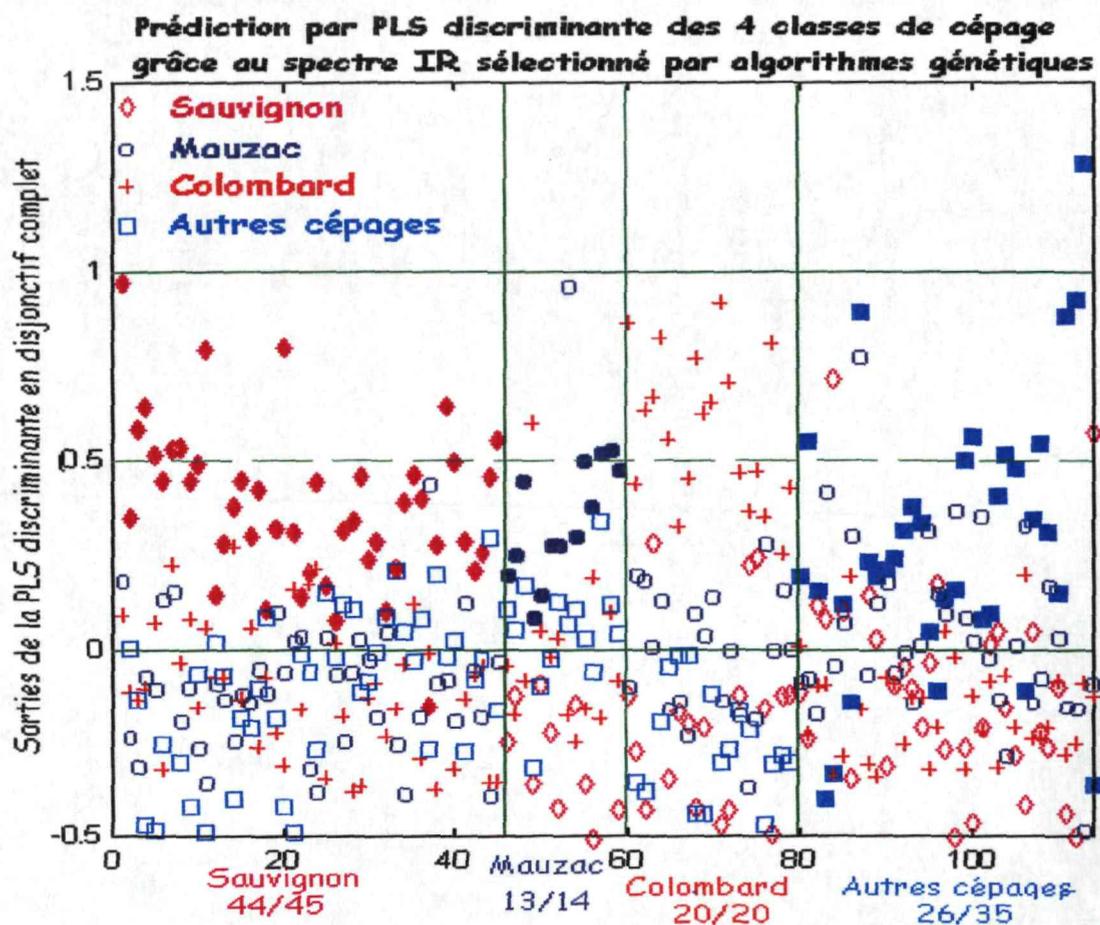
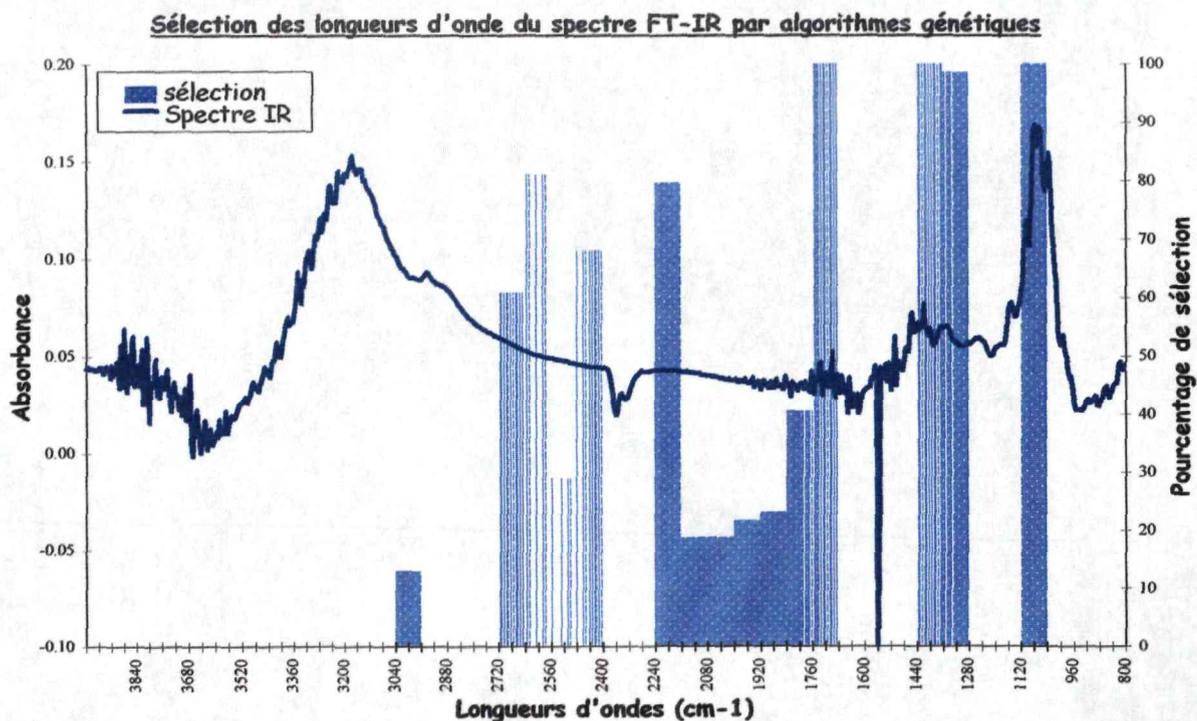


Figure 26 : prédiction des échantillons par PLS discriminante sur le spectre IR-TF filtré

Pour chaque échantillon (en abscisse) sont représentés en ordonnée les 4 prédictions pour chacune des classes ; l'échantillon est affecté à la classe pour laquelle la prédiction est la plus forte (la plus haute sur la Figure 26). Malgré le faible taux d'erreur, cette figure montre la difficulté de la discrimination due à la proximité des prédictions, dépassant rarement 0.5, alors que les classes apprises sont constituées de 0 et de 1.

Les longueurs d'onde sélectionnées fournissant le meilleur résultat sont détaillées dans la Figure 27 (AG12 avec la méthode de sélection ②). Les algorithmes génétiques ont permis d'éliminer les zones très bruitées du spectre. Une partie du premier pic est sélectionnée ( $1000-1200\text{ cm}^{-1}$ ), correspondant aux glucides ; par contre, l'artefact du  $\text{CO}_2$  atmosphérique ( $\sim 2300\text{ cm}^{-1}$ ) ainsi que l'énorme pic de l'eau ( $3000-3500\text{ cm}^{-1}$ ) sont soigneusement exclus de la sélection.



**Figure 27 : Sélection des longueurs d'onde du spectre IR-TF**

### IV. 2.2.3. Spectromètre UV

La même démarche a été adoptée pour les spectres ultraviolets : une classification est établie à partir des spectres bruts, puis elle est améliorée par la sélection des longueurs d'onde par algorithmes génétiques.

#### (1) Classification sur les spectres bruts

La classification basée sur les spectres bruts présente 52.3% de bons classements, avec une forte proportion de prédiction de la classe Sauvignon, pour les échantillons des classes Colombard et Autres cépages (C.f. Tableau 27).

**Tableau 27 : matrice de confusion des prédictions PLS discriminante du spectre UV brut**

	CLASSES REELLES				TOTAL
	Sauvignon	Mauzac	Colombard	Autres cépages	
Nombre d'échantillons	44	14	14	37	109
CLASSES PREDITES	Sauvignon	Mauzac	Colombard	Autres cépages	% erreur
Sauvignon	26	3	7	14	48%
Mauzac	3	6	0	3	50%
Colombard	1	1	5	0	28.6%
Autres cépages	14	4	2	20	50%
% bons classements	59.9%	42.9%	35.7%	54.0%	52.3%
Taux d'erreur	40.1%	57.1%	64.3%	46.0%	47.7%

**(2) Sélection des longueurs d'onde**

Les résultats de sélection des longueurs d'onde grâce aux diverses méthodes (C.f. p IV-98) figurent dans le **Tableau 28**. De manière similaire aux spectres IR-TF, les meilleures techniques de sélection des longueurs d'onde au sein de la population finale sont les méthodes ① et ②.

**Tableau 28 : Résultats d'une PLS discriminante sur les différents spectres UV filtrés**

Ep : Erreur en prédiction (en %) sur 10 lots en validation croisée ; Ec : erreur en calibration ; Cp : nombre de composantes principales intermédiaires.

METHODES 	①			②			③			④		
	Ep	Ec	Cp									
Algorithmes génétiques 												
Ag1	34%	18%	12	38%	20%	13	38%	34%	7	41%	42%	5
Ag2	35%	28%	9	38%	28%	8	39%	29%	8	39%	28%	8
Ag3	29%	24%	10	25%	22%	11	32%	22%	11	36%	26%	10
Ag4	23%	21%	10	26%	21%	11	32%	25%	10	40%	33%	9
Ag5	36%	27%	11	39%	24%	12	36%	27%	11	39%	25%	10
Ag6	24%	23%	10	27%	21%	11	33%	25%	10	37%	26%	9
Ag7	37%	24%	10	39%	24%	10	39%	24%	10	41%	26%	9
Ag8	32%	20%	11	33%	17%	12	37%	23%	12	53%	34%	8
Ag9	40%	29%	9	39%	31%	8	39%	31%	8	39%	31%	8
Ag10	28%	17%	12	28%	21%	11	35%	22%	12	39%	31%	8
Ag11	40%	25%	10	41%	25%	12	37%	28%	12	41%	24%	10
Ag12	34%	29%	10	35%	26%	10	38%	29%	9	38%	37%	8
Ag13	34%	24%	14	34%	23%	13	34%	26%	11	42%	33%	7

METHODES 
∪ Ag

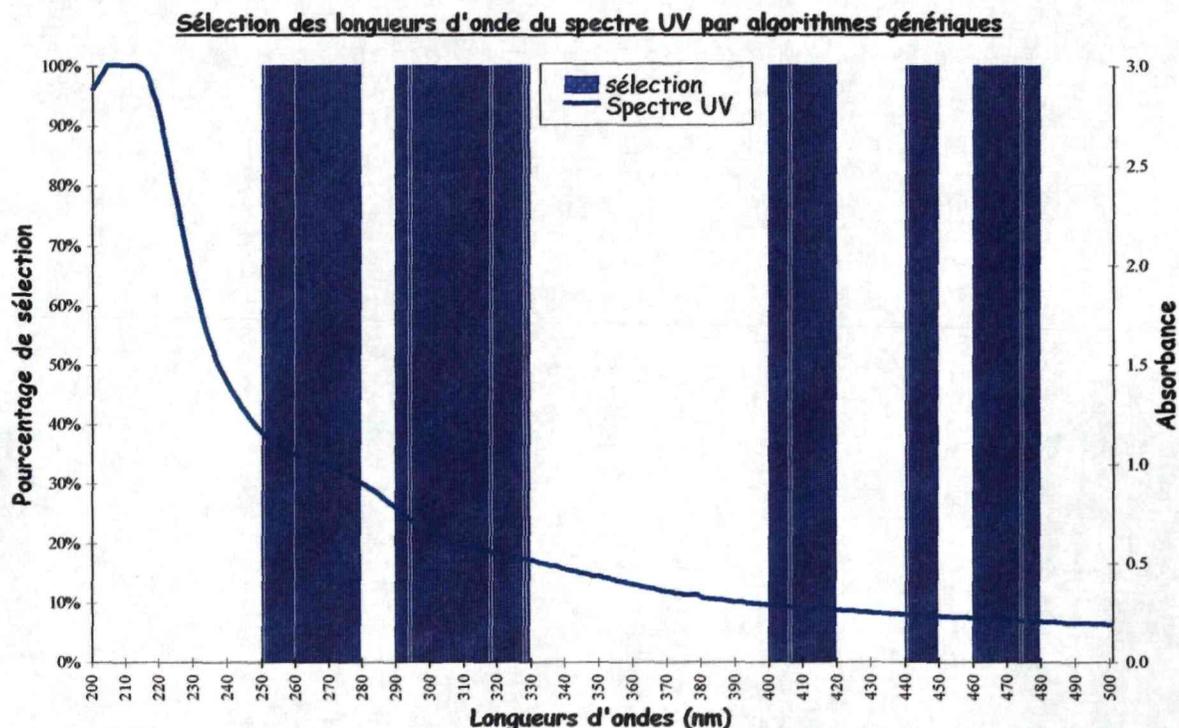
⑤			⑥			⑦		
Ep	Ec	Cp	Ep	Ec	Cp	Ep	Ec	Cp
24%	17%	15	37%	35%	6	∅	∅	∅

Les meilleurs résultats sont offerts par la quatrième exécution des algorithmes génétiques, en prélevant les longueurs d'onde sélectionnées par le meilleur chromosome (méthode ①). Ces résultats de classification des 4 classes de cépages sont détaillés dans le **Tableau 29**.

**Tableau 29 : matrice de confusion des prédictions par PLS discriminante du spectre UV filtré (AG4 et méthode ①)**

Nombre d'échantillons	CLASSES REELLES				TOTAL
	Sauvignon	Mauzac	Colombard	Autres cépages	% erreur
CLASSES PREDITES					
Sauvignon	36	1	4	5	21.7%
Mauzac	4	12	2	3	42.9%
Colombard	1	1	8	1	27.3%
Autres cépages	3	0	0	28	9.7%
% bons classements	81.8%	85.7%	57.1%	75.7%	77.1%
Taux d'erreur	18.2%	14.3%	42.9%	24.3%	22.9%

La sélection des longueurs d'onde par le meilleur chromosome est représentée dans la Figure 28. Le sommet du pic principal n'est pas sélectionné car la courbe sature pour certains échantillons. Etant donné que la différence analytique entre cépages n'est pas connue, il est difficile d'expliquer la pertinence de la sélection des longueurs d'onde, si ce n'est par l'élimination de zones bruitées ou redondantes.



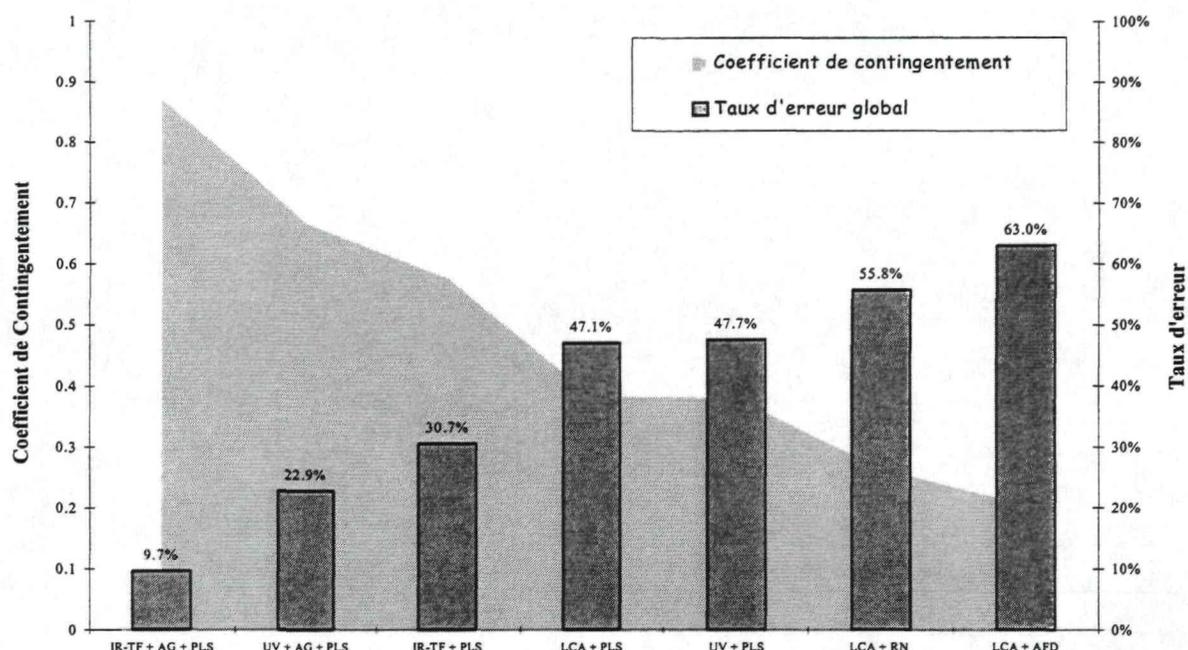
**Figure 28 : Sélection des longueurs d'onde du spectre UV**

#### IV. 2.2.4. Discussion et conclusions sur les classifications individuelles

##### (1) Comparaison des classifications

Afin d'estimer la performance des différentes classifications, le taux d'erreur global est analysé ainsi que le coefficient de contingement (C.f. pII-52). Ce dernier est calculé à l'aide de toute la matrice de confusion afin d'analyser la correspondance entre la discrimination établie par chaque capteur et la classification réelle des échantillons.

Les performances des divers capteurs sont très différentes, montrant que les capteurs d'arômes sont beaucoup moins aptes à discriminer les moûts de cépages blancs que le spectromètre IR-TF ou UV (C.f. Figure 29).



**Figure 29 : taux d'erreur et coefficients de contingement des différents systèmes de mesure**

IR-TF : spectromètre infrarouge ; UV : spectromètre ultraviolet ; LCA : capteurs d'arômes ; AFD : Analyse Factorielle Discriminante ; RN : réseaux de neurones ; PLS : *Partial Least Square* discriminante

Les divers traitements ou prétraitements ont une importance considérable sur les résultats de classification. En effet, le traitement des données des capteurs d'arômes est bien meilleur par PLS discriminante que par réseaux de neurones ou Analyse Factorielle Discriminante, comme le montrent le coefficient de contingement ainsi que le taux d'erreur. Donc, dans la suite des travaux sur la *fusion de niveau haut* qui combine les résultats de classification, nous n'introduisons dans le système que les résultats issus de la PLS discriminante.

Les prétraitements des spectres destinés à sélectionner les longueurs d'onde discriminantes par algorithmes génétiques sont extrêmement efficaces, permettant d'améliorer les résultats de classification d'un facteur 2 à 3 (C.f. **Tableau 24 & Tableau 26** pour l'IR-TF ; **Tableau 27 & Tableau 29** pour l'UV).

Ces mesures instrumentales peuvent être « comparées » à la discrimination par jury sensoriel. Malheureusement, l'analyse ne pourra pas être faite sur la même base car les 138 échantillons n'ont pas pu être dégustés.

Les 16 tests triangulaires permettent de conclure que la capacité de distinction des échantillons par le panel d'oenologues et celle des capteurs d'arômes ne sont pas corrélées. Le panel d'oenologues analyse avec la même performance les 8 échantillons bien classés par les capteurs d'arômes et les 8 autres, mal classés (**Annexe 6**) : 5 échantillons sur 8 identifiés avec un risque  $\alpha < 5\%$ , soit 37.5% d'erreur. Il semble relativement difficile de différencier les échantillons de cépages différents uniquement par olfaction.

Il est encore plus difficile de déterminer l'appartenance aux cépages : 2 dégustateurs ne se sont pas prononcés sur le cépage, 5 ont moins de bonnes réponses qu'un tirage aléatoire (entre 3 et 18% de bonnes réponses contre 25% pour une classification aléatoire) et le meilleur obtient 34% d'identification correcte ! *Il apparaît donc impossible d'identifier les cépages par la simple olfaction de moûts congelés (et décongelés)*. Ceci peut expliquer en partie que les capteurs d'arômes soient peu performants pour la classification.

## (2) Méthodologie de sélection des longueurs d'onde

La sélection des longueurs d'onde par algorithmes génétiques apparaît être un prétraitement des spectres indispensable, introduisant une amélioration considérable de la discrimination des classes de cépage. Différentes méthodes pour constituer le *spectre « filtré »* final à partir de la population finale ont été développées et testées. Il en ressort qu'il est nécessaire d'exécuter plusieurs fois (5 à 10 fois) les algorithmes génétiques afin d'obtenir les meilleurs résultats de classification. Les 2 filtres les plus efficaces sont celui correspondant au chromosome le plus performant ou celui correspondant au taux de sélection de variables calculé sur les chromosomes d'une exécution de l'algorithme.

Cette technique est relativement lourde en calcul, donc très longue. Afin de réduire les temps d'exécution, la population initiale peut être réduite. La pauvreté de la diversité initiale doit alors être compensée par des croisements doubles et des taux de mutation assez forts (de l'ordre de 1%) en début de reproduction afin d'explorer de nombreuses combinaisons de gènes ; puis, ces taux doivent être réduits pour permettre la convergence de la population. De plus, les variables à sélectionner sont générées à partir de larges fenêtres qui peuvent être réduites au cours de l'étude.

Lors de la fusion multi-capteur, les *spectres bruts* ainsi que les *spectres filtrés* seront employés, afin de comparer leurs performances.

## (3) Corrélation aux analyses chimiques

Lors de discriminations variétales ou qualitatives, il est extrêmement important de s'assurer que la discrimination n'est pas basée sur un artefact, qui n'est pas caractéristique des classes mais d'un facteur dû à l'échantillonnage. Dans notre application, de nombreux paramètres expérimentaux peuvent influencer la discrimination des moûts, tels que la maturité des raisins lors de la vendange, l'effet opérateur lors de la génération des échantillons, ou encore la dérive des capteurs au cours du temps. Ces deux derniers effets ne peuvent pas induire de discrimination artificielle car chaque opérateur a produit des échantillons dans chacune des classes et les échantillons ont été mesurés dans un ordre aléatoire. Quant à la maturité, elle peut être contrôlée par les mesures analytiques, en particulier du sucre et de l'acidité totale.

Le sucre et l'acidité totale sont relativement homogènes, quelle que soit la classe, même si les échantillons de Colombard sont légèrement moins mûrs que les autres échantillons, i.e. moins sucrés et plus acides (C.f. Figure 30).

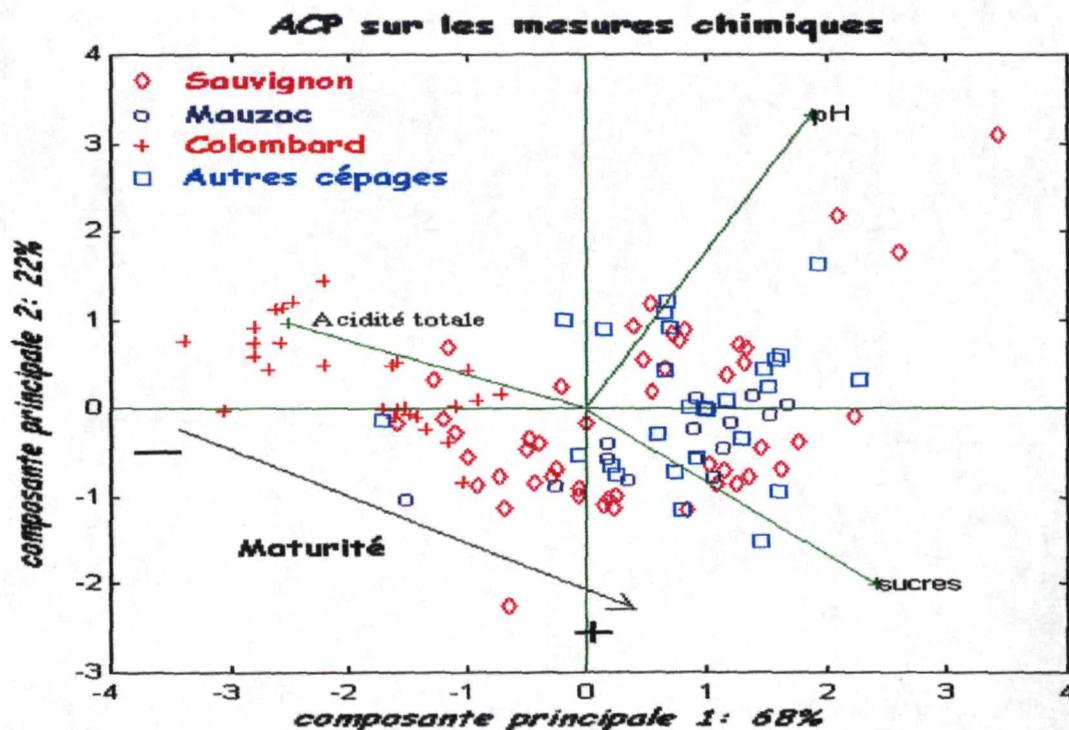


Figure 30 : Analyse en Composantes Principales sur les analyses chimiques

La classification par AFD confirme la distinction des échantillons de Colombard par rapport aux autres (C.f. Tableau 30), fournissant 30.8% d'erreur en calibration et 34.2% en validation croisée « *leave-one-out* ». La discrimination basée sur les analyses chimiques est meilleure que celle réalisée par les capteurs d'arômes !

Tableau 30 : matrice de confusion de classification des moûts par AFD sur les analyses chimiques

Nombre d'échantillons	CLASSES REELLES				TOTAL
	Sauvignon	Mauzac	Colombard	Autres cépages	117
CLASSES PREDITES $\uparrow$					% erreur $\uparrow$
Sauvignon	29	0	2	9	27.5%
Mauzac	7	11	0	6	54.2%
Colombard	4	0	25	0	13.8%
Autres cépages	9	3	0	12	50.0%
% bons classements	59.2%	78.6%	92.6%	44.4%	65.8%
Taux d'erreur $\uparrow$	40.8%	21.4%	7.4%	55.6%	34.2%

L'intensité de la mesure aromatique du jus de raisin n'est pas corrélée au taux de sucre ou à l'acidité ( $r^2 = 0.08$  et  $0.06$ , respectivement).

De plus, la corrélation entre les mesures spectrométriques et chimiques a été analysée. Le spectromètre IR-TF est couramment utilisé pour le dosage des sucres dans les produit agro-alimentaire. La prédiction du taux de sucre par les spectres infrarouge bruts est correcte, avec un SEP<sup>21</sup> de 6g.l<sup>-1</sup>, sachant que la gamme du taux de sucre des échantillons varie entre 108 et 221 g.l<sup>-1</sup>. Cependant, après la sélection des longueurs d'onde par algorithmes génétiques, le SEP est dégradé jusqu'à 12g.l<sup>-1</sup>, alors que l'erreur de discrimination des cépages est divisée par 3 ! Ceci signifie que les longueurs d'onde aptes à prédire le taux de sucre ont été partiellement éliminées car elles dégradaient la discrimination des cépages. La classification de moûts de raisin blancs est donc largement indépendante du taux de sucre des échantillons, qui introduit en réalité du bruit dans les spectres. La sélection des longueurs d'onde du spectre UV améliore en revanche la prédiction du sucre de 24 à 12g.l<sup>-1</sup>.

Quant à l'acidité totale, sa prédiction par les spectres IR-TF est légèrement altérée par la sélection des longueurs d'onde par algorithmes génétiques, le SEP passant de 0.59 à 0.65g.l<sup>-1</sup> d'eq. H<sub>2</sub>SO<sub>4</sub>, pour une gamme allant de 2.54 à 8.48g.l<sup>-1</sup> d'eq. H<sub>2</sub>SO<sub>4</sub>. La sélection des longueurs d'onde du spectre UV améliore les résultats de prédiction de l'acidité totale, passant de 1.24 à 0.68g.l<sup>-1</sup> d'eq. H<sub>2</sub>SO<sub>4</sub>. Les longueurs d'onde caractérisant les cépages doivent partiellement coïncider avec celles employées pour la quantification de l'acidité totale.

---

<sup>21</sup> SEP : « Standard Error of Prediction » =  $SEP = \sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / n - 1}$  ; avec  $n$  le nombre total d'échantillons,  $y_i$  la valeur réelle et  $\hat{y}_i$  la valeur estimée de l'échantillon  $i$ .

### IV. 3. FUSION MULTI-CAPTEUR

Après avoir comparé les résultats des classifications individuelles des 3 capteurs employés, leurs données sont traitées de façon simultanée grâce à 2 méthodes de *fusion multi-capteur*. La première méthode est une *fusion de niveau bas*, sur les signaux bruts ou filtrés. La seconde est une *fusion de niveau haut*, basée sur une approche bayésienne (pIV-116). Ces 2 méthodes seront appliquées aux capteurs d'arômes et aux spectres IR-TF et UV et leurs performances seront comparées (pIV-127) afin de déterminer la plus adaptée à la fusion multisensorielle.

#### IV. 3.1. Méthode de fusion multi-capteur de niveau bas

##### IV. 3.1.1. Méthodes

La *fusion de « niveau bas »* est destinée à fusionner les signaux commensurables (C.f. pII-45). Au cours de notre étude, cette technique est appliquée aux spectromètres IR-TF et UV.

Les spectres bruts concaténés fournissent un spectre commun, noté « **spectre IR&UV** » contenant 1101 longueurs d'onde, 301 provenant de la mesure UV et 800 de l'IR-TF.

La méthode de sélection des longueurs d'onde est appliquée au spectre concaténé, donc « a posteriori », fournissant un « **spectre IR&UV post.** » Les variables sont regroupées en 55 fenêtres contenant 20 points chacune, afin de réduire les temps de calcul. La population initiale contient 48 chromosomes et 30% des variables sont sélectionnées ; ce taux est relativement stable au cours des générations, soient 360 points (18 fenêtres) environ par chromosome. Le critère d'arrêt est l'obtention de plus de 40% de chromosomes uniques dans la population finale. La sélection finale des longueurs d'onde est effectuée grâce aux méthodes ① et ② (C.f. pIV-98), car elles ont donné les meilleurs résultats sur les spectres individuels (C.f. pIV-108).

Mais, plus le nombre de variables est élevé, plus l'exécution des algorithmes génétiques est lourde. Donc, la concaténation est également effectuée sur les spectres préalablement filtrés, c'est-à-dire dont les longueurs d'onde ont été sélectionnées par algorithmes génétiques. Ce spectre commun est appelé « **spectre IR&UV pre** », car les longueurs d'onde sont choisies « a priori », individuellement pour chaque capteur.

La méthode de classification multivariée appliquée à ces spectres est une PLS discriminante, avec validation croisée par blocs ou « *leave-one-out* ».

### IV. 3.1.2. Résultats

La fusion multi-capteur de niveau bas est appliquée uniquement aux données spectrométriques. Dans un premier temps, les résultats issus de la concaténation des spectres bruts sont exposés. Dans un second temps, la fusion des spectres filtré par les algorithmes génétiques est détaillée : les spectres filtrés après la concaténation (a posteriori) et ceux filtrés individuellement, avant concaténation (a priori).

#### (1) Fusion des spectres bruts

La fusion par PLS discriminante appliquée aux spectres bruts concaténés fournit des résultats de classification légèrement meilleurs que ceux du spectromètre IR-TF, avant toute sélection de longueurs d'onde : 28.0% d'erreur contre 30.7% en IR-TF seul (C.f. Tableau 31). Mais, le taux d'erreur de classification minimal est obtenu avec 20 composantes principales intermédiaires, ce qui considérable, pouvant entraîner une forte instabilité du modèle.

**Tableau 31 : matrice de confusion des prédictions par PLS discriminante sur le spectre commun IR-TF et UV**

Nombre d'échantillons	CLASSES REELLES				TOTAL
	Sauvignon	Mauzac	Colombard	Autres cépages	% erreur
CLASSES PREDITES					
Sauvignon	35	0	3	9	25.5%
Mauzac	1	10	1	3	33.3%
Colombard	1	1	10	1	23.1%
Autres cépages	7	3	0	22	31.3%
% bons classements	79.5%	71.4%	71.4%	62.9%	72.0%
Taux d'erreur	20.5%	28.6%	28.6%	37.1%	28.0%

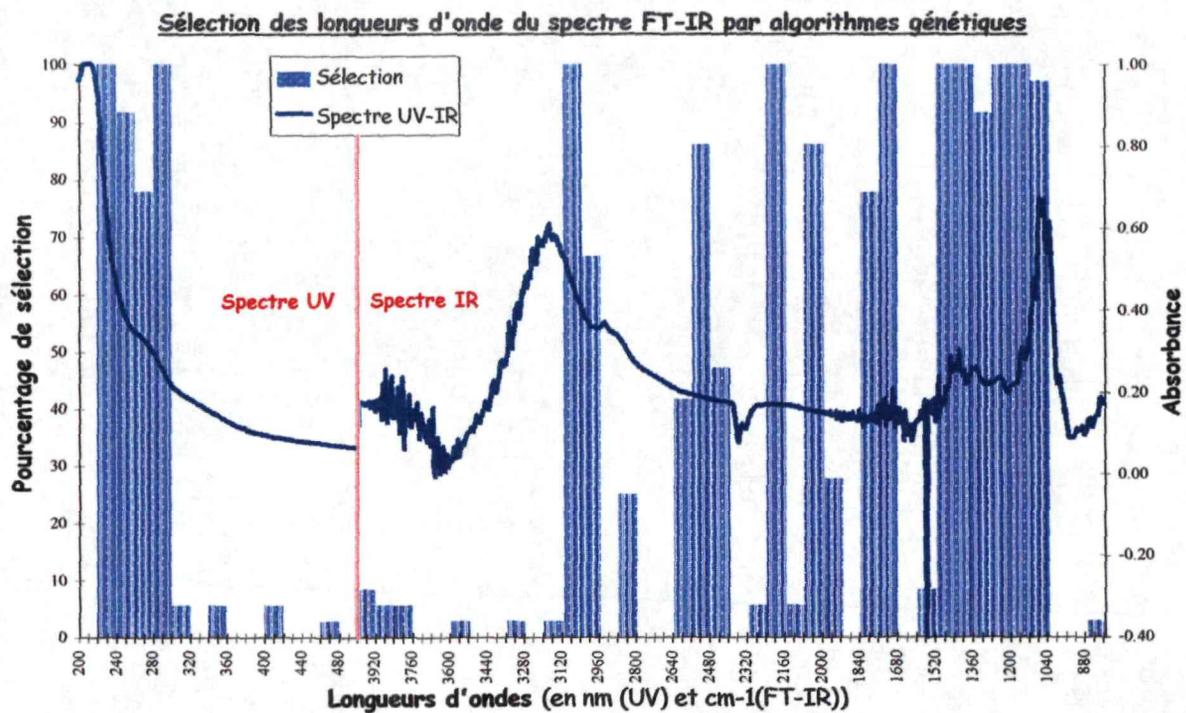
Taux d'erreur UV	40.1%	57.1%	64.3%	46.0%	47.7%
Taux d'erreur IR-TF	24.4%	71.4%	20.0%	31.4%	30.7%

Il apparaît donc nécessaire de sélectionner les longueurs d'onde de ce spectre commun, afin d'améliorer la discrimination et de réduire le nombre de composantes principales employées.

#### (2) Fusion des spectres filtrés

##### (a) Sélection de longueurs d'onde a posteriori

Nous présentons le spectre final constitué par la convolution du spectre commun avec le spectre de taux de sélection finale des variables (méthode ②) : il contient 701 points (C.f. Figure 31).



**Figure 31 : Sélection des longueurs d'onde du spectre commun UV-IR-TF**

Les résultats de discrimination sont améliorés d'un facteur supérieur à 2, car le taux d'erreur passe de 28.0% à 13.1% avec le même nombre de composantes principales intermédiaires (C.f. **Tableau 32**). Cependant, la sélection de variables appliquée à un spectre aussi large nécessite un temps de calcul considérable, de l'ordre de 24h pour la convergence des algorithmes génétiques ! De plus, ces résultats n'atteignent pas les performances du spectromètre IR-TF après sélection des longueurs d'onde.

**Tableau 32: matrice de confusion des prédictions par PLS discriminante après sélection des longueurs d'onde a posteriori**

SPECTRE IR-UV POST.	CLASSES REELLES				TOTAL
	Sauvignon	Mauzac	Colombard	Autres cépages	
Nombre d'échantillons	44	14	14	35	107
CLASSES PREDITES	Sauvignon	Mauzac	Colombard	Autres cépages	% erreur
Sauvignon	43	1	0	4	10.4%
Mauzac	0	10	0	5	33.3%
Colombard	1	1	14	0	12.5%
Autres cépages	0	2	0	26	7.1%
% bons classements	97.7%	71.4%	100%	74.3%	86.9%
Taux d'erreur	2.3%	28.6%	0%	25.7%	13.1%

Taux d'erreur IRUV brut	20.5%	28.6%	28.6%	37.1%	28.0%
-------------------------	-------	-------	-------	-------	-------

## (b) Sélection des longueurs d'onde a priori

La PLS discriminante appliquée sur les spectres filtrés par algorithmes génétiques avant concaténation (contenant 420 points) améliore la classification par rapport à celle fournie par les spectres bruts, jusqu'à obtenir 18.7% d'erreur en prédiction, grâce à 17 composantes principales intermédiaires (C.f. Tableau 33). Cependant, si ce taux d'erreur est comparé à ceux des spectres individuels filtrés, alors nous pouvons en déduire que la fusion de niveau bas n'est pas efficace, puisqu'elle est moins performante que la discrimination basée sur le spectromètre IR-TF uniquement.

Tableau 33 : matrice de confusion des prédictions par PLS discriminante

SPECTRE IRUV PRE	CLASSES REELLES				TOTAL
Nombre d'échantillons	44	14	14	35	107
CLASSES PREDITES	Sauvignon	Mauzac	Colombard	Autres cépages	% erreur
Sauvignon	42	0	3	7	19.2%
Mauzac	0	9	0	3	25%
Colombard	0	2	11	0	15.4%
Autres cépages	2	3	0	25	16.7%
% bons classements	95.5%	64.3%	78.6%	71.4%	81.3%
Taux d'erreur	4.5%	35.7%	21.4%	28.6%	18.7%

Taux d'erreur UV filtré	18.2%	14.3%	42.9%	24.3%	22.9%
Taux d'erreur IR-TF filtré	2.2%	7.1%	0%	25.7%	9.6%

La fusion sur les spectres filtrés a priori est moins performante que la sélection des longueurs d'onde du spectre concaténé (18.7% d'erreur contre 13.1%); en effet, elle ne tient pas compte de la corrélation entre les longueurs d'onde des 2 spectromètres.

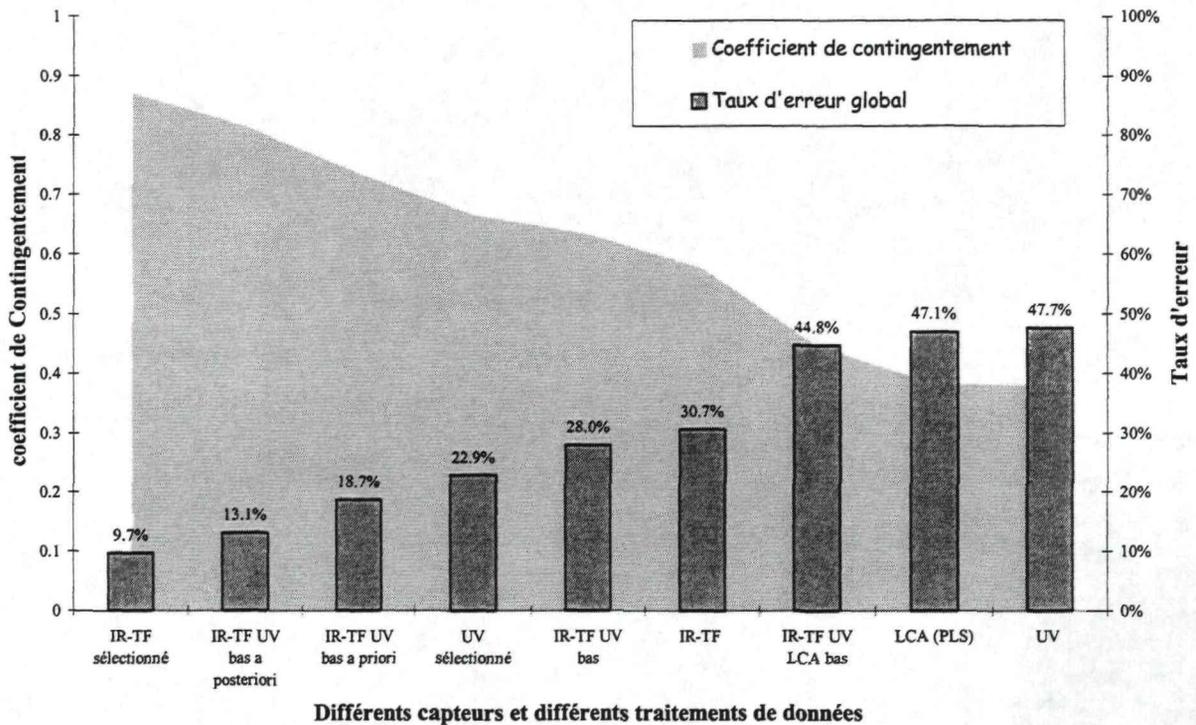
(3) Conclusion

La fusion multi-capteur de niveau bas améliore sensiblement le résultat sur les spectres bruts de chaque spectromètre, passant d'un taux d'erreur de 47.7% et 30.7%, respectivement pour les spectromètres UV et IR-TF, à 28.0%.

Comme pour les capteurs individuels, la sélection de variables améliore la fusion de niveau bas (qui devient alors une *fusion de niveau intermédiaire*). Le choix des longueurs d'onde par algorithmes génétiques s'avère une méthode extrêmement efficace. Cependant, l'application de cette méthode à des courbes contenant un grand nombre de variables (>1000) impose une charge de calcul très importante, induisant des temps de calcul de l'ordre de 24 à 48h par exécution d'un algorithme génétique (sur un PC Pentium 166 MHz) !

Nous avons donc essayé de contourner cette difficulté en effectuant la sélection individuellement pour chaque spectromètre ; malheureusement, cette démarche de sélection a priori est moins efficace que

celle appliquée au spectre concaténé, fournissant un taux d'erreur de 18.7% contre 13.1% pour la sélection a posteriori (C.f. Figure 32).



**Figure 32 : Taux d'erreur et coefficient de contingement des différents systèmes de mesure**

IR-TF : spectromètre infrarouge ; UV : spectromètre ultraviolet ; LCA : capteurs d'arômes ; « sélectionné » : filtré par algorithmes génétiques ; PLS : *Partial Least Square* discriminante ; « bas » : fusion de niveau bas ; « bas a posteriori » : fusion des spectres filtré a posteriori ; « bas a priori » : fusion des spectres filtré a priori.

Nous n'avons pas détaillé la fusion de niveau bas entre les spectres IR-TF et UV filtrés a posteriori et les courbes des capteurs d'arômes, mais celle-ci est exposée en [Annexe 7](#). Lors de la fusion des courbes des capteurs d'arômes avec celles des spectromètres (malgré la sélection préalable des longueurs d'onde des spectres !), le taux d'erreur augmente jusqu'à 44.8% (C.f. Figure 32). La concaténation de courbes contenant de très nombreuses variables, dont certaines parties peuvent être très bruitées, peut donc au contraire dégrader la discrimination.

### IV. 3.2. Fusion multi-capteur de niveau haut

Devant la lourdeur de la technique de fusion de niveau bas, nous avons repris et appliqué une démarche de fusion de niveau haut, combinant les classifications individuelles et non les données brutes. Les méthodes de fusion multi-capteur par l'erreur minimale et le coût minimal de Bayes ont été développées dans la partie bibliographique (pII-46).

#### IV. 3.2.1. Méthodes

##### (1) Fusion par erreur minimale de Bayes

La fusion de niveau haut est appliquée aux spectres IR-TF et UV bruts (C.f. Annexe 8), et également sur les spectres filtrés par algorithmes génétiques. Nous ne développerons que cet exemple.

Dans un 1<sup>er</sup> temps, nous considérerons que les capteurs sont indépendants ; alors, la règle de décision de Bayes d'erreur minimale suit l'Equation 11 (C.f. pII-49) et nécessite le calcul des vraisemblances a priori  $L_k^i(x)$  ; celles-ci sont regroupées dans la *matrice de causalité*  $M^i$  (C.f. Tableau 5, pII-49), calculée grâce aux matrices de confusion de chaque capteur.

La matrice de causalité des *spectres IR-TF filtrés* par algorithmes génétiques, calculée à partir de la matrice de confusion (Tableau 26, pIV-102), est présentée dans le Tableau 34. De même, celle du spectre UV filtré est détaillée dans le Tableau 35 (calculée à partir du Tableau 29). Les résultats de classification par PLS discriminante basée sur les courbes du capteurs d'arômes LCA1000 modifié (matrice de confusion du Tableau 23) fournissent la matrice de causalité du Tableau 36.

Tableau 34 : Matrice de causalité du spectromètre IR-TF

CLASSES PREDITES	CLASSES REELLES (107 ECHANTILLONS)			
	Sauvignon	Mauzac	Colombard	Autres cépages
Echantillons	44	14	14	35
$P_k$	0.41	0.13	0.13	0.33
$\hat{S}a$	0.98	0	0	0.09
$\hat{M}a$	0.02	0.86	0	0.17
$\hat{C}o$	0	0.14	1	0
$\hat{A}u$	0	0	0	0.74

Tableau 35 : Matrice de causalité du spectromètre UV

CLASSES PREDITES	CLASSES REELLES			
	Sauvignon	Mauzac	Colombard	Autres cépages
$\hat{S}a$	0.82	0.07	0.29	0.11
$\hat{M}a$	0.09	0.86	0.14	0.09
$\hat{C}o$	0.02	0.07	0.57	0.03
$\hat{A}u$	0.07	0	0	0.77

**Tableau 36 : Matrice de causalité du LCA avec traitement par PLS discriminante**

CLASSES PREDITES	CLASSES REELLES			
	Sauvignon	Mauzac	Colombard	Autres cépages
$\hat{S}a$	0.61	0.93	0.43	0.37
$\hat{M}a$	0	0.07	0	0
$\hat{C}o$	0.18	0	0.50	0.11
$\hat{A}u$	0.20	0	0.07	0.51

**(2) Fusion par risque minimum de Bayes**

La fusion par règle d'affectation de risque minimum de Bayes ajoute un coût de classification par rapport à la règle d'affectation d'erreur minimale de Bayes. Ce coût est déterminé en fonction de l'objectif de la discrimination et de l'expertise qui en résulte.

Le principal objectif de la discrimination de moûts de cépages blancs est de ne pas mélanger au cépage Sauvignon tout autre cépage de raisins blancs. Donc, le coût  $C(\hat{S}a|\bar{S}a)$  de classement d'un échantillon NON Sauvignon ( $\prod \bar{S}a$ ) comme un cépage Sauvignon ( $\prod S_a$ ) est très élevé ; par contre, affecter un Sauvignon à une autre classe de cépage est beaucoup moins grave, donc le coût  $C(\bar{S}a|S_a)$  est moins élevé. De même, confondre des classes autres que celle du Sauvignon n'est pas très important, donc le coût  $C(\bar{S}a|S_a)$  est encore moins élevé.

Nous définissons donc la fonction de coût  $C(k,h)$ , conforme à l'Equation 29, qui peut s'exprimer sous la forme de la matrice de coût  $\mathbf{C}$  (C.f. Tableau 37).

$$C: (k,h) \mapsto \left\{ \begin{array}{l} C(k|k) = 0, \quad \forall k \in G \\ C(h|k) = 1, \quad \forall h \neq k, (h,k) \in (\bar{S}a \times \bar{S}a) \\ C(h|S_a) = 10, \quad \forall h \in \bar{S}a \\ C(\hat{S}a|k) = 100, \quad \forall k \in \bar{S}a \end{array} \right\}, \text{ avec } \bar{S}a \in \{Ma, Co, Au\}.$$

**Equation 29 : définition de la fonction de coût**

**Tableau 37 : Matrice de Coût**

CLASSES PREDITES $\hat{\phantom{a}}$	CLASSES REELLES			
	Sauvignon (Sa)	Mauzac (Ma)	Colombard (Co)	Autres cépages (Au)
$\hat{S}a$	$C(\hat{S}a S_a)=0$	$C(\hat{S}a Ma)=100$	$C(\hat{S}a Co)=100$	$C(\hat{S}a Au)=100$
$\hat{M}a$	$C(\hat{M}a S_a)=10$	$C(\hat{M}a Ma)=0$	$C(\hat{M}a Co)=1$	$C(\hat{M}a Au)=1$
$\hat{C}o$	$C(\hat{C}o S_a)=10$	$C(\hat{C}o Ma)=1$	$C(\hat{C}o Co)=0$	$C(\hat{C}o Au)=1$
$\hat{A}u$	$C(\hat{A}u S_a)=10$	$C(\hat{A}u Ma)=1$	$C(\hat{A}u Co)=1$	$C(\hat{A}u Au)=0$

La règle d'affectation du risque moyen minimal suit l'Equation 16 (C.f. p II-51), sous l'hypothèse d'indépendance des capteurs :

$$\forall X \hat{Y} = \arg \min_{k \in G} \left\{ \sum_{h=1}^g \left( C(k|h) * p_h * \prod_{i=1}^N \Pr(X^i|h) \right) \right\}$$

Par exemple, si le capteur 1 affecte à l'objet  $x$  la classe Sauvignon (Sa) et le capteur 2 la classe « Autres cépages » (Au), alors la règle d'affectation pour deux capteurs suit l'Equation 30 :

$$\hat{Y}(x) = \arg \min \left\{ \begin{aligned} & C(\hat{S}a|Sa) * [p_{Sa} * (\Pr(\hat{S}a^1|Sa) * \Pr(\hat{A}u^2|Sa))] + C(\hat{S}a|Ma) * [p_{Ma} * (\Pr(\hat{S}a^1|Ma) * \Pr(\hat{A}u^2|Ma))] \\ & + C(\hat{S}a|Co) * [p_{Co} * (\Pr(\hat{S}a^1|Co) * \Pr(\hat{A}u^2|Co))] + C(\hat{S}a|Au) * [p_{Au} * (\Pr(\hat{S}a^1|Au) * \Pr(\hat{A}u^2|Au))] \\ & \dots \\ & C(\hat{A}u|Sa) * [p_{Sa} * (\Pr(\hat{S}a^1|Sa) * \Pr(\hat{A}u^2|Sa))] + C(\hat{A}u|Ma) * [p_{Ma} * (\Pr(\hat{S}a^1|Ma) * \Pr(\hat{A}u^2|Ma))] \\ & + C(\hat{A}u|Co) * [p_{Co} * (\Pr(\hat{S}a^1|Co) * \Pr(\hat{A}u^2|Co))] + C(\hat{A}u|Au) * [p_{Au} * (\Pr(\hat{S}a^1|Au) * \Pr(\hat{A}u^2|Au))] \end{aligned} \right\}$$

$$\hat{Y}(x) = \arg \min \left\{ \begin{aligned} & \left[ \begin{array}{cccc} C(\hat{S}a|Sa), C(\hat{S}a|Ma), C(\hat{S}a|Co), C(\hat{S}a|Au) \\ \dots \\ \dots \\ C(\hat{A}u|Sa), C(\hat{A}u|Ma), C(\hat{A}u|Co), C(\hat{A}u|Au) \end{array} \right] * \left[ \begin{array}{c} p_{Sa} * (\Pr(\hat{S}a^1|Sa) * \Pr(\hat{A}u^2|Sa)) \\ p_{Ma} * (\Pr(\hat{S}a^1|Ma) * \Pr(\hat{A}u^2|Ma)) \\ p_{Co} * (\Pr(\hat{S}a^1|Co) * \Pr(\hat{A}u^2|Co)) \\ p_{Au} * (\Pr(\hat{S}a^1|Au) * \Pr(\hat{A}u^2|Au)) \end{array} \right] \end{aligned} \right\}$$

$$\hat{Y}(x) = \arg \min \left\{ \mathbf{C} * \left[ \begin{array}{c} p_{Sa} * (\Pr(\hat{S}a^1|Sa) * \Pr(\hat{A}u^2|Sa)) \\ p_{Ma} * (\Pr(\hat{S}a^1|Ma) * \Pr(\hat{A}u^2|Ma)) \\ p_{Co} * (\Pr(\hat{S}a^1|Co) * \Pr(\hat{A}u^2|Co)) \\ p_{Au} * (\Pr(\hat{S}a^1|Au) * \Pr(\hat{A}u^2|Au)) \end{array} \right] \right\}$$

Equation 30 : règle d'affectation de risque minimum de Bayes appliquée à la discrimination de 4 classes de cépage<sup>22</sup>.

La matrice de coût est déterminée a priori, en fonction des objectifs et des contraintes de la classification de cépages. Cependant, la définition d'un coût de classification trop élevé dans la classe Sauvignon pourra aboutir à ne plus affecter aucun échantillon à la classe Sauvignon car le risque de se tromper est trop grand. Une démarche itérative de réglage des coûts peut alors être nécessaire, afin d'estimer le risque optimal qui minimise l'erreur de prédiction.

<sup>22</sup> Si tous les coûts sont égaux, alors la règle de risque minimal est égale à la règle de l'erreur minimale de Bayes

**IV. 3.2.2. Résultats**

Les meilleurs résultats individuels de discrimination sont fournis par les spectres IR-TF et UV filtrés (C.f. Figure 32, pIV-115). Nous nous intéresserons donc dans un premier temps à la fusion des données spectrométriques ; ces résultats seront ultérieurement fusionnés à ceux issus des capteurs d'arômes (pIV-122).

**(1) Fusion des spectromètres par erreur minimale de Bayes**

Les résultats de la fusion par erreur minimale de Bayes sur les spectres bruts ne présentent pas d'amélioration par rapport à ceux obtenus grâce aux données IR-TF seules, fournissant 30.8% d'erreur contre 30.7% d'erreur pour le spectromètre IR-TF seul (C.f. Annexe 8). Ceci peut s'expliquer par le fort taux d'erreur du spectromètre UV, obtenant 47.7% d'erreur grâce aux spectres bruts.

La fusion la plus intéressante est basée sur les meilleurs résultats des 2 spectromètres, c'est-à-dire sur la discrimination issue des spectres filtrés par algorithmes génétiques. Si l'on reprend l'exemple appliqué à la discrimination des cépages de l'Equation 11 (pII-49), alors la règle d'erreur minimale de Bayes affecte l'échantillon à la classe Sauvignon, compte tenu que le capteur 1 (IR-TF) l'a classé comme un Sauvignon (Sa) et le capteur 2 (UV) comme un « Autres cépages » (Au) (C.f. Equation 31) :

$$\hat{Y}(x) = \arg \max_{g \in \{1,2,3,4\}} \left\{ \begin{array}{l} P_{Sa} * (\Pr(\hat{S}a^1|Sa) * \Pr(\hat{A}u^2|Sa)) \\ P_{Ma} * (\Pr(\hat{S}a^1|Ma) * \Pr(\hat{A}u^2|Ma)) \\ P_{Co} * (\Pr(\hat{S}a^1|Co) * \Pr(\hat{A}u^2|Co)) \\ P_{Au} * (\Pr(\hat{S}a^1|Au) * \Pr(\hat{A}u^2|Au)) \end{array} \right\} = \arg \max_{g \in \{1,2,3,4\}} \left\{ \begin{array}{l} 0.41 * 0.98 * 0.07 \\ 0.13 * 0 * 0 \\ 0.13 * 0 * 0 \\ 0.33 * 0.09 * 0.77 \end{array} \right\} = \arg \max_{g \in \{1,2,3,4\}} \left\{ \begin{array}{l} 0.028 \\ 0 \\ 0 \\ 0.023 \end{array} \right\} = 1 \Leftrightarrow Sauvignon$$

**Equation 31 : règle d'affectation d'erreur minimale de Bayes appliquée à la discrimination de 4 classes de cépage.**

Quand les 2 capteurs ne sont pas en accord, la règle d'affectation donne principalement raison à la prédiction issue des mesures du spectromètre IR-TF, alors que la classification du spectromètre UV n'est respectée que dans un seul cas (1 case grisée sur 12 dans Tableau 38). Ceci s'explique par la différence entre les performances individuelles des 2 capteurs, le spectromètre IR-TF obtenant 9.7% d'erreur de discrimination alors que l'UV fournit 22.9% d'erreur.

**Tableau 38 : matrice d'affectation d'erreur minimale pour les deux capteurs IR-TF et UV**  
La case vide (∅) correspond à un cas qui ne s'est jamais présenté au cours de l'apprentissage.

CLASSES PREDITES par le capteur IR-TF	CLASSES PREDITES PAR LE CAPTEUR UV			
	$\hat{S}a$	$\hat{M}a$	$\hat{C}o$	$\hat{A}u$
$\hat{S}a$	Sa	Sa	Sa	Sa
$\hat{M}a$	Ma	Ma	Ma	Au
$\hat{C}o$	Co	Co	Co	∅
$\hat{A}u$	Au	Au	Au	Au

La fusion basée sur l'erreur minimale de Bayes améliore nettement le résultat de classification, obtenant 6.5% d'erreur totale, soient 7 échantillons mal classés sur 107 (C.f. Tableau 39).

Tableau 39 : résultats de la fusion par erreur minimale de Bayes des spectres IR-TF & UV filtrés

Nombre d'échantillons	CLASSES REELLES				TOTAL
	44	14	14	35	107
CLASSES PREDITES	Sauvignon	Mauzac	Colombard	Autres cépages	% erreur
Sauvignon	43	0	0	3	6.5%
Mauzac	1	12	0	1	14.3%
Colombard	0	2	14	0	12.5%
Autres cépages	0	0	0	31	0%
% bons classements	97.7%	85.7%	100%	88.6%	93.5%
Taux d'erreur par erreur minimale de Bayes	2.3%	14.3%	0%	11.4%	6.5%

TAUX D'ERREUR IR-TF	2.2%	7.1%	0%	25.7%	9.65%
Taux d'erreur UV	18.2%	14.3%	42.9%	24.3%	22.9%

(2) Fusion des spectromètres par risque minimum de Bayes

Malgré l'amélioration substantielle du taux d'erreur global grâce à la fusion par la règle d'affectation d'erreur minimale de Bayes, il subsiste 1 échantillon de Sauvignon affecté au cépage Mauzac, et surtout, 3 échantillons d'autres cépages attribués au Sauvignon (C.f. Tableau 39).

Si nous appliquons un coût de classification (C.f. Equation 29, p IV-117) à l'exemple de la page IV-119, où les deux spectromètres IR-TF et UV affectent un échantillon respectivement à la classe Sauvignon (Sa) et la classe « Autres cépages » (Au), alors la règle d'affectation du risque minimum attribue cet échantillon à la classe « Autres cépages » (C.f. Equation 32) et non pas à la classe « Sauvignon », comme c'était le cas avec la règle d'erreur minimale de Bayes.

$$\hat{Y}(x) = \arg \min_{g \in \{1,2,3,4\}} \left\{ \begin{matrix} P_{Sa} * (\Pr(\hat{S}a^1|Sa) * \Pr(\hat{A}u^2|Sa)) \\ P_{Ma} * (\Pr(\hat{S}a^1|Ma) * \Pr(\hat{A}u^2|Ma)) \\ P_{Co} * (\Pr(\hat{S}a^1|Co) * \Pr(\hat{A}u^2|Co)) \\ P_{Au} * (\Pr(\hat{S}a^1|Au) * \Pr(\hat{A}u^2|Au)) \end{matrix} \right\} = \arg \min_{g \in \{1,2,3,4\}} \left\{ \begin{matrix} \begin{bmatrix} 0 & 100 & 100 & 100 \\ 10 & 0 & 1 & 1 \\ 10 & 1 & 0 & 1 \\ 10 & 1 & 1 & 0 \end{bmatrix} * \begin{bmatrix} 0.41 * 0.98 * 0.07 \\ 0.13 * 0 * 0 \\ 0.13 * 0 * 0 \\ 0.33 * 0.09 * 0.77 \end{bmatrix} \\ \begin{bmatrix} 0 & 100 & 100 & 100 \\ 10 & 0 & 1 & 1 \\ 10 & 1 & 0 & 1 \\ 10 & 1 & 1 & 0 \end{bmatrix} * \begin{bmatrix} 0.028 \\ 0 \\ 0 \\ 0.023 \end{bmatrix} \end{matrix} \right\} = \arg \min_{g \in \{1,2,3,4\}} \left\{ \begin{matrix} 2.3 \\ 0.303 \\ 0.303 \\ 0.28 \end{matrix} \right\} = 4 \Leftrightarrow \text{Autres cépages}$$

Equation 32 : règle d'affectation de risque minimum de Bayes appliquée à la discrimination de 4 classes de cépage par les spectromètres IR-TF et UV

La matrice d'affectation du risque minimal est détaillée dans le Tableau 40. Seule une affectation a changé par rapport à la règle d'affectation d'erreur minimale de Bayes, correspondant à l'exemple

traité ci-dessus (case quadrillée dans le Tableau 40). L'introduction d'un coût d'attribution à la classe Sauvignon élevé restreint le nombre d'affectations à cette classe.

**Tableau 40 : règle d'affectation du risque minimum pour les deux capteurs IR-TF et UV**  
(Seule la case en noir qui a changé par rapport à la règle d'erreur minimale)

CLASSES PREDITES par le capteur IR-TF	CLASSES PREDITES par le capteur UV			
	$\hat{S}a$	$\hat{M}a$	$\hat{C}o$	$\hat{A}u$
$P_k$	0.44	0.13	0.13	0.33
$\hat{S}a$	Sa	Sa	Sa	Au
$\hat{M}a$	Ma	Ma	Ma	Au
$\hat{C}o$	Co	Co	Co	∅
$\hat{A}u$	Au	Au	Au	Au

Le résultat de la fusion par risque minimal de Bayes est détaillé dans la matrice de confusion suivante (C.f. Tableau 41). Le taux d'erreur global est identique à celui obtenu grâce à la règle d'erreur minimale, mais l'ajout d'un coût permet de ne plus affecter d'échantillons d'autres cépages au cépage Sauvignon.

Cependant, en évitant ces erreurs (case grisée du Tableau 41), la règle d'affectation provoque des erreurs contraires, c'est-à-dire qu'elle attribue la classe « Autres cépages » à 3 échantillons de Sauvignon (case noircie du Tableau 41).

**Tableau 41 : résultats de la fusion par la règle du risque minimum de Bayes entre les spectres IR-TF et UV filtrés**

Nombre d'échantillons	CLASSES REELLES				TOTAL
	Sauvignon	Mauzac	Colombard	Autres cépages	% erreur
Sauvignon	40	0	0	0	0%
Mauzac	1	12	0	1	14.3%
Colombard	0	2	14	0	12.5%
Autres cépages	3	0	0	34	8.1%
% bons classements	90.9%	85.7%	100%	97.1%	93.5%
Taux d'erreur par risque minimum de Bayes	9.1%	14.3%	0%	2.9%	6.5%

Taux d'erreur IR-TF	2.2%	7.1%	0%	25.7%	9.65%
Taux d'erreur UV	18.2%	14.3%	42.9%	24.3%	22.9%

Il apparaît donc que 6 échantillons sont classés « Sauvignon » par le spectromètre IR-TF et « Autres cépages » par le spectromètre UV, alors que 3 d'entre eux sont des échantillons de Sauvignon et 3 appartiennent réellement à la classe « autres cépages ».

Il est donc impossible par cette méthode d'obtenir 100% de bonne discrimination pour ces 2 classes. il semble donc judicieux d'affiner les règle d'affectation de Bayes, en y introduisant les degrés

d'appartenance aux classes, afin de pouvoir moduler l'affectation des échantillons. Cette démarche sera développée en conclusion (pIV-125).

### (3) Fusion des spectromètres et des capteurs d'arômes par erreur minimale de Bayes

La fusion par erreur minimale de Bayes est appliquée à la classification par PLS discriminante basée sur les courbes du capteurs d'arômes LCA1000 modifié et aux résultats de la discrimination des spectromètres IR-TF et UV filtrés.

La matrice d'affectation, fonction des matrices de causalité des 3 capteurs, est représentée (dépliée en 2D) dans Tableau 42.

**Tableau 42 : règle d'affectation d'erreur minimale pour les capteurs d'arômes, IR-TF et UV**

Les cases grisées correspondent aux classifications qui diffèrent par rapport à l'affectation basée uniquement sur les spectromètres.

CLASSES PREDITES par le LCA1000 $\hat{p}$		CLASSES PREDITES par le spectromètre IR-TF $\hat{p}$				
		$\hat{S}a$	$\hat{M}a$	$\hat{C}o$	$\hat{A}u$	
S p e c t	$\hat{S}a$	$\hat{S}a$	Sa	Ma	Co	Au
		$\hat{M}a$	∅	Ma	Ma	∅
		$\hat{C}o$	Sa	Sa**	Co	Au
		$\hat{A}u$	Sa	Au	Co	Au
r o m è t	$\hat{M}a$	$\hat{S}a$	Sa	Ma	Ma*	Au
		$\hat{M}a$	∅	Ma	Ma	∅
		$\hat{C}o$	Sa	Au	Co	Au
		$\hat{A}u$	Sa	Au	Co	Au
r e	$\hat{C}o$	$\hat{S}a$	Sa	Ma	Co	Au
		$\hat{M}a$	∅	Ma	Ma	∅
		$\hat{C}o$	Sa	Au	Co	Au
		$\hat{A}u$	Sa	Au	Co	Au
U V	$\hat{A}u$	$\hat{S}a$	Sa	Au	Co	Au
		$\hat{M}a$	∅	∅	∅	∅
		$\hat{C}o$	Sa	Au	∅	Au
		$\hat{A}u$	Sa	Au	∅	Au

Cette fusion incluant la contribution des capteurs d'arômes apporte une amélioration sensible du taux d'erreur global, qui diminue de 6.5% à 4.7% (soit de 2 échantillons) grâce au bon classement de TOUS les Sauvignons et TOUS les Mauzac (C.f. Tableau 43).

Ces 2 échantillons maintenant bien classés proviennent des changements d'affectation par rapport à la fusion des spectromètres seuls : celle-ci n'est pas due à la pertinence de la classification par capteurs d'arômes (car il ne fournit pas la classe correcte !), mais au désaccord qu'il crée entre les 2 spectromètres, entraînant la domination du choix du spectromètre UV (C.f. cases grisées du Tableau 42). Ce changement d'affectation entraîne le mauvais classement d'un échantillon de Colombard (affectation au Mauzac : \* dans le Tableau 42) et le bon classement d'un Sauvignon (\*\*) et de 2 Mauzac (\*).

Cependant, 3 échantillons appartenant à la classe « autres cépages » restent attribués à la classe Sauvignon, comme pour la fusion des spectromètres IR-TF et UV par erreur minimale de Bayes (C.f. Tableau 39, pIV-120).

**Tableau 43 : résultats de la fusion par erreur minimale de Bayes entre les capteurs d'arômes et les spectromètres IR-TF & UV.**

Nombre d'échantillons	CLASSES REELLES				TOTAL
	44	14	14	35	107
CLASSES PREDITES	Sauvignon	Mauzac	Colombard	Autres cépages	%erreur
Sauvignon	44	0	0	3	6.4%
Mauzac	0	14	1	1	12.5%
Colombard	0	0	13	0	0%
Autres cépages	0	0	0	31	0%
% bons classements	100%	100%	92.9%	88.6%	95.3%
Taux d'erreur FUSION erreur minimale de Bayes	0%	0%	7.1%	11.4%	4.7%
Taux d'erreur IR-TF	2.2%	7.1%	0%	25.7%	9.65%
Taux d'erreur UV	18.2%	14.3%	42.9%	24.3%	22.9%
Taux d'erreur LCA1000	32.7%	94.1%	43.3%	48.7%	47.1%

#### (4) Fusion des spectromètres et des capteurs d'arômes par risque minimal de Bayes

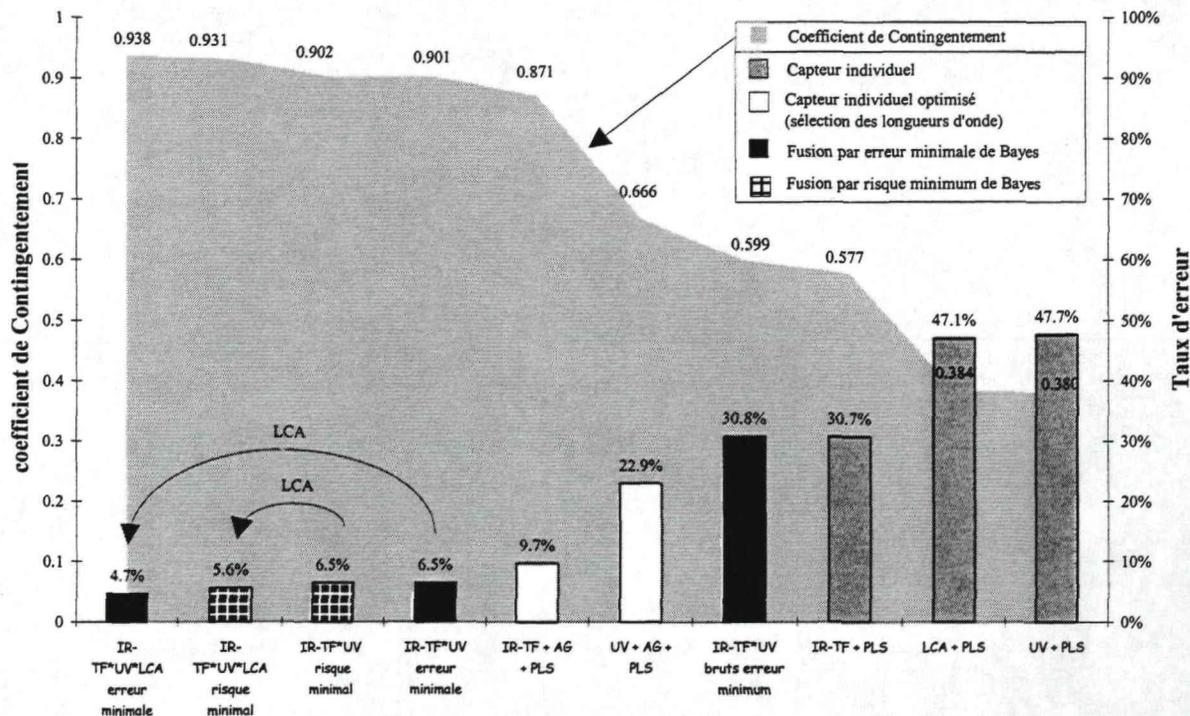
L'introduction des capteurs d'arômes dans la fusion par risque minimum de Bayes diminue légèrement le taux d'erreur de 6.5 à 5.6%, grâce au bon classement de tous les échantillons de Mauzac. L'introduction d'un coût de classification induit une contrainte qui freine l'affectation des échantillons ambigus à la classe Sauvignon, mais qui favorise la non-reconnaissance des échantillons réels de Sauvignon. Comme pour la fusion des spectromètres, 4 échantillons de Sauvignon sont attribués à d'autres classes, mais aucun échantillon n'est attribué de façon erronée à la classe Sauvignon.

**Tableau 44 : résultats de la fusion par risque minimum de Bayes entre les capteurs d'arômes et les spectromètres IR-TF et UV**

Nombre d'échantillons	CLASSES REELLES				TOTAL
	44	14	14	35	107
CLASSES PREDITES	Sauvignon	Mauzac	Colombard	Autres cépages	% erreur
Sauvignon	40	0	0	0	0%
Mauzac	0	14	1	1	12.5%
Colombard	0	0	13	0	0%
Autres cépages	4	0	0	34	10.5%
% bons classements	90.1%	100%	92.9%	97.1%	94.4%
Taux d'erreur par risque minimum de Bayes	9.9%	0%	7.1%	2.9%	5.6%
TAUX D'ERREUR IR-TF	2.2%	7.1%	0%	25.7%	9.65%
Taux d'erreur UV	18.2%	14.3%	42.9%	24.3%	22.9%
Taux d'erreur LCA1000	32.7%	94.1%	43.3%	48.7%	47.1%

**(5) Conclusion et discussion****(a) Synthèse des performances**

La fusion de niveau haut a été réalisée en généralisant la classification par erreur et risque minimum de Bayes. Les performances des différents systèmes de fusion sont comparées dans la Figure 33.



**Figure 33 : taux d'erreur et coefficient de contingement des différentes classifications**

IR-TF: spectromètre infrarouge; UV: spectromètre ultraviolet; LCA: capteurs d'arômes; « sélectionné »: filtré par algorithmes génétiques; PLS: *Partial Least Square* discriminante; « +AG »: spectres filtrés par algorithmes génétiques.

La fusion Bayésienne améliore nettement le résultat individuel des capteurs. La performance globale des 2 démarches (erreur minimale de Bayes ou risque minimum de Bayes) est similaire (1 échantillon mal classé d'écart faisant passer le taux d'erreur de 4.7 à 5.6%). La règle d'affectation de risque minimum de Bayes impose une contrainte de coût supplémentaire qui oriente la classification en fonction des objectifs de la discrimination.

Ainsi, grâce à la détermination experte d'un coût, cette démarche ne fait aucune erreur d'attribution à la classe Sauvignon d'échantillons appartenant à d'autres cépages de raisins blancs. Cependant, l'adjonction d'un coût élevé d'attribution à la classe Sauvignon limite la reconnaissance des échantillons ambigus de Sauvignon, qui sont alors affectés à d'autres classes de cépage (C.f. **Tableau 41**, pIV-121 & **Tableau 44**, pIV-123).

D'autre part, malgré le fort taux d'erreur présenté par les capteurs d'arômes (47.1%), le résultat de leur discrimination apporte une légère amélioration à la fusion des spectromètres IR-TF et UV (C.f. **Figure**

33), alors que l'adjonction des signaux des capteurs d'arômes à ceux des spectromètres lors de la fusion de niveau bas dégradait la classification (C.f. Annexe 7). Ceci correspond à 1 échantillon de Sauvignon et 2 de Mauzac improprement affectés par les spectromètres IR-TF et UV et qui ont été bien ré-affectés par l'ajout du LCA, et 1 Colombard mal classé en Mauzac. En effet, le capteur d'arômes affecte presque systématiquement les Mauzac à la classe Sauvignon, alors qu'il le fait 2 fois moins pour les Colombard. Le conflit IR-TF/UV sur l'attribution d'échantillons ambigus Mauzac/Colombard est donc régie par le LCA, avec le gain d'un échantillon bien classé (+2 Mauzac - 1 Colombard).

(b) Proposition d'améliorations de la fusion bayésienne

Au cours de notre étude, il est apparu que les techniques de fusion Bayésiennes employées dans cette étude sont un cas particulier appliqué à la fusion de classifications booléennes. Or, l'étape de classification par les capteurs individuels peu donner pour chaque individu un degré d'appartenance à chaque classe. La décision finale individuelle correspond à attribuer 1 au degré d'appartenance le plus élevé et 0 aux autres.

Il est apparu avantageux, par exemple lors la fusion des spectromètres (pIV-122), d'améliorer la fusion par raisonnement de Bayes en tenant compte des degrés d'appartenance - et non plus des valeurs binaires -, afin de prendre en considération la « confiance » associée à la classification de chaque échantillon.

Nous proposons donc d'inclure les degrés d'appartenance dans le calcul des règles d'affectation. La règle d'affectation de l'erreur minimale de Bayes, est détaillée dans l'Equation 34 et étendue à l'utilisation des degrés d'appartenance dans l'Equation 35.

Soit  $D^i(X)$ , le vecteur des degrés d'appartenance aux  $g$  classes de l'échantillon  $X$  attribué par le capteur  $i$  :  $D^i(X) = \{x_1^i, \dots, x_g^i\}$ . Notons  $x_j^i = \text{argmax}\{x_1^i, \dots, x_g^i\}$  la classe  $j$  choisie par le capteur  $i$ , dans le cas où les degrés d'appartenance sont transformés en une classification booléenne en attribuant l'échantillon à la classe pour laquelle le degré d'appartenance est maximal.

Rappelons que les probabilités a priori qu'a le capteur  $i$  d'affecter un échantillon  $X$  à toute classe  $j$  sachant qu'il appartient à la classe  $k$  sont rassemblées dans la matrice de causalité  $M^i$  :

$$M^i = \{\Pr(x_j^i | k) = L_k^i(x)\}$$

La règle d'affectation d'erreur minimale de Bayes correspond à l'Equation 11, pII-49 :

$$\forall X \hat{Y} = \arg \max_{k \in G} \left\{ p_k * \prod_{i=1}^N \Pr(X^i | k) \right\} = \arg \max_{k \in G} \left\{ \begin{matrix} p_1 * \Pr(X^1 | 1) * \dots * \Pr(X^N | 1) \\ \dots \\ p_g * \Pr(X^1 | g) * \dots * \Pr(X^N | g) \end{matrix} \right\}$$

$$= \arg \max_{k \in G} \left\{ \begin{matrix} p_1 * \left[ \prod_{i=1}^N M^i \right] \\ \dots \\ p_g * \left[ \prod_{i=1}^N M^i \right] \end{matrix} \right\} = \arg \max_{k \in G} \left\{ \left[ \prod_{i=1}^N M^i \right] * P \right\}$$

en posant  $P = \begin{bmatrix} p_1 & 0 & 0 \\ 0 & \dots & 0 \\ 0 & 0 & p_g \end{bmatrix}$ , la matrice des probabilité a priori des classes.

matrice de causalité du capteur 1      matrice de causalité du capteur N

Multiplication terme à terme

**Equation 33 : règle d'affectation d'erreur minimale de Bayes sous forme matricielle**

Cette équation a été appliquée aux classifications booléennes, en considérant que chaque capteur  $i$  n'affectait l'échantillon qu'à une classe unique  $j$ . La probabilité a priori était alors réduite à :

$\Pr(X^i | k)^i = [0, \dots, 0, \Pr(x_j^i | k), 0, \dots, 0]$ . Alors, l'Equation 33 devient :

$$\hat{Y} = \arg \max_{k \in G} \left\{ \left[ \prod_{i=1}^N (A^i(X) * M^i) \right] * P \right\}, \text{ où } A^i(X) \text{ est le vecteur d'affectation booléen de l'échantillon } X \text{ à}$$

Classe  $j$

la classe  $j$  par le capteur  $i$  :  $A^i(X) = [0, \dots, 0, 1, 0, \dots, 0]$ .

**Equation 34 : fusion par la règle d'affectation d'erreur minimale de Bayes pour des classifications booléennes**

L'Equation 34 se simplifie énormément lors de son application, grâce à la présence des nombreux 0 de  $A^i$  ; un exemple a été présenté dans l'Equation 12, pII-50.

Par contre, si on prend en considération les degrés d'appartenance pour toutes les classes  $D^i(X)$ , alors la probabilité a priori d'un échantillon  $X$  pour le capteur  $i$  est :  $\Pr(X^i | k)^T = \{ \Pr(x_1^i | k), \dots, \Pr(x_g^i | k) \}$

En remplaçant les attributions booléennes par les degrés d'appartenance, l'Equation 34 devient :

$$\forall X \hat{Y} = \arg \max_{k \in G} \left\{ \left[ \prod_{i=1}^N (D^i(X) * M^i) \right] * P \right\}$$

**Equation 35 : fusion par la règle d'affectation d'erreur minimale de Bayes pour les classifications en degrés d'appartenance**

De plus, au cours du développement théorique de la fusion, nous avons fait le choix d'un opérateur d'agrégation des vraisemblances conditionnelles des capteurs (C.f. l'Equation 8, pII-48). Notre sélection s'est portée sur un opérateur très strict, celui de l'intersection, qui revient à effectuer le produit des probabilités, sous l'hypothèse d'indépendance des capteurs.

Si nous voulons analyser l'effet de l'opérateur sur le résultat de la fusion, alors divers opérateurs peuvent être intégrés dans le calcul de la règle d'affectation d'erreur minimale de Bayes en suivant l'Equation 36, généralisation de l'Equation 35 :

$$\forall X \hat{Y} = \arg \max_{k \in G} \left\{ \left[ \otimes (D^i(X) * M^i) \right] * P \right\}$$

où  $\otimes$  est l'opérateur d'agrégation choisi

**Equation 36 : fusion par la règle d'affectation d'erreur minimale de Bayes pour les classifications en degrés d'appartenance et le choix d'un opérateur d'agrégation**

Il en est de même pour la fusion par la règle du risque minimal de Bayes. Celle-ci se généralise en fonction des degrés d'appartenance et de différents opérateurs d'agrégation sous la forme de l'Equation 37:

$$\forall X \hat{Y} = \arg \min_{k \in G} \left\{ C * \left( \left[ \otimes [D^i(X)]^i * M^i \right] * P \right) \right\}$$

où  $C$  est la matrice de coût.

**Equation 37 : fusion par la règle d'affectation de risque minimum de Bayes pour les classifications en degrés d'appartenance et le choix d'un opérateur d'agrégation**

Ces propositions de généralisation de la fusion basée sur la théorie Bayésienne seront développées dans le prolongement de la thèse. Il pourra être utile d'estimer les probabilités conditionnelles des différents capteurs, afin de pouvoir appliquer la fusion Bayésienne à des signaux non indépendants. De plus, dans le cas de la PLS discriminante, les sorties prédites ne sont pas de vraies degrés d'appartenance ; il est donc nécessaire d'établir au préalable une méthode de transformation des sorties en degrés d'appartenance.

### IV. 3.3. Pertinence de la fusion

#### IV. 3.3.1. Estimateurs de classification

Le taux d'erreur et le coefficient de contingement de toutes les classifications sont comparés dans la Figure 34. La classification individuelle de chaque capteur montre que le spectromètre IR-TF est le capteur le plus approprié pour la discrimination des moûts de cépages blancs.

Les capteurs d'arômes sont les moins aptes à classer les échantillons, confirmant que la fraction aromatique des moûts n'est pas suffisamment développée pour permettre une caractérisation concrète des cépages. De plus, la mesure des capteurs d'arômes est relativement peu sensible (seuil de détection de l'ordre du ppm) et extrêmement influencée par les conditions expérimentales, induisant une forte dérive à court terme des mesures. Ces causes peuvent empêcher toute discrimination, si les différences aromatiques entre les cépages sont infimes.

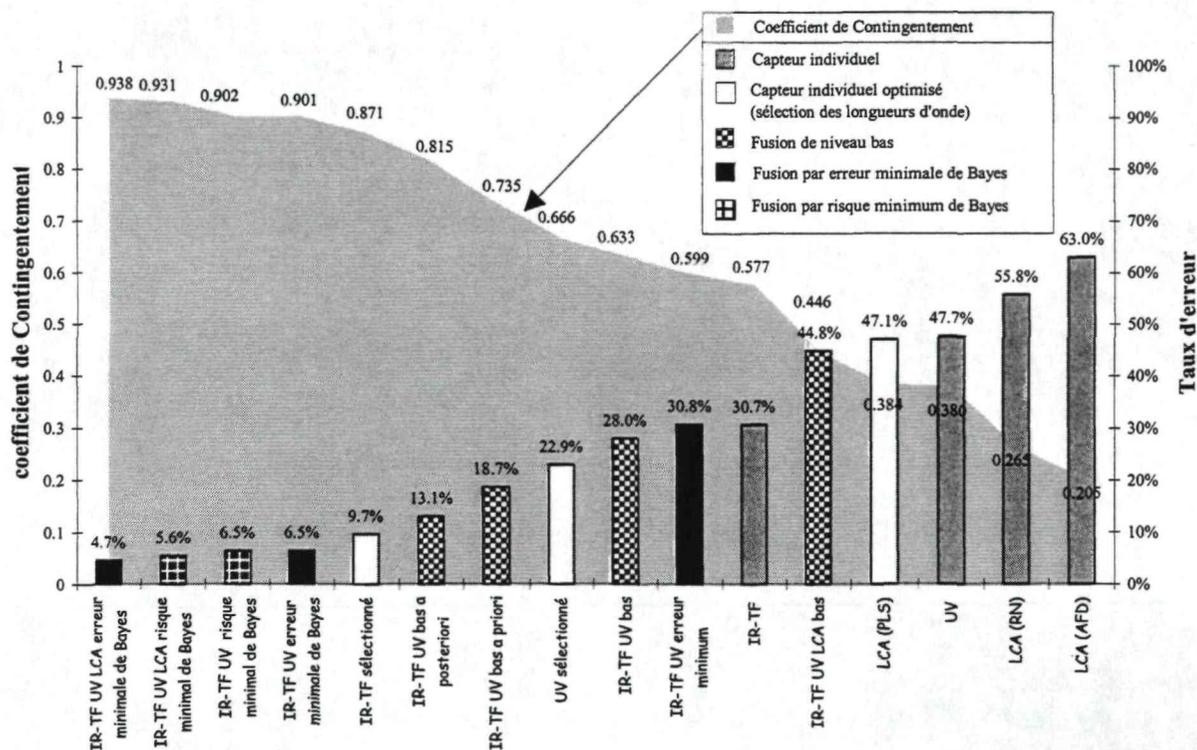


Figure 34 : taux d'erreur et coefficient de contingement pour les différents capteurs et de la méthodes de fusion

Quelle que soit la méthode de fusion, son application fournit des résultats meilleurs que ceux du spectromètre UV et des capteurs d'arômes pris individuellement. Les spectres IR-TF dont les longueurs d'onde ont été sélectionnées sont extrêmement performants ; le taux d'erreur obtenu est meilleur que ceux fournis par les fusions de niveau bas.

Par contre, les fusions de niveau haut sont beaucoup plus performantes que celle de niveau bas ; elles bénéficient directement de l'amélioration des discriminations induite par la sélection des longueurs d'onde et introduisent une non-linéarité grâce aux *degrés de confiance* attribués aux capteurs par classe sous forme de probabilité a priori. Il est donc préférable d'effectuer la fusion des classifications et non des données brutes.

Les deux démarches de fusion de Bayes présentent des performances équivalentes, mais la règle d'affectation de risque minimum de Bayes permet d'effectuer une classification sous contrainte, afin d'orienter la discrimination grâce à *l'expertise*. Dans certains cas, en particulier pour du contrôle qualité lié à la sécurité, il peut être extrêmement profitable de respecter des contraintes lors de la classification afin de ne pas se tromper sur certaines affectations cruciales, même si en contre partie le taux d'erreur augmente légèrement.

### IV. 3.3.2. Redondance des capteurs

La pertinence de la fusion s'exprime au travers du taux d'erreur et du coefficient de contingement, mais elle s'explique par la redondance et la complémentarité des capteurs. Celles-ci sont analysées par le calcul des corrélations entre variables, visualisées sur le cercle des corrélations. De plus, l'analyse de la corrélation entre les capteurs permet de valider l'hypothèse d'indépendance, supposée lors de la fusion Bayésienne.

Dans un premier temps, les corrélations sont calculées en tenant compte des spectres bruts (C.f. Figure 35).

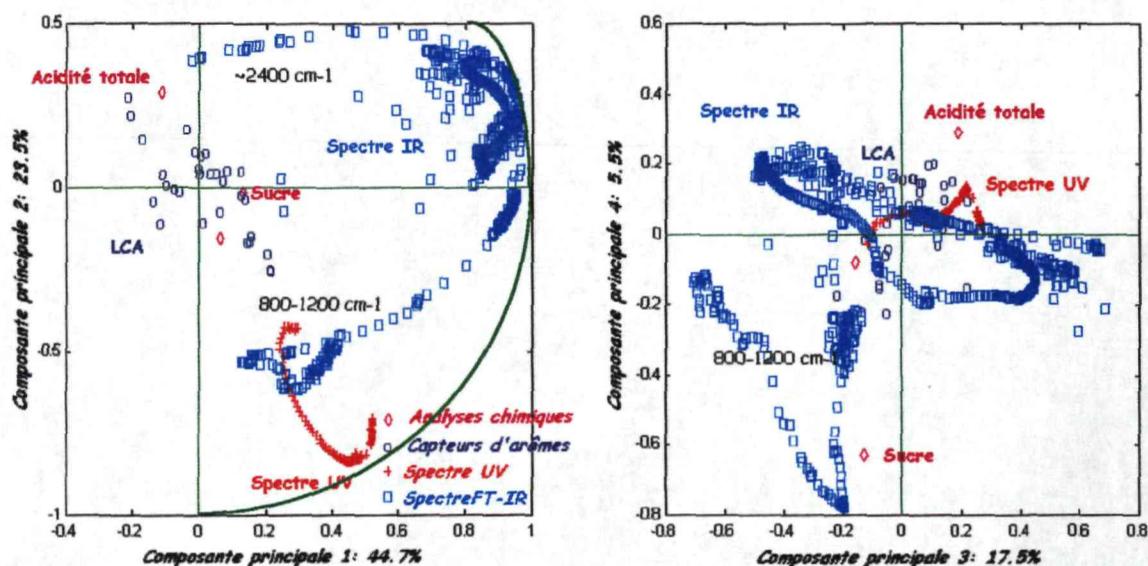


Figure 35 : corrélation entre les analyses chimiques (sucre, Acidité totale et pH (◆)), les descripteurs extraits du LCA1000 (O), les spectres UV (+) et IR-TF bruts (■).

Le premier axe est essentiellement construit par les données du spectromètre IR-TF, alors que le second axe est principalement basé sur le spectre UV. Cependant, une partie du spectre IR-TF (longueurs d'onde comprises entre 800 et 1200  $\text{cm}^{-1}$ ) est très corrélée avec le spectre UV, participant à la construction du second axe. Le troisième axe est aussi construit par le spectre IR-TF, alors que les données chimiques et aromatiques se répartissent sur la quatrième composante principale. Le spectre IR-TF étant très large (800 points), il domine la construction de cette ACP et, quelle que soit la variable, elle est corrélée à quelques longueurs d'onde du spectre IR-TF.

Cependant, au cours de l'étude de la classification des cépages, les longueurs d'onde des 2 spectres ont été sélectionnées par algorithmes génétiques, afin d'améliorer la discrimination en éliminant le bruit et les variables redondantes.

Lors de la fusion, 2 stratégies de sélection des longueurs d'onde ont été adoptées : la sélection « a priori », individuellement pour chaque capteur, qui a été employée pour la fusion de niveau bas « a

priori » et de **niveau haut**, et la sélection de longueurs d'onde « a posteriori » dans le spectre obtenu par concaténation des spectres UV et IR-TF .

La corrélation entre les différentes variables après la sélection des longueurs d'onde a priori est représentée par la **Figure 36**. Le premier axe est essentiellement construit par les longueurs d'onde du spectromètre IR-TF comprises entre 1040 et 1440  $\text{cm}^{-1}$  (C.f. spectre sélectionné **Figure 27**, pIV-104) ainsi que par celles du spectre UV supérieures à 290nm (C.f. spectre sélectionné **Figure 28**, pIV-106) . Ces deux parties de spectres semblent donc être redondantes, les composantes principales 3 et 4 ne les séparant pas non plus.

Il existe donc une corrélation très limitée entre les 2 spectromètres, ce qui permet de justifier l'hypothèse d'indépendance sur laquelle repose la combinaison des probabilités a priori des différents capteurs lors de la fusion Bayésienne.

Le second axe est principalement basé sur le spectre IR-TF, grâce aux longueurs d'onde comprises entre 2160 et 2720  $\text{cm}^{-1}$ . Les analyses chimiques ainsi que les mesures des capteurs d'arômes ont peu d'influence sur le premier plan principal, mais caractérisent plutôt le troisième axe (acidité totale, sucre ainsi que les longueurs d'onde entre 250-280 nm du spectre UV) et le quatrième axe (descripteurs du LCA1000).

La sixième composante principale (1.6% d'inertie), non figurée ici, est entièrement construite sur les descripteurs des capteurs d'arômes, indiquant le peu de corrélation entre ces données et toutes les autres. L'hypothèse d'indépendance des capteurs d'arômes et des spectromètres supposée lors de la fusion Bayésienne est donc pleinement justifiée.

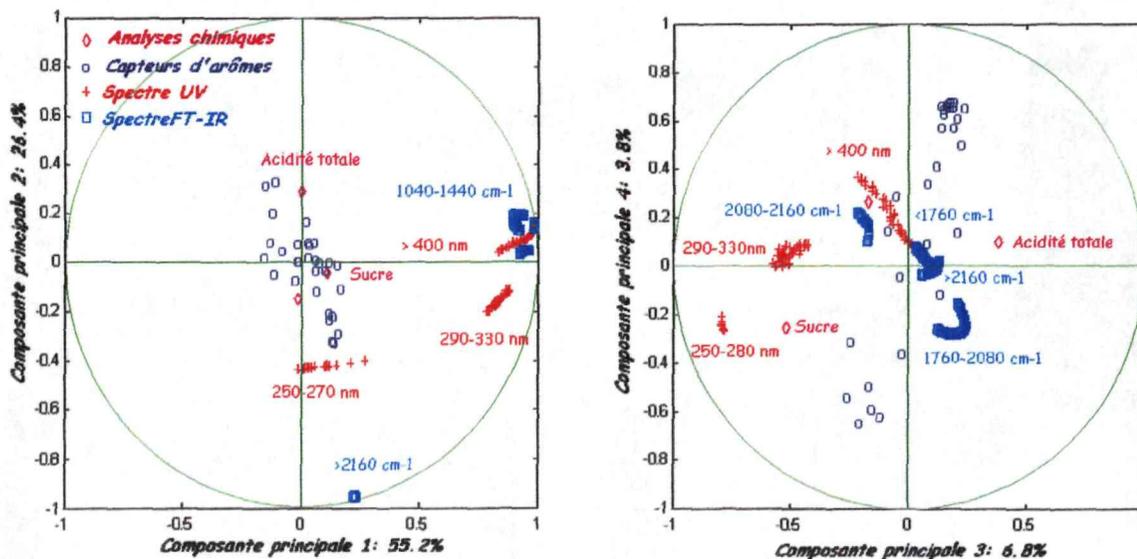


Figure 36 : corrélation entre les analyses chimiques (sucre, Acidité totale et pH), les descripteurs extraits du LCA1000, les spectres UV et IR-TF sélectionnés a priori

La sélection des longueurs d'onde a posteriori devrait éliminer les longueurs d'onde redondantes entre le spectre IR-TF et UV (C.f. spectre sélectionné Figure 31, pIV-113). Cependant, la Figure 37 montre que la sélection a posteriori est moins stricte que celle a priori et conserve les parties redondantes des spectres : les longueurs d'onde UV supérieures à 400nm et les IR-TF inférieures à 1050  $\text{cm}^{-1}$ . Le quatrième axe est largement bâti par les données chimiques alors que le cinquième est exclusivement construit par les descripteurs des capteurs d'arômes.

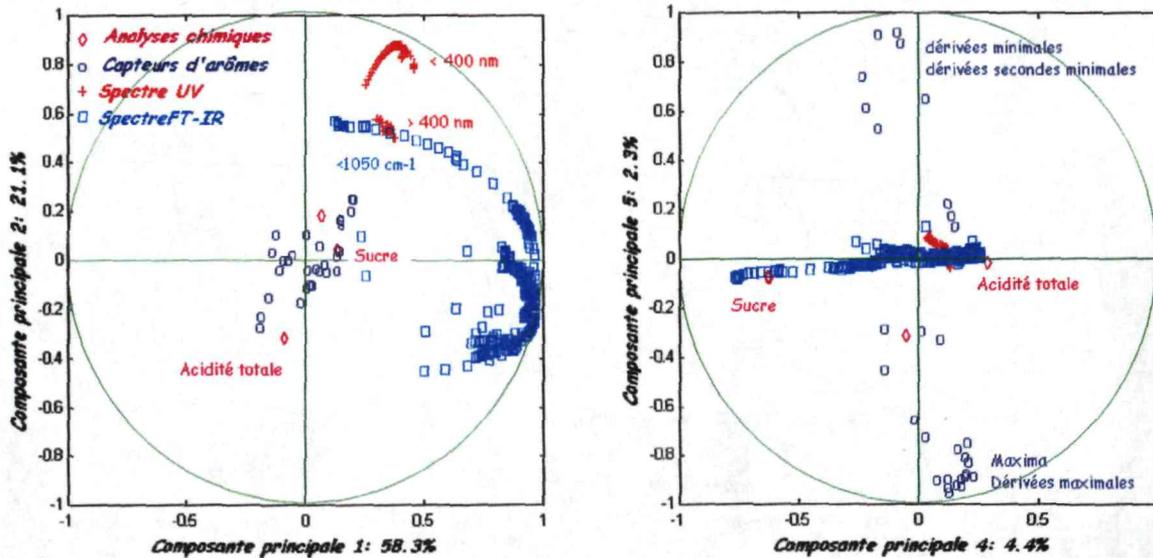


Figure 37 : corrélation entre les analyses chimiques (sucre, Acidité totale et pH), les descripteurs extraits du LCA1000, les spectres UV et IR-TF sélectionnés a posteriori

La redondance entre les mesures aromatiques, chimiques et spectrales est presque inexistante. Ces 3 types de mesures sont donc extrêmement complémentaires et adaptées à la fusion de données, malgré quelques longueurs d'onde redondantes entre les spectres UV et IR-TF.

## IV. 4. CONCLUSION

La discrimination de cépages de moûts de raisins blancs est donc possible, grâce au concours des capteurs d'arômes, spectromètres IR-TF et UV. La classification uniquement basée sur les capteurs d'arômes est insuffisante (47.1% d'erreur) et nécessite l'adjonction d'autres capteurs, aptes à mesurer différentes caractéristiques du moût en phase liquide.

Ce système multi-capteur dédié à la mesure organoleptique est remarquablement efficace pour la classification variétale des moûts de raisins blancs, grâce à l'application successive de *méthodes d'optimisation* et de *fusion*.

Les capteurs d'arômes ont été préalablement optimisés, grâce à l'application des méthodologies proposées dans le chapitre III. Néanmoins, une méthode développée ultérieurement, la *PLS discriminante*, s'est avérée extrêmement performante, même si elle doit être explorée plus avant.

Les résultats des spectromètres sont aussi amplement améliorés, en s'appuyant sur la technique des *algorithmes génétiques* dédiée à la sélection des longueurs d'onde : les résultats du spectromètre infrarouge passent de 30.7% à 9.7% d'erreur et de 47.7% à 22.9% d'erreur pour le spectromètre UV.

Lors de cette application, nous avons montré que l'exécution des algorithmes génétiques devait être répétée une dizaine de fois afin d'obtenir les meilleurs résultats, car cette démarche est non déterministe. Les différentes méthodes proposées pour le choix final des longueurs d'onde ont révélé que la meilleure sélection de variables était effectuée par le chromosome le plus performant ou par la convolution du spectre brut avec le taux de sélection des longueurs d'onde dans la population finale.

La fusion a été effectuée à deux niveaux : celui des données brutes (*niveau bas*) et celui des classifications (*niveau haut*).

Nous avons montré que la fusion de niveau haut est la plus appropriée pour la classification de cépages de raisins blancs : 4.7% d'erreur contre 13.1% d'erreur pour la fusion de niveau bas.

La *fusion de niveau bas* prend en considération l'intégralité du signal issu de chaque capteur, mais celle-ci est partiellement entachée de bruit et contient de fortes redondances. Il est donc extrêmement bénéfique d'extraire l'information pertinente de chacun des capteurs (fusion de niveau intermédiaire), et en allant plus loin, d'établir une classification individuelle pour chaque capteur (fusion de niveau haut). *Ceci est généralisable à toute problématique de discrimination, basée sur des signaux de capteurs.*

De plus, la fusion de niveau haut permet d'introduire dans la classification un « *diagnostic* » sur les performances des capteurs, grâce aux probabilités a priori calculées grâce aux classifications

individuelles et une « *expertise* » sous la forme de *coûts d'affectation*, grâce à la règle d'affectation de risque minimum de Bayes.

Nous avons proposé d'affiner ces méthodes, grâce à la prise en compte de « *degrés d'appartenance* » aux classes des échantillons et non une classification booléenne. En outre, nous avons indiqué que *l'opérateur d'agrégation* de la fusion, par défaut l'intersection, peut être modifié en fonction de la connaissance des capteurs et des objectifs de la classification.

La fusion multi-capteur par la démarche de Bayes fournit donc des résultats de classification excellents (4.7%), meilleurs que chaque capteur pris séparément, tirant partie de chacun, malgré la disparité de leurs performances individuelles.

# *Conclusion et perspectives*

## V. CONCLUSIONS ET PERSPECTIVES

Cette thèse traite le problème de la classification à partir de données multivariées issues de capteurs sensoriels (capteurs d'arômes et spectromètres).

Cette problématique générale se décompose en 2 thèmes :

⊗ la fusion multisensorielle ; par rapport aux méthodes, dites « boîtes noires », couramment employées pour la fusion, une approche tenant compte du pouvoir discriminant de chaque capteur et/ou de connaissances expertes est proposée.

⊗ l'optimisation de l'acquisition et du traitement des données ; en particulier, une démarche méthodologique expérimentale visant à optimiser la discrimination aromatique par capteurs d'arômes est élaborée et validée.

Le domaine d'application choisi est la classification variétale des moûts de raisins blancs.

La démarche expérimentale s'est focalisée sur 3 phases décisives de la **mesure des capteurs d'arômes**, soulevées lors de la revue bibliographique.

❶ L'extraction de l'information des signaux des capteurs d'arômes est généralement limitée au maximum de l'adsorption. Ce critère, conduisant à de fortes redondances entre les capteurs ( $r^2 > 0.8$ ), a été jugé insuffisant. Des descripteurs de forme des signaux ont donc été extraits et sélectionnés, grâce à des indices de performance calculés sur l'ensemble du système multi-capteur.

Par rapport à l'approche par modélisation, l'avantage de cette méthode est d'éviter la connaissance a priori de la forme des courbes et d'être peu sensible à la précision du début de l'acquisition

Les descripteurs sélectionnés grâce aux données du système multi-capteur constituent le sous-ensemble des descripteurs pertinents ; mais, le choix des descripteurs pour chaque capteur doit être réalisé au sein de ce sous-ensemble, grâce à des techniques de sélection de variables employées dans le troisième paragraphe, afin de choisir une combinaison idéale de descripteurs (par exemple, il n'est pas nécessaire de tenir compte de tous les maxima d'adsorption, étant donné leur forte corrélation).

D'autre part, une autre approche, encore jamais proposée dans la bibliographie des capteurs d'arômes a été proposée et validée : la PLS discriminante. Celle-ci a fourni des résultats meilleurs que ceux des systèmes de classification classiques basés sur les descripteurs, mais la robustesse de ce modèle devra être testée plus complètement.

② Après avoir montré l'importance et l'influence des conditions expérimentales sur la mesure, nous avons proposé d'appliquer la méthodologie des plans d'expériences pour quantifier l'effet de facteurs extérieurs sur la qualité de la mesure. Celle-ci est estimée grâce à des indices de performance, similaires à ceux employés lors de la sélection des descripteurs.

Cette démarche peut s'adapter à tout appareillage, en sélectionnant les paramètres expérimentaux pouvant être modifiés dans le système étudié. Nous avons identifié deux types de facteurs : ceux qui peuvent être optimisés une seule fois pour toutes les autres expérimentations, car ils ne sont pas liés à la nature de l'échantillon analysé (le mode d'injection et ses paramètres, la température de la cellule de mesure, etc.) ; et ceux qui sont corrélés à l'échantillon mesuré et qui doivent être adaptés à chaque application (température de l'espace de tête, volume, barbotage, hygrométrie, etc.).

Les plans d'expériences développés lors de nos applications sont des plans d'expériences complets à 2 niveaux. Si le nombre de facteurs est élevé, alors des plans d'expériences fractionnaires permettent de réduire le nombre d'expérimentations ; de même, si l'effet d'un facteur est non monotone, alors d'autres plans permettant de modéliser de manière non linéaire les sorties (plans centrés composites par exemple) doivent être réalisés. Etant donné que les critères étudiés sont complexes, il est important de prouver la monotonie des effets grâce à des répétitions au centre du domaine.

③ Peu de travaux ont porté sur le choix des capteurs d'arômes, et les utilisateurs choisissent généralement en fonction des spécifications peu informatives des constructeurs (C.f. Tableau 4, pII-40). Nous avons donc proposé d'user des techniques de sélection de variables pour choisir les capteurs de manière optimale.

Ces méthodes peuvent être exploitées quelles que soient les techniques d'extraction de l'information employées. Malgré la lourdeur des algorithmes génétiques, cette méthode semble être la plus adaptée à la sélection de capteurs.

Lors de cette étude, nous avons volontairement éludé un point méthodologique important : la correction à court et long terme de la dérive des capteurs d'arômes. Nous avons constaté une dérive aléatoire sur les mesures de moûts de raisins, effectuées durant 4 jours. La compensation de la dérive est un sujet de recherche qui sera développé dans le prolongement de cette thèse. En effet, ce sujet est extrêmement lourd : les seules techniques de correction de dérive proposées permettent de compenser la dérive des mêmes échantillons dont la mesure est répétée au cours du temps et non la mesure d'échantillons divers représentatifs de la dispersion des classes (*HOLMBERG et al.*, 1996, 1997 & *BAZZO et al.*, 1998) ; de plus, la correction est d'autant plus difficile que la distinction entre les classes est faible.

Un autre point important révélé par ce travail est la caractérisation précise des capteurs employés, d'un point de vue de leur seuil de détection, sensibilité et sélectivité. Aucune donnée fiable n'a été publiée sur les caractéristiques métrologiques des transducteurs, dans des conditions expérimentales données, pour des molécules volatiles présentes dans les espaces de tête de produits agro-alimentaires.

Il apparaît nécessaire de mettre en place des bancs d'essai normalisés, par exemple grâce au couplage avec une CPG, grâce auxquels les différents capteurs pourront être évalués, par le calcul de caractéristiques métrologiques issues de mesures d'échantillons de référence.

Nous avons également soulevé le problème de la mesure aromatique des boissons alcoolisées. Malgré le grand nombre de travaux de discrimination de bières, de vins et autre whisky, il est apparu que les discriminations ne peuvent pas être basées sur les arômes sans traitement préalable des échantillons.

Cette thèse a servi de base au montage de diverses collaborations de recherche destinées à lever ce verrou technologique avec les équipes de l'INRA-IPV (*Institut des Produits de la Vigne : laboratoire des Arômes et des Substances Naturelles*) de Montpellier et Pech Rouge, de l'Université Montpellier II (*Laboratoire de Génie Biologique et Sciences des Aliments*), de l'ENSC.M (*Laboratoire des Matériaux et Procédés Membranaires*), la société ALPHAMOS (*constructeur de capteurs d'arômes toulousain*), et les CHAIS BEUCAUROIS, SA (Beucaire, F), qui gèrent l'embouteillage des boissons alcoolisées pour le groupe CASINO (avec environ 2400 dégustations/expert/an), au sein d'un projet tout d'abord financé par la DRAF<sup>26</sup> : « *Détection des défauts organoleptiques majeurs dans les vins* », puis par la combinaison de la DGAL<sup>27</sup> : « *étude des verrous technologiques liés au développement industriel de capteurs pour la détection de défauts d'arômes dans les boissons alcoolisées* », de TRIAL<sup>28</sup> : « *conception d'un outil d'aide à la dégustation des vins* » et de l'ACTIA<sup>29</sup> : « *évaluation de la qualité des huiles d'olives vierges, des moûts de raisin ou des vins et des fruits, par mise en œuvre des capteurs d'arômes* ». Les résultats de ces travaux de recherche seront publiés fin 1999.

---

<sup>26</sup> DRAF : Direction Régionale de l'Agriculture et de la Forêt. Contrat Etat-Région, filière viti-vinicole, campagne 1995-1996, n°B 524/1567

<sup>27</sup> DGAL : Direction Générale de l'Alimentation, n°R98/05, Oct.97-Oct.99

<sup>28</sup> TRIAL : Transfert Innovation Agro-alimentaire . n° 05-97, Oct.97-Mars99

<sup>29</sup> ACTIA : Association de Coordination Technique pour l'Industrie Agro-alimentaire, n°97.30, Oct.97-Oct.99

Le second thème de la thèse était de proposer et de valider des méthodes de classification à partir de données multivariées fortement redondantes. L'application consiste à discriminer les cépages de moûts de raisins blancs grâce au concours de divers capteurs : les capteurs d'arômes préalablement optimisés grâce à la démarche expérimentale précédente et les spectromètres IR-TF et UV ; ces derniers ont également été optimisés par la sélection des longueurs d'onde grâce à la méthode des algorithmes génétiques.

Cette technique étant non déterministe, nous avons proposé et testé différentes méthodes de choix des longueurs d'onde, parmi les diverses populations finales de chromosomes résultant d'une dizaine d'exécution de ces algorithmes ; les méthodes les plus performantes sont le choix du meilleur chromosome ou la convolution du spectre avec le taux de sélection des longueurs d'onde dans la population finale d'une exécution d'un algorithme génétique.

Ce filtrage des spectres, grâce à la sélection de zones discriminantes, s'est avéré extrêmement efficace et devrait être systématiquement appliqué aux mesures spectrométriques. Cependant, cette méthode est assez longue et doit être réitérée une dizaine de fois, ces algorithmes étant non déterministes. Les temps de calculs peuvent être minimisés, grâce à un réglage adéquat des paramètres, tels que les fenêtres de lissage, une population moins nombreuse dont la faible diversité est compensée par de forts taux de mutation et de croisement, qui sont ensuite réduits pour faciliter la convergence, une validation croisée sur peu de blocs, etc.

La fusion multi-capteur a été réalisée à 2 niveaux. La fusion de niveau bas s'est révélée moins performante que celle de niveau haut, malgré l'amélioration apportée par le filtrage des spectres. La redondance et le bruit contenu dans les signaux peut même dégrader fortement la classification, comme par exemple lors de la concaténation des courbes des capteurs d'arômes.

La fusion de niveau haut, s'appuyant sur la théorie de Bayes, présente l'avantage de tenir compte des probabilités a priori issues des classifications individuelles, sans avoir recours à des lois de distribution probabilistes. De plus, la règle d'affectation de risque minimum de Bayes permet d'introduire un coût de classification, fixé grâce à l'*expertise* découlant du problème analysé. Ce coût peut éventuellement être fixé de manière itérative, grâce à une initialisation basée sur l'*expertise*, et une évolution en fonction du taux d'erreur de prédiction des différentes classes.

En outre, nous avons proposé de généraliser l'application de ces règles à des classifications par degré d'appartenance et à d'autres opérateurs d'agrégation. Ces méthodes devraient être testées prochainement, mais les sorties de la PLS discriminante ne sont pas des degrés d'appartenance proprement dit ; l'application de cette méthode nécessite donc une réflexion préalable visant à transformer les sorties comprises entre -0.5 et 1.5 en degrés d'appartenance aux classes.

Nous préconisons donc l'emploi de cette méthode de fusion de niveau haut pour tout système multi-capteur visant à établir une classification à partir de données fortement redondantes.

Sur un plan méthodologique, ce travail de thèse aura donc contribué à :

☒ une meilleure connaissance des capteurs d'arômes, notamment dans le domaine de l'optimisation de leur fonctionnement ;

☒ l'étude et la validation de différentes méthodes de classification et de fusion multisensorielle impliquant des données fortement redondantes.

Sur un plan pratique, elle a été la base d'une collaboration avec des équipes de recherche multidisciplinaires et des industriels ou utilisateurs finaux, dans le but de résoudre certains problèmes de la filière viti-vinicole.

En outre, grâce à la banque de données qu'elle a permis de mettre en place et aux problèmes qu'elle a soulevé, elle ouvre la voie à d'autres problématiques de recherche telles que la prédiction de la qualité des vins en terme de typicité aromatique, en fonction du potentiel aromatique mesuré dans les moûts (arômes et surtout précurseurs d'arômes).

# *Valorisation de la recherche*

## VI. VALORISATION DE LA RECHERCHE

### PUBLICATION DANS DES REVUES A COMITE DE LECTURE

ROUSSEL, S., FORSBERG, G., STEINMETZ, V., GRENIER, P., BELLON-MAUREL, V., 1998. Optimisation of Electronic Nose Measurements. Part I : Output Feature Selection. *J. Food Eng.*, 37 (2), pp 207-222.

ROUSSEL, S., FORSBERG, G., GRENIER, P., BELLON-MAUREL, V., 1998. Optimisation of Electronic Nose Measurements. Part II : Influence of Experimental Parameters. *J. Food Eng.*, accepté en juillet 1998.

PRIVAT, E., ROUSSEL, S., GRENIER, P., BELLON-MAUREL, V., 1998. Techniques de désalcoolisation en vue de la discrimination de boissons alcoolisées par capteurs d'arômes. *Sci. Al.*, 18 (5), pp 459-470.

### PUBLICATION DANS DES REVUES TECHNIQUES

ROUSSEL, S., GRENIER, P., BELLON-MAUREL, V., 1998. Les nez artificiels hument le vin. *La vigne*, 88, pp 60-61.

### PARTICIPATION AUX COLLOQUES

#### Présentations orales (conférencier invité)

ROUSSEL, S., 1998. Optimisation of gas-sensor array measurements: methodology of feature selection and influence of experimental parameters. *In: Formation Inter-entreprises ADRIA*. Paris : pp 1-21.

#### Poster

ROUSSEL, S., GRENIER, P., BELLON-MAUREL, V., 1997. Optimisation of electronic nose measurements: study of the experimental parameter influence and choice of the signal features. *In: Proceedings of the 4th symposium of Olfaction and Electronic Nose*. Nice, F.

ROUSSEL, S., GRENIER, P., BELLON-MAUREL, V., 1998. Methodology for SnO<sub>2</sub>-gas sensor selection using stepwise multivariate analysis. *In: Sensoral 98: international workshop on Sensing Quality of Agricultural Products*. Montpellier : pp 371-384.

# *Bibliographie*

## VII. BIBLIOGRAPHIE

ABE, H., KANAYA, S., TAKAHASHI, Y., SASAKI, S., 1988. Extended studies of the automated odor-sensing system based on plural semiconductor gas sensors with computerized pattern recognition techniques. *Anal. Chim. Acta*, 215, pp 155-168.

ADECHY, M., 1998. Comparison of multi-sensor array instruments for rancidity detection of edible oils. In: *Formation Inter-entreprises ADRIA*. Paris : 1p.

AFNOR, 1993. *Normes statistiques. Tome 3 : traitement de résultats de mesures*. 6ème édition : AFNOR (ed.).

AISHIMA, T., 1991. Discrimination of liquor aromas by pattern recognition analysis of responses from a gas sensor array. *Anal. Chim. Acta*, 143, pp 293-300.

ALEXANDER, P.W., DIBENEDETTO, L.T., HIBBERT, D.B., 1998. A field-portable gas analyzer with an array of six semiconductor sensors. Part 2: Identification of beer samples using artificial neural networks. *Field Anal. Chem. Tech.*, 2 (3), pp 145-153.

ALPHAMOS, 1997. *Application notes*. Toulouse, F : ALPHAMOS (ed.).

ALTHAINZ, P., GOSCHNICK, J., EHRMANN, S., ACHE, H.J., 1996. Multisensor microsystem for contaminants in air. *Sens. & Actu. B*, 33, pp 72-76.

AMOORE, J.E., 1970. *Molecular basis of odours*. Springfield, IL : C C Thomas (ed.).

AROMASCAN, I., 1995. *Plaquette technique*. Crewe : AROMASCAN plc (ed.).

AUSTIN, G., RUSSELL, I., 1996. A Gas-Sensor-Based On-Line Ethanol Meter for Breweries. *J. of Am. Soc. Brew. Chem.*, 54 (4), pp 212-215.

AYARI, I., HATON, J., 1995. Cadre pour la fusion multi-capteur. In: *Troisièmes journées francophones sur l'intelligence artificielle distribuée et les systèmes multi-agents*. Chambéry, F : pp 1-9.

BAGUELIN, M., VIDIE, L., 1998. Etude comparative des spectroscopies UV et IR-TF : application aux effluents industriels et matrices agro-alimentaires. Rapport de stage de recherche. *Ecole Nationale Supérieure des Techniques Industrielles et des Mines d'Alès*.

BALDET, P., 1999. Reconnaissance moléculaire à l'aide de matériaux denses contenant des cyclodextrines. *Th. : chimie des matériaux*, Université Montpellier II.

- BAO, F., NIU, M., ZHAO, X., 1996. A theoretical study on the gas sensitive characteristics of semiconductive ceramic sensors. *Jpn J. Appl. Phys.*, 35. Part I (8), pp 4216-4219.
- BARTLETT, P.N., ARCHER, P., LING-CHUNG, S.K., 1989. Conducting polymer gas sensors. Part I : fabrication and characterisation. *Sens. & Actu. B*, 19, pp 125-140.
- BARTLETT, P.N., GARDNER, J.W., 1992. Odour sensors for an electronic nose. In : *Sensors and Sensory Systems for an Electronic Nose*. Pays Bas : Kluwer Academic Publishers (ed.), pp 31-51.
- BAZZO, S., LOUBET, F., TAN, T.T., HEWITT-JONES, J.D., ENGELEN-CORNAX, C.E.M., QUADT, J.F.A., 1998. Quality control of edible oil using an electronic nose. *Seminars in food analysis*, 3, pp 1-10.
- BELLON, V., 1993. Fermentation control using ATR and FT-IR spectrometer. *Sens & Actu B*, 12, pp 57-64
- BELLON, V., VIGNEAU, J.L., LECLERCQ, M., 1993. Feasibility and Performances of a New, Multiplexed, Fast and Low-Cost Fiber-Optic NIR Spectrometer for the On-Line Measurement of Sugar in Fruits. *Appl. Spec.*, 47 (7), pp 1079-1083.
- BENADY, M., SIMON, J.E., CHARLES, D.J., MILES, G.E., 1995. Fruit Ripeness Determination by Electronic Sensing of Aromatic Volatiles. *Trans. ASAE*, 38 (1), pp 251-257.
- BERDAGUE, J.-L., 1995. Caractérisation instrumentale de la qualité des aliments par analyse rapide des composés volatils, (1) et (2). *Communication INRA de Theix. laboratoire Flaveurs*, 4 p.
- BERDAGUE, J.-L., TALOU, T., 1993. Exemples d'application aux produits carnés des senseurs de gaz à semiconducteurs. *Sci. Al.*, 13, pp 141-148.
- BORJESSON, T., EKLOV, T., JONSSON, A., SUNDGREN, H., SCHNURER, J., 1996. Electronic nose for odor classification of grains. *Anal.Tech. Instr.*, 73 (4), pp 457-461.
- BOURROUNET, B., TALOU, T., GASET, A., 1995. Application of a multi-gas-sensor device in meat industry for boar-taint detection. *Sens. & Actu. B*, 26-27, pp 250-254.
- BOURROUNET-DUBREUIL, B., 1998. Conception et développement d'un système à multiples capteurs de gaz. Application en agro-industrie. *Thèse de l'INPT, TOULOUSE*.
- BOX, G.E.P., HUNTER, J.S., 1957. Multi-factor designs for exploring response surface. *Ann. Math. Stat.*, 28, pp 195-241.
- BRESSOLETTE-HENRY, A., 1995. Relation entre descripteurs sensoriels et paramètres instrumentaux : application à l'étude de la texture de produits croustillants peu différenciés. *Th.: Sciences alimentaires*, ENSIA, Massy.

- BREUIL, P., 1989. Elaboration de couches minces de dioxyde d'étain sensibles à l'action des gaz. Performances électriques et mécanismes réactionnels. *Th.: Génie des procédés*, Ecole Nationale Supérieure des Mines de Saint-Etienne, Saint-Etienne.
- BRODA, S., ZELLNER, S., HANKE, A., SCHALLER, R., HABEGGER, R., SCHNITZLER, W.H., 1998. Use of an "electronic nose" for fast characterizing of volatile aroma compounds of carrot varieties. *Deutsche Lebensmittel Rundschau*, 94 (3), pp 77-80.
- BRONS, A., RABATEL, G., ROS, F., SEVILA, F., TOUZET, C., 1993. Plant grading by vision using neural networks and statistics. *Comp. & Elect. in Agric.*, 9, pp 25-39.
- BUZON, J., ROUSSEL, J.C., 1981. Spectrométrie d'absorption dans l'infrarouge. In : *Techniques de l'ingénieur ; traité d'Analyse et Caractérisation*. Schiltigheim, F : ISTRAL (ed.), pp P2855,1-10.
- CADET, F., OFFMANN, B., 1997. Direct Spectroscopic sucrose Determination of Raw Sugar Cane Juices. *J. Agric. Food Chem.*, 45, pp 166-171.
- CARAUX, G., LECHEVALLIER, Y., 1996. Règles de décision de Bayes et méthodes statistiques de discrimination. *Rev. d'Intell. Artif.*, 10 (2-3), pp 219-284.
- CHANG, S.C., 1980. Oxygen chemisorption on tin oxide: correlation between electrical conductivity and EPR measurements. *J. Vac. Sci. Technol.*, 17, pp 366-376.
- CHATILA, R., 1996. Fusion de données. Du numérique au symbolique: raisonnement sur l'incertain. In: *Support de cours de stage: "Pour les systèmes multicapteurs: la fusion de données"*. Polytechnique, Palaiseau, F : pp 1-38.
- CHTIOUI, Y., 1997. Reconnaissance automatique des semences par vision artificielle basée sur des approches statistiques et connexionnistes. *Th.: Sciences pour l'ingénieur de Nantes*, Université de Nantes. IRESTE, Nantes.
- CHTIOUI, Y., BERTRAND, D., BARBA, D., 1998. Feature selection by a genetic algorithm. Application to seed discrimination by artificial vision. *J. Sci. Food and Agric.*, 76 (1), pp 77-86.
- COMON, P., 1991. Classification supervisée par réseaux multi-couches. *Traitement du signal*, 8 (6), pp 387-407.
- CORCORAN, P., 1993. Electronic odour sensing systems. *Elec. & Com. Eng. J.*, 5 (5), pp 303-308.
- DANZART, M., 1997. Les principales classes de plans d'expériences et leurs domaines d'application. *Cahiers de Massy*, pp 11-18.
- DAVENEL, A., GRENIER, P., FOCH, B., BOUVIER, J., VERLAQUE, P., POURCIN, J., 1991. Filter, Fourier Transform Infrared, and Arometry, for Following Alcoholic Fermentation in Wines. *Journal of Food Science*, 56 (6).

- DAVAL, G., LEVETEAU, J., MACLEOD, P., 1980. Analyse topographique de l'électro-olfactogramme chez la grenouille. *J. Physiol. Paris*, 76, pp 559-567.
- DAVIDE, F., DI NATALE, C., D'AMICO, A., 1993. Sensor arrays figures of merit: definitions and properties. *Sens. & Actu. B*, 13-14, pp 113-122.
- DEFENEZ, M., WILSON, R.H., 1995. Mid-Infrared spectroscopy and chemiometrics for determining the type of fruit used in jam. *Journal Sci Food Agric*, 67, pp 461-467.
- DE LA PINSONNAIS, M., 1998. Characterisation of boiled taste in milk. In: *Formation Inter-entreprises ADRIA*. Paris : pp 1-2.
- DELWICHE, M., 1987. Theory of fruit firmness sorting by impact forces. *Trans. ASAE*, 30 (4), pp 1160-1166.
- DEVROYE, L., GYÖRFI, L., LUGOSI, G., 1996. *A Probabilistic Theory of Pattern Recognition*. New-York : Springer Verlag (ed.).
- DI NATALE, C., DAVIDE, F., D'AMICO, A., NELLI, P., GROPELLI, S., SBERVEGLIERI, G., 1996. An electronic nose for the recognition of the vineyard of a red wine. *Sens. & Actu. B*, 33, pp 83-88.
- DI NATALE, C., DAVIDE, F., D'AMICO, A., 1995a. A self-organizing system for pattern classification: time varying statistics and sensor drift effects. *Sens. & Actu. B*, 26-27, pp 237-241.
- DI NATALE, C., DAVIDE, F., D'AMICO, A., SBERVEGLIERI, G., NELLI, P., FAGLIA, G., 1995b. Metal oxide semiconductor gas sensors array as a tool for the analysis of wine vapours. In: *Proc. 8th Europ. Conf. on Food Chem*. Vienne, Au : pp 131-134.
- DI NATALE, C., DAVIDE, F., FAGLIA, G., NELLI, P., 1995c. Study of the effect of the sensor operating temperature on SnO<sub>2</sub>-based sensor-array performance. *Sens. & Actu. B*, 23, pp 187-191.
- DI NATALE, C., MACAGNANO, A., D'AMICO, A., DAVIDE, F., 1997a. Electronic-nose modelling and data analysis using a self-organizing map. *Meas. Sci. & Technol.*, 8 (11), pp 1236-1243.
- DI NATALE, C., MACAGNANO, A., DAVIDE, F., D'AMICO, A., PAOLESSE, P., BOSCHI, T., FACCIO, M., FERRI, G., 1997b. An electronic nose for food analysis. *Sens. & Actu. B*, 44, pp 521-526.
- DI NATALE, C., MACAGNANO, A., PAOLESSE, R., TARIZZO, E., MANTINI, A., D'AMICO, A., 1998. Can an electronic nose replace sensoral analysis ? In: *Sensoral 98: international workshop in Sensing Quality of Agricultural Products*. Montpellier : pp 329-340.
- DIKINSON, T.A., WHITE, J., KAUER, J.S., WALT, D.R., 1996. A chemical-detecting system based on a cross-reactive optical sensor. *Nature*, 382, pp 697-700.

- DROMIGNY, A., ZHU, Y.M., 1997. Improving the Dynamic Rang of Real-Time X-Ray Imaging Systems via Bayesian Fusion. *J. Nondestruct. Eval.*, 16 (3), pp 147-160.
- DUBOURDIEU, D., 1995. Les arômes du sauvignon : valorisation par la vinification. In: *Actes d'EUROVITI 95*. Montpellier, F : pp 25-31.
- DUBUISSON, B., 1990. *Diagnostic et reconnaissance des formes*. Paris : Hermès (ed.).
- DULL, G.G., 1978. Non destructive quality evaluation of agricultural products : a definition and practical approach. *J. Food Protection*, 41 (1), pp 50-53.
- DUPUIT, E., 1997. Etude exploratoire des caractéristiques organoleptiques du vin par spectrométrie moyen infrarouge par transformée de Fourier. Stage de DEA: chimie et physico-chimie moléculaire organique. Option: spectrométries et physico-chimie structurale. *CEMAGREF*.
- DUPUY, N., DUPONCHEL, L., HUVENNE, J., SOMBRET, B., LEGRAND, P., 1996. Classification of edible fats and oils by principal component analysis of Fourier transform infrared spectra. *Food Chem.*, 57 (2), pp 245-251.
- EDAN, Y., PASTERNAK, H., GUEDALIA, D., OZER, N., SHMULEVITCH, I., RACHMANI, D., FALLIK, E., GRINBERG, S., 1994. Multi-sensor quality classification of tomatoes. In: *Am. Soc. Agric. Eng. : Intern. Summer Meeting*. Kansas city, US : pp 1-19.
- ELKÖV, T., JOHANSSON, G., WINQUIST, F., LUNDSTRÖM, I., 1998. Monitoring sausage fermentation using an electronic nose. *J. Sci. Food and Agric.*, 76 (4), pp 525-532.
- ELKÖV, T., MARTENSSON, P., LUNDSTRÖM, I., 1997a. Enhanced selectivity of MOSFET gas sensors by systematical analysis of transient parameters. *Anal. Chim. Acta*, 353 (2-3), pp 291-300.
- ELKÖV, T., SUNDGREN, H., LUNDSTRÖM, I., 1997b. Distributed chemical sensing. *Sens. & Actu. B*, 45 (1), pp 71-77.
- ELLIOTMARTIN, R.J., MOTTRAN, T.T., GARDNER, J.W., HOBBS, P.J., BARTLETT, P.N., 1997. Preliminary investigation of breath sampling as a monitor of health in dairy cattle. *J. Agri. Engng. Res.*, 67 (4), pp 267-275.
- EMA, K., YOKOYAMA, M., NAKAMOTO, T., MORIIZUMI, T., 1989. Odour-sensing system using a quart-resonator sensor array and neural-network pattern recognition. *Sens. & Actu. B*, 18, pp 291-296.
- ESCUDIER, J., 1990. Evaluation de quelques techniques adaptées à la désalcoolisation des vins. *Rev. OEnol.*, 575, pp 57-60.
- ETIENNE, M., LUCON, M., STEINMETZ, V., PACHOT, A., FLEURENCE, J., 1997. The multi-gas sensor technology: a new tool for the quality evaluation of salmon. In: *Proc. of the final*

meeting of the EU concerted action: "Methods to determine the freshness of fish in research and industry". Nantes : pp 119-126.

ETIEVANT, P., ISSANCHOU, S., 1987. Le goût. *La recherche*, 18 (193), pp 1344-1353.

FEARN, T., DAVIES, T., 1997. Genetic Algorithms: the evolutionary solution to an old problem. *Spectr. Europe*, 9 (6), pp 25-27.

FENELON, J., 1981. *Qu'est-ce que l'analyse de données?* Paris : LEFONEN (ed.).

FIGARO ENGINEERING, I., 1990. *The basis of FIGARO gas sensor and SnO<sub>2</sub> gas sensor specification sheets*. Osaka, J : FIGARO Eng. Inc. (ed.).

FILDES, M., 1992. Sensor fusion for manufacturing. *Sensors*, 9 (1), pp 11-19.

FIS INTELLIGENT SENSORS, I., 1994. *FIS product review - Reliability of FIS gas sensors*. Osaka, J : FIS inc. (ed.).

FORSBERG, G., 1997. Méthode pour l'optimisation de systèmes multi-capteur: les "nez électroniques". Mémoire de fin d'étude d'ingénieur. *CEMAGREF*.

FRANCOIS, J., 1996. Systèmes multi-capteur : évaluation de la robustesse et de l'état de fonctionnement. Rapport de DEA. *CEMAGREF*.

FREUND, M.S., S, L.N., 1995. A chemically diverse conducting polymer-based "electronic nose". *Proc. Natl. Acad. Sci. USA*, 92, pp 2652-2656.

GALLINARI, P., GASCUEL, O., 1996. Statistique, apprentissage et généralisation. Application aux réseaux de neurones. *Rev. d'Intell. Artif.*, 10 (2-3), pp 285-344.

GARDNER, J., PEARCE, T., FRIEL, S., BARTLETT, P., BLAIR, N., 1994. A multisensor system for the flavour monitoring using an array of conducting polymers and predictive classifiers. *Sens. & Actu. B*, 18 (1-3), pp 240-243.

GARDNER, J.W., 1991. Detection of vapours and odours from a multisensor array using pattern-recognition techniques. Part 1. Principal Component and Cluster Analysis. *Sens. & Actu. B*, 4, pp 109-116.

GARDNER, J.W., BARTLETT, P.N., 1994a. A brief history of electronic noses. *Sens. & Actu. B*, 18-19, pp 211-220.

GARDNER, J.W., BARTLETT, P.N., 1994b. Intelligent ChemSADs for Artificial Odor-Sensing of Coffees and Lager Beers. *In: Olfaction and Taste XI*. Toulouse : pp 690-693.

GARDNER, J.W., BARTLETT, P.N., 1995. Application of conducting polymer technology in microsystems. *Sens. & Actu. A*, 51, pp 57-66.

- GARDNER, J.W., BARTLETT, P.N., 1996. Performance definition and standardization of electronic noses. *Sens. & Actu. B*, 33 (1-3), pp 60-67.
- GARDNER, J.W., CRAVEN, M., DOW, C., HINES, E.L., 1998. The prediction of bacteria type and culture growth phase by an electronic nose with a multi-layer perceptron network. *Meas. Sci. & Technol.*, 9 (1), pp 120-127.
- GARDNER, J.W., HINES, E.L., 1997. Pattern Analysis Techniques. In : *Handbook of Biosensors and Electronic Noses : Medicine, Food, and the Environment*. Frankfurt, D : CRC Press (ed.), pp 933-651.
- GARDNER, J.W., HINES, E.L., TANG, H.C., 1992. Detection of vapours and odours from a multisensor array using pattern-recognition techniques. Part 2. Artificial neural networks. *Sens. & Actu. B*, 9, pp 9-15.
- GARNIER, P., 1987. Les polymères conducteurs. *La Recherche*, 18, pp 704-715.
- GIBSON, T.D., PROSSER, O., HULBERT, J.N., MARSHALL, R.W., CORCORAN, P., LOWERY, P., RUCK-KEENE, E.A., HERON, S., 1997. Detection and simultaneous identification of micro-organisms from headspace samples using an electronic nose. *Sens. & Actu. B*, 44, pp 413-422.
- GÖPEL, W., 1998. Chemical imaging of gas mixtures and aromas: recent trends. In: *Formation Inter-entreprises ADRIA*. Paris : pp 1-17.
- GRANDVALET, Y., 1995. Injection de bruit dans les perceptrons multicouches. *Th.: Contrôle des systèmes*, Université Technologique de Compiègne, Compiègne.
- HINES, E.L., GARDNER, J.W., 1994. An artificial neural emulator for an odour sensor array. *Sens. & Actu. B*, 18-19, pp 661-664.
- HIVERT, B., HOUMMADY, M., MIELLE, P., MAUVAIS, G., HENRIOUD, J., HAUDEN, D., 1995. A fast and reproducible method for gas sensor screening to flavour compounds. *Sens. & Actu. B*, 26-27, pp 242-245.
- HOBBS, P.J., MISSELBROOK, T.H., PAIN, B.F., 1995. Assessment of odours from LivestockWastes by Photoionisation Detector, an Electronic Nose, Olfactometry and Gas Chromatography-Mass Spectrometry. *J. Agric. Engng Res.*, 60, pp 137-144.
- HODGINS, D., 1997. The electronic nose: sensor array-based instruments that emulate the human nose. In : *Techniques for analysing food aroma.* : Food Sci. Techn. (ed.), pp 331-371.
- HOFMANN, T., SCHIEBERLE, P., KRUMMEL, C., FREILING, A., BOCK, J., HEINERT, L., KOHL, D., 1997. High resolution gas chromatography selective odorant measurement by multisensor

- array (HRGC/SOMSA) : a useful approach to standardise multisensor arrays for use in the detection of key food odorants. *Sens. & Actu. B*, 41 (1-3), pp 81-87.
- HOLMBERG, M., DAVIDE, F., DI NATALE, C., D'AMICO, A., WINQUIST, F., LUNDSTRÖM, I., 1997. Drift counteraction in odour recognition applications: lifelong calibration method. *Sens. & Actu. B*, 42 (3), pp 185-194.
- HOLMBERG, M., GUSTAFSSON, F., HORNSTEN, E.G., WINQUIST, F., NILSSON, L.E., LJUNG, L., LUNDSTRÖM, I., 1998. Bacteria classification based on feature extraction from sensor data. *Biotech. Tech.*, 12 (4), pp 319-324.
- HOLMBERG, M., WINQUIST, F., LUNDSTRÖM, I., GARDNER, J.W., HINES, E.L., 1995. Identification of paper quality using a hybrid electronic nose. *Sens. & Actu. B*, 26-27, pp 246-249.
- HOLMBERG, M., WINQUIST, F., LUNDSTRÖM, I., DAVIDE, F., DI NATALE, C., D'AMICO, A., 1996. Drift counteraction for an electronic nose. *Sens. & Actu. B*, 35-36, pp 528-535.
- HÖRCHNER, U., KALIVAS, J., 1995. Simulated annealing based optimization algorithms : fundamentals and wavelength selection applications. *J. of Chemometrics*, 9, pp 283-308.
- HULBERT, J.N., 1998. The Bloodhound TM BH114 sensory array system. In: *Formation Inter-entreprises ADRIA*. Paris : pp 1-6.
- JEANJEAN, N., PALLUAU, M., PREVOSTO, B., 1995. Le nez artificiel. Technologies et exemples d'application. Synthèse bibliographique. *CEMAGREF-ENGREF*.
- JOHNSON, S.R., SUTTER, J.M., ENGELHARDT, H.L., JURIS, P.C., WHITE, J., KAUER, J.S., DIKINSON, T.A., WALT, D.R., 1997. Identification of multiple analytes using an optical sensor array and pattern recognition neural networks. *Anal. Chem.*, 69, pp 4641-4648.
- JONDA, S., FLEISCHER, M., MEIXNER, H., 1996. Temperature control of semiconductor metal-oxide gas sensors by means of fuzzy logic. *Sens. & Actu. B*, 34, pp 396-400.
- JOU, K.D., HARPER, W.J., 1998. Pattern recognition of Swiss cheese aroma compounds by SPME/GC and an electronic nose. *Milchwissenschaft Milk Science International*, 53 (5), pp 259-263.
- JOURDAN, F., 1997. Architecture du système olfactif périphérique. In : *Sciences cognitives, diversité des approches*. Paris : M.B. Gordon, Hermès (ed.), pp 41-46.
- KALMAN, E.L., WINQUIST, F., LUNDSTRÖM, I., 1997. Technical note: a new pollen detection method based on an electronic nose. *Atmos. Env.*, 31 (11), pp 1715-1719.
- KELLER, P.E., KOUZES, R.T., KANGAS, L.J., 1994. Three neural network based sensor systems for environmental monitoring. In: *IEEE electro94 Conference*. BOSTON, MA : pp 377-382.

- KEMSLEY, E.K., ZHUO, L., HAMMOURI, M.K., WILSON, R.H., 1993. Quantitative analysis of sugars solutions using infrared spectroscopy. *Food Chemistry*, 44, pp 299-304.
- KOHONEN, T., 1990. The Self-Organizing Map. *Proc. of IEEE*, 78 (9), pp 1464-1480.
- KONDOH, J., SHIOKAWA, S., 1994. New application of shear horizontal Surface Acoustic Wave sensors to identifying fruit juices. *Jpn J. Appl. Phys.*, 33 (1), pp 3095-3099.
- KRAMER, A., 1998. The NST MOSFET based hybrid sensor array. In: *Formation Inter-entreprises ADRIA*. Paris : 1p.
- KRZANOWSKI, W., 1987. Selection of variables to preserve multivariate data structure using principal component analysis. *Appl. Stat.*, 36, pp 22-33.
- KVAAL, K., McEWAN, A., 1996. Analysing complex sensory data by non-linear artificial neural networks. In : *Multivariate Analysis of Data in Sensory Science*. : T. Noes and E. Risvik (ed.), pp 103-132.
- LAI, Y.W., KEMSLEY, K.E., WILSON, R.H., 1994. Potential of Fourier Transform Infrared Spectroscopy of the Authentication of vegetable Oils. *J. Agric. Food Chem.*, 42 (5), pp 1154-1159.
- LAURET, F., 1982. La qualité des fruits et le système de production-distribution-consommation. *Bulletin d'information CEMAGREF*, 289, pp 68-72.
- LEARDI, R., BOGGIA, R., TERRILE, M., 1994. Application of a genetic algorithm to feature selection under full validation conditions and outlier detection. *J. Chem.*, 8, pp 65-79.
- LLOBET, E., BREZMES, J., VILOANOVA, X., SUEIRAS, J.E., CORREIG, X., 1997. Qualitative and quantitative analysis of volatile organic compounds using transient and steady-state responses of a thick-film tin oxide gas sensor array. *Sens. & Actu. B*, 41 (1-3), pp 13-21.
- LUCAS, Q., 1997. Analyses de vins et solutions alcoolisées sur FOX 4000 et FOX 5000. *Rapport d'analyses*, 31 p.
- LUDWIG, L., BERK, F., GRAUEL, A., 1996. Using Evolutionary Algorithms for Structural Optimisation of an Artificial Neural Network performing the Analysis of Electronic Nose Data. *Z. Angew Math. Mech.*, 76, pp 499-500.
- LUNDSTRÖM, I., PETERSSON, L., 1996. Chemical sensors with catalytic metal gates. *J. Vac. Sci. Technol. A*, 14 (3), pp 1539-1545.
- LUO, R.C., KAY, M.G., 1989. Multisensor integration and fusion in intelligent systems. *IEEE Trans. on syst., man and cyber.*, 19 (5), pp 901-931.

- MACLEOD, p., SAUVAGEOT, F., 1986. Bases neurophysiologiques de l'évaluation sensorielle des produits alimentaires. *Les cahiers de l'ENSBANA*.
- MAEKAWA, T., TAMAKI, J., MIURA, N., YAMAZOE, N., MATSUSHIMA, S., 1992. Development of SnO<sub>2</sub>-based ethanol gas sensor. *Sens. and Actu. B*, 9, pp 63-69.
- MALINGREY, B., 1987. Spectrométrie d'absorption dans l'ultraviolet et le visible. In : *Techniques de l'ingénieur ; traité d'Analyse et Caractérisation*. Schiltigeim, F : ISTRAL (ed.), pp P2795,1-13.
- MARCO, S., PARDO, A., DAVIDE, F., DI NATALE, C., A., D.A., HIERLEMANN, A., MITROVIC, J., SHWEIZER, M., WEIMAR, U., GÖPEL, W., 1996. Different strategies for the identification of gas sensing system. *Sens. & Actu. B*, 34, pp 213-223.
- MARQUIS, F., 1997. De l'échantillon à la mesure: peut-on comparer l'offre matérielle? In: *Formation Inter-entreprises ADRIA*. Quimper : pp 1-28.
- MARTENS, H., NAES, T., 1984. *Multivariate calibration. Chemometrics, mathematics and statistics in chemistry*. Kowalski, Reided Publishing company (ed.).
- MAUNG, K., 1941. Measurement of association in a contingency table with special references to the pigmentation of hair and eye color of Scottish school children. In : *Annals of Eugenics*.
- MAY, I.P., BYFIELD, M.P., LINDSTROM, M., WUNSCH, L.F., 1997. Chiral Discrimination using a quartz crystal microbalance and comparison with gas chromatographic retention data. *Chirality*, 9, pp 225-232.
- MIELLE, P., 1994a. Contrôle rapide de la qualité aromatique d'un produit alimentaire par capteurs d'arôme. *Communication INRA de Dijon*, 2 p.
- MIELLE, P., 1994b. Projet pifomètre: du détecteur de gaz au capteur d'arôme. *Document interne INRA Dijon*, 13 p.
- MIELLE, P., 1996a. "Electronic noses": towards the objective of instrumental characterisation of food aroma. *Trends in Food Sci. & Technol.*, 7 (Special issue on Flavour Perception.), pp 432-438.
- MIELLE, P., 1996b. Managing dynamic thermal exchanges in commercial semiconducting gas sensors. *Sens. and Actu. B*, 34 (1-3), pp 533-538.
- MIELLE, P., 1998. Une technique de pointe au service du contrôle de la qualité aromatique. *Le technoscope de Biofutur*, 174 (99), pp 2-10.
- MIELLE, P., HIVERT, B., 1994. Le "nez électronique": nécessité ou gadget technologique? *Dossier d'information. INRA Dijon*, 13 p.

- MISSELBROOK, T.H., HOBBS, P.J., PERSAUD, K.C., 1997. Use of an Electronic Nose to Measure Odour concentration Following Application of Cattle Slurry to Grassland. *J. Agri. Engng. Res.*, 66, pp 213-220.
- MOLIMARD, P., 1997. Les relations analyse sensorielle et analyse instrumentale. In: *Formation Inter-entreprises ADRIA*. Quimper : pp 18-26.
- MOLTO, E., SELFA, E., PONS, R., FORNES, I., 1996. Non destructive measuring of firmness using impact sensors. In: *EurAgEng 96*. Madrid : pp 1-8.
- MOY, L., 1993. Mesures d'odeurs à l'aide d'un nez électronique. (Odors measurement with electronic nose). *Caoutchoucs & plastiques*, 1993 (728), pp 54-57.
- MOY, L., TALOU, T., COLLET, P., 1991a. Capteurs à semi-conducteur et détection d'odeurs. Nouvelle méthode de mesure. *Parfums & cosmétiques*, 2, pp 58-60.
- MOY, L., TALOU, T., DUBOURDIEU, D., 1991b. Utilisation de capteurs à semi-conducteur pour la détection d'odeurs. Une nouvelle méthode de mesure. *Bios*, 22, pp 61-65.
- MUIR, D.D., HUNTER, E.A., BANKS, J.M., 1997. Aroma of cheese. 2. Contribution of aroma to the flavour of Cheddar cheese. *Milchwissenschaft*, 52 (2), pp 85-88.
- MÜLLER, R., 1991. High selectivity obtainable with nonselective chemosensors. *Sens. & Actu. B*, 4, pp 35-39.
- NAKAMOTO, T., FUKUDA, A., MORIIZUMI, T., 1993. Perfume and flavour identification by odour-sensing system using quartz-resonator sensor array and neural-network pattern recognition. *Sens. & Actu. B*, 10, pp 85-90.
- NAKAMOTO, T., FUKUDA, A., MORIIZUMI, T., ASAKURA, Y., 1991. Improvement of identification capability in an odor-sensing system. *Sens. & Actu. B*, 3(3), pp 221-226.
- NAMDEV, P.K., ALROY, Y., SINGH, V., 1998. Sniffing out trouble: use of an electronic nose in bioprocess. *Biotechnol. Prog.*, 14, pp 75-78.
- NANTO, H., KAWAI, T., SOKOOSHI, H., USUDA, T., 1993. Aroma identification using quartz-resonator in conjunction with pattern recognition. *Sens. & Actu. B*, 13-14, pp 718-120.
- NANTO, H., TSUBAKINO, S., IKEDA, S., ENDO, F., 1995. Identification of aromas from wine using quartz-resonator gas sensors in conjunction with neural-network analysis. *Sens. & Actu. B*, 24-25, pp 794-796.
- NARENDRA, P., FUGUNAGA, K., 1977. A branch and bound algorithm for feature subset selection. *IEEE Trans. Comp.*, C-26, pp 917-922.

- NEAVES, P.I., HATFIELD, J.V., 1995. A new generation of integrated electronic noses. *Sens. & Actu. B*, 26-27, pp 223-231.
- NICOLAS, C., 1998. Perspectives et difficultés d'application des capteurs gaz en contrôle qualité dans les IAA. In: *Formation Inter-entreprises ADRIA*. Paris : pp 1-12.
- NICOLAS, C., BARROS, A.S., RUTLEDGE, D.N., HOSSENLOPP, J., TRYSTRAM, G., EMONET, C., 1998a. Detecting information in gas sensor responses using analysis of variance. *Analisis*, 26 (3), pp 135-141.
- NICOLAS, C., CAREL, P., HOSSENLOPP, J., TRYSTRAM, G., RUTLEDGE, D.N., 1998b. Contribution of the gas sensors for the characterisation of products. In: *Sensoral 98: international workshop in Sensing Quality of Agricultural Products*. Montpellier : pp 357-370.
- NICOLAS, C., TRYSTRAM, G., HOSSENLOPP, J., CAREL, P., EMONET, C., 1997. Perspectives et difficultés d'application des capteurs d'odeurs en contrôle qualité dans les industries agro-alimentaires. In: *Journées Automatique, Agriculture et Agro-alimentaire*. CEMAGREF-Clermont-Ferrand : pp 55-64.
- OLAFSDOTTIR, G., MARTINSDOTTIR, E., JONSSON, E.H., 1997. Rapid gas sensor measurements to determine spoilage of capelin. *J. Agric. Food Chem.*, 45, pp 2654-2659.
- OZER, N., ENGEL, B.A., SIMON, J.E., 1995. Fusion classification techniques for fruit quality. *Trans. ASAE*, 38 (6), pp 1927-1934.
- PALAGOS, B., 1997. Critères d'évaluation des modèles. Document interne. CEMAGREF.
- PAPADOPOULOS, C.A., VLACHOS, D., AVARITSIOTIS, J.N., 1996. Comparative study of various metal-oxide-based gas-sensor architectures. *Sens. & Actu. B*, 32, pp 61-69.
- PEARCE, T.C., GARDNER, J.W., FRIEL, S., BARTLETT, P.N., BLAIR, N., 1993. Electronic Nose for Monitoring the Flavour of Beers. *Analyst*, 118, pp 371-377.
- PENZA, M., VASANELLI, L., 1997. SAW NO<sub>x</sub> gas sensor using WO<sub>3</sub> thin-film sensitive coating. *Sens. & Actu. B*, 41, pp 31-36.
- PERROT, N., TRYSTRAM, G., LEGUENNEC, D., GUELY, F., 1996. Sensor fusion for real time quality evaluation of biscuit during baking. Comparison between Bayesian and fuzzy approaches. *J. Food Eng.*, 29 (3-4), pp 301-315.
- PERSAUD, K., DODD, G.H., 1982. Analysis of discrimination mechanisms in the mammalian olfactory system using a model nose. *Nature (London)*, 299, pp 352-355.

PERSAUD, K., KHAFFAF, S.M., HOBBS, P.J., SNEATH, R.W., 1996. Assessment of conducting polymer odour sensors for agricultural malodour measurements. *Chem. Senses*, 21 (5), pp 495-505.

PICQUE, D., CATTENOZ, T., PERRET, B., CORRIEU, G., 1997. Caractérisation d'un système à base de semi-conducteurs pour l'analyse de composés volatils. In: *Récents progrès en Génie des Procédés*. Paris : pp 13-18.

PIJOLAT, C., SAUVAN, M., DURAND, J., 1997. Amélioration de la sélectivité de micro-capteurs de gaz SnO<sub>2</sub>. Application à l'automatisation des procédés. In: *Récents progrès en Génie des Procédés*. Paris : pp 37-42.

PING, W., YI, T., HAIBAO, X., FARONG, S., 1997. A novel method for diabetes diagnosis based on electronic noses. *Bios. & Bioelec.*, 12 (9-10), pp 1031-1036.

PLACKETT, R.L., BURMAN, J.P., 1946. The design of optimum multifactorial experiments. *Biometrika*, 33, pp 305-325.

POIRIER, J., 1989. Estimateurs et tests d'hypothèses. In : *Techniques de l'ingénieur, Traité Mesures et Contrôle*. Schiltigheim, F : ISTRAL (ed.), pp R250,1-14

PRIVAT, E., 1997. Elimination de l'éthanol et concentration des arômes pour la caractérisation du vin par le nez artificiel. Rapport de fin d'étude d'ingénieur ISIM en Sciences et Techniques des Industries Alimentaires. *ISIM-CEMAGREF*.

PRIVAT, E., ROUSSEL, S., GRENIER, P., BELLON-MAUREL, V., 1998. Techniques de désalcoolisation en vue de la discrimination de boissons alcoolisées par capteurs d'arômes. *Sci. Al.*, 18(5), pp 459-470.

RAKOV, W., 1998. Smart atmosphere monitoring by electronic nose: SamSelect-Electronic nose for industrial deployment. In: *Formation Inter-entreprises ADRIA*. Paris : pp 1-3.

RAZUNGLES, A., 1985. Contribution à l'étude des caroténoïdes du raisin. Teneur et localisation dans la baie, évolution en cours de maturation. Th.: *Oenologie*, Ecole Nationale Supérieure d'Agronomie de Montpellier, Montpellier.

ROBIE, D., 1997. Comparison of two electronic noses - Application to tobacco industry. In: *Proc. of the 4th symposium of Olfaction and Electronic Nose*. Nice, F : 1p.

ROBILLARD, B., 1997. Forces et faiblesses d'un "nez électronique". In: *Formation Inter-entreprises ADRIA*. Quimper : pp 1-13.

ROBILLARD, B., 1998. Deux recherches d'applications particulières en oenologie. In: *Formation Inter-entreprises ADRIA*. Paris : pp 1-13.

- ROCHA, S., DELGADILLO, I., FERRER CORREIA, A., BARROS, A., WELLS, P., 1998. Application of an Electronic Aroma Sensing System to Cork Stopper Quality Control. *J. Agric. Food Chem.*, 46 (1), pp 145-151.
- ROMAIN, A.-C., NICOLAS, J., ANDRE, P., 1997. In situ measurement of olfactive pollution with inorganic semiconductors : limitations due to humidity and temperature influence. In: *Proc. of the 4th symposium of Olfaction and Electronic Nose*. Nice, F. : pp 30.
- ROMEDER, J., 1973. *Méthodes et programmes d'analyse discriminante*. Paris : Dunod (ed.).
- ROMEYER, F.M., 1984. Les composés phénoliques du raisin *Vitis vinifera*. Evolution au cours de la maturation du fruit et conséquences technologiques. *Th.: Oenologie*, Ecole Nationale Supérieure d'Agronomie de Montpellier, Montpellier.
- ROUSSEL, S., FORSBERG, G., GRENIER, P., BELLON-MAUREL, V., 1998a. Optimisation of Electronic Nose Measurements. Part II: Influence of Experimental Parameters. *J. Food Eng.*, accepté.
- ROUSSEL, S., FORSBERG, G., STEINMETZ, V., GRENIER, P., BELLON-MAUREL, V., 1998b. Optimisation of Electronic Nose Measurements. Part I: Output Feature Selection. *J. Food Eng.*, 37 (2), pp 207-222.
- ROUSSEL, S., GRENIER, P., BELLON-MAUREL, V., 1998c. Methodology for SnO<sub>2</sub>-gas sensor selection using stepwise multivariate analysis. In: *Sensoral 98: international workshop on Sensing Quality of Agricultural Products*. Montpellier : pp 371-384.
- RUMELHART, D.E., HINTON, G.E., WILLIAMS, R.J., 1986. Learning internal representations by error propagation. In : *Parallel Distributed Processing*. Cambridge : MIT Press. (ed.).
- SADO, G., SADO, M., 1991. *Les plans d'expériences. De l'expérimentation à l'assurance qualité*. Paris : AFNOR Technique (ed.).
- SARIG, Y., 1998. utilization of the olfactory characteristics of fruit and vegetables as a potential method for determining their ripeness and readiness for harvest. In: *Sensoral 98: international workshop in Sensing Quality of Agricultural Products*. Montpellier : pp 385-430.
- SAVITZKY, A., GOLAY, M.J.E., 1964. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.*, 36, pp 1627-1637.
- SCHIFFMAN, S.S., FERMANI, B.G., TROY NAGLE, H., 1996. Analysis of medication off-odors using an electronic nose. *Chem. Sens.*, 22, pp 119-128.
- SEASHOLTZ, M.B., KOWALSKI, B., 1993. The parsimony principle applied to multivariate calibration. *Anal. Chim. Acta*, 277, pp 165-177.

- SEBAG, M., SCHOENAUER, M., 1996. Contrôle d'un algorithme génétique. *Rev. d'Intell. Artif.*, 10 (2-3), pp 389-428.
- SEKI, K., FURUYAMA, T., KAWASUMI, T., SAKURAI, Y., ISHII, H., KAJIKAWA, K., OUCHI, Y., MASUDA, T., 1997. UV Photoelectron Spectroscopy of Substituted Polyacetylenes. *J. Phys. Chem. B*, 101, pp 9165-9169.
- SHIERS, V., ADECHY, M., 1998. Use of multi-sensor array devices to attempt to predict shelf-lives of edible oils. *Seminars in Food Analysis*, 3, pp 43-52.
- SHURMER, V.H., GARDNER, J.W., 1992. Odour discrimination with an electronic nose. *Sens. & Actu. B*, 8, pp 1-11.
- SHURMER, V.H., GARDNER, J.W., CHAN., 1989. The application of discrimination techniques to alcohols and tobaccos using oxide sensors. *Sens. & Actu. B*, 18, pp 361-371.
- SIEDLECKI, W., SKLANSKY, J., 1988. On automatic feature selection. *Intern. J. of Pat. Recog. and Artif. Intell.*, 2 (2), pp 197-220.
- SIMON, J.E., HERTZONI, A., BORDELON, B., MILES, G.E., CHARLES, D.J., 1996. Electronic Sensing of Aromatic Volatiles for Quality Sorting of Blueberries. *J. Food Sci.*, 61 (5), pp 967-969 & 972.
- SMELT, A.J., 1998. Optimisation of a multi gas-sensor system and its use for quality assessment of whiting. In: *Formation Inter-entreprises ADRIA*. Paris : pp 1-10.
- SMELT, A.J., SCHEERMAN, P., LUTEN, J.B., 1997. Optimisation of a multi gas-sensor system, training of the artificial neural network and use of the electronic nose system for quality assessment of whiting. In: *Proc. of the final meeting of the EU concerted action: "Methods to determine the freshness of fish in research and industry"*. Nantes : pp 110-118.
- STEINMETZ, V., 1997. Fusion multisensorielle appliquée en temps réel aux décisions qualitatives sur les produits agro-alimentaires. *Th.: Génie des procédés bio-industriels*, ENGREF, Montpellier.
- STEINMETZ, V., BELLON, V., 1995. Sensor fusion for quality control of agricultural products. In: *2nd proc. of EurAgEng workshop on Artif. Intelli. in Agric.* Wageningen, Pays-Bas : pp 255-261.
- STEINMETZ, V., CROCHON, M., TALOU, T., BOURROUNET, B., 1995. Sensor fusion for fruit quality assessment: application to melon. In: *proceedings of the International Conference ASAE: "harvest & post-harvest technologies for fresh fruit and vegetables "*. Mexico : pp 9p.
- STOLL, D., MITROVICS, J., 1998. *Moses II User's Manual and application notes*. Tübingen : Lennartz electronic GmbH (ed.).

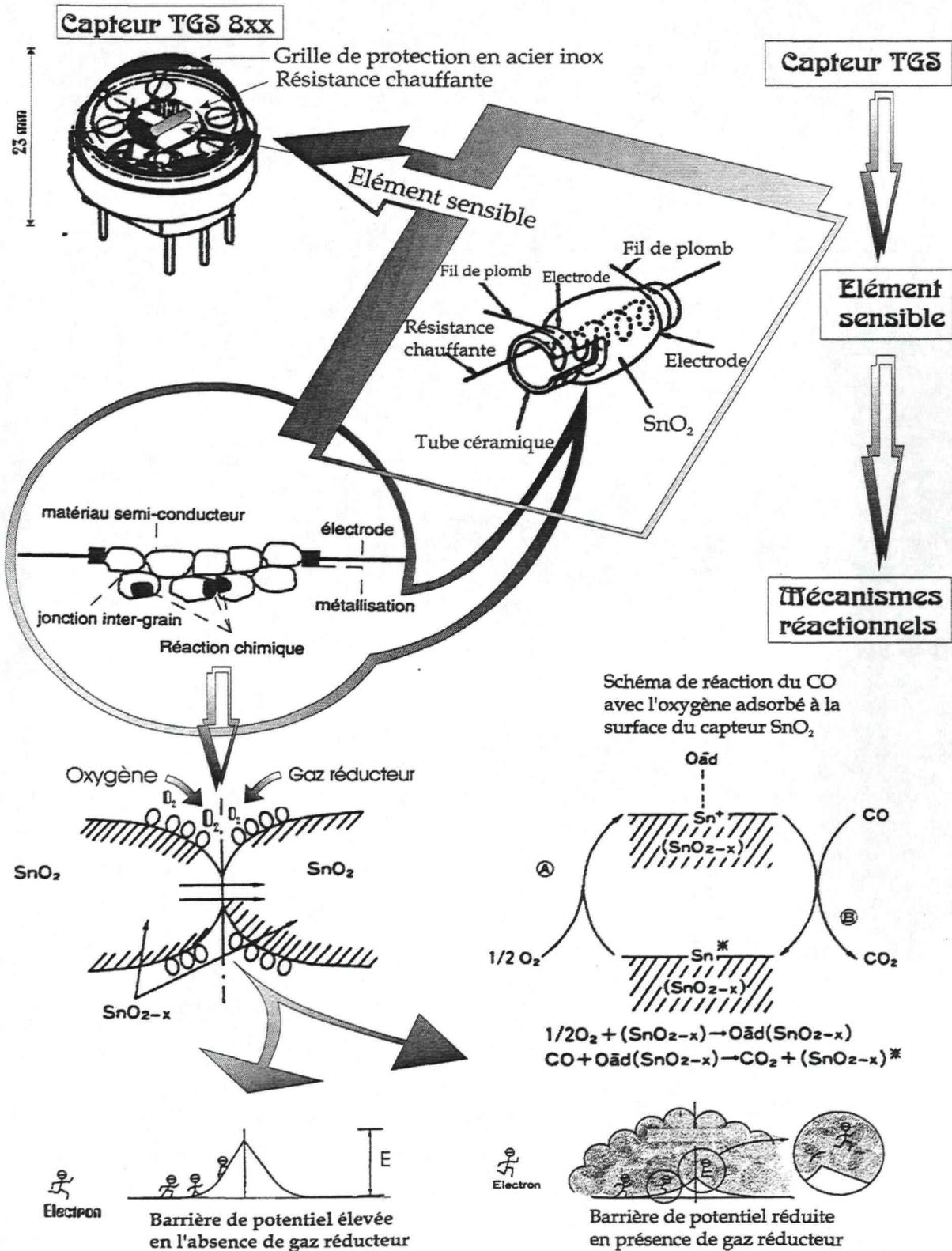
- SUZUKI, K., TAKADA, T., 1995. Highly sensitive odour sensors using various SnO<sub>2</sub> thick films. *Sens. & Actu. B*, 24-25, pp 773-776.
- TALOU, T., 1992. Un " nez électronique " pour l'industrie. *La recherche*, 246 (23), pp 1058-1059.
- TAN, T.T., LOUBET, F., LABRECHE, S., AMINE, H., 1997. Quality control of coffee using the FOX4000 electronic nose. In: *ISIE 97, IEEE Intern. Symp. on Indust. Elec.* Guimaraes, P : pp 140-146.
- TJOMB, P., 1993. Un nez électronique pour maîtriser son procédé. *RLA*, 504, pp 60-62.
- TOMINAGA, T., MURAT, M.L., DUBOURDIEU, D., 1998. Development of a method for analyzing the volatile thiols involved in the characteristic aroma of wines made from *Vitis vinifera* L. cv. Sauvignon Blanc. *J. Agric. Food Chem.*, 46 (3), pp 1044-1048.
- THOMPSON, J., 1990. Performance index. *Univ. of California, Davis*.
- VAN DEN BROEK, W., WIENKE, D., MELSEN, W.J., BUYDENS, L.M.C., 1997. Optimal wavelength range selection by a genetic algorithm for discrimination purposes in spectroscopic infrared imaging. *Appl. Spec.*, 51 (8), pp 1210-1217.
- VAN DE VOORT, F.R., 1992. Fourier Transform Infrared Spectroscopy applied to food analysis. *Food Research International*, 25, pp 397-403.
- VAN DE VOORT, F.R., ISMAIL, A., SEDMAN, J., 1995. A rapid, automated method for determination of cis and trans content of fats and oils by Fourier transform infrared spectroscopy. *JAOCS*, 72 (8), pp 873-880.
- VARDIN, R.-J., VIDAL, B., 1991. Etude par spectroscopie UV du chromophore 3,4-méthylènedioxybenzène dans différents environnements moléculaires et dans des produits naturels. *Th.*: Institut National Polytechnique de Toulouse.
- VENTURI, G., 1996. Algorithmes génétiques et apprentissage. *Rev. d'Intell. Artif.*, 10 (2-3), pp 345-388.
- VERNAT-ROSSI, V., 1998. Capteurs d'arômes: construire une base de données pertinente. In: *Formation Inter-entreprises ADRIA*. Paris : pp 1-21.
- VERNAT-ROSSI, V., VERNAT, G., BERDAGUE, J., 1996. Discrimination de produits agro-alimentaires par senseurs de gaz à semi-conducteurs en atmosphère ambiante de laboratoire. Différentes approches de traitement du signal. *Analysis*, 24 pp 309-315.
- VILANOVA, X., LLOBET, E., ALCUBILLA, R., SUEIRAS, J.E., CORREIG, X., 1996. Analysis of the conductance transient in thick-film tin oxide gas sensors. *Sens. & Actu. B*, 31, pp 175-180.

- VISSER, F.R., TAYLOR, M., 1998. Improved performance of the Aromascan A32S electronic nose and its potential for detecting aroma differences in dairy products. *J. of Sensory Studies*, 13 (1), pp 95-120.
- VLACHOS, D., PAPADOPOULOS, C., AVARITSIOTIS, J., 1996a. On the electronic interaction between additives and semiconducting oxide gas sensor. *Appl. Phys. Lett.*, 69 (5), pp 650-652.
- VLACHOS, D., PAPADOPOULOS, C., AVARITSIOTIS, J., 1996b. A technique for suppressing ethanol interference employing Seebeck effect devices with carrier concentration modulation. *In: Eurosensors X. Leuven, B* : pp 39-42.
- VLACHOS, D., SKAFIDAS, P., AVARITSIOTIS, J., 1995. The effect of humidity on tin-oxide thick-film gas sensors in the presence of reducing and combustible gases. *Sens. & Actu. B*, 24-25, pp 491-494.
- WALMSLEY, A., HASHWELL, S., METCALFE, E., 1991. Methodology for the selection of suitable sensors for incorporation into a gas sensor array. *Anal. Chim. Acta*, 242, pp 31-36.
- WHITE, J., KAUER, J., DICKINSON, T., WALT, D., 1996. Rapid analyte recognition in a device based on optical sensors and olfactory system. *Anal. Chem.*, 68 (13), pp 2191-2202.
- WIDE, P., WINQUIST, F., DRIANKOV, D., 1997. An air-quality sensor system with fuzzy classification. *Meas. Sci. Technol.*, 8, pp 138-146.
- WILSON, D.M., DEWEERTH, S.P., 1997. Signal processing for improving gas sensor response time. *Sens. & Actu. B*, 41 (1-3), pp 63-70.
- WISE, B.M., GALLAGHER, N.B., 1997. *PLS\_Toolbox. A chemometrics tutorial*. Manson, WA : Eigenvector Research, Inc. (ed.).
- YANG, J.H., HONAVAR, V., 1998. Feature subset selection using a genetic algorithm. *IEEE Intell. Syst. & Their Appl.*, 13 (2), pp 44-49.
- YEA, B., KONISHI, R., OSAKI, T., SUGAHARA, K., 1994. The discrimination of many kinds of odor species using fuzzy reasoning and neural networks. *Sens. & Actu. A*, 45, pp 159-165.
- ZIMMERMANN, D., LECLERCQ, C., 1995. Le nez électronique: une technologie intéressante pour la caractérisation des malts spéciaux. *BIOS Boissons Conditionnement* , 257, pp 29-34.
- ZINI, M., 1997. Nez électronique: vers des applications concrètes. *RIA* , 568, pp 62-63.
- ZÜRCHER, E., KUCERA, L.J., 1997. Electronic odour testing of wood and wood products. *In: Plant Biomechanics 1997*. Reading, GB : pp 61-64.

# *Annexes*

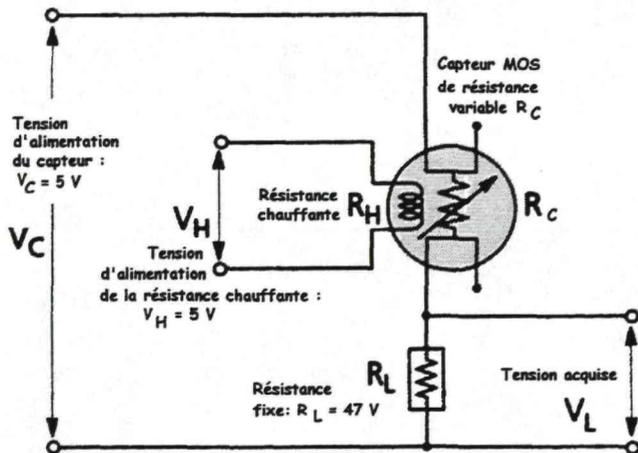
# VIII. ANNEXES

## Annexe 1 : schéma de fonctionnement d'un capteur en oxyde métallique semi-conducteur



## Annexe 2 : méthode d'acquisition de la résistance d'un capteur MOS

La variation de résistance due à l'adsorption de composés volatils est généralement acquise grâce à un pont diviseur de tension, comme le montre la **Figure 1** : la variation de tension acquise est inversement proportionnelle au changement de résistance. Celle-ci est une fonction non linéaire de la concentration en molécules volatiles (logarithmique ou  $C^n$  avec  $n < 1$ ) (*SHURMER et al.*, 1989) (*DAVIDE et al.*, 1993), avec une approximation de linéarité pour de faibles concentrations et le respect du principe de superposition.



La tension acquise est  $V_L$ .

La résistance de l'élément sensible MOS  $R_C$  peut alors être calculée grâce à la relation suivante :

$$R_C = R_L * \left( \frac{V_C}{V_L} - 1 \right)$$

Donc, lors de l'adsorption de molécules volatiles, la résistance  $R_C$  du capteur diminuant, la tension acquise  $V_L$  augmente.

Figure 1 : schéma d'acquisition de la résistance d'un capteur MOS FIGARO

**Annexe 3 : répartition des échantillons de moûts de raisins blancs**

REFERENCE	PROVENANCE	CAPTEURS D'AROMES LCA1000	SPECTROMETRE IR-TF	SPECTROMETRE UV	SUCRES (g/l)	ACIDITE TOTALE	PH
BOUBJ GA SA	Gaillac (81)	SA1	SA1	SA1	165	4.60	3.22
BOUGB GA SA	Gaillac (81)	SA2	SA2	SA2	141	5.41	3.21
JLFAV GA SA	Gaillac (81)	SA3	SA3	SA3	165	4.96	3.13
LAGAN GA SA	Gaillac (81)	SA4	SA4	SA4	160	5.32	3.15
AUQUE GA SA	Gaillac (81)	SA5	SA5	SA5	163	5.32	3.18
VAISS GA SA	Gaillac (81)	SA6	SA6	SA6	190	3.61	3.31
RECH GA SA	Gaillac (81)	SA7	SA7	SA7	158	5.77	3.19
ROUQU GA SA	Gaillac (81)	SA8	SA8	SA8	162	5.05	3.25
MERCI GA SA	Gaillac (81)	SA9	SA9	SA9	150	5.68	3.21
TERAL GA SA	Gaillac (81)	SA10	SA10	SA10	155	4.60	3.22
REYNES GA SA	Gaillac (81)	SA11	SA11	SA11	167	4.96	3.19
SEGUR GA SA	Gaillac (81)	SA12	SA12	SA12	145	3.97	3.22
RIVIER GA SA	Gaillac (81)	SA13	SA13	SA13	163	4.42	3.20
VALAX GA SA	Gaillac (81)	SA14	SA14	SA14	140	6.04	3.18
ESSAI TAILLE ESCABE GA SA	Gaillac (81)	SA15	SA15	SA15	182	4.42	3.21
377 SD SA	S <sup>t</sup> Sardos (81)	SA16	SA16	SA16	158	4.78	3.48
VIRE9 CA SA	Gaillac (81)	SA17	SA17	SA17	202	3.70	3.44
SAU-N-140	Nîmes (30)	SA18	SA18	SA18	166	5.40	3.54
SAU-N-141.1	Nîmes (30)	SA19	SA19	SA19	190	4.01	3.34
SAU-N-141.2	Nîmes (30)	SA20	SA20	SA20	195	3.70	3.40
SAU-N-145	Nîmes (30)	SA21	SA21	SA21	174	6.35	3.38
SAU-N-146	Nîmes (30)	SA22	SA22	SA22	204	6.40	3.04
SAU-N-1471	Nîmes (30)	SA23	SA23	SA23	150	6.8	3.29
SAU-N-1471-b	Nîmes (30)	SA24	SA24	SA24	150	7.00	3.35
SAU-N-1472	Nîmes (30)	SA25	SA25	SA25	175	4.85	3.41
SAU-N-1472-h	Nîmes (30)	SA26	SA26	SA26	180	4.20	3.48
SAU-148.1	Nîmes (30)	SA27	SA27	SA27	171	4.80	3.5
SAU-148.2	Nîmes (30)	SA28	SA28	SA28	171	4.85	3.52
SAU-N-150	Nîmes (30)	SA29	SA29	SA29	192	3.30	3.34
SAU-N-151	Nîmes (30)	SA30	SA30	SA30	180	2.95	4.00
SAU-N-152	Nîmes (30)	SA31	SA31	SA31	190	3.75	3.78
SAU-N-153	Nîmes (30)	SA32	SA32	SA32	182	4.70	3.80
SAU-N-155	Nîmes (30)	SA33	SA33	SA33	183	4.45	3.54
SAU-N-156	Nîmes (30)	SA34	SA34	SA34	187	4.90	3.55
SAU-N-157	Nîmes (30)	SA35	SA35	SA35	176	5.60	3.33
SAU-N-158	Nîmes (30)	SA36	SA36	SA36	182	5.45	3.47
SAU-PER-1	Pech Rouge (11)	SA37			187	4.98	3.22
SAU-PER-2	Pech Rouge (11)	SA38	SA38		187	5.62	3.24
SAU-PER-3	Pech Rouge (11)	SA39	SA39	SA39	200	3.48	3.38
SAU-PER-4	Pech Rouge (11)	SA40	SA40	SA40	192	3.82	3.34
SAU-PER-5	Pech Rouge (11)	SA41	SA41	SA41	203	4.70	3.28
SAU-PER-6	Pech Rouge (11)	SA42	SA42	SA42	180	4.45	3.22
SAU-PER-8	Pech Rouge (11)	SA43	SA44	SA44	192	5.30	3.25
SAU-PER-9	Pech Rouge (11)	SA44	SA45	SA45	180	5.01	3.21
SAU-PER-10	Pech Rouge (11)	SA45	SA46	SA46	214	4.04	3.53
BAGA4 CA SA	Gaillac (81)	SA46	SA47	SA47	197	3.79	3.33
BOUGM GA SA	Gaillac (81)	SA47			165	4.69	3.44
297 SD SA	S <sup>t</sup> Sardos (81)	SA48			167	4.69	3.5
917 SD SA	S <sup>t</sup> Sardos (81)	SA49			187	4.60	3.52
<b>SAUVIGNON</b>		<b>49</b>	<b>45</b>	<b>44</b>	<b>175.8</b>	<b>4.80</b>	<b>3.36</b>
<b>Ecart-type</b>					<b>17.8</b>	<b>0.89</b>	<b>0.19</b>

REFERENCE	PROVENANCE	CAPTEURS D'AROMES LCA1000	SPECTROMETRE IR-TF	SPECTROMETRE UV	SUCRES (g/l)	ACIDITE TOTALE	PH
575R1 SX MA	Gaillac (81)	MA1	MA1	MA1	167	3.41	3.40
899R1 SX MA	Gaillac (81)	MA2	MA2	MA2	157	3.32	3.24
739R1 SX MA	Gaillac (81)	MA3	MA3	MA3	173	3.50	3.36
740R1 SX MA	Gaillac (81)	MA4	MA4	MA4	158	3.59	3.27
741R1 SX MA	Gaillac (81)	MA5	MA5	MA5	170	3.68	3.24
738R1 SX MA	Gaillac (81)	MA6	MA6	MA6	160	4.05	3.18
BAGA2 CA MA	Cahors (46)	MA7	MA7	MA7	138	4.70	3.05
738 R2 SX MA	Gaillac (81)	MA8	MA8	MA8	178	3.39	3.32
739 R2 SX MA	Gaillac (81)	MA9	MA9	MA9	180	3.01	3.30
740 R2 SX MA	Gaillac (81)	MA10	MA10	MA10	183	3.01	3.46
898 R2 SX MA	Gaillac (81)	MA11	MA11	MA11	175	2.54	3.42
899 R2 SX MA	Gaillac (81)	MA12	MA12	MA12	172	2.92	3.44
741 R2 SX MA	Gaillac (81)	MA13	MA13	MA13	183	3.48	3.36
575 R2 SX MA	Gaillac (81)	MA14	MA14	MA14	175	3.10	3.39
<b>MAUZAC</b>		<b>14</b>	<b>14</b>	<b>14</b>	<b>169.2</b>	<b>3.41</b>	<b>3.32</b>
<b>Ecart-type</b>					<b>12.4</b>	<b>0.53</b>	<b>0.11</b>

REFERENCE	PROVENANCE	CAPTEURS D'AROMES LCA1000	SPECTROMETRE IR-TF	SPECTROMETRE UV	SUCRES (g/l)	ACIDITE TOTALE	PH
EHEM GE CO	Gers (32)	CO1	CO1		131	8.48	3.25
ENHER GE CO	Gers (32)	CO2	CO2	CO2	167	6.17	3.16
MOUTON GE CO	Gers (32)	CO3	CO3	CO3	125	8.01	3.10
DESR2 GE CO	Gers (32)	CO4	CO4	CO4	151	7.78	3.30
BRUR2 GE CO	Gers (32)	CO5	CO5	CO5	141	8.16	3.25
EFTM1 GE CO	Gers (32)	CO6	CO6		118	7.78	3.19
EFFGR GE CO	Gers (32)	CO7	CO7		126	7.97	3.19
EFFG2 GE CO	Gers (32)	CO8	CO8		108	7.11	3.28
EFDFL GE CO	Gers (32)	CO9	CO9	CO9	109	8.45	3.17
EFVER GE CO	Gers (32)	CO10	CO10	CO10	113	7.68	3.27
EFNO2 GE CO	Gers (32)	CO11	CO11	CO11	114	7.88	3.23
EFDF2 GE CO	Gers (32)	CO12	CO12		116	7.78	.28
EFNOU GE CO	Gers (32)	CO13	CO13	CO13	121	8.16	3.22
EFVE2 GE CO	Gers (32)	CO14	CO14	CO14	109	6.91	3.33
608 GE CO	Gers (32)	CO15	CO16	CO16	158	6.34	3.29
609 GE CO	Gers (32)	CO16	CO17	CO17	155	6.72	3.33
605 GE CO	Gers (32)	CO17	CO18	CO18	158	6.63	3.27
695 GE CO	Gers (32)	CO18	CO19	CO19	138	6.72	3.28
553 GE CO	Gers (32)	CO19	CO21	CO21	143	6.43	3.21
693 GE CO	Gers (32)	CO20	CO22		158	6.82	3.22
606 GE CO	Gers (32)	CO21			150	6.91	3.23
607 GE CO	Gers (32)	CO22			151	6.63	3.22
625 GE CO	Gers (32)	CO23			155	7.30	3.22
626 GE CO	Gers (32)	CO24			145	6.34	3.21
551 GE CO	Gers (32)	CO25			158	6.05	3.31
692 GE CO	Gers (32)	CO26			151	7.30	3.22
694 GE CO	Gers (32)	CO27			151	5.76	3.19
<b>COLOMBARD</b>		<b>27</b>	<b>20</b>	<b>14</b>	<b>137.8</b>	<b>7.20</b>	<b>3.13</b>
<b>Ecart-type</b>					<b>18.8</b>	<b>0.78</b>	<b>0.57</b>

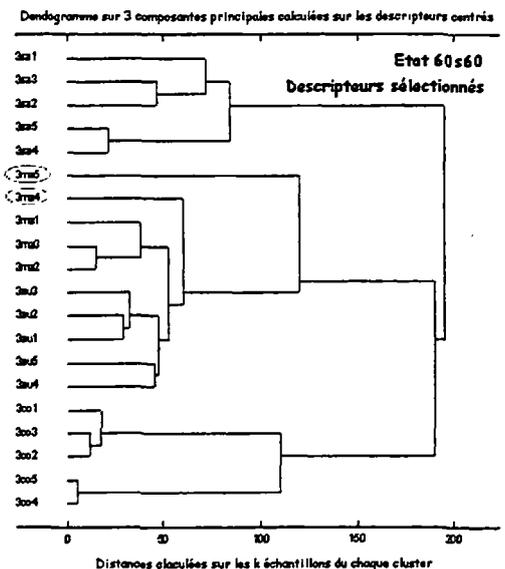
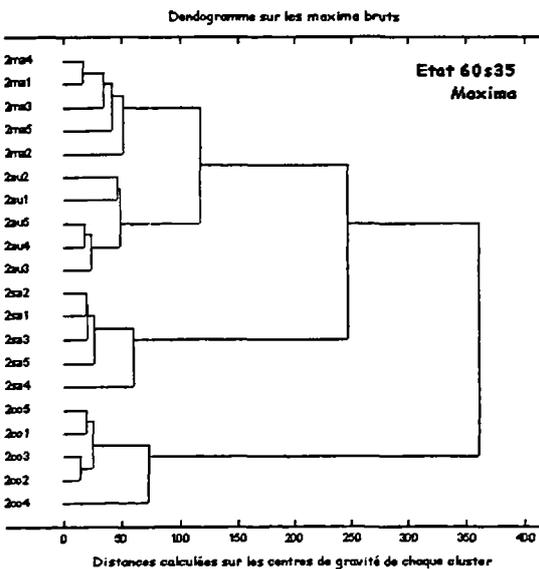
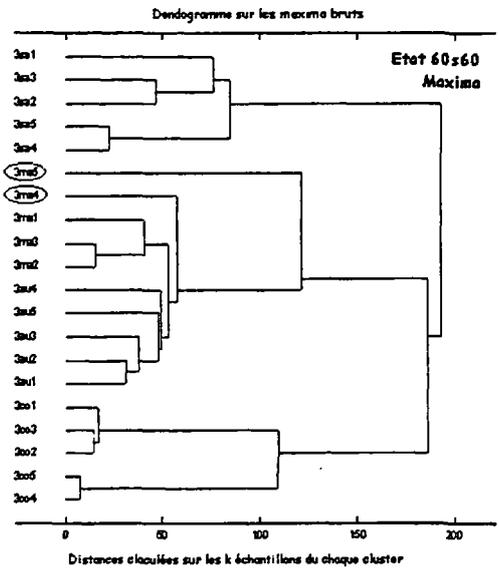
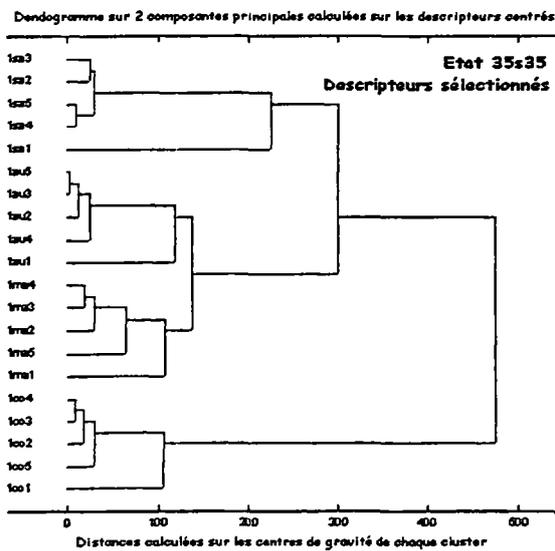
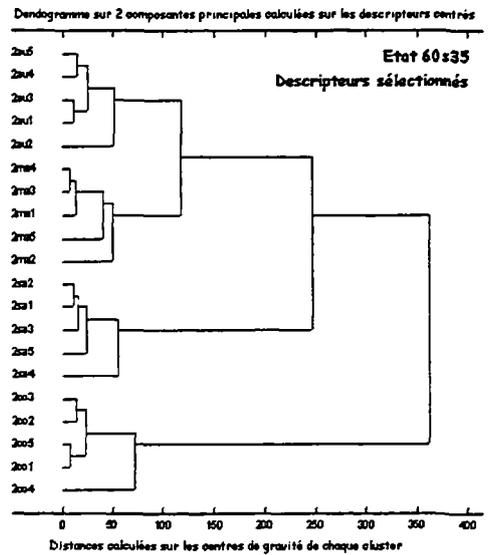
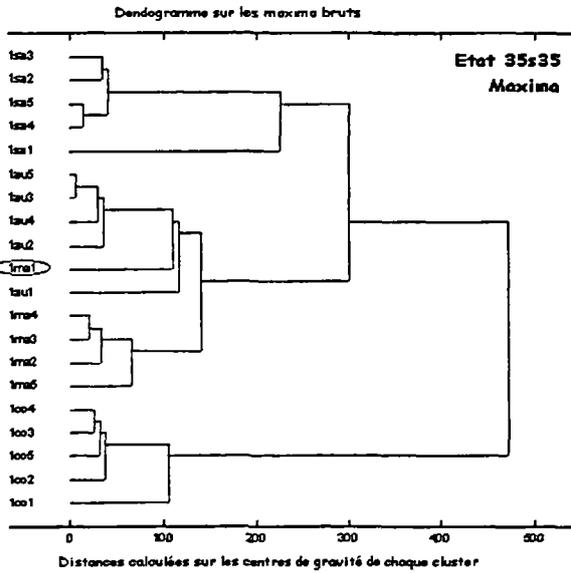
REFERENCE	PROVENANCE	CAPTEURS D'AROMES LCA1000	SPECTROMETRE IR-IF	SPECTROMETRE UV	SUCRES (g/l)	ACIDITE TOTALE	PH
MBLR1 SE LL	Gaillac (81)	AU1	AU1	AU1	123	3.40	3.39
096 SD CA	S <sup>t</sup> Sardos (81)	AU2	AU2	AU2	192	4.20	3.5
610 SD MU	S <sup>t</sup> Sardos (81)	AU3	AU3	AU3	155	4.29	3.53
BAGA1 CA CA	Cahors (46)	AU4	AU4	AU4	199	4.51	3.27
BAGA3 CA MU	Cahors (46)	AU5	AU5	AU5	155	3.41	3.41
BAGA5 CA SE	Cahors (46)	AU6	AU6	AU6	162	3.50	3.24
BAGA6 CA VI	Cahors (46)	AU7	AU7	AU7	178	4.51	3.26
VIRE7 CA CA	Cahors (46)	AU8	AU8	AU8	175	4.15	3.34
VIRE8 CA MU	Cahors (46)	AU9	AU9	AU9	157	4.05	3.49
VIRE10 CA SE	Cahors (46)	AU10	AU10	AU10	165	3.22	3.38
VIRE11 CA VI	Cahors (46)	AU11	AU11	AU11	178	3.96	3.41
SEL EGOUT RA LL 203	Gaillac (81)	AU12	AU12	AU12	178	3.68	3.43
SEL 46% RA LL	Gaillac (81)	AU13	AU13	AU13	178	3.50	3.50
SEL 67% RA LL	Gaillac (81)	AU14	AU14	AU14	178	3.40	3.53
SEL 72% RA LL	Gaillac (81)	AU15	AU15	AU15	178	3.41	3.52
Pinot blanc-PER	Pech Rouge (11)	AU16	AU16	AU16	209	3.90	3.36
Aranel-PER	Pech Rouge (11)	AU17	AU17	AU17	221	4.40	3.29
Vermentino-PER	Pech Rouge (11)	AU18	AU19	AU19	130	5.57	3.16
Roussanne-PER	Pech Rouge (11)	AU19	AU20	AU20	192	3.96	3.40
Arriloba-PER	Pech Rouge (11)	AU20	AU21	AU21	192	2.79	3.55
Chenin-PER	Pech Rouge (11)	AU21	AU22	AU22	152	4.76	3.45
Riesling-PER	Pech Rouge (11)	AU22	AU23	AU23	183	3.92	3.30
CHA-PER-E	Pech Rouge (11)	AU23	AU24	AU24	192	4.42	3.35
CHA-PER-T	Pech Rouge (11)	AU24	AU25	AU25	184	4.33	3.42
Perdea-PER	Pech Rouge (11)	AU25	AU26	AU26	169	4.75	3.26
VION 600 B	Nîmes (30)	AU26	AU27	AU27			
VION 600 A	Nîmes (30)	AU27	AU28	AU28			
VION 600 C	Nîmes (30)	AU28	AU29	AU29			
VION 600 D	Nîmes (30)	AU29	AU30	AU30			
363	Nîmes (30)	AU30	AU31	AU31			
364	Nîmes (30)	AU31	AU32	AU32			
365	Nîmes (30)	AU32	AU33	AU33			
366	Nîmes (30)	AU33	AU34	AU34			
367	Nîmes (30)	AU34	AU35	AU35			
MBLR3 SE LL	Gaillac (81)	AU35	AU38	AU38	145	3.46	3.49
315 SD SE	S <sup>t</sup> Sardos (81)	AU36		AU36	167	3.22	3.68
<b>Autres Cépages</b>		<b>36</b>	<b>35</b>	<b>36</b>	<b>173.6</b>	<b>3.95</b>	<b>3.40</b>
<b>Ecart-type</b>					<b>22.2</b>	<b>0.61</b>	<b>0.12</b>

### Annexe 4 : Echantillons de référence des 4 classes de cépages pour la sélection des capteurs.

SAUVIGNON			MAUZAC		
	REFERENCE	N°		REFERENCE	N°
1	JLFAV GASA	SA3	1	575R1 SX MA	MA1
2	LAGANGASA	SA4	2	899R1 SX MA	MA2
3	TERAL GASA	SA10	3	739R1 SX MA	MA3
4	RIVIERGASA	SA13	4	740R1 SX MA	MA4
5	SA-N-140	SA18	5	741R1 SX MA	MA5
6	SA-N-1472	SA25	6	738R1 SX MA	MA6
7	SA-N-152	SA31	7	738R2 SX MA	MA8
8	SA-PER-4	SA40	8	739R2 SX MA	MA9
9	SA-PER-9	SA44	9	740R2 SX MA	MA10
10	BAGA4	SA46	10	575R2 SX MA	MA14

COLOMBARD			AUTRES CEPAGES		
	REFERENCE	N°		REFERENCE	N°
1	DESR2 GE CO	CO4	1	096 SD CA	AU2
2	BRUR2 GE CO	CO5	2	610 SD MU	AU3
3	EFTEM1 GE CO	CO6	3	BAGA5 CA SE	AU6
4	EFFGR GE CO	CO7	4	VIRE11 CA VI	AU11
5	EFFG2 GE CO	CO8	5	SEL EGOUT RA LL	AU12
6	EFDFL GE CO	CO9	6	PER-ROUSSANE	AU19
7	EFNO2 GE CO	CO11	7	PER-CHENIN	AU21
8	EFDF2 GE CO	CO12	8	PER-RIESLING	AU22
9	EFNOU GE CO	CO13	9	CHA-PER-T	AU24
10	EFVE2 GE CO	CO14	10	PER-PERDEA	AU25

**Annexe 5 : classifications hiérarchiques ascendantes des 5 répétitions d'échantillon de moûts de cépages différents, dans 3 états différents (conditions expérimentales) et grâce à 2 sous-ensembles de descripteurs (colonne 1 : maxima ; colonne2 : maxima et descripteurs n°10, 14, 20 et 24)**



## Annexe 6 : Organisation des dégustations

Les tests des postes 1 à 8 correspondent à des échantillons correctement discriminés par les capteurs d'arômes, et les postes 9 à 16 correspondent à des échantillon ambigus.

Le nombre d'échantillons correctement discriminés par le panel sensoriel comportant 8 oenologues est reporté dans les quatrième colonnes. Les cases noircies correspondent aux échantillons non significativement discriminés. Les case en gris foncé correspondent à une différence significative entre les 2 échantillons testés, avec un risque  $\alpha < 5\%$ , celles en gris clair à un risque de  $\alpha < 1\%$  et celles en blanc à  $\alpha < 0.1\%$ .

**Tableau 1 : échantillons des tests triangulaires. (Numérotation : C.f. Annexe 2)**

ECHANTILLONS BIEN DISCRIMINES PAR LES CAPTEURS D'AROMES				Nombre d'échantillons correctement discriminés	ECHANTILLONS MAL DISCRIMINES PAR LES CAPTEURS D'AROMES				Nombre d'échantillons correctement discriminés
Echantillons	A	B	C		Echantillons	A	B	C	
Poste 1	SA44	SA44	CO7	6/8	Poste 9	SA30	SA30	CO14	7/8
Poste 2	SA11	CO15	CO15	5/8	Poste 10	SA30	SA44	SA30	5/8
Poste 3	SA10	AU3	SA10	8/8	Poste 11	SA31	SA31	SA11	3/8
Poste 4	SA8	AU12	AU12	5/8	Poste 12	SA31	SA31	AU12	6/8
Poste 5	AU24	CO8	AU24	4/8	Poste 13	CO13	CO5	CO13	5/8
Poste 6	MA9	CO16	MA9	8/8	Poste 14	SA11	SA11	CO20	7/8
Poste 7	MA6	MA6	SA14	7/8	Poste 15	AU18	AU32	AU32	8/8
Poste 8	AU14	MA3	MA3	6/8	Poste 16	AU16	SA44	SA44	6/8

## Annexe 7 : résultat de la fusion de niveau bas entre le spectre IR-TF/UV filtré a posteriori et les courbes du LCA 1000

La courbe commune contient 4306 variables, 3605 points provenant des courbes des 5 capteurs concaténées du LCA et 701 points du spectre IR&UV. Cependant, les résultats de classification sont dégradés par rapport à ceux basés uniquement sur les spectres IR-TF et UV, aboutissant à 44.8% d'erreur en prédiction.

**Tableau 2 : matrice de confusion des prédictions en validation croisée par bloc (10 lots) d'une PLS discriminante appliquée aux spectres IR-TF & UV sélectionnés a posteriori et courbes des 5 MOS du LCA (17 lvs et données centrées)**

	CLASSES REELLES				TOTAL
	Sauvignon	Mauzac	Colombard	Autres cépages	% erreur
Nombre d'échantillons	45	14	20	35	114
CLASSES PREDITES					
Sauvignon	16	3	3	5	40.7%
Mauzac	13	10	2	2	73.0%
Colombard	7	1	8	3	57.9%
Autres cépages	8	0	1	25	26.5%
% bons classements	36.4%	71.4%	57.1%	71.4%	55.2%
Taux d'erreur	63.6%	28.6%	42.9%	28.6%	44.8%

### **Annexe 8 : résultat de la fusion par erreur minimale de Bayes des classifications issues des spectres IR-TF et UV bruts.**

La matrice de causalité du capteur IR-TF est calculée à l'aide de la matrice de confusion issue de la classification par les spectres bruts (Tableau 25). Celle du spectromètre UV est calculée à partir de la matrice de confusion du Tableau 28.

**Tableau 3 : Matrice de causalité du spectromètre IR-TF**

CLASSES PREDITES	CLASSES REELLES (107 ECHANTILLONS)			
	Sauvignon	Mauzac	Colombard	Autres cépages
Echantillons	44	14	14	35
$P_k$	0.41	0.13	0.13	0.33
$\hat{S}a$	0.68	0.29	0.29	0.23
$\hat{M}a$	0	0.29	0	0.03
$\hat{C}o$	0.09	0.14	0.71	0
$\hat{A}u$	0.23	0.29	0	0.74

**Tableau 4 : Matrice de causalité du spectromètre UV**

CLASSES PREDITES	CLASSES REELLES			
	Sauvignon	Mauzac	Colombard	Autres cépages
$\hat{S}a$	0.59	0.21	0.50	0.34
$\hat{M}a$	0.07	0.43	0	0.09
$\hat{C}o$	0.02	0.07	0.36	0
$\hat{A}u$	0.32	0.29	0.14	0.57

Les affectations sont calculées grâce à l'équation du minimum de l'erreur de Bayes et sont détaillées dans le tableau suivant.

**Tableau 5 : règle d'affectation d'erreur minimale pour les deux capteurs IR-TF et UV**

CLASSES PREDITES par le capteur IR-TF	CLASSES PREDITES par le capteur UV			
	$\hat{S}a$	$\hat{M}a$	$\hat{C}o$	$\hat{A}u$
$\hat{S}a$	Sa	Sa	Co	Sa
$\hat{M}a$	Ma	Ma	Ma	Ma
$\hat{C}o$	Co	Ma	Co	Co
$\hat{A}u$	Au	Au	Ma	Au

Cette règle d'affectation des échantillons fournit la classification suivante :

**Tableau 6 : résultats de la fusion de niveau haut entre les spectromètres IR-TF & UV**

	CLASSES REELLES				TOTAL
Nombre d'échantillons	44	14	14	35	107
CLASSES PREDITES	Sauvignon	Mauzac	Colombard	Autres cépages	
Sauvignon	30	4	1	8	30.2%
Mauzac	1	5	0	1	28.6%
Colombard	4	1	13	0	27.8%
Autres cépages	9	4	0	26	33.3%
% bons classements	68.2%	35.7%	92.9%	74.3%	69.2%
Taux d'erreur FUSION par erreur minimale de Bayes	31.8%	64.3%	7.1%	25.7%	30.8%

Taux d'erreur UV	40.1%	57.1%	64.3%	46.0%	47.7%
Taux d'erreur IR-TF	24.4%	71.4%	20.0%	31.4%	30.7%

La fusion des capteurs n'améliore pas le résultat global de classification par rapport à celui du spectromètre IR-TF ; par contre, la discrimination du cépage Colombard est nettement améliorée (7.1% d'erreur).

---

## Optimisation des capteurs d'arômes et fusion multisensorielle appliquée à la caractérisation de produits agro-alimentaires

---

Cette thèse traite le problème de la classification à partir de données multivariées issues de capteurs sensoriels : capteurs d'arômes et spectromètres.

Le premier axe consiste à optimiser de nouveaux systèmes de mesure aromatique, les « nez électroniques » ou capteurs d'arômes, grâce à la conception d'une démarche méthodologique expérimentale comprenant 3 étapes :

l'extraction de descripteurs de forme des signaux et leur sélection grâce à des critères de performance définis pour les systèmes multi-capteurs ; l'application de plans d'expériences afin de déterminer l'influence de facteurs extérieurs sur la qualité de la mesure et les conditions expérimentales optimales ; enfin, l'adaptation de techniques de sélection de variables (méthodes pas à pas et algorithmes génétiques) au choix des transducteurs à insérer dans le système multi-capteur d'arômes.

Ces méthodologies ont été appliquées à des solutions modèles et validées sur des échantillons naturels : des moûts de raisins blancs de cépages différents.

Le second axe de la thèse est l'application de la fusion multi-capteur à la classification supervisée à partir de données multivariées redondantes. Elle est employée pour la discrimination de cépages de moûts de raisins blancs, grâce aux mesures aromatiques préalablement optimisées et aux analyses spectrométriques dans l'infrarouge et ultraviolet.

Deux méthodes de fusion multisensorielle sont proposées et comparées : la fusion de niveau bas, combinant les signaux bruts, et la fusion de niveau haut, tenant compte des classifications individuelles et de l'expertise, grâce à l'erreur ou au risque minimum de Bayes. Les cépages sont correctement discriminés par la technique de fusion Bayésienne (4.7% d'erreur), montrant ainsi l'efficacité des capteurs ainsi que des méthodes d'optimisation et de fusion.

Ces démarches pourront être appliquées à toute problématique de discrimination basée sur la combinaison de données multivariées issues de différents capteurs.

**Mots clés :** capteurs d'arômes ; nez électronique ; fusion multi-capteur ; spectrométrie IR-TF ; spectrométrie UV ; agro-alimentaire ; moûts de raisin.

---

## Aroma sensor optimisation and multi-sensor fusion applied to food product characterisation

---

This Ph.D. work is dealing with classification problem using multivariate data provided by sensory devices : aroma sensors and spectrometers.

The first axis is the optimisation of new aroma measurement devices, « electronic noses » or aroma sensors, by designing an experimental methodological approach, divided in three stages :

extraction of shape features from the signals and their selection relatively to performance indexes designed for the multi-sensor array ; the application of experimental design methods, in order to assess the external parameter influence on measurement quality and determine the most suitable experimental conditions ; the use of feature selection techniques (stepwise methods and genetic algorithms) to choose the transducers to be included in the aroma multi-sensor array.

These methodologies have been applied to synthetic solutions and validated on natural samples : different varieties of white grape musts.

The second axis of this thesis is to apply multi-sensor fusion techniques to supervised classification using multivariate redundant data. These methods are used for the discrimination of must varieties, carrying out aroma measurements previously optimised and infrared and ultraviolet spectrometric analysis.

Two multi-sensor fusion techniques have been compared : low level fusion, combining the sensor rough signals, and high level fusion taking into account the sensor individual classifications, using Bayes minimum error or risk. The varieties were correctly discriminated, by performing Bayesian fusion (error rate of 4.7%), and consequently showing the efficiency of the sensor used and of the optimisation and fusion methods.

These approaches can be applied to any discrimination problem, dealing with complex sensor combination.

**Key words :** aroma sensor ; electronic nose ; spectrometry FT-IR ; spectrometry UV ; multi-sensor fusion ; food products ; grape must.