

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE  
UNIVERSITÉ DU QUÉBEC

MÉMOIRE PRÉSENTÉ À  
L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

COMME EXIGENCE PARTIELLE  
À L'OBTENTION DE LA  
MAÎTRISE EN GÉNIE ÉLECTRIQUE  
M.Ing.

PAR  
YOUSOUF CHERAWALA

PRÉDICTION DE LA TENEUR EN ANTHOCYANE DU FRUIT DE LA  
CANNEBERGE PAR VISION PAR ORDINATEUR

MONTRÉAL, LE 27 JUIN 2007

© droits réservés de Youssef Chherawala

CE MÉMOIRE A ÉTÉ ÉVALUÉ  
PAR UN JURY COMPOSÉ DE :

M. Richard Lepage, directeur de mémoire  
Département de génie de la production automatisée à l'École de technologie supérieure

M. Robert Sabourin, président du jury  
Département de génie de la production automatisée à l'École de technologie supérieure

M. Gilles Doyon, examinateur externe  
Agriculture et Agroalimentaire Canada, Centre de Recherche et de Développement sur  
les Aliments (CRDA), Saint-Hyacinthe

IL A FAIT L'OBJET D'UNE SOUTENANCE DEVANT JURY ET PUBLIC

LE 21 JUIN 2007

À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

# **PRÉDICTION DE LA TENEUR EN ANTHOCYANE DU FRUIT DE LA CANNEBERGE PAR VISION PAR ORDINATEUR**

Youssouf Chherawala

## **SOMMAIRE**

L'objectif de ce travail est d'établir une méthode non destructive, s'appuyant sur la vision par ordinateur, pour prédire le taux d'anthocyane de fruits de canneberge. La littérature montre qu'il existe une relation entre la couleur externe des fruits, leur calibre et le Tacy. La caméra numérique commerciale utilisée a permis d'acquérir quatre photos avec brassage aléatoire pour chaque échantillon de canneberges afin d'extraire leurs couleurs et leurs surfaces. La couleur canneberge dans l'espace colorimétrique ICCLAB (CIELAB encodé sur 8 bits) possède la forme d'un boomerang. L'analyse par composantes principales suivie d'une régression linéaire a permis de modéliser la couleur canneberge sous forme d'une échelle couleur. Le modèle prédictif du Tacy établi par régression linéaire multiple utilise la variable colorimétrique  $a^*$  de l'espace ICCLAB et la surface des fruits de canneberges. En considérant un terme d'erreur type identique à celui de l'analyse chimique ( $\pm 3$ ), ce modèle permet d'obtenir 100 % de prédictions réussies sur les données de la saison de récolte 2005 en enlevant les données aberrantes, et 77 % sur les données de 2006.

# **PREDICTION OF THE ANTHOCYANIN CONTENT OF THE FRUIT OF CRANBERRY BY COMPUTER VISION**

Youssef Chherawala

## **ABSTRACT**

The objective of this work is to build a nondestructive method based on computer vision to predict the anthocyanin content of cranberry fruits. The literature shows that there is a relation between the external color of the fruits, their size and Tacy. A digital camera is used to acquire four snapshots with random mixing for each cranberries sample in order to extract their colors and their surfaces. The cranberry color in the ICCLAB colorimetric space (CIELAB coded on 8 bits) has the shape of a boomerang. A principal component analysis followed-up by a linear regression allowed to build a cranberry color scale. The predictive model of Tacy was built using multiple linear regressions. It uses the colorimetric variable  $a^*$  from ICCLAB colorspace and the surface of the cranberry fruits. By considering a standard error of the mean similar to the chemical analysis ( $\pm 3$ ), this model provides 100% of successful predictions on the data from the 2005 harvest season by removing the outliers, and 77% of successful predictions on the data of 2006.

## REMERCIEMENTS

Je tiens à remercier mon directeur de maîtrise, Richard Lepage, Ph.D, pour son soutien et sa grande disponibilité tout au long de mon projet. Je tiens également à remercier Gilles Doyon, agronome, Ph.D et professeur associé à l'ÉTS, pour ses conseils précieux sur les besoins de l'industrie agro-alimentaire. Enfin, ce projet n'aurait pas pu être mené à terme sans la coopération de Mr Jean-François Sylvain, directeur technique d'Atoka Canneberges et de son équipe, qui m'ont gracieusement accueilli dans leur laboratoire lors de la phase expérimentale de mes travaux.

## TABLE DES MATIÈRES

	Page
SOMMAIRE .....	i
ABSTRACT .....	ii
REMERCIEMENTS.....	iii
TABLE DES MATIÈRES .....	iv
LISTE DES TABLEAUX.....	vii
LISTE DES FIGURES .....	xi
LISTE DES ABRÉVIATIONS ET SIGLES .....	xvi
INTRODUCTION.....	1
CHAPITRE 1 Évaluation non destructive de la canneberge .....	3
1.1 Introduction.....	3
1.2 Méthodes d'évaluation de la qualité de la canneberge .....	4
1.2.1 Analyse destructive.....	4
1.2.2 Corrélation couleur/pigment .....	7
1.3 Évaluation des fruits et légumes par vision par ordinateur.....	9
1.3.1 Dispositif d'acquisition d'image .....	9
1.3.2 Traitement de l'image.....	13
1.3.3 Méthodes d'analyse de données .....	15
1.4 Conclusion .....	17
CHAPITRE 2 La couleur.....	18
2.1 Introduction.....	18
2.2 Aspect physique .....	18
2.3 Aspect physiologique.....	19
2.4 Colorimétrie .....	21
2.4.1 Lois de Grassmann .....	21
2.4.2 Expérience d'égalisation des couleurs.....	22
2.4.3 Illuminants standards de la CIE.....	23
2.4.4 Les systèmes de primaires.....	26
2.4.5 Espace couleur uniforme.....	34
2.4.6 Distances couleur.....	39
2.5 Acquisition d'image couleur .....	41
2.5.1 Choix de l'illuminant.....	42
2.5.2 Balance des blancs.....	43
2.6 La gestion des couleurs.....	47
2.6.1 La correction gamma .....	47
2.6.2 Le gamut .....	50

2.6.3	Gestion des couleurs ICC (International Color Consortium).....	52
2.6.4	Profils ICC .....	54
2.7	Conclusion .....	58
CHAPITRE 3	Acquisition et traitement de l'image.....	59
3.1	Introduction.....	59
3.2	Dispositif d'acquisition.....	59
3.3	Traitement d'image.....	64
3.3.1	Élimination des reflets spéculaires .....	64
3.3.2	Segmentation.....	69
3.4	Analyse de la variance (ANOVA).....	77
3.5	Conclusion .....	79
CHAPITRE 4	Couleur canneberge.....	81
4.1	Introduction.....	81
4.2	Base de données .....	81
4.3	Espaces couleur .....	83
4.3.1	Analyse par composantes principales.....	86
4.4	Comparaison des espaces ICCLAB et de l'ACP.....	89
4.5	Modélisation de la couleur canneberge .....	93
4.5.1	Échelle couleur canneberge.....	104
4.6	Conclusion .....	107
CHAPITRE 5	Prédiction du Tacy .....	109
5.1	Introduction.....	109
5.2	Modélisation de la surface de la canneberge.....	109
5.2.1	Modèle sphérique (1).....	109
5.2.2	Modèle ellipsoïdal (2).....	110
5.2.3	Modèle sphéroïdal (3).....	111
5.2.4	Caractérisation de la canneberge dans les photos par ses informations morphologiques .....	111
5.3	Prédiction de l'anthocyane.....	116
5.3.1	Modèles de régression linéaire.....	116
5.3.2	Validation des modèles de régression.....	122
5.3.3	Choix du modèle prédictif.....	125
5.3.4	Discussion .....	149
5.3.5	Cas des mélanges extrêmes .....	150
5.3.6	Validation avec les données de la récolte 2006.....	156
5.3.6.1	Base de données 2006.....	156
5.3.6.2	Modèle prédictif .....	157
5.3.6.3	Validation du modèle hybride 2006 .....	160
5.4	Tacy numérique.....	166
5.5	Conclusion .....	167
CONCLUSION	.....	168

RECOMMANDATIONS.....	170
ANNEXE 1 Spécifications des illuminants fluorescents .....	171
ANNEXE 2 Mire couleur GretagMacbeth 24 cases .....	175
ANNEXE 3 Analyse de variabilité typique entre les 4 photos d'un même échantillon ou lot.....	178
ANNEXE 4 Couleur moyenne des fruits composant les échantillons standard de la saison de récolte 2005 représentée dans l'espace ICCLAB.....	188
ANNEXE 5 Photos de lots typiques de la saison de récolte 2005.....	191
ANNEXE 6 Photos de lots typiques de la saison de récolte 2006.....	205
ANNEXE 7 Prédiction du Tacy pour les échantillons de la base de données 2006..	210
BIBLIOGRAPHIE.....	215

## LISTE DES TABLEAUX

		Page
Tableau I	Illuminants standards type lumière du jour .....	24
Tableau II	Illuminants standards .....	25
Tableau III	Coordonnées xyz d'illuminants standard .....	33
Tableau IV	Coordonnées xyz des primaires ITU-R BT.709 et du D <sub>65</sub> .....	55
Tableau V	Caractérisation de l'illuminant .....	60
Tableau VI	Paramètre de réglage des caméras numériques .....	62
Tableau VII	Analyse de la variance .....	78
Tableau VIII	Classes standard de la canneberge (saison 2005) .....	82
Tableau IX	Variance expliquée par les bandes du système ICCLAB sur tout l'espace.....	84
Tableau X	Coefficient de corrélation linéaire entre les bandes du système ICCLAB sur tout l'espace .....	85
Tableau XI	Variance exprimée par les bandes du système ICCLAB pour L* < 110 .....	85
Tableau XII	Coefficient de corrélation linéaire entre les bandes du système ICCLAB pour L* < 110 .....	86
Tableau XIII	Pourcentage de la variabilité totale exprimée par chaque composante principale.....	88
Tableau XIV	Coefficients de détermination des modèles.....	97
Tableau XV	Coefficients du modèle polynomial du 7 <sup>ème</sup> ordre (avec marge de confiance à 95%).....	97
Tableau XVI	Coefficients du modèle polynomial quadratique (avec marge de confiance à 95%).....	101
Tableau XVII	Modèle et coefficients.....	119
Tableau XVIII	Modèle et coefficients.....	121

Tableau XIX	Validation des modèles polynomiaux de prédiction du Tacy (incluant 2 séries de données aberrantes, et sans les mélanges extrêmes) .....	124
Tableau XX	Modèle et coefficients .....	125
Tableau XXI	Statistique de succès de prédiction avec une erreur type de $\pm 2$ ....	130
Tableau XXII	Statistique de succès de prédiction avec une erreur type de $\pm 3$ ....	131
Tableau XXIII	Statistique de succès de prédiction avec une erreur type de $\pm 2$ ....	132
Tableau XXIV	Statistique de succès de prédiction avec une erreur type de $\pm 3$ ....	133
Tableau XXV	Statistique de succès de prédiction avec une erreur type de $\pm 2$ ....	138
Tableau XXVI	Statistique de succès de prédiction avec une erreur type de $\pm 3$ ....	139
Tableau XXVII	Statistique de succès de prédiction avec une erreur type de $\pm 2$ ....	140
Tableau XXVIII	Statistique de succès de prédiction avec une erreur type de $\pm 3$ ....	141
Tableau XXIX	Validation des modèles polynomiaux de prédiction du Tacy (sans les 2 lots de mélanges extrêmes et ni les 2 séries de données aberrantes).....	146
Tableau XXX	Prédiction du Tacy pour les échantillons de la base de test (sans données aberrantes, ni mélanges extrêmes).....	148
Tableau XXXI	Prédiction du Tacy avec modèle hybride pour les échantillons de la base de test (sans données aberrantes, avec mélanges extrêmes) .....	155
Tableau XXXII	Statistique de succès de prédiction avec une erreur type de $\pm 3$ ....	162
Tableau XXXIII	Statistique de succès de prédiction avec une erreur type de $\pm 3$ pour Tacy laboratoire $\leq 50$ .....	164
Tableau XXXIV	Statistique de succès de prédiction avec une erreur type de $\pm 3$ pour Tacy laboratoire $> 50$ .....	165
Tableau XXXV	Spécificités des illuminants fluorescents standard .....	172
Tableau XXXVI	Spécificités des illuminants fluorescents à large bande .....	172
Tableau XXXVII	Spécificités des illuminants fluorescents à bande étroite .....	173
Tableau XXXVIII	Analyse de variabilité de l'échantillon standard A2 .....	179

Tableau XXXIX	Grand résumé de l'analyse de variabilité de l'échantillon standard A2.....	180
Tableau XL	ANOVA à une entrée de l'échantillon standard A2 .....	181
Tableau XLI	Analyse de variabilité de l'échantillon standard B2 .....	182
Tableau XLII	Grand résumé de l'analyse de variabilité de l'échantillon standard B2.....	183
Tableau XLIII	ANOVA à une entrée de l'échantillon standard B2 .....	184
Tableau XLIV	Analyse de variabilité de l'échantillon standard C2 .....	185
Tableau XLV	Grand résumé de l'analyse de variabilité de l'échantillon standard C2.....	186
Tableau XLVI	ANOVA à une entrée de l'échantillon standard C2 .....	187
Tableau XLVII	Photos des trois échantillons du lot n° 446 (29-oct-05) présentant 12% de fruits pourris.....	192
Tableau XLVIII	Échantillon type n°1 de la base de test (lot n°450, 27-oct-05).....	194
Tableau XLIX	Tacy chimique et Tacy prédit (lot n°450, 27-oct-05) .....	194
Tableau L	Échantillon type n°2 de la base de test (lot n°445, 28-oct-05).....	195
Tableau LI	Tacy chimique et Tacy prédit (lot n°445, 28-oct-05) .....	195
Tableau LII	Échantillon type n°3 de la base de test (lot n°449, 29-oct-05).....	196
Tableau LIII	Tacy chimique et Tacy prédit (lot n°449, 29-oct-05) .....	196
Tableau LIV	Échantillon type n°4 de la base de test (lot n°431, 26-oct-05).....	197
Tableau LV	Tacy chimique et Tacy prédit (lot n°431, 26-oct-05) .....	197
Tableau LVI	Photos des trois échantillons du lot n° 442 (29-oct-05) considérés comme des données aberrantes .....	198
Tableau LVII	Tacy chimique et Tacy prédit des 3 échantillons du lot n°442 (29-oct-05).....	199
Tableau LVIII	Photos des trois échantillons du lot n° 443 (27-oct-05) considérés comme des données aberrantes .....	200
Tableau LIX	Tacy chimique et Tacy prédit des 3 échantillons du lot n°443 (27-oct-05).....	201

Tableau LX	Échantillon du mélange extrême n° 1 (échantillon non commercial).....	202
Tableau LXI	Tacy chimique et Tacy prédit du mélange extrême n° 1 (échantillon non commercial).....	202
Tableau LXII	Échantillon du mélange extrême n° 2 (échantillon non commercial).....	203
Tableau LXIII	Tacy chimique et Tacy prédit du mélange extrême n° 2 (échantillon non commercial).....	203
Tableau LXIV	Échantillon typique de la saison 2006 n° 1 .....	206
Tableau LXXV	Tacy chimique et Tacy prédit de l'échantillon typique de la saison 2006 n° 1 .....	206
Tableau LXXVI	Échantillon typique de la saison 2006 n° 2 .....	207
Tableau LXXVII	Tacy chimique et Tacy prédit de l'échantillon typique de la saison 2006 n° 2.....	207
Tableau LXXVIII	Échantillon typique de la saison 2006 n° 3 .....	208
Tableau LXIX	Tacy chimique et Tacy prédit de l'échantillon typique de la saison 2006 n° 3.....	208
Tableau LXX	Échantillon typique de la saison 2006 n° 4 .....	209
Tableau LXXI	Tacy chimique et Tacy prédit de l'échantillon typique de la saison 2006 n° 4.....	209
Tableau LXXII	Prédiction du Tacy pour les échantillons de la base de données 2006 .....	211

## LISTE DES FIGURES

	Page
Figure 1	Spectre de la lumière visible..... 18
Figure 2	Courbe normalisée de la sensibilité spectrale des cônes S, M et L ..... 20
Figure 3	Trichromie et antagonisme dans la rétine ..... 21
Figure 4	Expérience d'égalisation des couleurs..... 23
Figure 5	Répartition spectrale d'illuminants standards de la CIE ..... 25
Figure 6	Spectres des illuminants fluorescents ..... 26
Figure 7	Fonctions colorimétriques du système de primaires RVB par l'observateur de référence CIE 1931 à 2° ..... 27
Figure 8	Fonctions colorimétriques du système de primaires XYZ de l'observateur de référence CIE 1931 à 2° ..... 29
Figure 9	Courbe d'émission du corps noir en fonction de la température de couleur et coordonnées d'illuminants standard dans le diagramme chromatique..... 33
Figure 10	Seuils différentiels de chromaticité obtenu par Macadam (grossissement ellipses ×10)..... 35
Figure 11	Espace colorimétrique CIELAB..... 36
Figure 12	Seuils différentiels dans l'espace a*b* ..... 38
Figure 13	La mire couleur <i>ColorChecker</i> de GretagMacbeth..... 46
Figure 14	Courbe de sensibilité aux différences de luminance..... 48
Figure 15	Correction gamma..... 50
Figure 16	Comparaison de gamuts dans le plan du diagramme de chromaticité A : spectre du visible, B : caméra 3 CCD, C : imprimante..... 51
Figure 17	Comparaison des gamuts Adobe RVB et sRVB ..... 58
Figure 18	Matériels d'acquisition ..... 60

Figure 19	Photos des 4 répétitions d'un échantillon de canneberges suivant le protocole.....	63
Figure 20	Photo n° 3 de l'échantillon précédent .....	64
Figure 21	Localisation des reflets spéculaire. (a) : Bande saturation du système TSL (b) : Position des reflets spéculaire.....	66
Figure 22	Position des reflets spéculaires détectés.....	66
Figure 23	Position des reflets spéculaires après correction manuelle .....	67
Figure 24	Approximation bilinéaire .....	68
Figure 25	Image après élimination des reflets spéculaires .....	69
Figure 26	Image binaire du seuillage du fond.....	70
Figure 27	Image binaire du seuillage du fond après fermeture et ouverture ...	71
Figure 28	Zone où la méthode de la ligne de partage des eaux est inefficace (en rouge) .....	72
Figure 29	Image sans reflet spéculaire filtrée par un filtre gaussien .....	73
Figure 30	Détection de contour par l'opérateur de Sobel.....	73
Figure 31	Réunion du masque binaire et des contours .....	74
Figure 32	Masque binaire pour la méthode de partage de la ligne des eaux ...	74
Figure 33	Résultat de la segmentation des canneberges.....	75
Figure 34	Résultat de la segmentation des canneberges après correction manuelle .....	75
Figure 35	Processus de segmentation des canneberges .....	76
Figure 36	Couleur moyenne des fruits standard dans l'espace couleur ICCLAB .....	83
Figure 37	Couleur canneberge exprimée suivant les deux premiers vecteurs propres de l'ACP.....	89
Figure 38	Diagramme de double projection sans les données .....	92
Figure 39	Diagramme de double projection avec les données.....	93

Figure 40	Régression de la couleur moyenne de chaque canneberge exprimée suivant les deux premières composantes principales de l'ACP ( $R^2 = 0,95$ ).....	98
Figure 41	Intervalle de confiance à 95 % de la fonction et des observations de la régression de la couleur canneberge moyenne par fruit.....	99
Figure 42	Intervalle de confiance à 99% de la fonction et des observations de la régression de la couleur canneberge moyenne par fruit.....	100
Figure 43	Régression de la couleur canneberge moyenne par photo exprimée suivant les deux premières composantes principales de l'ACP ( $R^2 = 0,99$ ).....	101
Figure 44	Intervalle de confiance à 95% de la fonction et des observations de la régression de la couleur canneberge moyenne par photo.....	102
Figure 45	Intervalle de confiance à 99% de la fonction et des observations de la régression de la couleur canneberge moyenne par photo.....	103
Figure 46	Échelle couleur continue .....	104
Figure 47	Échelle couleur discontinue, écart couleur = 10, fond: $L^* = 100, a^* = 0, b^* = 0$ (CIELAB) .....	105
Figure 48	Échelle couleur discontinue, écart couleur = 10, fond: $L^* = 50, a^* = 0, b^* = 0$ (CIELAB) .....	106
Figure 49	Erreur entre le modèle sphérique (1) et sphéroïdal (3) en fonction de l'élongation .....	113
Figure 50	Histogramme de l'élongation des canneberges.....	114
Figure 51	Pourcentage cumulé de l'élongation des fruits .....	115
Figure 52	Tacy prédit VS Tacy laboratoire (meilleur modèle sur la totalité de l'espace) .....	120

Figure 53	Tacy prédit VS Tacy laboratoire (meilleur modèle sur l'espace restreint).....	121
Figure 54	Écart de la vraie moyenne avec son erreur type .....	127
Figure 55	Corrélation entre le Tacy du laboratoire et le Tacy prédit .....	128
Figure 56	Intervalle de confiance à 95% des prédictions du Tacy après regroupement des 4 photos.....	129
Figure 57	Modèle d'erreur type de $\pm 2$ après regroupement des 4 photos.....	130
Figure 58	Modèle d'erreur type de $\pm 3$ après regroupement des 4 photos.....	131
Figure 59	Modèle d'erreur type de $\pm 2$ avec photos séparées .....	132
Figure 60	Modèle d'erreur type de $\pm 3$ avec photos séparées .....	133
Figure 61	Écart de la vraie moyenne avec son erreur type (sans données aberrantes) .....	135
Figure 62	Corrélation entre le Tacy du laboratoire et le Tacy prédit (sans données aberrantes).....	136
Figure 63	Intervalle de confiance à 95% des prédictions du Tacy après regroupement des 4 photos (sans données aberrantes) .....	137
Figure 64	Modèle d'erreur type de $\pm 2$ après regroupement des 4 photos (sans données aberrantes).....	138
Figure 65	Modèle d'erreur type de $\pm 3$ après regroupement des 4 photos (sans données aberrantes).....	139
Figure 66	Modèle d'erreur type de $\pm 2$ avec photos séparées (sans données aberrantes).....	140
Figure 67	Modèle d'erreur type de $\pm 3$ avec photos séparées (sans données aberrantes).....	141
Figure 68	Modèle prédictif du Tacy (vue 1) .....	142
Figure 69	Modèle prédictif du Tacy (vue 2) .....	143
Figure 70	Modèle prédictif du Tacy (vue 3) .....	144
Figure 71	Modèle prédictif du Tacy (vue 4) .....	145
Figure 72	Modélisation du Tacy laboratoire sur les données de 2006 .....	158

Figure 73	Intervalle de confiance à 95% des prédictions du Tacy avec regroupement des 4 photos.....	161
Figure 74	Modèle d'erreur type de $\pm 3$ avec regroupement des 4 photos .....	162
Figure 75	Modèle d'erreur type de $\pm 3$ avec regroupement des 4 photos .....	164
Figure 76	Modèle d'erreur type de $\pm 3$ avec regroupement des 4 photos .....	165
Figure 77	Spectres des illuminants fluorescents standards.....	173
Figure 78	Spectres des illuminants fluorescents à bande large.....	174
Figure 79	Spectres des illuminants fluorescents à bande étroite.....	174
Figure 80	Modèle de la mire couleur 24 cases de GretagMacbeth .....	176
Figure 81	Valeurs des cases de la mire couleur 24 cases GretagMacbeth ....	177
Figure 82	Couleur moyenne des fruits standard de canneberge dans l'espace ICCLAB.....	189
Figure 83	Couleur moyenne des fruits standard de canneberge suivant L* et a* .....	189
Figure 84	Couleur moyenne des fruits standard de canneberge suivant L* et b* .....	190
Figure 85	Couleur moyenne des fruits standard de canneberge suivant a* et b* .....	190

## LISTE DES ABRÉVIATIONS ET SIGLES

Tacy	Anthocyanes totales ( <i>Total Anthocyanin</i> )
R <sup>2</sup>	Coefficient de détermination
RVB	Espace colorimétrique : (Rouge, Vert, Bleu)
TSL	Espace colorimétrique : (Teinte, Saturation, Luminance)
CMJ	Espace colorimétrique : (Cyan, Magenta, Jaune)
$\lambda$	Longueur d'onde (lambda)
CIE	Commission internationale de l'éclairage
IRC	Indice de rendu des couleurs ( <i>Color Rendering Index</i> )
ANOVA	Analyse de la variance ( <i>ANalysis Of VAriance</i> )
ACP	Analyse par composantes principales
SD	Écart-type ( <i>Standard Deviation</i> )
SEM	Erreur type ( <i>Standard Error of the Mean</i> )
CV %	Coefficient de variation en % (SD/moyenne×100)
NS	Non significatif ( $\alpha = 0,001$ : seuil de signification ou niveau alpha)
RMSE	Racine de l'erreur quadratique moyenne ( <i>Root Mean Square Error</i> )
N, S	Nombre d'individus (objets, répétitions) dans une population ou pour un échantillon

## INTRODUCTION

La cueillette et la transformation de la canneberge sont des industries en pleine effervescence. La canneberge est un petit fruit de couleur plus ou moins rougeâtre que l'on retrouve principalement dans des compotes d'assaisonnement de viandes, dans divers jus et dans des préparations de gâteau ou de biscuit. Le prix de revient pour le producteur dépend du taux d'anthocyane dans le fruit. L'anthocyane est une molécule que l'on retrouve à la surface de la canneberge et qui donne le goût acidulé au fruit et qui est responsable de la couleur rouge du fruit. Le taux d'anthocyane est déterminé par analyse chimique selon une procédure standardisée et reconnue par l'industrie. L'analyse d'un échantillon de lot dure entre 40 et 45 minutes en fonction du nombre de répétitions.

L'objectif de ce travail est d'étudier une méthode non destructive, s'appuyant sur la vision par ordinateur, pour prédire le taux d'anthocyane des fruits de canneberge. Cette méthode doit à long terme remplacer la méthode d'extraction chimique de l'anthocyane, car cette dernière entraîne le rejet de produits chimiques dans l'environnement et nécessite beaucoup de temps.

L'hypothèse de départ suppose que l'anthocyane se situe directement sous l'épiderme du fruit, elle est donc visible de l'extérieur. La méthode de vision par ordinateur devra acquérir les images des fruits en utilisant une caméra commerciale, puis traiter les informations colorimétriques et morphologiques des fruits du lot afin de prédire le taux d'anthocyane avec un écart inférieur ou égale à celui obtenu en laboratoire d'usine industrielle.

Dans le CHAPITRE 1, les études passées relatant la relation entre la couleur des canneberges et leur teneur en anthocyane sont exposées, suivi de la présentation de l'utilisation de la vision par ordinateur dans l'industrie des fruits et légumes. Le

CHAPITRE 2 détaille la notion de la couleur, notamment ses aspects physiques et ses aspects perceptuels, et explique les divers processus utilisés pour une acquisition et une reproduction fidèle de la couleur. Le CHAPITRE 3 montre quant à lui le matériel et la méthode utilisés pour acquérir et segmenter les images des canneberges. Dans le CHAPITRE 4, la gamme des couleurs des canneberges est étudiée afin d'être modélisée sous forme d'une échelle couleur canneberge possédant une seule dimension, qui servira de référence colorimétrique. Enfin le CHAPITRE 5 présente la construction et la validation du modèle prédictif de l'anthocyane à partir des données des saisons de récolte 2005 et 2006.

## CHAPITRE 1

### Évaluation non destructive de la canneberge

#### 1.1 Introduction

La canneberge (*Vaccinium macrocarpon*) est un fruit très prisé en Amérique du Nord. Il pousse dans de petits arbustes, et possède une couleur rouge et un goût légèrement acidulé qui expliquent son attrait majeur. Les Amérindiens utilisaient la canneberge comme nourriture, plante médicinale et lors de cérémonies. La science redécouvre aujourd'hui les nombreuses propriétés médicinales de la canneberge. Ce fruit possède la particularité d'inhiber la capacité d'adhésion de certaines bactéries aux différentes parois du corps. La recherche sur le sujet montre que la canneberge est très efficace contre les maladies ayant une origine bactérienne, comme les infections urinaires, certains ulcères gastriques et aussi la plaque dentaire et les caries. La canneberge permet également de prévenir les maladies cardio-vasculaires, en empêchant l'accumulation des lipoprotéines LDL, aussi appelées mauvais cholestérol, dans les artères. De plus, la présence importante de vitamine C et d'antioxydants, éléments bien connus pour leur action contre le vieillissement, augmente l'attrait du grand public pour la canneberge, ce dernier étant soucieux de consommer des produits naturels et bons pour sa santé. Malgré une demande croissante pour les produits de la canneberge sur les marchés, on le trouve sous forme fraîche uniquement entre septembre et décembre, car sa période de récolte se limite aux mois de septembre et octobre. Deux méthodes de récolte différentes existent, suivant l'emploi futur des canneberges. La première consiste à inonder les champs, puis à détacher les fruits de l'arbuste à l'aide d'une batteuse spécifique. Les fruits sont ensuite récupérés et placés dans un camion pour leur livraison et sont destinés à être transformés. L'autre méthode de récolte consiste à détacher les fruits de l'arbuste à l'aide d'un peigne, sans utilisation d'eau. Ces fruits-là sont destinés à la vente comme produit frais. Après la récolte, la qualité des fruits doit être évaluée.

Dans l'industrie agro-alimentaire, pendant très longtemps, le classement des fruits en différentes catégories de qualité a été réalisé par un expert humain, i.e. œil et main. Or l'expert humain est soumis à la fatigue et à des changements d'humeur, qui peuvent influencer son jugement. De plus, deux experts ne trieront pas non plus les mêmes fruits de manière exactement identique. Le point de vue de l'expert humain est ainsi subjectif. Il est donc nécessaire d'utiliser des méthodes objectives pour le tri des fruits en différentes catégories, suivant leur calibre, leur gamme de couleurs, leur maturité ou encore suivant la quantité de dommage externe. Dans le cas des canneberges, c'est la quantité de pigment renfermé par les fruits qui est utilisée pour déterminer leur qualité. En effet, plus les fruits frais contiendront d'anthocyane, plus leur couleur sera rouge et uniforme et plus leurs produits transformés tels que les cocktails ou les sauces auront une belle couleur rouge. De plus, la présence importante d'anthocyane dans le fruit est le gage que ce dernier renferme toutes ses propriétés médicinales. Ainsi, la mesure de la teneur en anthocyane des fruits est utilisée comme un des critères pour payer les producteurs.

Nous verrons dans un premier temps la mesure de la teneur en pigment réalisé par une mesure chimique, puis des expériences sur la mesure de la teneur en pigment par une mesure de la couleur externe des fruits. Enfin, nous verrons les travaux utilisant la vision par ordinateur dans l'industrie agro-alimentaire, ainsi que les principes de cette première.

## **1.2 Méthodes d'évaluation de la qualité de la canneberge**

### **1.2.1 Analyse destructive**

La couleur rouge caractéristique des canneberges provient de la présence à la surface du fruit d'un pigment dénommé anthocyane. Un fruit contenant peu d'anthocyane possède

une couleur rosée à rouge pâle, tandis qu'un fruit en contenant beaucoup possède une couleur rouge foncé. La quantité de pigment renfermée par les canneberges dépend de divers facteurs, tels la variété des fruits, les conditions de croissance (ensoleillement), l'état physiologique de la plante et des fruits, la taille et la position des fruits sur la plante, ainsi que de l'application de produits chimiques [1,2]. La canneberge renferme un nombre élevé de types d'anthocyane, mais les 4 principaux pigments donnant la couleur rouge au fruit sont le galactoside de cyanidine, le galactoside de péonidine, l'arabinoside de cyanidine et l'arabinoside de péonidine [3,4]. La méthode employée dans l'industrie de la canneberge pour évaluer la teneur en anthocyane est par analyse chimique. Le principe est d'extraire le maximum de pigments des fruits sans les détruire ou les altérer. Pour cela, les pigments sont dans un premier temps extraits dans une solution chimique (HCl ou HCl / E-OH), puis la densité optique de cette solution est mesurée par un spectrophotomètre à la longueur d'onde du pic d'absorbance maximal de l'anthocyane. Enfin, la teneur totale en anthocyane (Tacy) en mg/100g de fruits est estimée linéairement à partir de la densité optique [5]. La relation entre la teneur en anthocyane et l'absorbance n'est linéaire que pour de faibles concentrations, il est donc nécessaire de diluer grandement la solution. De plus, le pH de la solution ainsi que la nature du solvant influencent la quantité extraite de pigment, il est ainsi préférable d'opérer à un pH faible et d'utiliser un alcool comme solvant pour une meilleure stabilité et extraction [2].

Durant la période de récolte en automne, le processus est le suivant : les fruits récoltés par flottaison aux champs sont transportés par camion jusqu'à l'usine de transformation. Ils sont alors déversés dans des bacs de tri, pour être par la suite lavés et débarrassés des saletés. Les canneberges sont alors échantillonnées au hasard. Une sélection de 1 Kg est mesurée pour les diverses analyses. Jusqu'à trois sous-lots de 100 g de canneberges sont préparés et mélangés avec une solution d'acide chlorhydrique HCl (0,2N) pendant trois minutes pour extraire le Tacy des fruits. 10 ml du mélange sont pipetés puis filtrés sur un entonnoir Buchner et du papier Whatman. On y rajoute 50 ml du HCl (0,2N). Le zéro est

ajusté sur le spectrophotomètre (Barnstead Thermolyne) avec une solution de HCl normalisée. L'absorbance est mesurée à 515 nm (pic d'absorbance maximum des solutions aqueuses, qui correspond à la zone de couleur verte du spectre électromagnétique du visible) et servira à estimer le Tacy en mg/100 g de fruits à l'aide d'une table de conversion [6] en industrie.

L'analyse chimique permet d'obtenir la teneur en anthocyane d'un lot de fruits de la manière la plus précise. Néanmoins, elle possède de nombreux inconvénients. Tout d'abord, cette méthode est relativement longue, elle dure en moyenne 40 à 45 min pour l'analyse d'un seul échantillon de 100 gr. Durant la période de récolte, une quantité considérable de lots arrive afin d'être évaluée. Cela ajoute une somme de travail importante pour le personnel du laboratoire à cause de la lenteur de la méthode. De plus, l'analyse chimique est un procédé onéreux et qui entraîne le rejet de produits polluants (acides) dans l'environnement.

L'analyse chimique est une méthode destructive, les canneberges servant à l'estimation du Tacy sont perdues pour la vente. Ainsi, seule une partie très limitée ou minimale d'un lot de canneberges est testée. Or la partie testée n'est pas systématiquement représentative du lot au complet, il est donc nécessaire de réaliser plusieurs tests afin de restreindre l'incertitude du résultat obtenu. Cela engendre inévitablement davantage de perte de matériel. La nature destructive de la méthode oblige la réalisation d'un compromis entre la quantité de matériel sacrifiée lors du test et la justesse de l'estimation du Tacy ou sa variabilité.

L'utilisation d'une méthode non destructive permettrait de résoudre ce conflit en utilisant les informations visuelles externes des canneberges.

### 1.2.2 Corrélation couleur/pigment

L'anthocyane étant le pigment responsable de la couleur des canneberges, il est donc fort probable qu'une corrélation existe entre la teneur en pigment du fruit et sa couleur. Cette hypothèse a d'abord été vérifiée sur les jus de canneberge [7]. La couleur de différents jus a été mesurée distinctement par les trois modèles de colorimètres suivants : le Hunterlab D25, le Gardner XL-10 et le GERS (*General Electric Recording Spectrophotometer*), dans les systèmes colorimétriques XYZ, CIELAB et TSL. Le coefficient de détermination ( $R^2$ ) exprime la qualité de l'estimation d'une régression entre deux variables, sa valeur est comprise entre 0 et 1, et plus sa valeur est élevée, plus la variable indépendante explique la variable dépendante. Les résultats montrent que le coefficient de détermination ( $R^2$ ) entre le Tacy et les couleurs mesurées des jus dans les différents systèmes colorimétriques est très élevé ( $R^2 > 0,90$ ), sauf pour la bande  $b^*$  du système CIELAB et la teinte du système TSL.

La mesure de la couleur directement à la surface de fruits frais est également possible, puisque l'anthocyane est située directement sous l'épiderme de la canneberge [8], même si la mesure s'avère plus délicate que dans le cas des jus. La couleur à la surface des fruits a été mesurée à l'aide d'un colorimètre Hunterlab par Francis en 1957 [9]. Les expériences furent réalisées pour trois variétés de canneberges, avec trois calibres différents (petit, moyen, gros) et trois couleurs (rouge pâle, rouge, rouge foncé). Les canneberges des lots expérimentaux sont au préalable triées par des experts humains. Les canneberges de différents calibres et variétés furent parfois mélangées entre elles dépendamment de l'expérience, mais chaque lot expérimental fut composé de canneberges possédant une même couleur. Les lots expérimentaux sont ensuite placés dans le récipient circulaire rotatif du colorimètre pour la mesure de la couleur externe, la rotation du récipient permet virtuellement de fusionner les couleurs individuelles des canneberges, et ainsi d'obtenir une mesure plus précise et représentative de l'ensemble du lot. Les résultats ont montré qu'il existe une forte corrélation entre la teneur en

anthocyane et les bandes  $L^*$  et  $a^*$  du système CIELAB ( $R^2 > 0,90$ ), et peu de corrélation avec la teinte (approximée par  $\arctan(a^*/b^*)$ ) et la bande  $b^*$  du CIELAB. Les résultats montrent également l'influence de la variété des canneberges et de leur taille sur la qualité de la régression. D'une manière générale, pour un poids de lot égal, plus le calibre des canneberges augmente, plus la teneur en anthocyane diminue. Des résultats très intéressants ( $R^2 > 0,95$ ) sont obtenus par régression multiple suivant la couleur et le calibre des fruits, même en combinant les variétés.

Une autre expérience montre l'importance de l'homogénéité des échantillons pour l'évaluation du Tacy à partir de la couleur externe des canneberges [10]. La couleur externe des canneberges fraîches fut mesurée à l'aide d'un spectrophotomètre de la compagnie Agrtron, spécialement conçu pour les canneberges. Cette dernière mesure la réflexion du rouge à 640 nm et celle du vert à 546 nm et donne une indication sur la couleur du fruit sous la forme du ratio vert/rouge. Les canneberges utilisées pour l'expérience provenaient de lots réels, contenant une large gamme de couleurs et de canneberges hétérogènes. Les résultats ont montré que dans le cas de lots homogènes, il existait une corrélation entre la couleur moyenne et le Tacy. Par contre, dans le cas des lots hétérogènes, il n'existait pas de corrélation entre la couleur moyenne et le Tacy. En effet, considérons deux lots hétérogènes ayant la même couleur moyenne, le premier composé de 70% de fruits rouge foncé et de 30% de fruits jaunes et le second composé de 80% de fruits rouges et de 20% de fruits jaunes. Même si les deux lots précédents possèdent la même couleur moyenne, leurs Tacy sont très différents, le premier possédant un Tacy de 0,81 exprimé en mg de Congo Red (CR) par gr, et le second de 0,23 mg CR/gr. Ainsi, il existe une relation univoque entre la couleur moyenne d'un lot et son Tacy uniquement si l'on considère des lots de couleur homogène.

Les résultats précédents montrent qu'il existe une forte corrélation entre la teneur en anthocyane et la couleur des canneberges, notamment la couleur à la surface. Néanmoins, les résultats des 2 dernières expériences ne sont pas suffisants pour être

applicables directement dans l'industrie. En pratique, les transformateurs reçoivent en provenance des producteurs des lots hétérogènes. En effet, ces lots contiennent des fruits de tous les calibres et de toutes les couleurs. De plus, les transformateurs doivent traiter un grand nombre de variétés différentes de canneberges. Les canneberges fraîches ne possédant pas en général une coloration uniforme, le problème de l'acquisition de la couleur globale d'un fruit se pose alors. Il est donc nécessaire de pouvoir évaluer avec exactitude la teneur en anthocyane de lots réels, et non uniquement de lots triés de manière homogène. Les colorimètres sont des appareils de mesure de la couleur de haute précision. Mais l'acquisition d'un colorimètre représente un investissement important (10.000 à 20.000 \$ CAN), qui rebute les transformateurs et qui ne résout pas le problème des lots hétérogènes.

Il apparaît donc qu'un protocole plus complet pour l'estimation de la teneur en anthocyane à partir de la couleur externe des fruits est nécessaire, et qu'il doit utiliser un dispositif d'acquisition de l'information couleur moins onéreux et simple d'utilisation, sans altérer les fruits.

### **1.3 Évaluation des fruits et légumes par vision par ordinateur**

#### **1.3.1 Dispositif d'acquisition d'image**

Dans le cadre de l'évaluation non destructive de fruits, les dispositifs d'acquisition d'images sont les plus appropriés [11]. L'acquisition de l'information se fait grâce à un capteur positionné à distance, assurant ainsi l'absence de contact physique avec la denrée alimentaire. Cela garantit la non-altération des aliments à la suite du test. Plusieurs types de procédés peuvent être employés dans le but d'acquérir une image :

- a. radiographie;
- b. imagerie par résonance magnétique (IRM);
- c. caméra numérique *charge-coupled device* (CCD);

d. caméra multispectrale.

La radiographie est une technique permettant de réaliser des clichés, par émission de rayons X, des structures internes d'un organisme vivant, traduisant l'opacité plus ou moins marquée des tissus ou organes par un niveau de gris plus ou moins élevé. La radiographie sert à évaluer la qualité interne des tissus des fruits, ou bien à détecter la présence de parasites au niveau des noyaux des fruits [12] par l'étude de l'opacité des zones du cliché. La tomographie au rayon X, qui permet de visualiser l'organe en 3D en superposant les clichés, est elle aussi utilisée [13].

L'IRM utilise le phénomène de résonance pour l'acquisition de l'image. Un champ magnétique est émis afin d'exciter les protons présents dans les molécules d'eau de tout organisme vivant. Une fois que les protons entrent en résonance, l'excitation est interrompue, et des antennes enregistrent l'énergie restituée par les protons. Les données recueillies sont ensuite traitées par ordinateur afin de former une image. Cette technique est utilisée pour détecter les tissus malades des fruits, à un stade où la maladie n'est pas encore visible sur la surface externe [14].

L'IRM et la radiographie permettent ainsi d'obtenir une image de l'intérieur d'un fruit ou légume. L'IRM possède l'avantage d'être non invasive, contrairement à la radiographie qui peut se révéler néfaste à haute dose. Néanmoins, ces deux techniques reposent sur des propriétés physiques différentes, l'utilisation de l'IRM ou de la radiographie dépend donc de la nature physique des symptômes à évaluer [15,16]. Les appareils servant aux mesures sont ceux utilisés dans le milieu médical. Ces méthodes sont employées notamment sur les pommes, les poires et les pêches.

Les caméras numériques CCD permettent d'obtenir des images couleur de la scène observée. Le principe d'acquisition de la couleur repose sur l'aspect trichromatique de la couleur, dérivé des lois de Grassmann [17] et de l'étude du système visuel humain. Il

existe deux types de caméra CCD permettant d'obtenir des images couleur en 2D. Les caméras matricielles mono CCD fournissent une acquisition trichromatique grossière en utilisant un unique capteur matriciel composé de photodiodes (éléments sensibles à la lumière). Chacune de ces dernières est recouverte par un filtre de couleur rouge, vert ou bleu. Les motifs de la mosaïque de répartition des filtres sont nombreux (filtre colonne, filtre de Bayer, filtre de Rockwell...) et varient selon le constructeur. De manière générale, il y a 2 fois plus de filtres verts que de filtres rouges et bleus, car l'œil humain présente un maximum de sensibilité aux longueurs d'ondes du vert. La couleur de chaque pixel de l'image est alors calculée à partir d'un quadruplet de photodiodes par une interpolation logicielle. Une telle opération n'est pas exempte de défaut, et présente parfois des aberrations chromatiques et engendre une perte de résolution. Les caméras matricielles triCCD possèdent quant à elles 3 capteurs CCD, un pour chacune des composantes RVB. La lumière est décomposée par des prismes ou des miroirs semi-réfléchissants, puis projetée sur chacune des 3 matrices distinctes. La qualité de résolution et de la sensibilité de la caméra dépendent surtout de celle du bloc optique. Les caméras CCD couleurs sont utilisées dans diverses applications de vision artificielle pour les fruits et légumes, telle la classification des olives [18] et des pommes [19] en fonction des défauts externes, ou encore la localisation de pommes pour guider un robot lors de la cueillette [20].

La caméra multispectrale permet d'obtenir des images de la scène autour de certaines longueurs d'onde spécifiques du spectre de la lumière. Chaque canal de l'image est formé de l'image de la scène à une longueur d'onde spécifique. Il y a autant de canaux que de longueurs d'onde sélectionnées. Les caméras multispectrales sont formées d'une caméra mono CCD à niveaux de gris pour acquérir les images, ainsi que d'un filtre à bande passante étroite placé devant l'objectif. Il est alors nécessaire de prendre autant de clichés qu'il y a de canaux. Un autre dispositif consiste à placer un spectrographe devant la caméra. Le spectrographe acquiert une ligne spatiale de la scène (composante x de l'image), dont il décompose le spectre suivant la composante y de l'image. L'image

ainsi décomposée est acquise par la caméra mono CCD. La caméra multispectrale est souvent utilisée aux longueurs d'onde de l'infrarouge. La caméra multispectrale est employée à la segmentation des fruits ou légumes du fond de l'image [21] en capturant l'image dans l'infrarouge. Néanmoins, les images multispectrales sont aussi utilisées pour mesurer la fermeté des fruits, leur degré de maturités, ainsi que la mesure du degré Brix (pourcentage de sucre contenu dans un jus) [22,23]. Les longueurs d'onde des filtres spectraux sont choisies soit en fonction de la longueur d'onde d'absorption des pigments caractéristiques des fruits, soit dans le domaine proche de l'infrarouge. En effet, les substances organiques telles que le sucre présentent des pics d'absorbance dans les domaines proches de l'infrarouge.

L'utilisation de caméras CCD permet d'obtenir une image de la surface externe des fruits et légumes, néanmoins une seule image n'est en général pas suffisante pour évaluer toutes leurs surfaces. En effets, la caméra ne peut capturer que l'information de la surface qui lui est visible, or la face cachée peut être relativement différente de la surface visible. Il est alors usuel d'utiliser soit plusieurs caméras [24], soit une unique caméra avec rotation des fruits et légumes afin d'obtenir les images de toute la surface [25,26].

L'illuminant utilisé pour éclairer la scène joue lui aussi un rôle fondamental dans le rendu de l'image. Il existe plusieurs types d'illuminant qui diffèrent en fonction de leur spectre d'émission et sont caractérisés par leur température couleur (Kelvin). Dans le contexte de la reproduction couleur pour les produits agro-alimentaire, l'usage de lampes spéciales reproduisant la lumière du jour ( $D_{65}$ ) est de plus en plus commun [27,28]. Selon le réglage de l'illuminant, l'aspect des caractéristiques d'intérêt de l'image peut changer radicalement, apparaissant soit d'une façon floue, soit distinctement [29]. Un dispositif adéquat d'éclairage permet également de réduire des défauts dans l'image comme la présence d'ombrage ou bien encore les réflexions spéculaires en fonction de la disposition des sources lumineuses [30]. Il est ainsi

recommandé de placer l'illuminant à un angle de  $45^\circ$  avec l'axe de l'objectif de la caméra, afin d'éviter la formation de reflets spéculaires dans l'image [31]. L'ensemble du dispositif d'acquisition, d'éclairage ainsi que les fruits et légumes sont isolés en étant placés dans un environnement cloisonné [25,27].

L'image ainsi acquise est ensuite transférée vers un ordinateur afin d'être traitée. Le transfert se fait soit par l'utilisation d'une carte d'acquisition dans le cas d'une caméra vidéo, soit directement dans le cas d'une caméra numérique.

### **1.3.2 Traitement de l'image**

L'image acquise par l'ordinateur est composée de pixels et se présente sous forme numérique. Elle est semblable à une matrice à deux dimensions, et il est alors possible de lui appliquer divers opérateurs mathématiques. La première phase de travail sur l'image est en fait une étape de prétraitement. Elle sert à corriger les défauts du système d'acquisition, comme les déformations géométriques, une mise au point incorrecte, la présence de bruit dans l'image ou bien encore une illumination inégale de la scène. Cette étape a pour but de faciliter la segmentation de l'image. La segmentation consiste à séparer le ou les objets sur lesquels sera portée l'analyse, du fond de l'image. Trois techniques sont généralement utilisées pour la segmentation : le seuillage, la détection de contour et la recherche de régions homogènes.

La technique de seuillage s'appuie sur une étude de l'histogramme de l'image. Ce dernier est un graphe qui décrit pour chaque valeur de pixel comprise dans la plage de codage de l'image, le nombre de pixels de l'image possédant cette valeur. Il existe un histogramme pour chacune des bandes de l'image. Les pics de l'histogramme représentent un ensemble de pixels possédant des valeurs similaires, ils sont le plus souvent caractéristiques d'un objet ou d'une surface homogène d'un objet. La

segmentation s'effectue donc en ne conservant que les pixels appartenant à un pic caractéristique de l'objet, en attribuant des seuils de séparations au niveau des vallées autour du pic. Les seuils peuvent être soit fixés manuellement par observation de l'histogramme, soit automatiquement par des méthodes statistiques, ou bien par analyse de la pente pour détecter les vallées.

La technique de détection de contours se définit comme la recherche de discontinuités locales entre les différentes zones de couleur, d'éclairement ou de texture de l'image. Les contours peuvent être trouvés par des approches classiques fondées sur des techniques différentielles pour détecter les variations brusques du signal. Ces détecteurs reposent sur la recherche d'un extremum de la dérivée première, ou d'un passage par zéros de la dérivée seconde. Ils emploient des filtres passe-haut pour détecter les hautes fréquences de l'image, c'est-à-dire les variations brusques, caractéristiques des contours. Les contours sont repérés par des masques de convolution de dimension  $2 \times 2$  ou  $3 \times 3$ . Les détecteurs de gradient les plus couramment utilisés sont les filtres de Sobel, de Roberts et de Prewitt. Le masque laplacien est quant à lui utilisé pour le calcul de la dérivée seconde.

Les méthodes analytiques constituent une nouvelle approche de la détection de contours. Elles sont fondées sur certains critères tentant de définir une bonne détection de contours. Trois critères ont été définis par Canny en 1986 [32]:

- a. garantir une bonne détection, c'est-à-dire une réponse forte même à de faibles contours;
- b. garantir une bonne localisation;
- c. assurer que pour un contour, il n'y aura qu'une seule détection.

Les filtres analytiques usuels sont les filtres de Canny et de Canny-Deriche.

Enfin, la segmentation d'une image par approche structurelle consiste à rechercher des zones homogènes. Chaque région est définie comme un ensemble de pixels connexes, possédant des attributs similaires (couleur, luminosité, texture...). Le processus de

formation des régions considère à la fois l'homogénéité des attributs des pixels appartenant à la même région, et la différence des attributs avec les pixels des régions connexes. Dans la technique de croissance de régions [33], les pixels appartenant de manière certaine à une région donnée sont appelés les germes, et serviront de régions initiales. Les pixels de l'image connexe à la région initiale sont ensuite associés à la région s'ils vérifient les critères d'homogénéité, entraînant la croissance de la région. Le procédé est itératif, et continue jusqu'à ce qu'aucun pixel connexe à la région ne vérifie les critères d'homogénéité. Une autre technique, celle de partage de régions [34], considère l'image entière comme région initiale. Les critères d'homogénéité sont testés sur la région, si la région les satisfait tous, alors elle est considérée comme une partition de l'image. Si ce n'est pas le cas, alors cette région est subdivisée en deux sous régions régulières. Chacune des sous régions subit le même traitement, jusqu'à ce que toutes les régions de l'image vérifient tous les critères d'homogénéité. Néanmoins, le procédé séquentiel engendré par les méthodes structurelles nécessite des temps de calcul excessifs. De plus, les critères d'homogénéité dépendent de la nature de l'image à traiter, il est donc difficile d'adapter ces algorithmes à des cas concrets.

Une fois l'image segmentée, une analyse de la couleur, de la texture, de la forme et de la dimension des objets est alors possible.

### **1.3.3 Méthodes d'analyse de données**

Les méthodes d'analyse de données sont employées pour simuler de manière artificielle la réflexion humaine, afin d'aider cette dernière à réaliser des jugements complexes de manière précise. L'entraînement de tels systèmes se fait généralement à partir d'exemples. Les techniques d'apprentissage permettent de trouver automatiquement des relations non triviales entre les caractéristiques des données à étudier. Dans le cas de l'industrie agro-alimentaire, les caractéristiques prises en compte sont souvent la

couleur, la forme, la taille et la texture. Elles sont ensuite reliées par apprentissage à d'autres critères, tels que la catégorie d'un fruit dans le cas d'un problème de classification [35] ou à une propriété chimique dans le cas d'un problème de prédiction [36]. Les deux principaux domaines dans lesquels les algorithmes d'apprentissage sont employés sont les réseaux de neurones artificiels et l'apprentissage statistique.

Les réseaux de neurones artificiels sont inspirés du système nerveux biologique. Leur faculté d'apprendre à partir d'exemples suscite un grand intérêt. Ils sont composés de neurones artificiels, disposés en une seule ou plusieurs couches. Le réseau représente ainsi une structure complexe d'interdépendance entre les différents neurones. Les réseaux de neurones sont capables de réaliser des tâches d'inspection à un niveau de performance proche de celui d'un expert humain. Le type de problème, classification ou prédiction, influence le type de réseau de neurones qui sera utilisé. Les réseaux de neurones les plus utilisés sont les perceptrons multicouches et les réseaux RBF (*Radial Basis Function*), mais la supériorité des réseaux SVM (*Support Vector Machine*) dans de nombreuses applications de classification les rend de plus en plus populaires [37]. Les réseaux de neurones sont utilisés dans des problèmes de classification de fruits et légumes pour leur simplicité et leur efficacité dans un très grand nombre d'applications [22,36,38-41].

L'apprentissage statistique utilise les propriétés statistiques des données d'entraînement. Il repose sur un modèle probabiliste, telle la théorie Bayésienne qui possède un cadre mathématique approfondi. Le classificateur de Bayes est souvent utilisé pour identifier les pixels appartenant à des zones endommagées des fruits et légumes [19,26]. L'analyse discriminante est une autre branche de l'apprentissage statistique. Elle crée à partir de combinaisons des variables existantes, une nouvelle variable qui maximise la différence entre les différentes classes dans un problème de classification, ou qui réduit l'intervalle de confiance dans un problème de prédiction.

## 1.4 Conclusion

L'évaluation de la teneur en anthocyane est un problème fondamental pour l'industrie de la canneberge, car elle détermine le prix de revient des lots. Les études faites sur la relation entre la teneur en anthocyane et les espaces colorimétriques indiquent que les bandes  $L^*$  et  $a^*$  de l'espace CIELAB sont fortement corrélées avec la teneur en anthocyane. Il est ainsi possible de prolonger ces études par une analyse de la couleur externe des canneberges dans l'espace CIELAB, à l'aide d'un système d'acquisition de vision par ordinateur. En effet, l'acquisition des données se fait par des capteurs optiques, il n'y a ainsi aucun contact avec les aliments, ce qui rend les systèmes de vision par ordinateur très attrayants pour l'industrie agroalimentaire. De plus, la chute de prix constante du matériel électronique, tant au niveau des appareils photographiques numériques que des ordinateurs, permet une démocratisation de la vision par ordinateur, la rendant accessible même par des usines de taille modeste. Ce type de système a déjà fait ses preuves dans l'évaluation de nombreux fruits et légumes. Dans le cas de l'évaluation de la canneberge, une bonne connaissance de la notion de couleur est nécessaire afin de préparer le protocole expérimental.

## CHAPITRE 2

### La couleur

#### 2.1 Introduction

La couleur suscite depuis l'antiquité la curiosité, non seulement des scientifiques, mais également des artistes. Nous verrons tout d'abord l'aspect physique de la perception de la couleur, puis l'aspect physiologique de cette dernière. Nous étudierons enfin la colorimétrie, qui est la science de la mesure et de la reproduction de la couleur.

#### 2.2 Aspect physique

La couleur d'un objet dépend de l'interaction entre la source lumineuse et la surface de l'objet. Newton montra, dès 1666, que la lumière blanche du soleil pouvait se décomposer grâce à un prisme en une multitude de rayonnements de couleurs. Elle se propage sous forme d'onde magnétique, dont le spectre visible s'étend de 380 nm à 780 nm (Figure 1). En fonction de la valeur de la longueur d'onde, la lumière nous apparaît sous différentes couleurs. La lumière peut être considérée comme un flot de particules énergétiques, dénuées de masse, défini par sa répartition spectrale d'énergie.



Figure 1 Spectre de la lumière visible

*(Tiré de [42])*

Lorsque la lumière rencontre un objet, elle est soit réfléchi, soit transmise, soit absorbée. Un rayon lumineux peut donc être réfléchi de manière spéculaire ou diffuse

par l'objet, ou être modifié par son passage au travers de l'objet. Dans le cas de l'absorption, aucun rayon lumineux n'est retransmis, il n'est donc pas possible d'évaluer la couleur. La couleur d'un objet est donc définie soit par son spectre de réflectance, soit par son spectre de transmittance.

Le signal spectral  $S(\lambda)$ , appelé stimulus, résultant de l'interaction d'une source lumineuse de spectre  $I(\lambda)$  avec un objet de réflectance  $R(\lambda)$  est défini de la manière suivante :

$$S(\lambda) = I(\lambda) \times R(\lambda) \quad (2.1)$$

La couleur perçue dépend ainsi de la source lumineuse et de l'objet donné, mais également du système visuel de l'observateur de la scène.

### 2.3 Aspect physiologique

Le système visuel humain, composé spécifiquement des yeux, du nerf optique et du cerveau, nous permet de distinguer les formes, et notamment les couleurs. L'œil humain comporte des photorécepteurs au niveau de la rétine. Il est possible de distinguer les bâtonnets, qui sont au nombre de 100 millions, et les cônes, qui sont eux au nombre de 4 à 7 millions. Les bâtonnets sont plus petits que les cônes, mais 10 fois plus sensibles. Ils sont situés au contour de la fovéa et sont associés à la perception de l'intensité de la lumière. Le pigment visuel des bâtonnets est la rhodopsine, qui absorbe la lumière de tout le spectre visible, et qui possède un maximum de sensibilité à 500 nm (lumière verte). Les bâtonnets sont activés uniquement en vision nocturne, car la rhodopsine ne se forme que dans l'obscurité.

Les cônes sont associés à la perception de la couleur et sont activés en vision diurne. Ils sont situés dans la fovéa. Trois types de cônes existent, dénommés L, S et M. Les cônes S ont une sensibilité maximale pour une longueur d'onde de 440 nm (couleur bleue), les cônes M à 545 nm (couleur verte) et les cônes L à 580 nm (couleur rouge). Les noms des

cônes L, M et S font référence aux gammes de longueurs d'onde auxquels ils sont sensibles (respectivement *Long*, *Medium* et *Short*). La présence de ces trois cônes est à l'origine de la notion de trivariance de la perception humaine, qui stipule que trois variables indépendantes sont suffisantes pour décrire les couleurs.

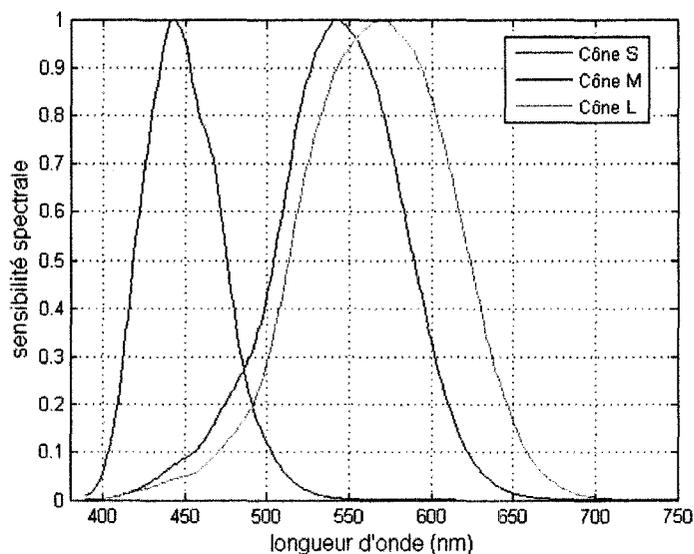


Figure 2 Courbe normalisée de la sensibilité spectrale des cônes S, M et L <sup>1</sup>

L'information visuelle issue des cônes et des bâtonnets n'est pas directement acheminée au cerveau par l'intermédiaire du nerf optique, elle fait au préalable l'objet d'un traitement rétinien. Les cellules ganglionnaires de la rétine reçoivent les influx des cônes et des bâtonnets dans une zone dénommée champ récepteur. L'information transmise au cerveau est en fait le contraste de lumière entre le centre et le pourtour d'un champ récepteur. L'activité des trois types de cônes, S, M et L, étant restreinte à trois régions distinctes du spectre, le bleu, le vert et le rouge, le contraste mesuré s'effectue sur des comparaisons chromatiques entre les différents types de cônes. Il existe trois types d'opposition [44]:

- a. Antagonisme vert-rouge (L – M);

<sup>1</sup> Données mise à disposition par [43]

- b. Antagonisme jaune-bleu ( $L + M - S$ );
- c. Antagonisme achromatique noir-blanc ( $L + M + S$ ).

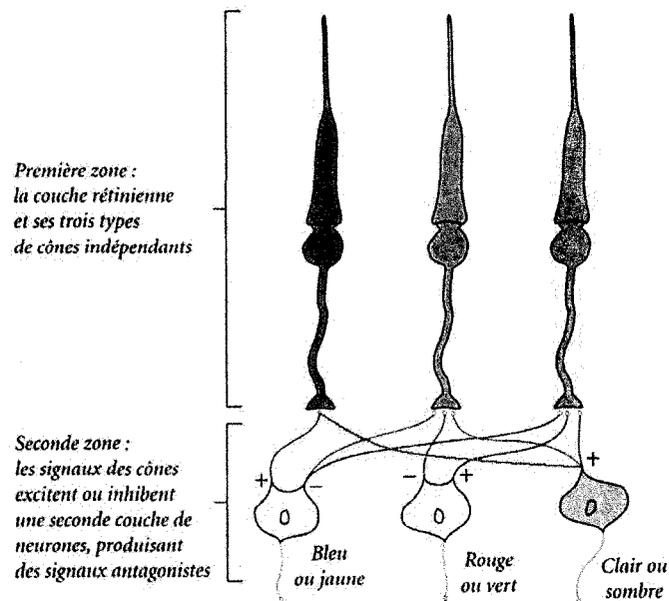


Figure 3 Trichromie et antagonisme dans la rétine  
(Tiré de [45])

## 2.4 Colorimétrie

### 2.4.1 Lois de Grassmann

Les lois de Grassmann sont à l'origine de la colorimétrie trichromatique. Elles ont permis d'établir le comportement additif et linéaire de la sensation de la couleur, indépendamment du spectre de la couleur. Elles sont établies à partir de 5 hypothèses. Soit A, B et C trois stimuli de couleur, qui vérifient les propriétés suivantes<sup>1</sup>:

- a. Réflexivité  $A \equiv A$  ;
- b. Symétrie si  $A \equiv B$  alors  $B \equiv A$  ;

<sup>1</sup>  $A \equiv B$  Signifie A est indifférentiable de B.

- c. Transitivité si  $A \equiv B$  et  $B \equiv C$  alors  $A \equiv C$  ;
- d. Additivité si  $A \equiv B$  alors  $A + C \equiv B + C$  ;
- e. Dilatation  $A \equiv B$   $k.A \equiv k.B$  avec  $k > 0$  .

Ces hypothèses ont permis à Grassmann d'établir trois lois de manière expérimentale. Premièrement, pour égaliser visuellement une couleur, trois sources lumineuses de longueurs d'onde distinctes sont nécessaires et suffisantes. Ensuite, il remarque que dans un mélange additif de lumières colorées, seules les couleurs perçues sont importantes, et non leur répartition spectrale. De plus, si dans un mélange additif de lumières colorées, une ou plusieurs lumières sont graduellement modifiées, la couleur résultante sera elle aussi graduellement modifiée.

#### 2.4.2 Expérience d'égalisation des couleurs

Les expériences d'égalisation s'appuient sur les lois de Grassmann et la théorie de la trivariance visuelle. L'objectif est d'associer à chaque stimulus visuel élémentaire, une mesure de ce stimulus couleur. L'expérience d'égalisation doit être menée selon des conditions précises. Elle consiste à faire comparer par un observateur dit standard un stimulus visuel à un stimulus de référence. Les stimuli sont projetés dans un champ visuel uniforme, entouré d'un fond uniforme. Les expériences menées indépendamment par Guild [46] et Wright [47] utilisaient trois couleurs primaires rouge, vert et bleu, qui servaient à générer les stimuli visuels. Des primaires quasi monochromatiques aux longueurs d'onde dominante de 630 nm, 543 nm et 460 nm ont été utilisées par Guild, et générées en filtrant une source à incandescence. Dans l'expérience de Wright, les primaires furent générées par un monochromateur aux longueurs d'onde respectives de 650 nm, 530 nm et 460 nm. Le champ visuel de l'observateur était de forme carrée, sous-tendant un angle de  $2^\circ$ , vu sur un fond noir. Durant ces expériences, l'ensemble des longueurs d'onde monochromatique du spectre du visible (380 à 780 nm) a été égalisé par une synthèse additive des trois primaires. Ces primaires satisfont donc aux lois de Grassmann [48], et ces expériences ont permis d'établir le système RVB CIE 1931.

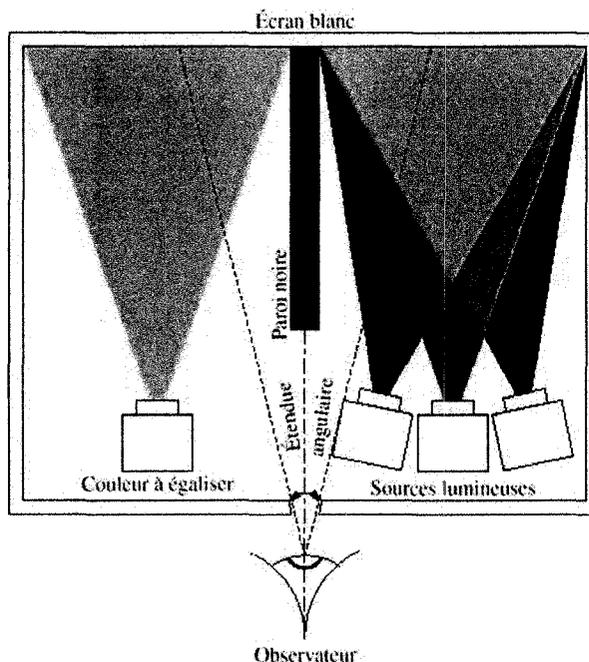


Figure 4 Expérience d'égalisation des couleurs

(Tiré de [49])

### 2.4.3 Illuminants standards de la CIE

La plus importante des spécifications de la couleur est celle développée par la Commission Internationale de l'Éclairage (CIE). Elle fournit une méthode standard pour décrire le stimulus d'une couleur, dans des conditions contrôlées d'éclairage et d'observation, établies à partir des connaissances des réactions de l'observateur moyen. La forte corrélation entre la couleur observée d'un objet et la source lumineuse qui l'éclaire oblige une catégorisation des différents illuminants en fonction de leur température couleur. Cette dernière tient ses origines dans les découvertes de Planck au sujet des corps noirs. Le corps noir est un concept théorique. Il absorbe toutes les radiations incidentes et émet un spectre continu caractéristique en fonction de sa température. Ainsi, la couleur du corps noir va du rouge (température basse) vers le bleu-blanc (température élevée). Une échelle de température de couleur a été créée à

partir de la couleur du corps noir sous différentes températures. Cette échelle s'exprime en Kelvin. Une comparaison entre l'échelle de température couleur et la couleur des sources lumineuses permet de caractériser chaque illuminant. La lumière du jour possède un spectre continu et est souvent prise comme référence. Plusieurs familles d'illuminant existent. Tout d'abord les illuminants de type D (pour *Daylight*) correspondent aux différents types de lumières naturelles du jour. Les illuminants de type tungstène sont normalisés suivant les illuminants A, B ou C. Les illuminants de type F (pour fluorescent) sont divisés en 12 sous-familles en fonction du type de phosphore utilisé. Ils sont employés pour simuler les lumières de type D. Il existe également l'illuminant E, qui est un illuminant théorique équiénergétique. Il convient de remarquer que parmi les sources artificielles, certaines possèdent un spectre continu, d'autres un spectre discontinu et enfin celles possédant un spectre de raies. Pour obtenir un bon rendu des couleurs sur toute la gamme du visible, il convient d'utiliser des illuminants possédant suffisamment d'énergie sur tout le spectre du visible afin d'éviter le phénomène de métamérisme. Le métamérisme est ce phénomène qui nous fait percevoir la même sensation colorée, sous certaines conditions d'éclairage, pour deux objets qui possèdent pourtant des spectres de réflectance différents.

Tableau I  
Illuminants standards type lumière du jour

Dénomination	Description	Température couleur
D <sub>50</sub>	Lumière du jour au lever ou coucher du soleil	5000 K
D <sub>55</sub>	Lumière en milieu de matinée ou d'après-midi	5500 K
D <sub>65</sub>	Lumière du jour à midi	6504 K
D <sub>75</sub>	Ciel nuageux	7500 K

Tableau II  
Illuminants standards

Dénomination	Description	Température couleur
A	Lampe tungstène, couleur orangée	2856 K
B	Lampe tungstène, couleur lumière du soleil	4874 K
C	Lampe tungstène, couleur ciel nordique	6774 K
E	Point blanc théorique	5400 K
F1 – F6	Fluorescent standard	
F7 – F9	Fluorescent à bande large	
F10 – F12	Fluorescent à trois bandes étroites	

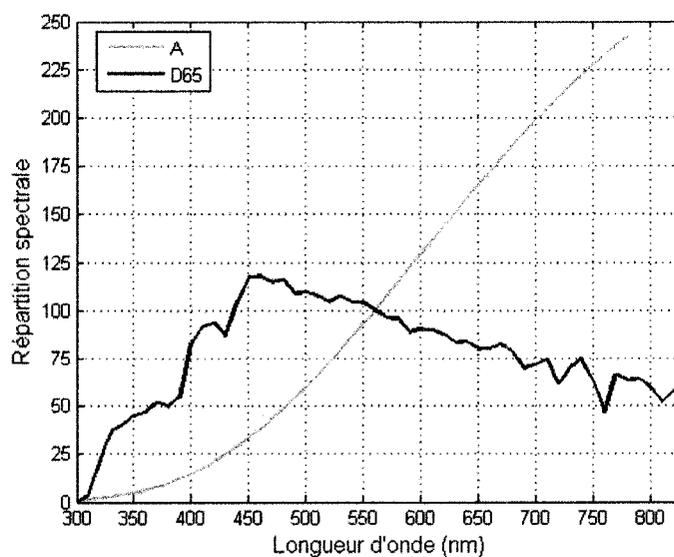


Figure 5 Répartition spectrale d'illuminants standards de la CIE <sup>1</sup>

<sup>1</sup> Données mises à disposition par [50]

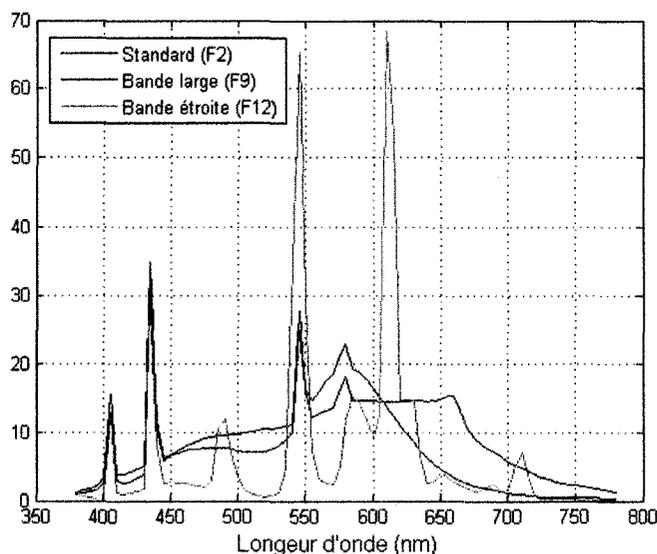


Figure 6 Spectres des illuminants fluorescents <sup>1</sup>

#### 2.4.4 Les systèmes de primaires

Le système de couleur RVB CIE 1931 s'inspire des résultats des expériences de Guild et de Wright. Ce système s'appuie sur trois primaires monochromatiques de longueur d'onde 700 nm pour le rouge, 546,1 nm pour le vert et 435,8 nm pour le bleu [51]. L'ensemble des couleurs du spectre du visible peut être construit par synthèse additive des trois primaires précédentes, pondérée par des coefficients. Ces coefficients, appelés fonctions colorimétriques ou composantes trichromatiques spectrales, ont été déterminés lors des expériences d'égalisation réalisées par des observateurs standards. Les fonctions colorimétriques sont liées à chacune des primaires et sont dénommées de la manière suivante :  $\bar{r}(\lambda)$ ,  $\bar{v}(\lambda)$  et  $\bar{b}(\lambda)$ .

<sup>1</sup> Données mises à disposition par [50]

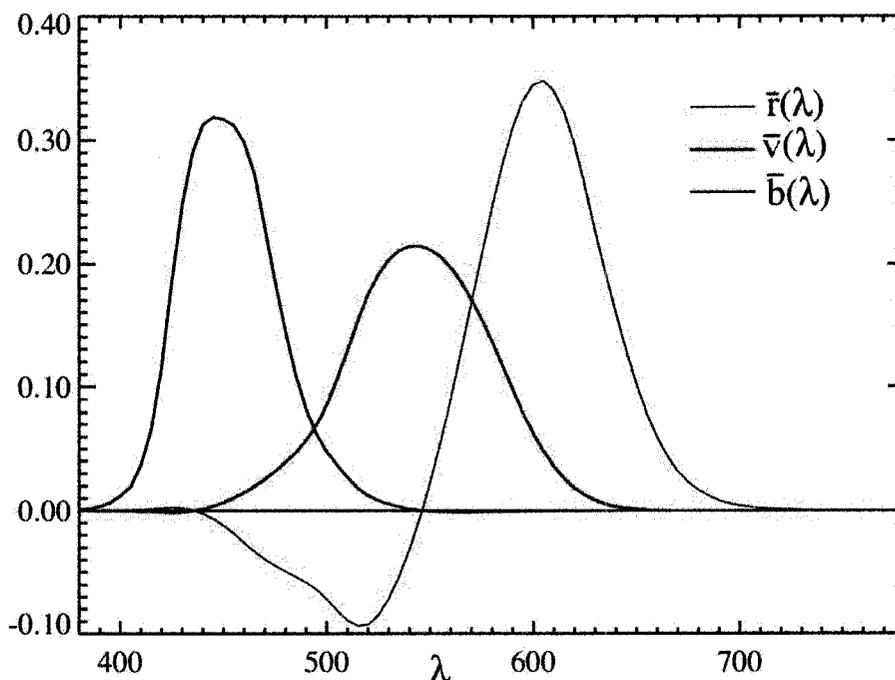


Figure 7 Fonctions colorimétriques du système de primaires RVB par l'observateur de référence CIE 1931 à 2° (Tiré de [52])

Ces primaires présentent la particularité d'être normalisées sur le domaine du visible :

$$\int_{\lambda_{\min}}^{\lambda_{\max}} \bar{r}(\lambda) d\lambda = \int_{\lambda_{\min}}^{\lambda_{\max}} \bar{v}(\lambda) d\lambda = \int_{\lambda_{\min}}^{\lambda_{\max}} \bar{b}(\lambda) d\lambda \quad (2.2)$$

Néanmoins, le système colorimétrique RVB CIE 1931 possède un inconvénient majeur, la fonction  $\bar{r}(\lambda)$  possède une zone négative, ce qui s'oppose au principe de synthèse additive. La zone négative avait été obtenue lors de l'expérience par ajout d'une source primaire rouge au niveau de la source monochromatique à égaliser. Cela n'est pas convenable en pratique, car toutes les couleurs ne peuvent être reproduites. C'est pourquoi un nouveau système colorimétrique a été introduit.

Pour pallier aux défauts du système colorimétrique RVB 1931, Judd va introduire un nouveau système de primaires, le XYZ CIE 1931 puis 1964, dit imaginaire, car ses

primaires n'ont pas de réalité physique. Le système XYZ possède les caractéristiques suivantes [51] :

- a. Aucune des fonctions colorimétriques  $\bar{x}(\lambda)$ ,  $\bar{y}(\lambda)$  et  $\bar{z}(\lambda)$  ne comportent de valeurs négatives;
- b. Le triangle délimité par les trois primaires XYZ englobe l'ensemble des couleurs pures (monochromatique);
- c. L'intégrale de chaque fonction sur le spectre visible est égale aux deux autres;
- d. La fonction colorimétrique  $\bar{y}(\lambda)$  décrit à elle seule l'efficacité lumineuse d'un stimulus.

Ainsi, chaque courbe est normalisée sur le spectre du visible :

$$\int_{\lambda_{\min}}^{\lambda_{\max}} \bar{x}(\lambda) d\lambda = \int_{\lambda_{\min}}^{\lambda_{\max}} \bar{y}(\lambda) d\lambda = \int_{\lambda_{\min}}^{\lambda_{\max}} \bar{z}(\lambda) d\lambda \quad (2.3)$$

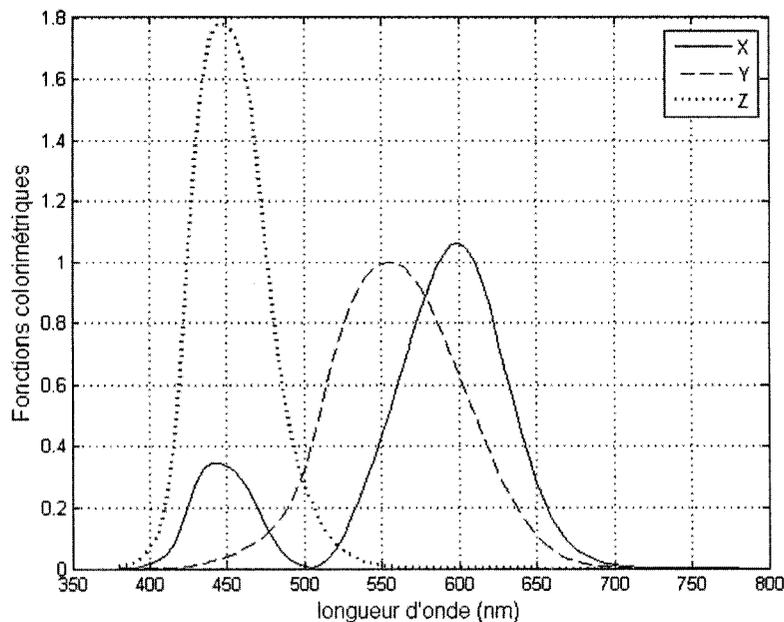


Figure 8 Fonctions colorimétriques du système de primaires XYZ de l'observateur de référence CIE 1931 à 2°<sup>1</sup>

Le transfert des coordonnées du système RVB au système XYZ se fait de façon linéaire au moyen d'une matrice de passage :

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 2,7690 & 1,7518 & 1,1300 \\ 1,0000 & 4,5907 & 0,0601 \\ 0,0000 & 0,0565 & 5,5943 \end{bmatrix} \begin{bmatrix} R \\ V \\ B \end{bmatrix}$$

où les colonnes de la matrice (2,7690; 1,0000; 0,0000), (1,7518; 4,5907; 0,0565) et (1,1300; 0,0601; 5,5943) sont respectivement les coordonnées XYZ des primaires rouge, vert et bleu.

<sup>1</sup> Données mises à disposition sur [50]

Afin de faciliter la visualisation de l'espace XYZ, le diagramme de chromaticité a été créé. Ce diagramme utilise les coordonnées normalisées suivantes :

$$\left\{ \begin{array}{l} x = \frac{X}{X+Y+Z} \\ y = \frac{Y}{X+Y+Z} \\ z = \frac{Z}{X+Y+Z} \end{array} \right. \quad (2.4)$$

La somme de ces trois coordonnées étant égale à 1, on leur préfère le triplet xyY. La visualisation des couleurs se fait alors dans un espace à deux dimensions suivant les coordonnées x et y, à luminance Y constante.

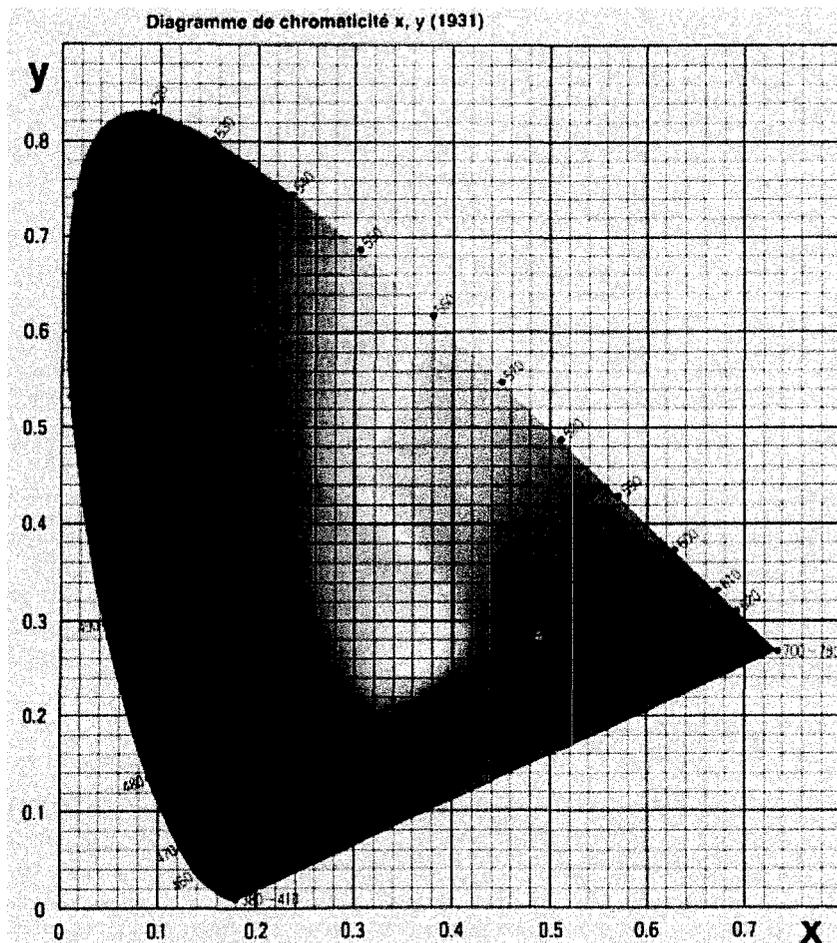


Figure 4 Diagramme de chromaticité

*(Tiré de [49])*

La courbe allant de 380 nm à 780 nm représente l'ensemble des couleurs pures. La droite passant du bleu (380 nm) au rouge (780 nm) s'appelle la droite des pourpres. L'ensemble des couleurs visibles est contenu entre la courbe des couleurs pures et la droite des pourpres. Les couleurs reproductibles par les trois primaires rouge (700 nm), vert (546,1 nm) et bleu (435,8 nm) sont contenues dans le triangle formé par ces primaires. Ce diagramme permet ainsi de connaître l'ensemble des couleurs reproductible, mais il possède également des propriétés géométriques intéressantes sur la perception des couleurs en fonction de l'illuminant.

Le diagramme de chromaticité possède de nombreuses propriétés entre les couleurs et l'illuminant de référence appelé blanc de référence. Le blanc de référence est lui-même défini par ses coordonnées dans le système  $xyY$  sur le diagramme de chromaticité. L'additivité est la principale caractéristique de ce diagramme. Toutes les couleurs contenues entre la courbe des couleurs pures et la droite des pourpres peuvent être obtenues par un mélange additif entre une couleur pure et un blanc de référence. Ainsi, le résultat d'un tel mélange se situera sur la droite qui relie la couleur pure au blanc de référence. Plus généralement, un mélange entre deux couleurs se situera sur la droite joignant ces deux points. Pour n'importe quelle couleur, il est possible de retrouver la longueur d'onde dominante, pour cela, il suffit de prolonger la droite passant par le blanc de référence et la couleur, jusqu'à atteindre la courbe des couleurs pures. Il est également possible de retrouver la couleur complémentaire d'une couleur du spectre en traçant une droite passant par la couleur en question et le blanc de référence, jusqu'à atteindre l'autre côté des couleurs pures. Grâce au diagramme de chromaticité, il est donc facile de prédire l'apparence d'une couleur en fonction du blanc de référence. Dans la pratique, il est donc nécessaire de pouvoir caractériser ce blanc de référence par ses coordonnées.

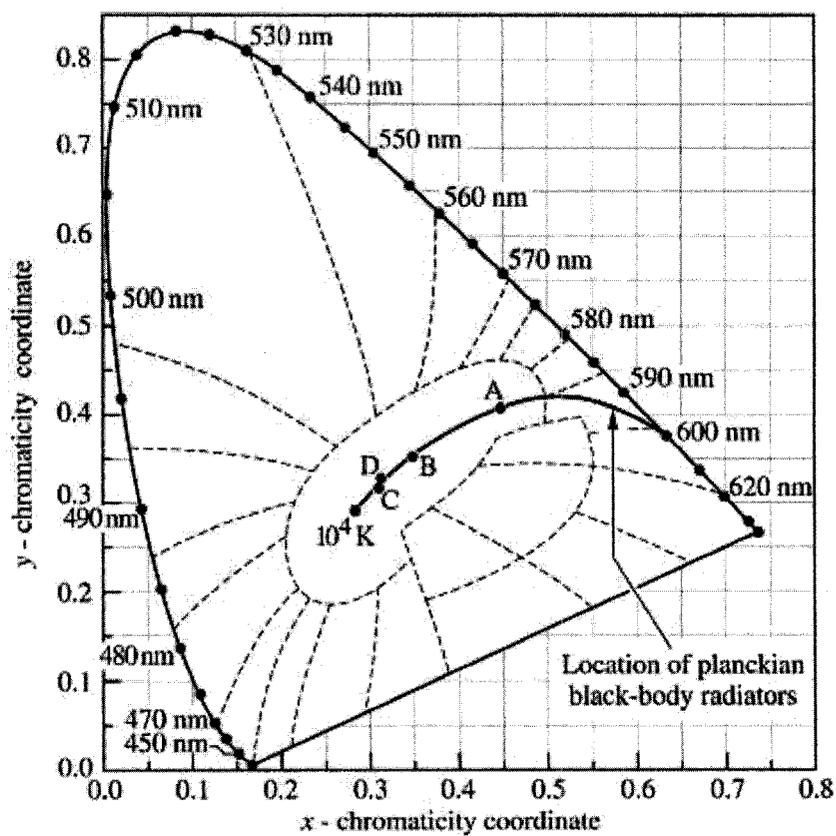


Figure 9 Courbe d'émission du corps noir en fonction de la température de couleur et coordonnées d'illuminants standard dans le diagramme chromatique  
(Tiré de [53])

Tableau III

Coordonnées xyz d'illuminants standard

Illuminant	x	y	z
A	0,448	0,407	0,145
C	0,310	0,316	0,374
D <sub>65</sub>	0,313	0,329	0,358
E	0,333	0,333	0,333

Le système de primaires XYZ CIE 1931 constitue la base de la colorimétrie. De ce fait, tous les systèmes de primaires et tous les espaces colorimétriques peuvent être reliés à ce système.

#### **2.4.5 Espace couleur uniforme**

Un bon espace couleur doit être fidèle à la perception humaine des couleurs. Les deux espaces colorimétriques précédents présentent justement un défaut majeur, les couleurs ne sont pas réparties selon la perception humaine. En effet, la distance entre deux couleurs n'est pas systématiquement proportionnelle avec la différence perçue par l'homme entre ces deux couleurs. Les expériences de Macadam ont démontré la non-uniformité de l'espace  $xyY$  [54]. L'expérience consistait à étudier l'aptitude du système visuel humain à égaliser les couleurs dans des conditions où la couleur des stimuli visuels variait selon plusieurs directions dans le plan  $xy$  autour d'une couleur de référence. Les zones sur lesquelles la variation des stimuli apparaissait visuellement identique à la couleur de référence ont été modélisées par des ellipses, et étudiées pour 24 positions différentes de la couleur de référence. Les différences d'orientations et de tailles des ellipses illustrent clairement la non-uniformité de l'espace  $xyY$ . Divers espaces ont alors été conçus afin de pallier la non-uniformité des systèmes colorimétriques précédents.

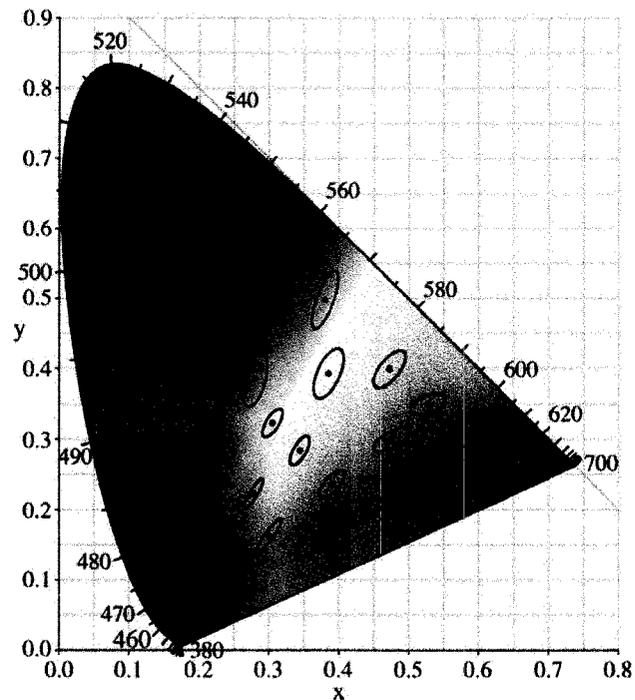


Figure 10 Seuils différentiels de chromaticité obtenu par Macadam  
(grossissement ellipses  $\times 10$ ) (Tiré de [55])

L'espace colorimétrique CIELAB se rapproche de la perception humaine des couleurs. Il a été normalisé par la CIE en 1976, et découle des travaux de Adams et Nikerson. Cet espace a pour objectif une uniformisation des sensations colorées. Il faut qu'à une même distance entre deux couleurs proches corresponde toujours une même différence perceptuelle, indépendamment de la position des couleurs. Cela équivaut à ce que toutes les couleurs indifférenciables à une autre par la perception humaine soient comprises dans un cercle de rayon unitaire. Le système CIELAB sépare notamment la notion de luminance (axe  $L^*$ ) de celle de chromaticité (axe  $a^*$  et  $b^*$ ). C'est aussi un espace à opposition de couleur. L'axe  $L^*$  représente donc la luminance, soit une opposition de couleur noir-blanc. Les axes  $a^*$  et  $b^*$  représentent quant à eux respectivement les oppositions de couleurs rouge-vert et bleu-jaune.

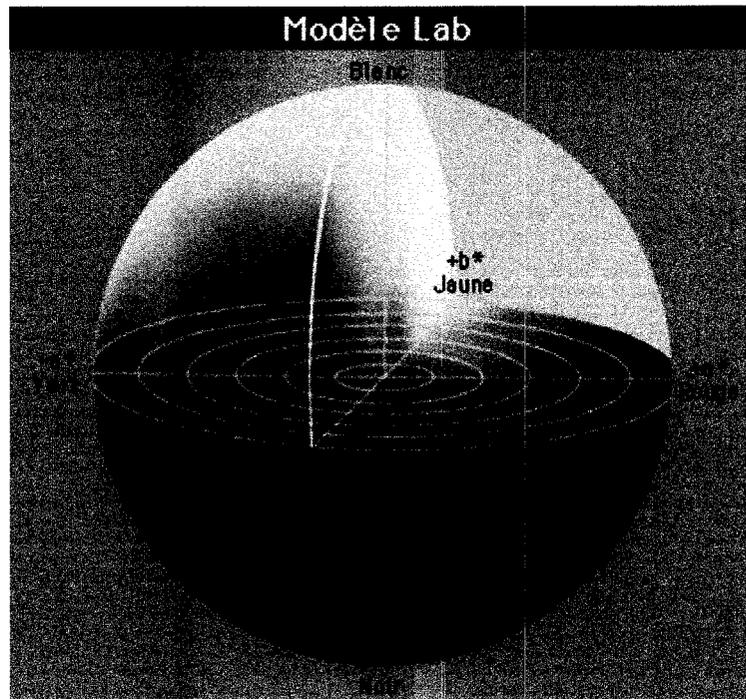


Figure 11 Espace colorimétrique CIELAB

(Tiré de [56])

La transformation non linéaire permettant de passer des coordonnées XYZ aux coordonnées CIELAB est donnée par les équations suivantes :

$$L^* = \begin{cases} 116 \times \left( \frac{Y}{Y_0} \right)^{\frac{1}{3}} - 16 & \text{pour } \frac{Y}{Y_0} > 0,008856 \\ 903,3 \times \frac{Y}{Y_0} & \text{pour } \frac{Y}{Y_0} \leq 0,008856 \end{cases}$$

$$a^* = 500 \times \left[ f\left( \frac{X}{X_0} \right) - f\left( \frac{Y}{Y_0} \right) \right]$$

$$b^* = 200 \times \left[ f\left( \frac{Y}{Y_0} \right) - f\left( \frac{Z}{Z_0} \right) \right]$$
(2.5)

Avec :

$$f(x) \begin{cases} x^{\frac{1}{3}} & \text{si } x > 0,008856 \\ 7,787x + 16/116 & \text{si } x \leq 0,008856 \end{cases} \quad (2.6)$$

Cette transformation se définit en fonction des coordonnées trichromatiques ( $X_0$ ,  $Y_0$ ,  $Z_0$ ) de l'illuminant de référence.

Afin de pouvoir évaluer la différence entre deux couleurs, la CIE a choisi en 1976 d'utiliser la distance euclidienne pour tous les espaces colorimétriques normalisés. Ainsi, l'écart de couleur CIE 1976 entre deux stimuli dans l'espace CIELAB, est défini de la manière suivante :

$$\Delta E_{a^*b^*}^* = \sqrt{\Delta L^{*2} + \Delta a^{*2} + \Delta b^{*2}} \quad (2.7)$$

Avec :

- a.  $\Delta L^*$  : l'écart de luminance;
- b.  $\Delta a^*$  : l'écart chromatique rouge-vert;
- c.  $\Delta b^*$  : l'écart chromatique bleu-jaune.

Il est également possible de calculer la distance couleur à partir des coordonnées cylindriques de l'espace CIELAB, de manière identique à la précédente, en utilisant les écarts de luminances ( $\Delta L_{a^*b^*}^*$ ), de chroma ( $\Delta C_{a^*b^*}^*$ ) et de teinte ( $\Delta H_{a^*b^*}^*$ ) :

$$\Delta E_{a^*b^*}^* = \sqrt{(\Delta L_{a^*b^*}^*)^2 + (\Delta C_{a^*b^*}^*)^2 + (\Delta H_{a^*b^*}^*)^2} \quad (2.8)$$

Avec :

$$\begin{cases} \Delta C_{a^*b^*}^* = C_2^* - C_1^* \\ \Delta h_{ab} = h_2 - h_1 \\ \Delta H_{a^*b^*}^* = 2\sqrt{C_1^* \cdot C_2^*} \cdot \sin(\Delta h_{ab}/2) \end{cases} \quad (2.9)$$

Et pour chacune des deux couleurs :

$$\begin{cases} C^* = \sqrt{a^{*2} + b^{*2}} \\ h = \text{atan2}(b^*, a^*) \end{cases} \quad (2.10)$$

Même si l'espace CIELAB est plus uniforme que les espaces RVB et XYZ, la figure suivante montre que l'objectif n'est pas entièrement atteint :

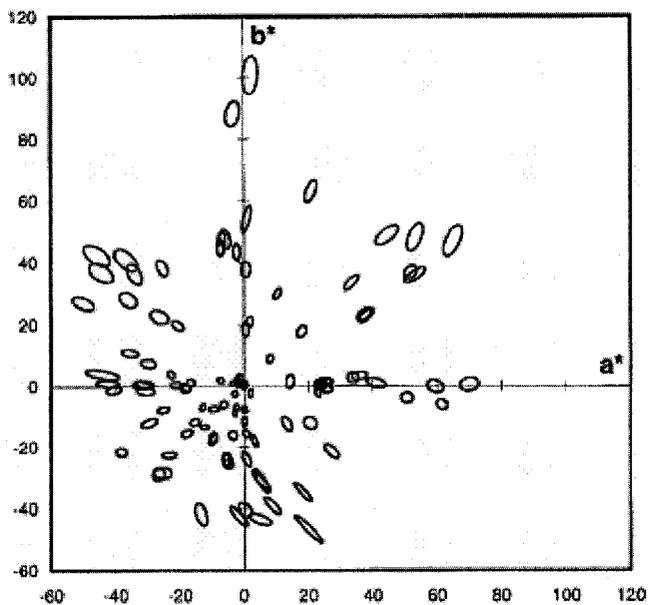


Figure 12 Seuils différentiels dans l'espace  $a^*b^*$   
(Tiré de [57])

En effet, même si les ellipses possèdent une uniformité dans la répartition de l'orientation et de leur forme, elles restent très différentes de cercles de rayon unitaire.

### 2.4.6 Distances couleur

Les formules de distance couleur ont pour objectif de modéliser les écarts entre les couleurs proches tout en étant fidèles à la perception humaine. Les travaux menés pour la création de l'espace CIELAB ont montré qu'il sera très difficile (voire illusoire) de créer un jour un espace parfaitement uniforme. Les travaux de recherche se sont ainsi concentrés sur le développement de nouvelles formules d'écarts de couleur dans l'espace CIELAB, plutôt que sur l'élaboration d'un nouvel espace perceptuellement uniforme. La plus grande demande pour ces formules provient de l'industrie textile, c'est donc elle qui définit en grande partie les problèmes que ces formules devront résoudre, ainsi que les formules qui seront adoptées comme standard. Trois formules sont couramment utilisées : la CMC (*Colour Measurement Committee*), la CIE 1994, et la CIEDE 2000. Elles s'appuient toutes sur l'espace couleur CIELAB, dont elles tentent de corriger les défauts. Ces formules intègrent des fonctions de pondération rendant la mesure de l'écart couleur conforme à la perception humaine, ainsi que des paramètres permettant de les adapter aux différents besoins de l'industrie en fonction de la nature de l'objet à mesurer (réflexivité, texture, couleur de fond...).

La formule CMC (1 : c) fut développée à l'origine pour l'industrie textile, et fut adoptée comme norme ISO pour cette dernière en 1995. Elle possède 2 paramètres : 1 pour la luminance et c pour le chroma, permettant de l'adapter aux besoins de différentes applications. La formule est la suivante [58] :

$$\Delta E = \sqrt{\left(\frac{\Delta L}{l \cdot S_L}\right)^2 + \left(\frac{\Delta C}{c \cdot S_C}\right)^2 + \left(\frac{\Delta H}{S_H}\right)^2} \quad (2.11)$$

Le couple (1 : c) prend en général les valeurs (1 :1) ou (2 :1). Les termes  $S_L$ ,  $S_C$  et  $S_H$  représentent les fonctions de pondérations respectivement de la luminance, du chroma et de la teinte, et ces fonctions sont assez complexes (non présentées).

La formule CIE 94 fut recommandée par la CIE en 1994 pour les essais sur le terrain [59]. La formule possède trois paramètres :  $k_L$ ,  $k_C$  et  $k_H$  respectivement pour la luminance, le chroma et la teinte. Ces paramètres sont généralement fixés chacun à une valeur égale à 1. La formule CIE 94 est la suivante :

$$\Delta E_{ab94}^* = \sqrt{\left(\frac{\Delta L^*}{k_L \cdot S_L}\right)^2 + \left(\frac{\Delta C_{a^*b^*}^*}{k_C \cdot S_C}\right)^2 + \left(\frac{\Delta H_{a^*b^*}^*}{k_H \cdot S_H}\right)^2} \quad (2.12)$$

Les fonctions de pondérations sont plus simples que celle du CMC, et d'une manière générale elles prennent les valeurs suivantes :

$$\begin{cases} S_L = 1 \\ S_C = 1 + 0,045 \cdot \Delta C_{a^*b^*}^* \\ S_H = 1 + 0,015 \cdot \Delta C_{a^*b^*}^* \end{cases} \quad (2.13)$$

La formule CIEDE 2000 fut publiée par la CIE en 2001 [60]. L'espace CIELAB possède une faible uniformité dans les régions de bleu [61]. Ainsi, la formule CIEDE 2000, qui est la plus élaborée de toutes, intègre une fonction de rotation  $R_T$ , améliorant ses performances dans les régions de bleu. La forme générale de la formule est la suivante [62] :

$$\Delta E_{00} = \sqrt{\left(\frac{\Delta L'}{k_L \cdot S_L}\right)^2 + \left(\frac{\Delta C'}{k_C \cdot S_C}\right)^2 + \left(\frac{\Delta H'}{k_H \cdot S_H}\right)^2 + R_T \left(\frac{\Delta C'}{k_C \cdot S_C}\right) \cdot \left(\frac{\Delta H'}{k_H \cdot S_H}\right)} \quad (2.14)$$

Les termes  $\Delta L'$ ,  $\Delta C'$  et  $\Delta H'$  ne sont pas calculés directement à partir des coordonnées CIELAB des couleurs, mais à partir d'une correction de ces dernières en fonction de leurs chroma. Le calcul entier de l'écart couleur CIEDE 2000 est long et complexe. L'implémentation de l'algorithme doit être précise pour éviter toute erreur dans les

calculs, même s'il existe des discontinuités dans la formule à cause de l'erreur engendrée par les imprécisions de calcul [63]. Cette formule possède également 3 paramètres ( $k_L, k_C, k_H$ ) fixés généralement aux valeurs (1, 1, 1) ou (2, 1, 1).

Ces trois formules sont actuellement utilisées par l'industrie textile pour le calcul de l'acceptabilité des couleurs. La formule CMC fut adoptée comme norme par l'industrie textile, mais dans le même temps, la formule CIE 94 permettait d'obtenir des résultats comparables avec des calculs moins lourds, plongeant ainsi l'industrie textile dans le doute pour le choix de la formule à adopter définitivement. De plus, les deux formules précédentes ne sont pas exemptes de défauts. La CIE décida alors d'établir une nouvelle formule, la CIEDE2000, afin de fournir un outil de mesure de la différence couleur beaucoup plus fiable pour l'industrie. Néanmoins, les études sur ces trois formules se succèdent, indiquant des résultats contradictoires en fonction de la base de test utilisée [64,65]. Aucune de ces formules ne se démarque nettement des autres, seul le futur indiquera laquelle aura été finalement retenue par l'industrie.

La différence de couleur commence à devenir imperceptible pour l'œil humain pour des distances couleur CIE 1976 inférieures à 2 unités dans le système CIELAB. Les formules de différence de couleurs, modélisant la sensibilité de l'œil en plaçant toutes couleurs de différence imperceptible dans un cercle unitaire, donneront donc une valeur de l'écart de couleur inférieure à celle fournie par la distance euclidienne. Il est recommandé d'utiliser les formules de différence de couleur pour une distance couleur inférieure à 5 unités. Pour les distances plus grandes, le calcul de la distance euclidienne est suffisant comme mesure de l'écart.

## **2.5 Acquisition d'image couleur**

L'acquisition d'image couleur nécessite le réglage et la calibration de l'appareil numérique, mais aussi de l'illuminant, afin de s'assurer de la fidélité des couleurs.

### 2.5.1 Choix de l'illuminant

L'illuminant joue un rôle important sur notre perception des couleurs. En effet, une même couleur peut nous sembler différentes si elle est vue à la lumière du jour où avec une lumière artificielle. Notre système de perception des couleurs est dépendant de la source lumineuse qui éclaire l'objet. L'œil humain possède néanmoins la capacité de s'adapter aux conditions d'éclairages différentes, ainsi, une feuille de papier blanche sera reconnue comme blanche dans la plupart des conditions d'éclairage. Malheureusement, les appareils de mesure de la couleur, tels les caméras numériques, ne sont pas encore aussi performants que le système visuel humain. C'est la raison pour laquelle les couleurs de certaines photos possèdent des dominantes rouges ou bleues si l'appareil n'est pas correctement paramétré.

Pour régler ce problème, la CIE a depuis longtemps normalisé les conditions d'observation de la scène. Elle énonce que les conditions de référence pour utiliser le système CIELAB pour décrire les différences de couleurs perçues nécessitent l'emploi d'un illuminant de type  $D_{65}$  avec un éclairement d'environ 1000 lux. De plus, depuis la démocratisation des appareils photo numérique, la visualisation des photos se fait de plus en plus sur les écrans d'ordinateur, leur point blanc (blanc de référence) fixé à  $D_{65}$  est donc devenu la norme pour les caméras numériques. Il est ainsi important d'utiliser un illuminant  $D_{65}$  pour éclairer la scène. Un illuminant possédant une température couleur corrélée proche de 6500 K n'est pas obligatoirement un bon illuminant, car une multitude de spectres peuvent correspondre à une même température couleur, dès que l'on sort du cadre d'étude du corps noir. En effet, la température couleur corrélée d'un illuminant est déterminée en projetant la couleur de l'illuminant sur la courbe d'émission du corps noir dans le diagramme de chromaticité [66]. Ainsi, un illuminant ayant une température couleur de 6504 K n'est pas forcément l'illuminant normalisé  $D_{65}$ . C'est pourquoi un autre critère, l'indice de rendu des couleurs (IRC) a été établi pour évaluer la capacité d'un illuminant à reproduire un corps noir ayant la même température

couleur. Plus précisément, le IRC indique la similitude entre la perception des couleurs sous un illuminant donné, et l'illuminant de référence qui est le corps noir ayant la même température couleur que l'illuminant à tester. Les valeurs de l'IRC s'échelonnent de 0 à 100, 0 étant un résultat exécrationnel et 100 représentant le résultat de l'illuminant de référence lui-même. C'est pourquoi une lampe à incandescence possède un IRC de 100, car c'est un corps noir par définition. L'IRC de fluorescents standards de type F2 est de 64, tandis que les fluorescents à large bande peuvent avoir un IRC jusqu'à 95. Il est préférable de ne pas utiliser les fluorescents à spectre discontinu ou présentant de nombreux pics (ANNEXE 1).

### 2.5.2 Balance des blancs

Comme il a été vu précédemment, l'illuminant éclairant la scène doit avoir comme blanc de référence le  $D_{65}$ . Malheureusement, il est difficile d'avoir à sa disposition en tout temps une source normalisée  $D_{65}$ . La caméra numérique doit donc être capable de s'adapter à différentes conditions d'éclairage, comme le fait le système visuel humain, afin que les couleurs de la photographie soient fidèles à celles de la scène originale. La caméra numérique utilise pour cela la balance des blancs. Elle effectue la correspondance entre le point blanc de l'illuminant de la scène et le point blanc du support de reproduction (papier ou écran d'ordinateur). Elle est effectuée en deux étapes. La première consiste à estimer l'illuminant éclairant la scène à partir d'une étude statistique des valeurs des pixels de l'image [67]. La seconde étape consiste, en utilisant l'estimation de l'illuminant, à égaliser les données des capteurs dans le format du point blanc du support de reproduction, en l'occurrence le  $D_{65}$ . À titre d'exemple, la méthode de balance des blancs GDT (*Generalized Diagonal Transform*) employée dans les caméras numériques sera présentée, car elle n'utilise que l'estimation du point blanc de l'illuminant de la scène, souvent l'unique information fournie par l'estimation [68,69]. La formule du GDT est la suivante [70] :

$$RVB_{out} = F_{out} \times D \times F_{in} \times RVB_{in} \quad (2.15)$$

Avec :

- a.  $RVB_{in}$  : vecteur  $3 \times 1$  des valeurs linéaires RVB du capteur de la caméra;
- b.  $F_{in}$  : matrice de passage  $3 \times 3$  vers un espace couleur intermédiaire [68];
- c.  $D$  : matrice diagonale  $3 \times 3$  effectuant la balance des blancs;
- d.  $F_{out}$  : matrice de passage  $3 \times 3$  de l'espace couleur intermédiaire vers RVB;
- e.  $RVB_{out}$  : vecteur  $3 \times 1$  des valeurs linéaires RVB.

La formule précédente est décrite pour les valeurs linéaires RVB. La matrice  $F_{in}$  permet de passer dans un espace colorimétrique intermédiaire (Bradford RVB ou XYZ par exemple), dans lequel la balance des blancs sera effectuée par la matrice  $D$  :

$$D = \begin{bmatrix} \frac{R_{observation}}{R_{scène}} & 0 & 0 \\ 0 & \frac{V_{observation}}{V_{scène}} & 0 \\ 0 & 0 & \frac{B_{observation}}{B_{scène}} \end{bmatrix} \quad (2.16)$$

dans laquelle le triplet  $(R, V, B)_{scène}$  représente le point blanc estimé [67] de l'illuminant de la scène et le triplet  $(R, V, B)_{observation}$  représente le point blanc de l'illuminant d'observation du support de reproduction de l'image. Enfin, la matrice  $F_{out}$  permet le retour dans l'espace RVB linéaire à partir de l'espace intermédiaire. La méthode GDT assure que le point blanc de la scène correspondra au point blanc de l'environnement d'observation. Néanmoins, cette méthode n'assure pas l'adaptation au changement d'illuminant pour le reste des couleurs.

L'appareil photo numérique propose en général toute une palette d'options permettant de capter les couleurs de manière optimale. Pour la balance des blancs, l'appareil photo est capable de la réaliser de manière automatique. Un ensemble de pré réglage du constructeur est également disponible, permettant d'adapter l'appareil à des conditions d'éclairage comme la lumière du jour, le ciel nuageux, l'illuminant type tungstène ou encore type fluorescent. Si les conditions de l'éclairage de la scène sont connues, il est préférable d'utiliser les réglages de l'appareil plutôt que la balance des blancs automatique. Si aucun des modes précédents ne donne de résultats satisfaisants, ou bien si un calibrage plus précis est souhaité, il est toujours possible d'effectuer une balance des blancs manuelle. Cette dernière consiste à placer un blanc de référence neutre devant les capteurs de l'appareil, afin qu'ils s'ajustent sur ce blanc.

L'appareil photo peut enregistrer l'image dans plusieurs formats, les plus répandus sont le TIFF (*Tagged Image File Format*) et le JPEG (*Join Photographic Experts Group*). Le format TIFF n'altère pas l'image, car il utilise un mode de compression sans perte, mais en contrepartie, il peut engendrer des fichiers de taille considérable. Le format JPEG utilise une compression avec perte, qui permet de réduire la taille des fichiers tout en conservant des qualités visuelles très bonnes, à condition de ne pas appliquer un taux de compression trop drastique. Les appareils photo haut de gamme proposent aussi le format RAW, qui enregistre l'image telle qu'elle a été enregistrée par les photosites du capteur optique, sans aucun traitement (réduction du bruit, balance des blancs...) de ses valeurs par le processeur interne de la caméra. L'image est alors vierge de toute altération. Il est alors possible pour l'utilisateur de traiter lui-même les paramètres (balance des blancs, contraste...) de cette image grâce à un logiciel approprié. Enfin, certains modèles de caméra proposent le choix entre le profil couleur sRVB (*standard RVB*) présent dans toutes les caméras, et le profil couleur Adobe RVB.

L'utilisation de mires couleur permet de calibrer ou de vérifier la calibration de l'appareil. Les mires en niveaux de gris permettent la calibration de la balance des

blancs en mode manuel, tandis que les mires en couleur permettent de vérifier la capacité d'acquisition fidèle de la couleur de l'appareil. La mire couleur la plus populaire est la *ColorChecker* 24 cases de GretagMacbeth. Elle est spécialement conçue pour la vérification de la fidélité des couleurs, et elle est couramment utilisée dans les industries de la photographie et plus généralement pour le calibrage couleur d'une chaîne graphique. Elle est également largement utilisée dans la recherche lors de test sur les caméras numériques. Les deux premières lignes de la mire représentent les couleurs rencontrées le plus communément dans la nature, comme les nuances de couleurs de la peau, les couleurs de feuillages ou du ciel. La troisième ligne contient les couleurs primaires des systèmes RVB et CMJ (Cyan, Magenta, Jaune). Enfin, la dernière ligne représente les différents niveaux de gris (couleur achromatique) allant du blanc au noir. Chacune des cases possède une couleur très précise, il est donc possible d'évaluer son appareil en comparant les valeurs de la photographie de la mire avec celles de la mire réelle (ANNEXE 2).

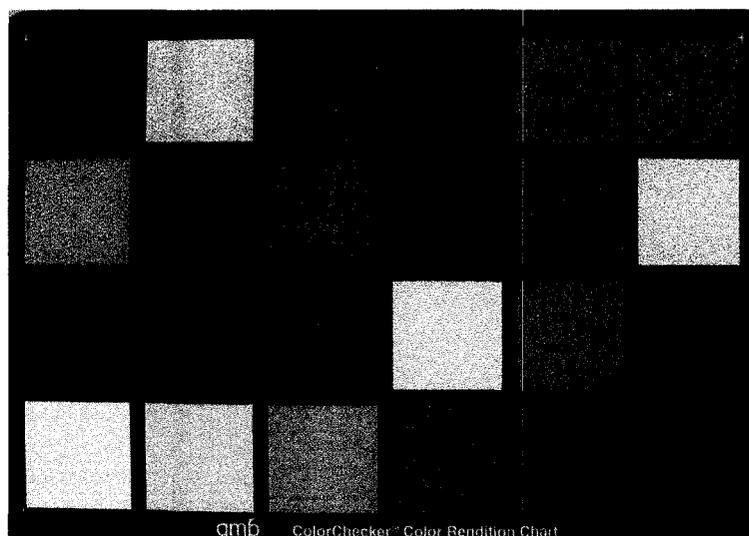


Figure 13 La mire couleur *ColorChecker* de GretagMacbeth

## 2.6 La gestion des couleurs

La gestion des couleurs est essentielle au bon fonctionnement de toute chaîne de reproduction couleur, dans laquelle l'image est généralement transférée d'une caméra numérique vers un ordinateur pour être travaillée graphiquement, puis transférée vers l'imprimante pour être imprimée. La gestion des couleurs consiste ainsi à assurer la fidélité des couleurs reproduites aux couleurs originales, quel que soit le média de représentation. En effet, de nombreux appareils permettent l'acquisition de couleur (scanner, caméra numérique, caméscope...), tandis que d'autres permettent la reproduction des couleurs (imprimante, moniteur à tube cathodique). Ces appareils, en fonction de leur type de technologie et de leur constructeur, possèdent un comportement différent vis-à-vis de la couleur. Par exemple, il y a de fortes chances que deux moniteurs de marque différente affichent deux couleurs différentes pour un même triplet RVB. Il est donc nécessaire que tous ces appareils puissent communiquer l'information couleur de manière fidèle entre eux, tout en rendant leurs différences intrinsèques transparentes.

### 2.6.1 La correction gamma

Le facteur gamma représente la non-linéarité de la reproduction de l'intensité lumineuse. Cette non-linéarité est liée non seulement aux phénomènes physiques internes aux différents appareils, mais également à la perception humaine.

L'œil humain ne présente pas une réponse linéaire face à l'intensité lumineuse. Les expériences initiées par Weber ont permis de caractériser la sensibilité de l'œil aux différences de luminance [71]. Il est possible d'observer (Figure 14) une discontinuité entre  $10^{-2}$  cd.m<sup>-2</sup> et  $10^{-3}$  cd.m<sup>-2</sup>, due au passage des cônes aux bâtonnets, et qu'à partir de  $1$  cd.m<sup>-2</sup> la courbe est régie par le logarithme de la luminance :

$$\log \Delta L = \text{constante} + \log L \quad (2.17)$$

Ainsi, cette courbe met en évidence le fait que la vision humaine est plus sensible au contraste de luminance pour les faibles valeurs de luminance que pour les valeurs élevées.

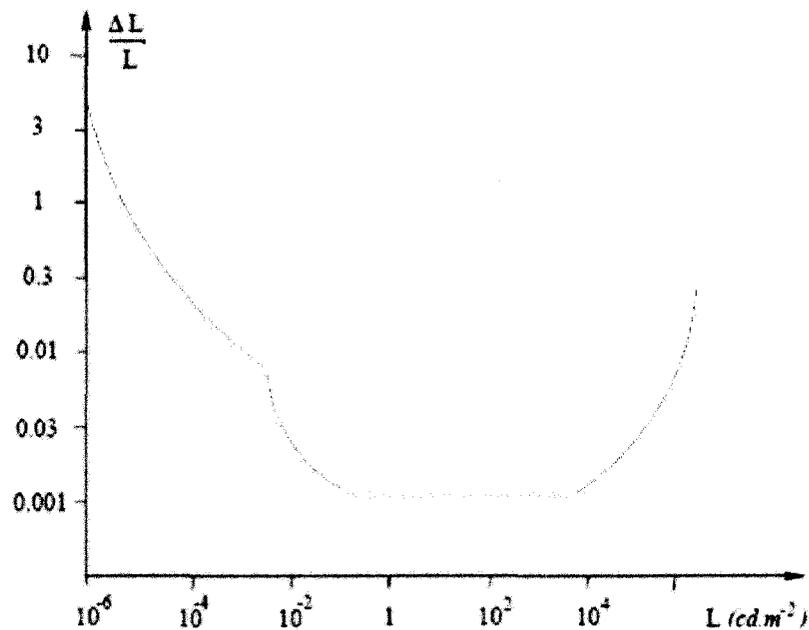


Figure 14 Courbe de sensibilité aux différences de luminance  
(Tiré de [72])

Les appareils photo numériques possèdent une réponse linéaire par rapport à l'intensité lumineuse. Les valeurs de chaque pixel sont le plus communément codées sur 8 bits, voire 10 ou 12 bits pour les appareils les plus performants. Pour un codage linéaire correct de la couleur, il est nécessaire d'utiliser 12 à 14 bits par composante pour satisfaire les sensations visuelles de l'œil. Or pour un codage sur 8 bits, un codage linéaire est nettement insuffisant. Il est alors nécessaire de recourir à la correction gamma, corrigeant la luminance de l'image :

$$S = E^\gamma \quad (2.18)$$

Avec :

- a.  $E$  : le signal à corriger;
- b.  $S$  : le signal corrigé;
- c.  $\gamma$  : le facteur de correction.

Pour un codage sur 8 bits, un facteur  $\gamma = 0,45$  est généralement employé dans les appareils photo numériques, privilégiant ainsi les faibles niveaux de luminance auquel l'oeil est plus sensible.

Les moniteurs à tube cathodique (notés CRT pour *Cathod Ray Tube*) ne possèdent pas non plus de réponse linéaire. En effet, la luminance reproduite à l'écran n'est pas directement proportionnelle à la tension à l'entrée du tube cathodique. La fonction de transfert d'un écran peut être modélisée par une correction gamma d'un facteur qui est communément de  $\gamma = 2,2$  pour les systèmes de type PC. Par coïncidence, la fonction de transfert des moniteurs compense parfaitement la correction gamma imposée pour satisfaire l'œil ( $0,45 \times 2,2 = 1$ ), lui permettant ainsi d'afficher les couleurs de façon linéaire.

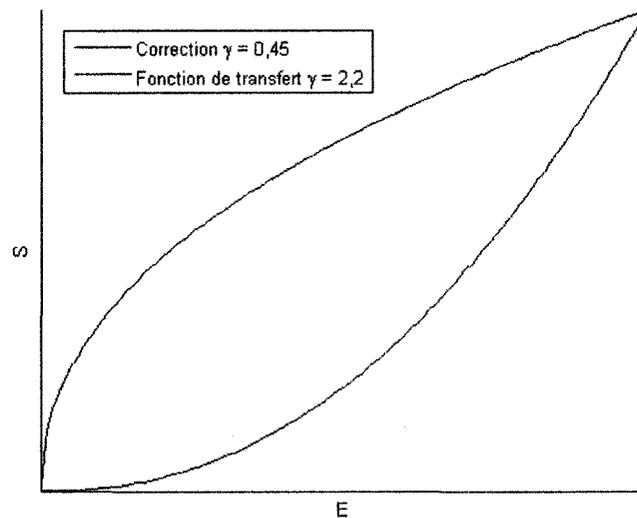


Figure 15 Correction gamma

### 2.6.2 Le gamut

Le gamut est un terme anglais qui désigne l'ensemble des couleurs qu'un appareil est en mesure d'identifier. Le gamut de l'œil est l'ensemble du spectre du visible. Les caméras numériques ne sont capables de capter qu'une partie limitée de l'ensemble des couleurs que l'œil humain peut discerner. Les moniteurs CRT ne peuvent afficher qu'un nombre de couleurs encore plus restreint, tandis qu'une imprimante en reproduira encore moins. Le gamut de chacun de ces appareils doit être parfaitement identifié, afin d'assurer la constance de la chaîne couleur. La caractérisation et la calibration de tels appareils se font en utilisant des mires couleur, propres à chacun des appareils.

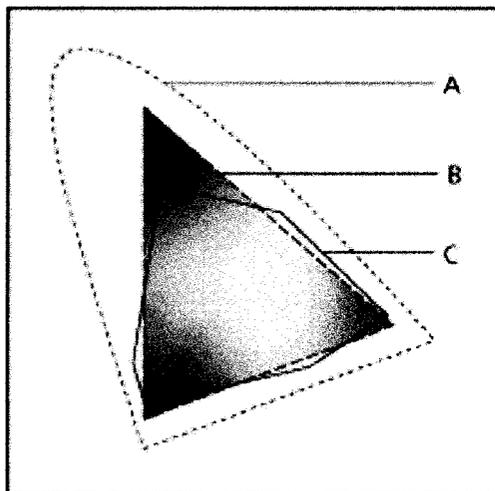


Figure 16 Comparaison de gamuts dans le plan du diagramme de chromaticité

A : spectre du visible, B : caméra 3 CCD, C : imprimante

*(Tiré de [73])*

Le comportement colorimétrique de chacun des appareils, une fois défini, est stocké dans un fichier appelé profil. Néanmoins, le problème de communication des couleurs entre les différents appareils n'est toujours pas résolu. En effet, la différence entre les gamuts des divers appareils soulève le problème des couleurs situées hors du gamut d'un appareil. Il est alors nécessaire de trouver le moyen d'imbriquer un gamut large dans un autre plus restreint. De plus, les différentes industries ne travaillent pas avec le même point blanc, le point blanc des moniteurs est fixé au  $D_{65}$  tandis que le point blanc de l'imprimerie est le  $D_{50}$ . Les primaires sont aussi différentes, les appareils se servant de la synthèse additive de couleur utilisent les primaires RVB tandis que les appareils se servant de la synthèse soustractive utilisent les primaires CMJ. La couleur exacte de ces primaires varie aussi en fonction des technologies de chaque constructeur. L'ensemble de ces points doit être résolu pour permettre une gestion des couleurs efficace.

### 2.6.3 Gestion des couleurs ICC (International Color Consortium)

L'ICC (*International Color Consortium*) a été fondé en 1993 par un ensemble de huit industriels<sup>1</sup> du monde de la couleur, dans le but de standardiser la gestion des couleurs entre les différentes plateformes et les différents périphériques. Les travaux de l'ICC sont les plus répandus dans le marché de la couleur. La solution de l'ICC est d'utiliser un référentiel commun, le PCS (*Profile Connection Space*), qui permet aux profils des différents périphériques de se connecter entre eux. Ce profil utilise un codage des couleurs indépendant des matériels en utilisant au choix les espaces colorimétriques CIEXYZ ou CIELAB. Le PCS est fondé sur l'observateur standard CIE 1931, c'est-à-dire en utilisant comme point blanc le D<sub>50</sub>, avec un angle d'observation de 2°. Le profil d'un périphérique, qui définit l'ensemble des caractéristiques colorimétriques de l'appareil, est alors utilisé par le module de gestion de couleur CMM (*Color Management Module*) afin d'effectuer la transition entre l'espace couleur propre de l'appareil et le PCS.

L'ICC définit trois types de périphérique présent dans une chaîne couleur :

- a. Les périphériques d'entrées (caméra, scanners...);
- b. Systèmes d'affichage;
- c. Périphérique de sortie (imprimante, presse...).

Pour les périphériques d'entrées, le CMM effectue le passage de l'espace couleur du périphérique vers le PCS, tandis que pour les systèmes d'affichage et les périphériques de sortie, le CMM effectue le passage du PCS vers l'espace couleur du périphérique.

---

<sup>1</sup> Adobe Systems Incorporated, Agfa-Gevaert N.V., Apple Computer Inc., Eastman Kodak Company, FOGRA-Institute (Honorary), Microsoft Corporation, Silicon Graphics Inc., Sun Microsystems Inc., Taligent Inc.

Afin de résoudre les problèmes de variété de primaires, de passage entre les points blancs des différents périphériques, et de mise en correspondance entre les gamuts, l'ICC propose quatre solutions :

- a. Le mode de rendu colorimétrie perceptive;
- b. Le mode de rendu colorimétrie relative;
- c. Le mode de rendu saturation;
- d. Le mode de rendu colorimétrie absolue.

Le mode de rendu colorimétrie perceptive conserve les relations de distance et de contraste entre les couleurs lors du passage à un gamut plus petit. Le gamut de départ est compressé dans le gamut d'arrivée. Ainsi, même les couleurs du gamut de départ situées à l'extérieur du gamut d'arrivée sont prises en compte. Cela a comme effet de produire un résultat agréable à l'œil en terme de sensation colorée. Néanmoins, en cas de passage vers un gamut différent, les couleurs ne seront plus les mêmes que dans l'image originale. Ce mode de rendu effectue le changement du point blanc lorsque cela est nécessaire.

Le mode de rendu colorimétrie relative calque à l'identique les couleurs du gamut de départ dans le gamut d'arrivée. Si des couleurs sont situées hors du gamut d'arrivée, alors la couleur la plus proche du gamut d'arrivée leur sera attribuée. Ce mode de rendu effectue aussi le changement du point blanc lorsque cela est nécessaire. Cela signifie que ce mode établit une correspondance entre les blancs et les noirs des deux espaces et il décale les autres couleurs en se fondant sur la progression de leurs valeurs numériques. Ce mode conserve les couleurs originales d'une manière plus fidèle que le mode de rendu colorimétrie perceptive.

Le mode de rendu saturation conserve la saturation des couleurs. Il est à utiliser lorsque la conservation des teintes est moins importante que celle de leur éclat. Il est privilégié

dans des images destinées à la communication, comme des affiches publicitaires. Ce mode de rendu effectue aussi le changement du point blanc lorsque cela est nécessaire.

Le mode de rendu colorimétrie absolue conserve les couleurs entre les gamuts à l'identique. Les couleurs situées hors du gamut d'arrivée sont remplacées par la couleur la plus proche de ce dernier. Ce mode de rendu ne modifie pas le point blanc, mais cela peut donner une dominante colorée au blanc après impression.

#### **2.6.4 Profils ICC**

Les profils ICC de chaque appareil sont la pièce maîtresse de la solution de gestion de couleur proposée par l'ICC. L'utilisateur a le choix, en fonction des disponibilités, d'utiliser le profil fourni par le constructeur, de créer lui-même un profil ou bien d'utiliser des profils génériques comme l'Adobe RVB ou le sRVB.

Les profils fournis par le constructeur sont propres à chaque modèle. Néanmoins, ces profils ne tiennent pas compte des variabilités entre les différents appareils d'un même modèle.

Si l'utilisateur travaille dans des conditions très contrôlées, de type studio, il peut alors être intéressant pour lui de créer les profils pour ses appareils lui-même. Pour cela, une mire couleur est utilisée pour calibrer correctement l'appareil, mais également pour compenser ses défauts. Dans le cas des appareils photo, une photo de la mire couleur est prise, dans des conditions identiques à celles de son utilisation. L'image est alors analysée par un logiciel, afin de produire un fichier profil effectuant la correspondance entre les valeurs captées de la mire et leurs valeurs réelles. L'utilisateur peut aussi effectuer l'analyse de la photo manuellement. Le format RAW se prête bien à la création de profil personnalisé. Dans le cas des moniteurs CRT et des imprimantes, une charte type est fournie sous format informatique et également papier. L'utilisateur doit alors

faire correspondre visuellement l'affichage sur l'écran ou la reproduction sur papier de la charte au format informatique, avec la charte au format papier en modifiant les paramètres de son périphérique. Pour moins de subjectivité, l'utilisateur peut se servir d'un colorimètre. Tous ces profils ne sont valides que pour les réglages et les conditions d'éclairage dans lesquels ils ont été créés. Toute modification ultérieure de l'un de ces paramètres rend le profil invalide. Il est alors nécessaire de créer un nouveau profil pour chaque condition d'utilisation.

L'espace Adobe RVB fut créé en 1998 par Adobe, avec comme objectif de remplacer l'ensemble des profils des différents périphériques. Il intègre le  $D_{65}$  comme point blanc, et un gamma de 2,2. Son gamut est suffisamment large pour englober les gamuts des écrans, scanners et imprimantes. Il est surtout utilisé par les professionnels de l'image pour son gamut étendu.

L'espace sRVB est issu des travaux communs de Microsoft et de Hewlett Pakard dans le milieu des années 90 pour trouver un espace colorimétrique indépendant du matériel. Cet espace est parfaitement compatible avec le PCS de l'ICC, et il est utilisé pour toutes les applications Internet dans lesquelles les systèmes de navigation ne sont pas en mesure de gérer les profils. L'espace sRVB est fondé sur les caractéristiques des moniteurs CRT, avec des conditions d'observation identiques à celles d'un bureau de travail. Les primaires de cet espace sont celles définies par la norme ITU-R BT.709 pour les phosphores des moniteurs CRT avec comme point blanc le  $D_{65}$  et un gamma de 2,2 :

Tableau IV  
Coordonnées xyz des primaires ITU-R BT.709 et du  $D_{65}$

	Rouge	Vert	Bleu	$D_{65}$
x	0,6400	0,3000	0,1500	0,3127
y	0,3300	0,6000	0,0600	0,3290
z	0,0300	0,1000	0,7900	0,3583

Pour la conversion de l'espace sRVB vers le PCS, il faut dans un premier temps appliquer une correction gamma de 2,2 pour linéariser le triplet non linéaire sRVB en triplet R'V'B'. Ensuite, ce triplet est converti dans l'espace XYZ par la matrice dite sRVB, et finalement, l'adaptation chromatique est réalisée pour le passage du point blanc D<sub>65</sub> au D<sub>50</sub> grâce à la matrice de Bradford, dont les coefficients sont établis à partir de la sensibilité spectrale des cônes.

Le détail de l'algorithme de conversion est le suivant. Le triplet sRVB codé sur 8 bits est normalisé de la manière suivante :

$$sRVB = sRVB_{8bits} / 255,0 \quad (2.19)$$

Ensuite, la correction gamma est appliquée pour linéariser le triplet sRVB normalisé. Une correction gamma de 2,4 au lieu de 2,2 est utilisée pour corriger l'offset présent dans certains moniteurs :

$$\left\{ \begin{array}{l} \text{Pour } RVB \leq 0,03928 : \\ R'V'B' = RVB/12,92 \\ \text{Sinon:} \\ R'V'B' = ((RVB + 0,055)/1,055)^{2,4} \end{array} \right. \quad (2.20)$$

Le passage aux coordonnées XYZ est ensuite effectué en utilisant la matrice sRVB:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix}_{D65} = \begin{pmatrix} 0,4124 & 0,3576 & 0,1805 \\ 0,2126 & 0,7152 & 0,0722 \\ 0,0193 & 0,1192 & 0,9505 \end{pmatrix} \cdot \begin{pmatrix} R' \\ V' \\ B' \end{pmatrix} \quad (2.21)$$

L'adaptation chromatique de changement d'illuminant est effectuée par l'utilisation de la matrice de Bradford :

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix}_{D50} = \begin{pmatrix} M_{BFD} \end{pmatrix}^{-1} \cdot \begin{pmatrix} r_{D50}/r_{D65} & 0 & 0 \\ 0 & v_{D50}/v_{D65} & 0 \\ 0 & 0 & b_{D50}/b_{D65} \end{pmatrix} \cdot \begin{pmatrix} M_{BFD} \end{pmatrix} \cdot \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}_{D65} \quad (2.22)$$

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix}_{D50} = \begin{pmatrix} 1,0479 & 0,0229 & -0,0502 \\ 0,0296 & 0,9904 & -0,0171 \\ -0,0092 & 0,0151 & 0,7519 \end{pmatrix} \cdot \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}_{D65} \quad (2.23)$$

Avec :

- $M_{BFD}$  : matrice d'adaptation chromatique de Bradford;
- $(r, v, b)_{D50}$  : réponses des cônes au blanc  $D_{50}$  dans l'espace de Bradford;
- $(r, v, b)_{D65}$  : réponses des cônes au blanc  $D_{65}$  dans l'espace de Bradford.

Après avoir effectué la correction gamma, l'ensemble de ces étapes peut être résumé au calcul matriciel suivant :

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix}_{D50} = \begin{pmatrix} 0,4361 & 0,3851 & 0,1431 \\ 0,2225 & 0,7169 & 0,0606 \\ 0,0139 & 0,971 & 0,7141 \end{pmatrix} \cdot \begin{pmatrix} R' \\ V' \\ B' \end{pmatrix} \quad (2.24)$$

Le profil sRVB doit être utilisé pour un écran dont le point blanc est fixé à  $D_{65}$ , situé dans un environnement dont la lumière ambiante émane d'un illuminant  $D_{50}$ . Le sRVB possède un gamut moindre que l'Adobe RVB. Néanmoins, il est le profil par défaut de tous les appareils photo numériques, et il s'impose également pour les moniteurs CRT et les imprimantes.

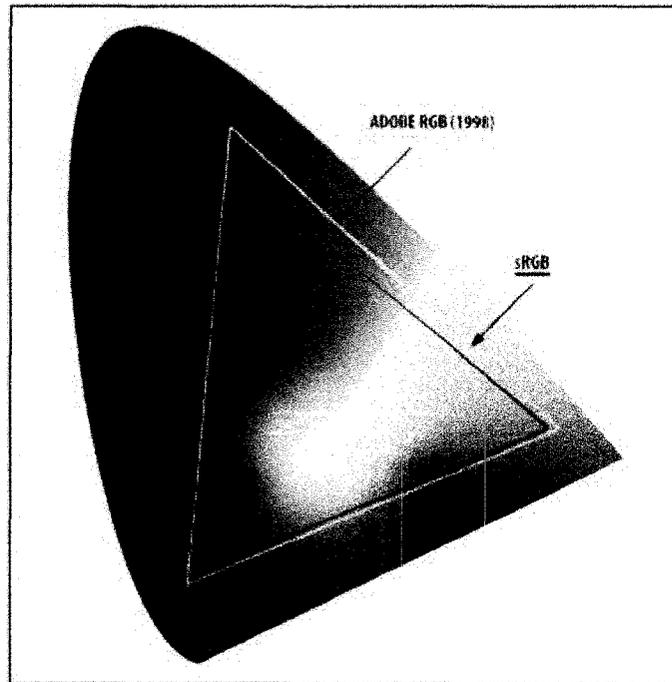


Figure 17 Comparaison des gamuts Adobe RVB et sRVB

*(Tiré de [74])*

## 2.7 Conclusion

La colorimétrie est une science très complexe qui implique à la fois des phénomènes purement physiques, mais également la physiologie et la psychologie humaine. La compréhension de tous les mécanismes de la perception de la couleur est encore loin d'être acquise. L'utilisation des nouvelles technologies dans la reproduction de la couleur a contribué à simplifier les tâches, mais en contrepartie, de nouveaux problèmes liés aux périphériques numériques sont apparus dans la gestion des couleurs. Néanmoins, des solutions commencent à apparaître pour une gestion des couleurs efficace. L'ensemble de ces aspects devra être pris en compte lors de l'acquisition des données couleur des canneberges par un appareil photo numérique.

## CHAPITRE 3

### Acquisition et traitement de l'image

#### 3.1 Introduction

La qualité d'un lot de canneberges sera évaluée par un système de vision par ordinateur. Cette section détaille le dispositif d'acquisition des images des fruits lors de la collecte de données, ainsi que le protocole de prise d'images. Les différentes étapes de traitements de l'image pour la segmentation des canneberges seront également décrites. Enfin, la répétabilité de la méthode sera testée et validée par une analyse statistique typique ainsi qu'une analyse de la variance à une entrée (ANOVA).

#### 3.2 Dispositif d'acquisition

Le matériel d'acquisition des images se compose d'une caméra numérique, d'un trépied ainsi que d'un récipient pour contenir les canneberges. La caméra CANON *Power Shot A-70* a été choisie pour la saison 2005, car elle est bon marché et offre de bonnes performances colorimétriques avec le profil sRVB [27]. La caméra CANON *Power Shot A-620* a été choisie pour la saison 2006. Le récipient utilisé est un plat blanc de marque Rubbermaid (22,9 x 15,2 x 5,1 cm). La couleur du fond, dans les conditions d'observation des couleurs du système CIELAB, doit être d'un gris uniforme et achromatique, de luminance  $L^* = 50$ . Une plaque de couleur grise ( $L^* = 59,80$ ,  $a^* = -0,40$ ,  $b^* = -2,6$  en  $D_{65}, 2^\circ$ ) mesurée avec un colorimètre Hunter Lab, se rapprochant des conditions précédentes, a été placée dans le fond du récipient. Cela a pour but de minimiser l'effet de contraste simultané, qui stipule que deux couleurs mises côte à côte apparaissent plus contrastées que si elles étaient placées séparément. Ainsi, la couleur du fond ne doit pas influencer la perception de la couleur des canneberges d'un observateur humain. Le fond gris a également été surélevé de 4 cm afin de limiter l'ombre portée par

les bords. Enfin, le fond gris limite les réflexions spéculaires colorées du fond sur les fruits.

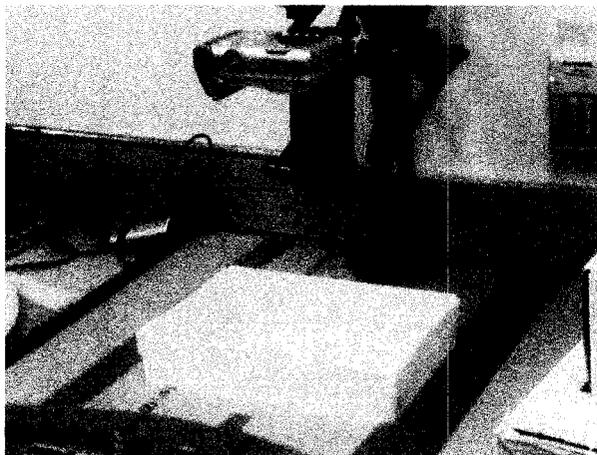


Figure 18 Matériels d'acquisition

L'illuminant utilisé pour éclairer la scène est celui présent dans les locaux du laboratoire. La température couleur de la lumière a été mesurée à l'aide du thermocolorimètre IIF de Minolta, et l'intensité lumineuse à l'aide d'un posemètre de marque Sekonic.

Tableau V  
Caractérisation de l'illuminant

Marque	: Philips
Référence	: F32T8/TL841/ALTO
Type	: Fluorescent similaire à F11
Température couleur	: 4100 K (4290 K mesuré)
IRC	: 86
Intensité lumineuse	: 80 Foot Candle

Le réglage des paramètres de chacun des appareils numériques est effectué en deux étapes : ajustement de la balance des blancs, puis recherche du couple temps d'exposition / ouverture optimal. Ces appareils offrent une large gamme de réglage de balance des blancs prédéfinis, notamment le réglage fluorescent, mais également les modes de réglage automatique et manuel. Les trois réglages donnent des résultats convenables. Néanmoins, le réglage automatique dépend des éléments présents dans la scène et peut être modifié d'une photo à la suivante, il ne permet aucun contrôle de la part de l'utilisateur. Le réglage fluorescent est établi en fonction d'un éclairage fluorescent standard, qui ne correspond pas forcément à notre illuminant. La balance des blancs en mode manuel a donc été choisie, en prenant comme blanc de référence une feuille de papier blanc d'imprimante au format lettre classique. Le couple temps d'exposition / ouverture a été initialement réglé à partir d'une table fournissant les valeurs du couple en fonction de l'intensité lumineuse. Ces réglages ont ensuite été testés sur la mire couleur *ColorChecker* 24 cases de GretagMacbeth.

Les valeurs sous l'illuminant  $D_{50}$  avec l'observateur à  $2^\circ$  fournies par GretagMacbeth dans le système CIELAB de la dernière ligne de la mire ont été comparées avec celles obtenues par les caméras réglées au préalable. Le réglage du couple temps d'exposition / ouverture a été affiné jusqu'à se rapprocher au plus près des valeurs fournies par GretagMacbeth. Les paramètres de réglage des caméras sont présentés ci-dessous (Tableau VI). Le flash a été désactivé afin de réduire les réflexions spéculaires à la surface des canneberges, et la résolution maximale a été choisie afin de conserver le maximum de détail. Les appareils numériques n'offrent que le format JPEG pour la sauvegarde de l'image, or ce format utilise un algorithme de compression qui altère l'image si le niveau de compression est trop important. La qualité de l'image est réglée sur Super Fin, pour atténuer l'effet de la compression. Enfin, ces appareils utilisent le profil sRGB, qui est désormais inclus dans tous les appareils numériques.

Tableau VI  
Paramètre de réglage des caméras numériques

Paramètre	Ajustement A70 (2005)	Ajustement A620 (2006)
Mode	Manuel	Manuel
Temps d'exposition	1/30 s.	1/15 s.
Ouverture	3,2	2,8
Sensibilité ISO	50	50
Objectif	5,4 *16,2 mm	5,4 *16,2 mm
Focale	5,4 mm	7,3 mm
Taille de l'image	2048*1536 pixels	2048*1536 pixels
Qualité de l'image	Super fin	Super fin
Flash	Désactivé	Désactivé
Balance des blancs	Manuel	Manuel
Mode champs autofocus	automatique	automatique
Zoom	non	non
Retardateur	2 s.	Mesure à distance

La couleur des canneberges est souvent non homogène. Or une caméra n'est capable de capturer uniquement la surface des fruits visible à son objectif, la moitié de la surface des fruits reste donc cachée. Pour résoudre ce problème, un protocole spécifique est établi, dans lequel 4 clichés (répétitions) de chaque échantillon de canneberges sont pris, avec un brassage aléatoire entre chaque cliché. Durant le brassage, la face exposée de chaque fruit est aléatoirement modifiée, permettant ainsi de découvrir un maximum de surface des canneberges.

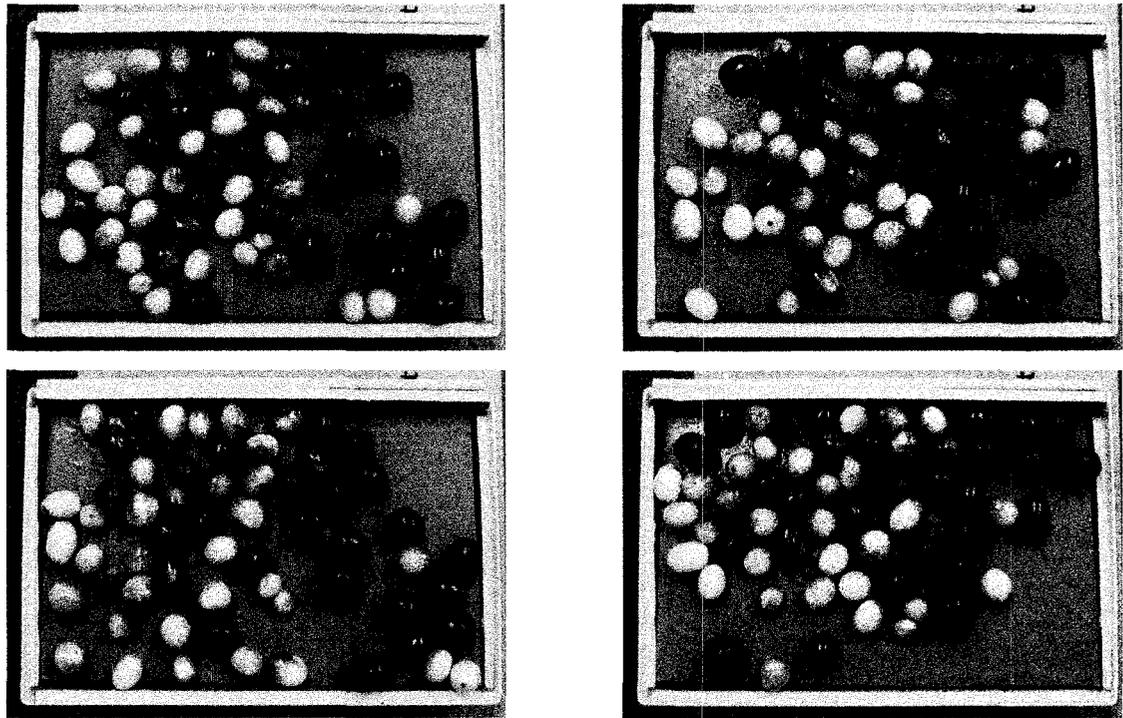


Figure 19 Photos des 4 répétitions d'un échantillon de canneberges suivant le protocole

### 3.3 Traitement d'image

Les photos de chaque échantillon de canneberges sont segmentées afin d'extraire les informations sur les fruits de l'image. Les canneberges possèdent une surface naturellement brillante, et elles arrivent en plus bien souvent humides. Cela provoque la présence de reflets spéculaires à la surface des fruits dans les photos. Ces reflets doivent être supprimés pour ne pas influencer la mesure de la couleur de la canneberge, mais également pour faciliter la segmentation. Lors de l'étape de prétraitement, les zones de reflets seront remplacées par la couleur originale de la canneberge, puis les canneberges seront segmentées individuellement.

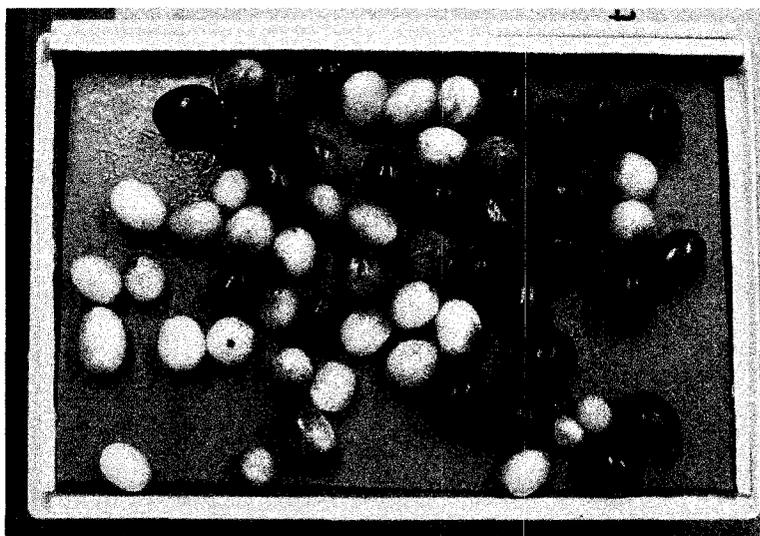


Figure 20 Photo n° 3 de l'échantillon précédent

#### 3.3.1 Élimination des reflets spéculaires

En fonction de leurs orientations, certaines parties de la surface des canneberges agissent comme des miroirs, et réfléchissent la couleur blanche de la source lumineuse, masquant la couleur de la surface originale. Ces réflexions doivent être éliminées afin de faciliter

la segmentation des fruits. Les réflexions spéculaires sont localisées dans les images aux centres des fruits, et possèdent une surface relativement limitée. L'espace colorimétrique permettant la meilleure distinction entre les fruits et les reflets a été déterminé après visualisation des images dans les espaces de représentation suivants : RVB, CIELAB et TSL (Teinte, Saturation, Luminance). L'espace TSL exprime les couleurs de manière subjective, se rapprochant du vocabulaire des peintres avec les notions de teintes, saturation et d'intensité lumineuse. Les reflets se distinguent le mieux dans la bande de saturation du système TSL. La saturation caractérise le côté pâle ou vif de la couleur, en d'autres mots, il mesure la proportion de mélange entre la couleur pure et le blanc. Les reflets sont très proches du blanc pur, ils possèdent ainsi une saturation très faible. De plus, les zones de reflets représentent un espace restreint de l'image. Les reflets sont donc localisés par un seuillage sur la bande de saturation du système TSL et par un filtrage sur le nombre de pixels. À cause de la nature de la surface des canneberges, il est possible de remarquer que les reflets sur les fruits rouges sont légèrement plus saturés et plus grands que les reflets sur les fruits foncés. Pour une plus grande précision, ces deux types de reflets sont segmentés séparément afin que d'autres objets entrant dans l'intervalle séparant ces reflets ne soient pas pris en compte. Les zones de seuillage de ces deux types de reflets sont ensuite réunies par l'opérateur logique OU binaire afin de localiser la position des reflets. Ces zones seront ensuite corrigées dans l'image originale.

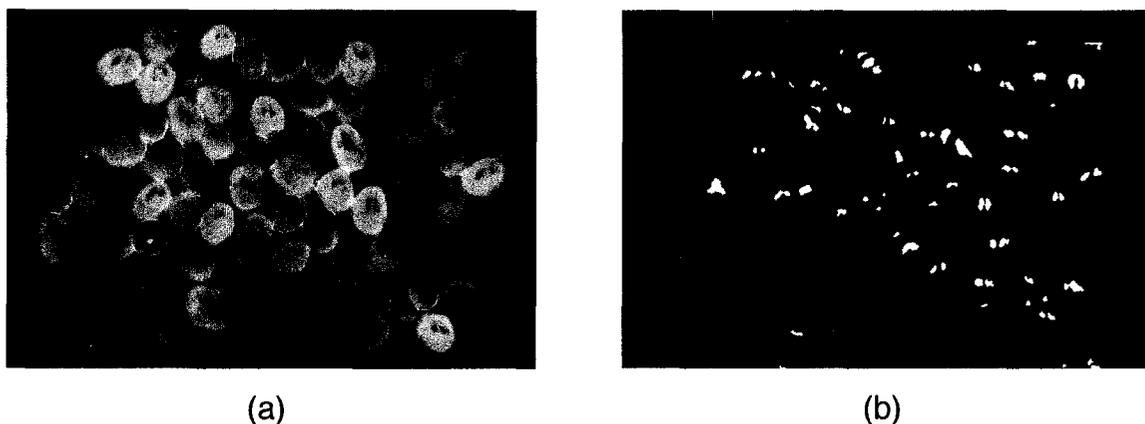


Figure 21 Localisation des reflets spéculaire.

(a) : Bande saturation du système TSL (b) : Position des reflets spéculaire

Néanmoins, certaines parties du fond gris possèdent un niveau de saturation très similaire à celui de certains reflets, ils sont ainsi considérés comme des reflets, tandis que certains autres reflets ne sont pas détectés. Une correction manuelle est donc nécessaire pour localiser tous les reflets, et enlever les fausses détections.

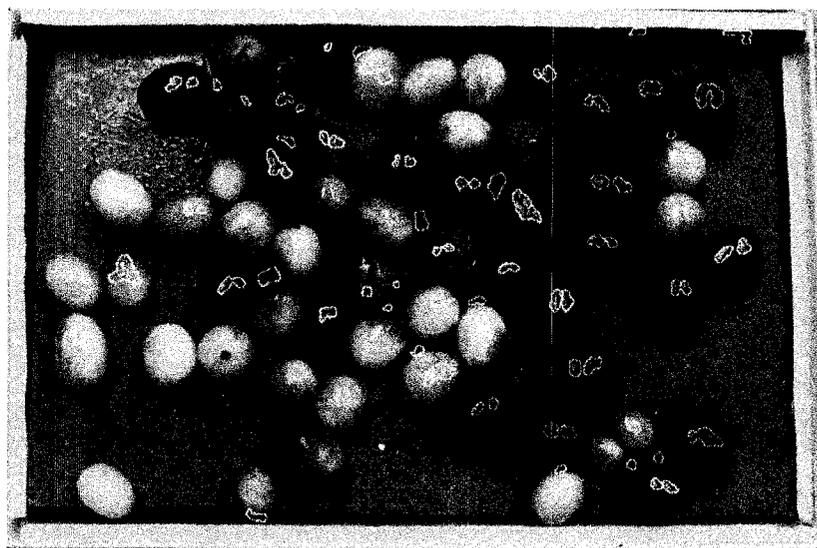


Figure 22 Position des reflets spéculaires détectés

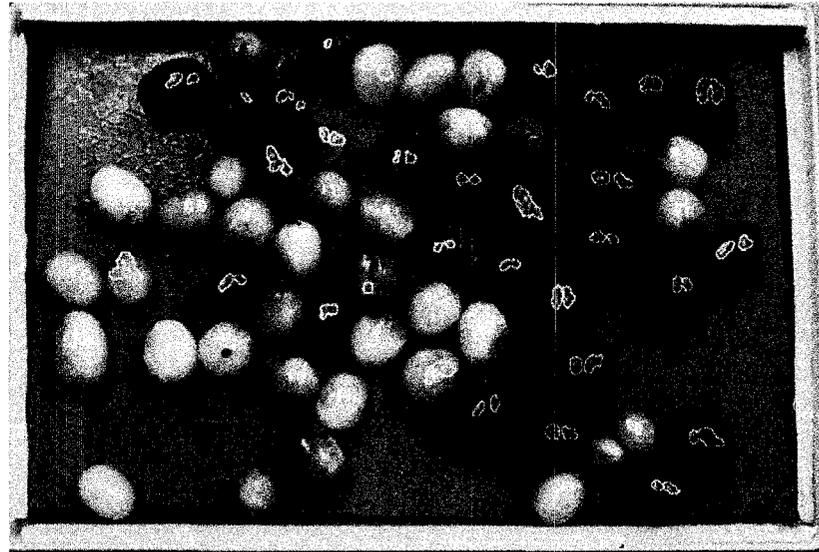


Figure 23 Position des reflets spéculaires après correction manuelle

Une méthode d'approximation bilinéaire est utilisée pour remplir les zones de reflets spéculaires à partir des couleurs présentes aux contours des reflets, tout en évitant de créer des discontinuités au niveau de la transition avec la zone de reflet. Cette méthode consiste, pour chaque pixel  $i$  de la zone de reflet, à utiliser les quatre pixels les plus proches situés hors de la zone de reflet. La nouvelle valeur du pixel  $i$  sera alors fonction des valeurs de ces quatre pixels, pondérées par des coefficients inversement proportionnels à leur distance au pixel  $i$ . Ainsi, plus un pixel sélectionné sera proche du pixel  $i$  par rapport aux trois autres, plus il contribuera à la nouvelle valeur de  $i$ , tandis qu'un pixel éloigné contribuera peu. Pour des raisons de simplicité algorithmique, les pixels les plus proches sont recherchés uniquement selon les directions horizontales et verticales. La figure suivante résume la recherche des pixels les plus proches :

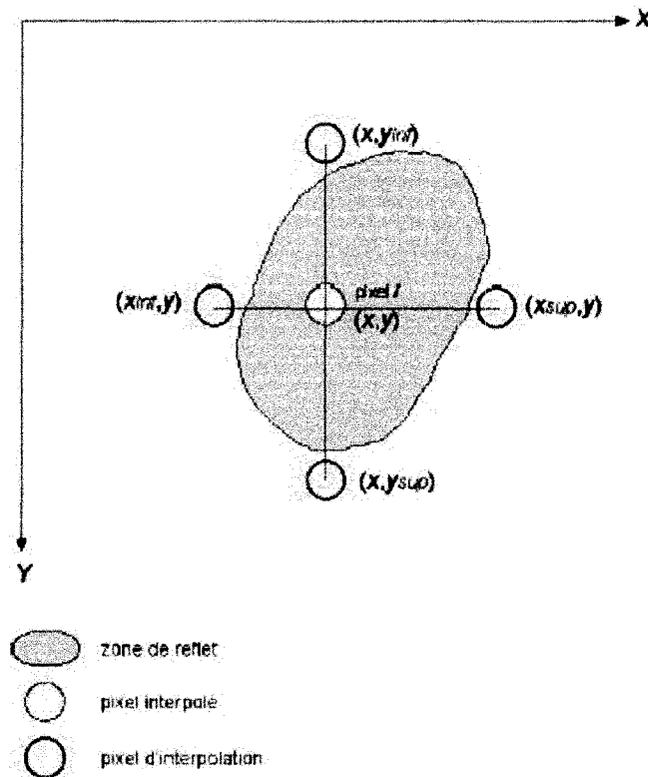


Figure 24 Approximation bilinéaire

Les poids associés à chacun des pixels avoisinants sont inversement proportionnels à leur distance au pixel de reflet à corriger :

$$\left\{ \begin{array}{ll} p_{haut} = \frac{1}{|y - y_{inf}|} & p_{bas} = \frac{1}{|y - y_{sup}|} \\ p_{gauche} = \frac{1}{|x - x_{inf}|} & p_{droites} = \frac{1}{|x - x_{sup}|} \end{array} \right. \quad (3.1)$$

La nouvelle valeur du triplet CIELAB affectée au pixel  $i$  s'obtient de la manière suivante :

$$val_{pixel\_i} = \frac{p_{haut} \times val_{haut} + p_{bas} \times val_{bas} + p_{gauche} \times val_{gauche} + p_{droite} \times val_{droite}}{p_{haut} + p_{bas} + p_{gauche} + p_{droite}} \quad (3.2)$$

Avec  $val$  correspondant aux bandes  $L^*$ ,  $a^*$  et  $b^*$  du CIELAB.

Le résultat du remplissage des reflets spéculaires sur les fruits est illustré dans la figure ci-dessous :

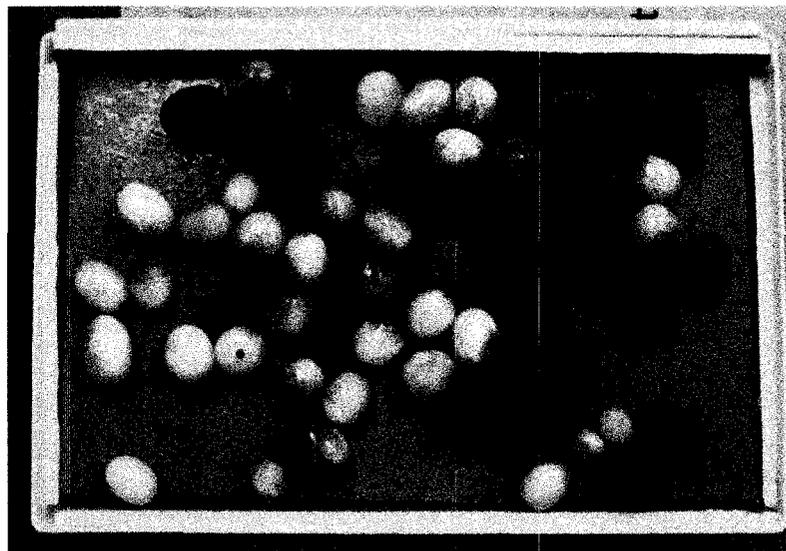


Figure 25 Image après élimination des reflets spéculaires

### 3.3.2 Segmentation

La segmentation a pour objectif d'isoler chacune des canneberges individuellement pour en extraire certaines caractéristiques. Dans un premier temps, les canneberges sont isolées en utilisant la formule de calcul de l'écart couleur CIEDE 2000 (voir CHAPITRE 2), puis elles seront isolées individuellement par l'opérateur morphologique de la ligne de partage des eaux.

Le fond possède une couleur uniforme, il est donc plus aisé d'identifier le fond plutôt que les canneberges dont la gamme de couleurs est très vaste. Néanmoins, à cause de l'éclairage non uniforme, de l'ombrage dû aux canneberges, de la présence d'eau provoquant des reflets spéculaires et de traces de jus de canneberge sur le fond, la couleur de ce dernier varie, et se confond parfois avec la couleur des canneberges. Les techniques classiques de seuillage et de croissance de région donnent des résultats mitigés quelque soit l'espace colorimétrique. La formule de calcul de l'écart couleur CIEDE 2000 a été retenue pour segmenter le fond gris dans l'espace CIELAB. Pour cela, la valeur typique du fond gris a été choisie comme valeur de référence du fond. Ensuite la distance CIEDE 2000 entre la valeur référence du fond et chacun des pixels de l'image a été calculée. Finalement, les pixels dont la distance couleur avec la couleur référence est inférieure à un certain seuil sont considérés comme étant des pixels appartenant au fond gris. Le même processus a aussi été appliqué pour isoler plus spécifiquement les zones ombragées du fond (Figure 26). L'opérateur morphologique de fermeture est appliqué au masque binaire des canneberges afin de combler les trous éventuels au sein d'un fruit, suivi de l'opérateur d'ouverture pour uniformiser les contours (Figure 27).

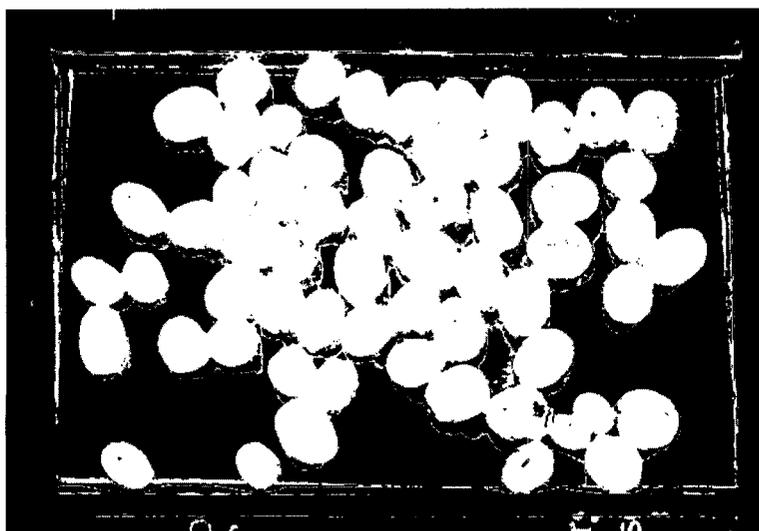


Figure 26 Image binaire du seuillage du fond

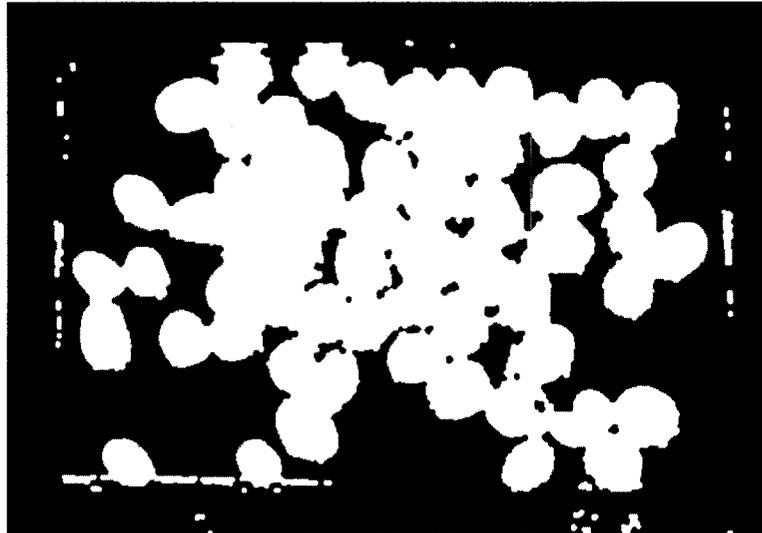


Figure 27 Image binaire du seuillage du fond après fermeture et ouverture

Les canneberges sont isolées du fond, mais elles restent accolées les unes aux autres. L'opérateur morphologique de la ligne de partage des eaux permet de séparer des objets convexes en contact dans une image [75,76]. Pour cela, l'opposé de la fonction distance est appliqué au masque binaire des objets. Cette nouvelle image est visualisée comme une surface en relief, l'altitude de chaque pixel étant déterminée par sa valeur respective. Les minima de ce relief sont considérés comme des puits à partir desquels le relief sera inondé. Il est alors possible de définir la ligne de partage des eaux comme étant la crête formant la limite entre deux bassins versant.

Dans notre cas, les fruits isolés du fond ne peuvent être directement séparés à l'aide de l'opérateur de partage des eaux en partant de leur masque binaire (Figure 27). En effet, la méthode de partage des eaux est inefficace dans les zones où des blocs de plusieurs fruits totalement collés entre eux apparaissent après le seuillage :

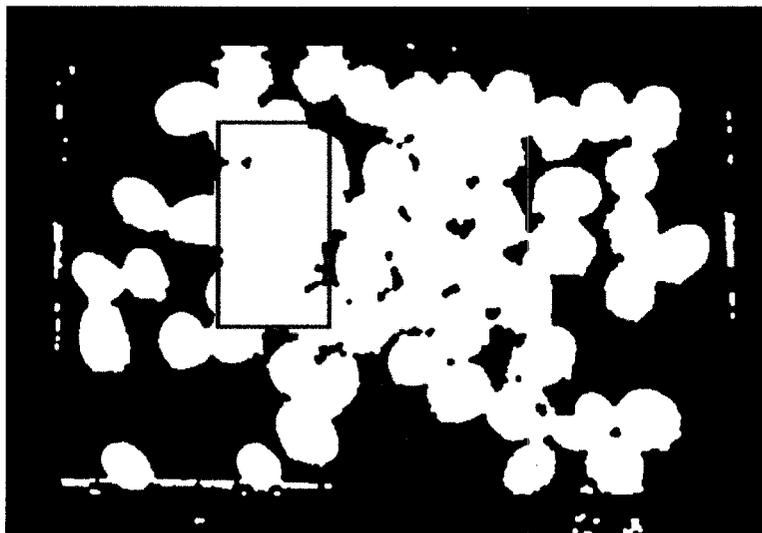


Figure 28 Zone où la méthode de la ligne de partage des eaux est inefficace (en rouge)

De l'information sur les contours des fruits est alors ajoutée au masque binaire précédent. Pour cela, l'image sans reflet spéculaire est filtrée par un filtre gaussien, afin de réduire le bruit, mais surtout pour effectuer un filtrage au niveau de la résolution des canneberges pour éliminer leur texture (Figure 29). L'opérateur de détection de contour de Sobel est appliqué sur chacune des bandes de l'image filtrée, dans le système colorimétrique RVB. Les fausses arêtes sont nombreuses dans l'image des contours, un seuil est ainsi fixé sur la magnitude des arêtes, afin de ne conserver que les vrais contours (Figure 30). Il en résulte des contours incomplets, mais qui complètent le masque binaire des canneberges.

Les informations sur les contours sont ajoutées au masque binaire des canneberges (Figure 31). Une ouverture permet d'uniformiser les contours, ce dernier masque sera utilisé pour la méthode de la ligne de partage des eaux (Figure 32).



Figure 29 Image sans reflet spéculaire filtrée par un filtre gaussien

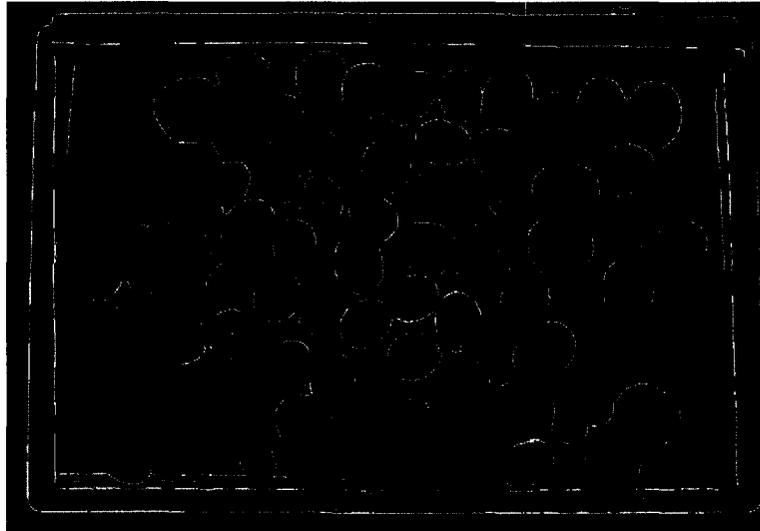


Figure 30 Détection de contour par l'opérateur de Sobel

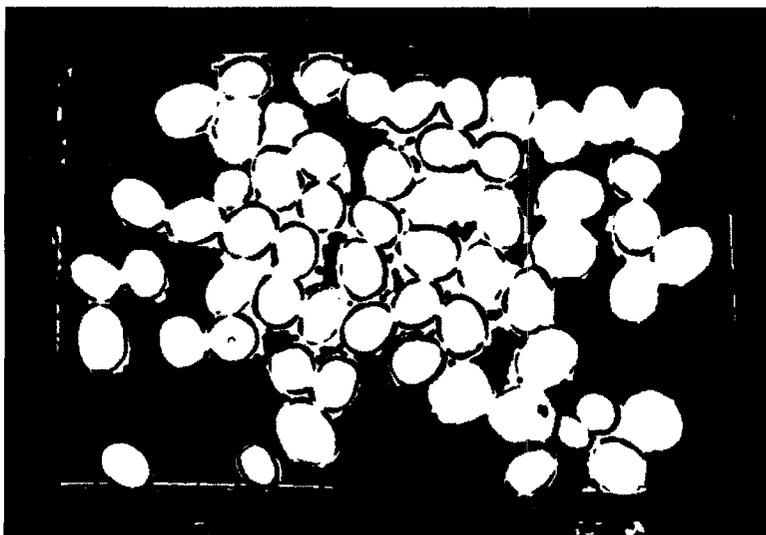


Figure 31 Réunion du masque binaire et des contours

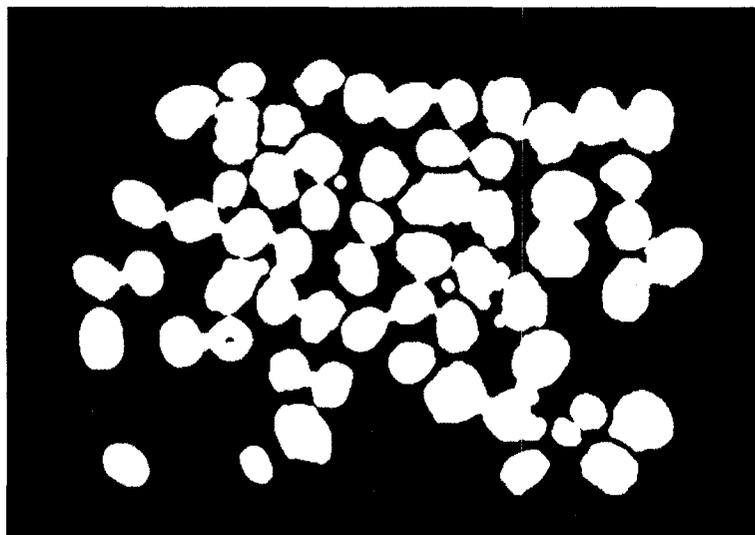


Figure 32 Masque binaire pour la méthode de partage de la ligne des eaux

Après application de la méthode de la ligne de partage des eaux, les objets ayant un nombre de pixels trop restreint pour être des canneberges sont filtrés, et les trous au sein des fruits sont comblés à l'aide d'un opérateur morphologique. Certains fruits peuvent ne pas être séparés à la suite de l'algorithme de partage des eaux, tandis que certains

autres peuvent être mal segmentés, une correction manuelle de la segmentation est alors nécessaire.

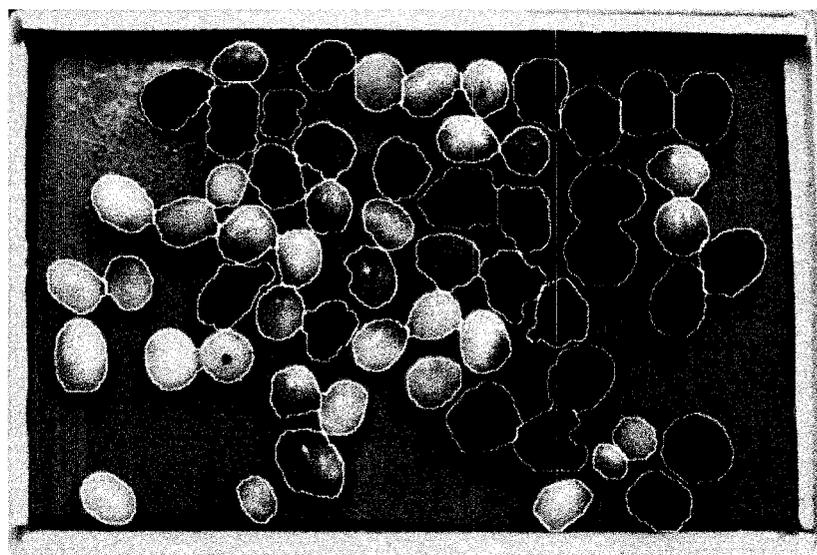


Figure 33 Résultat de la segmentation des canneberges

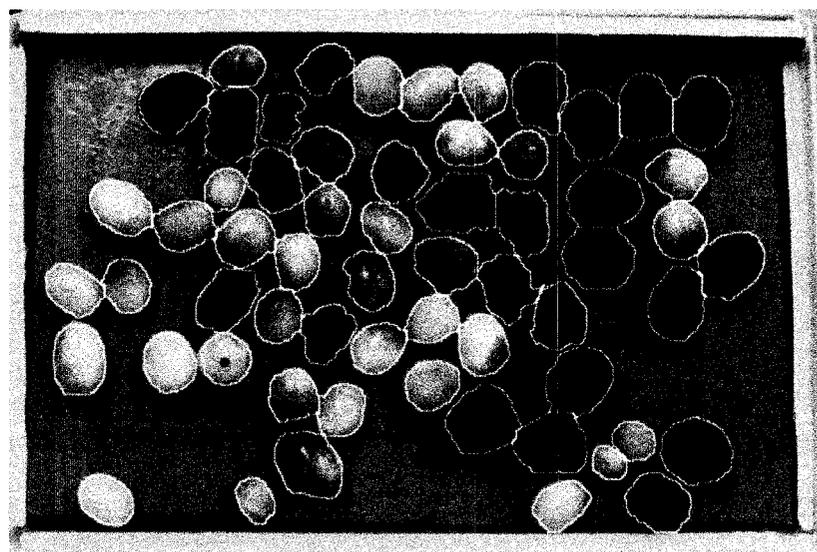


Figure 34 Résultat de la segmentation des canneberges après correction manuelle

La figure suivante résume l'ensemble du processus de segmentation des canneberges :

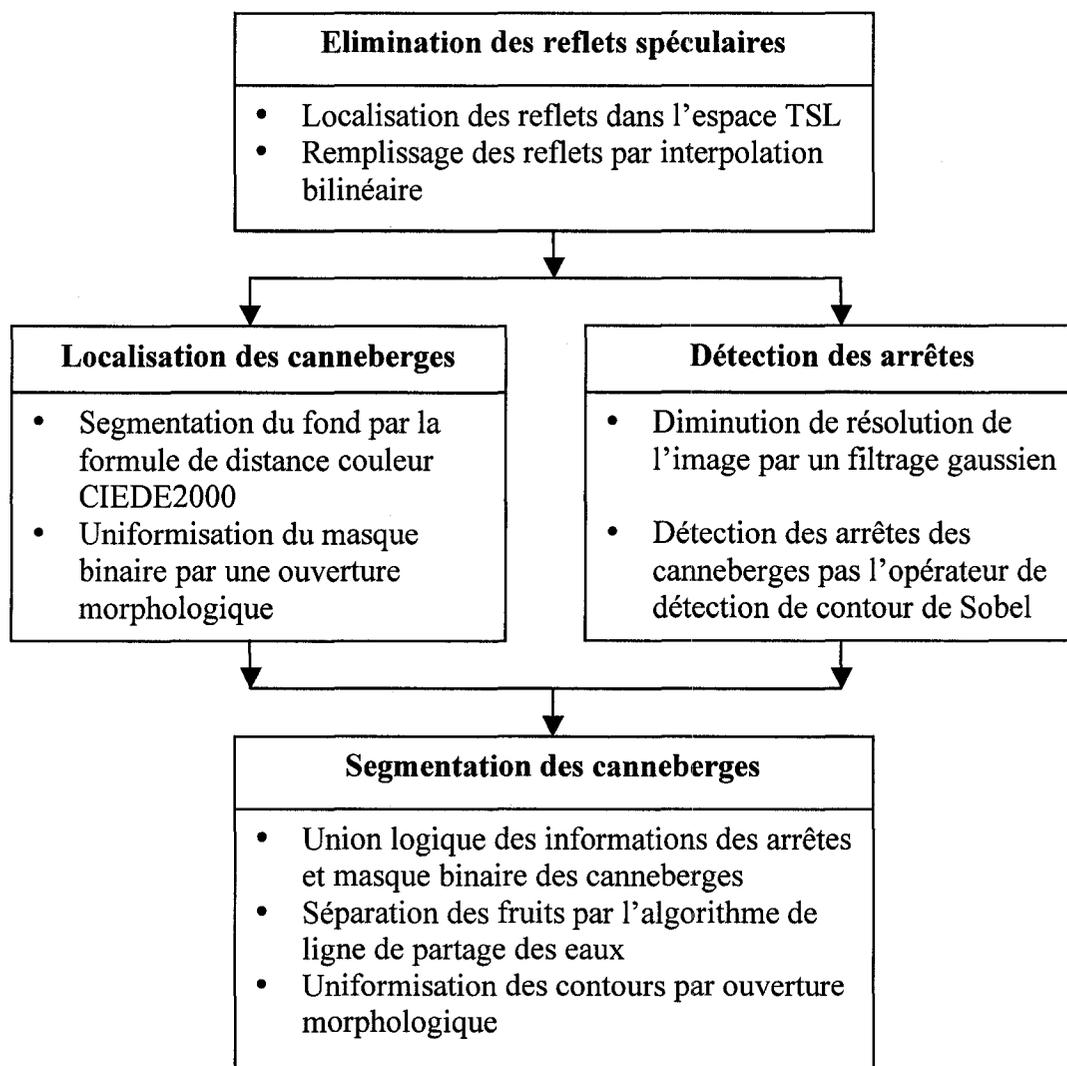


Figure 35 Processus de segmentation des canneberges

Les informations colorimétriques et de dimension de chaque canneberge peuvent être désormais extraites.

### 3.4 Analyse de la variance (ANOVA)

Chaque lot échantillonné de canneberges est photographié 4 fois, chaque photographie est appelée répétition. Avant de pouvoir regrouper les informations issues des 4 photographies, l'influence du brassage entre chaque répétition doit être au préalable étudiée statistiquement, afin de s'assurer de la répétabilité de la méthode. On obtient une mesure de la variabilité.

L'analyse de la variance à une entrée (ANOVA) est une méthode statistique permettant de comparer les moyennes de plusieurs groupes de populations ou des effets de traitements. Le test d'ANOVA à une entrée n'est applicable qu'aux populations remplissant certaines conditions : les populations doivent être indépendantes entre elles, elles doivent être normalement distribuées et elles doivent posséder la même variance. L'hypothèse nulle  $H_0$ , qui suppose que les valeurs moyennes des populations sont toutes égales, est testée lors de l'ANOVA.

Pour cela, la variance est distinguée en deux groupes : la variance totale et la variance intragroupe (MCE). La variance totale est calculée à partir des échantillons de toutes les populations réunies, tandis que la variance intragroupe est calculée individuellement pour chacune des populations. La variance intergroupe (MCF) est alors déterminée en soustrayant la variance totale par l'ensemble des variances intragroupes.

Si l'hypothèse nulle  $H_0$  est vérifiée, la variance intergroupe et la variance intragroupe devraient être proche. Cela est évalué en calculant le rapport F entre la variance intergroupe et la variance intragroupe, puis en comparant F à la distribution de Fisher correspondante, dépendamment des nombres de degrés de liberté, et du risque  $\alpha$  fixé arbitrairement par l'expert. Si F est supérieur, cela signifie que la probabilité que  $H_0$  soit valide est inférieure au risque  $\alpha$ . Il est alors possible de déduire que l'hypothèse de départ  $H_0$  n'est pas vérifiée à un risque  $\alpha$ , c'est-à-dire qu'au moins l'une des populations

possède une moyenne différente des autres. Si par contre  $F$  est inférieur, cela signifie que l'hypothèse  $H_0$  est vérifiée à un risque  $\alpha$ . Le terme  $p$ , généralement fourni par les logiciels de statistique, indique la probabilité d'une distribution avec un  $F$  donné d'être une distribution de Fisher. Si  $p$  est inférieur à  $\alpha$ , alors  $H_0$  n'est pas vérifiée à un risque  $\alpha$  [77].

Tableau VII  
Analyse de la variance

Source de variation	degrés de liberté	Sommes des carrés	Moyennes des carrés (variance)	F
facteur	$k - 1$	SCF	$MCF = SCF/(k - 1)$	$MCF/MCE$
erreur	$n - k$	SCE	$MCE = SCE/(n - k)$	-
total	$n - 1$	$SCT = SCF + SCE$	-	-

$$SCF = \sum_{i=1}^k (\bar{Y}_i - \bar{Y})^2 \quad (3.3)$$

$$SCE = \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 \quad (3.4)$$

Avec :

- a.  $k$  : nombre de population ;
- b.  $n$  : nombre total d'individus de toutes les populations ;
- c.  $n_i$  : le nombre d'individus de la population  $i$  ;
- d.  $Y_{ij}$  :  $j^{\text{ème}}$  individu de la population  $i$  ;
- e.  $\bar{Y}_i$  : valeur moyenne de la population  $i$  ;
- f.  $\bar{Y}$  : valeur moyenne de tous les individus de toutes les populations.

Dans le cas des échantillons de canneberges, chacune des photos représente une population de canneberges. Les quatre répétitions de mesure d'un même échantillon de fruits permettent d'étudier l'influence du brassage aléatoire sur les fruits, et donc de vérifier la répétabilité des mesures.

Une analyse de la variabilité typique (moyenne, SD, SEM, CV%) ainsi que l'ANOVA à une entrée ont été effectuées sur trois échantillons dits standards, car les fruits possèdent une couleur homogène à leur surface et également entre eux. Ces trois échantillons, dénommés A2, B2 et C2 sont composés de fruits de taille moyenne et de couleur respective : rouge foncé, rouge et rosé. L'analyse de variabilité s'est portée sur chacune des composantes couleur des fruits dans l'espace CIELAB en utilisant un profil sRVB, ainsi que sur la taille des fruits caractérisée par leur nombre de pixels dans l'image. Les résultats (ANNEXE 3) montrent que les coefficients de variation entre les moyennes des quatre répétitions sont inférieurs à 3%, ce qui est bien en dessous des 10% maximum généralement toléré dans un milieu industriel. De plus, l'ANOVA, effectuée avec le logiciel Origin v7.0383, montre qu'avec un risque  $\alpha$  de 0,001, il n'y a pas de différence significative entre les moyennes des différentes répétitions. La répétabilité du protocole est ainsi vérifiée, et il est possible de regrouper les informations des quatre répétitions.

### **3.5 Conclusion**

Ce chapitre a dans un premier temps présenté une méthode d'acquisition de l'image des fruits de canneberges, avec brassage aléatoire, permettant de découvrir un maximum de surface des fruits. Ensuite, la méthode de traitement de l'image utilisée sur chacune des photos a permis de segmenter les canneberges individuellement. Néanmoins, une correction manuelle du résultat de la segmentation est souvent nécessaire. La segmentation a permis d'extraire les informations colorimétriques et de dimension des canneberges à partir des photos de l'échantillon. Enfin, la répétabilité de la méthode a été validée par une analyse statistique typique montrant que le coefficient de variation

des moyennes de la couleur et de la taille des fruits de chaque répétition est inférieur à 3%, ainsi que par une ANOVA à une entrée montrant que les moyennes ne sont pas significativement différentes avec un risque  $\alpha$  de 0,001. Il est donc possible de regrouper les informations de chacune des répétitions afin de restreindre l'incertitude de mesure. Les attributs des fruits ainsi extraits permettront de prédire le Tacy des échantillons de 100 gr de canneberges.

## **CHAPITRE 4**

### **Couleur canneberge**

#### **4.1 Introduction**

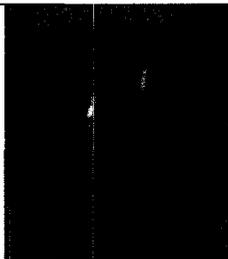
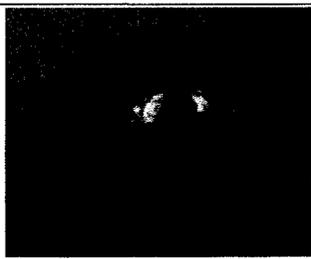
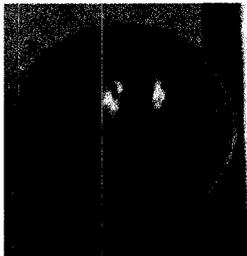
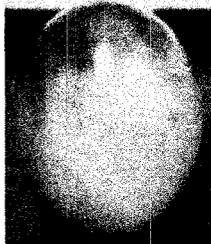
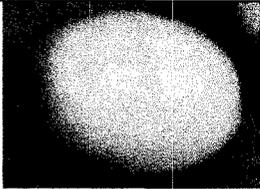
La forme de la couleur canneberge dans son espace de représentation doit bien être comprise afin de générer une échelle couleur canneberge au format papier comme outil de référence colorimétrique, mais également pour pouvoir ultérieurement relier la couleur au Tacy. Dans cette section, la procédure permettant la création de l'échelle couleur canneberge sera vue en détail. Tout d'abord, la base de données utilisée sera présentée. Elle sera dans un premier temps étudiée dans l'espace couleur ICCLAB. Ensuite une réduction de la dimension de l'espace de représentation sera effectuée par une analyse par composantes principales (ACP), puis une régression linéaire par méthode des moindres carrés permettra de modéliser la couleur canneberge. Enfin les différentes formes de visualisation de l'échelle couleur canneberge seront présentées.

#### **4.2 Base de données**

La base de données est constituée à partir de canneberges triées au préalable par un expert humain selon leur grosseur et leur couleur lors de la saison de cueillette 2005. Quatre catégories de grosseurs (petit, moyen, gros, très gros), ainsi que 4 catégories de couleurs (rouge foncé, rouge, rosée, jaune-vert) ont été créées, conduisant à un total de 16 classes. La grosseur des fruits est évaluée indirectement en fonction de la masse des fruits, mesurée à l'aide d'une balance, tandis que l'évaluation de la couleur est réalisée directement par l'expert. Néanmoins en pratique, toutes les classes n'ont pu être reconstituées, à cause d'un nombre insuffisant de canneberges pour atteindre les 100 gr d'un lot. Ainsi, les fruits de couleurs rosées et ceux de couleur jaune-vert sont classés indépendamment de leurs grosseurs, conduisant à un total de 10 classes (Tableau VIII).

Les fruits appartenant à ces classes seront considérés comme les standards, ils sont représentatifs de l'ensemble des canneberges existantes. À partir des 4 répétitions pour chacune des classes, un total de 3000 canneberges est disponible. La variété des lots standards est connue, mais n'est pas étudiée dans le présent travail.

Tableau VIII  
Classes standard de la canneberge (saison 2005)

		Taille			
		petit	moyen	gros	très gros
Couleur	rouge foncé				
	rouge				
	rosée				
	vert/jaune				

### 4.3 Espaces couleur

L'espace ICCLAB est le codage de l'espace CIELAB adaptant les plages de chaque bande couleur sur l'intervalle 8 bits [0 ; 255], tel que défini par l'*International Color Consortium* (ICC). L'espace couleur retenu pour le projet est l'espace couleur ICCLAB. La base de données d'étude est composée de l'ensemble des couleurs moyennes des fruits appartenant aux lots dits standards, créés durant la saison de cueillette 2005. Cette dernière est ainsi constituée de 3000 échantillons de couleurs de canneberges (Figure 36 et ANNEXE 4). La figure suivante nous montre la forme de la variabilité de la couleur.

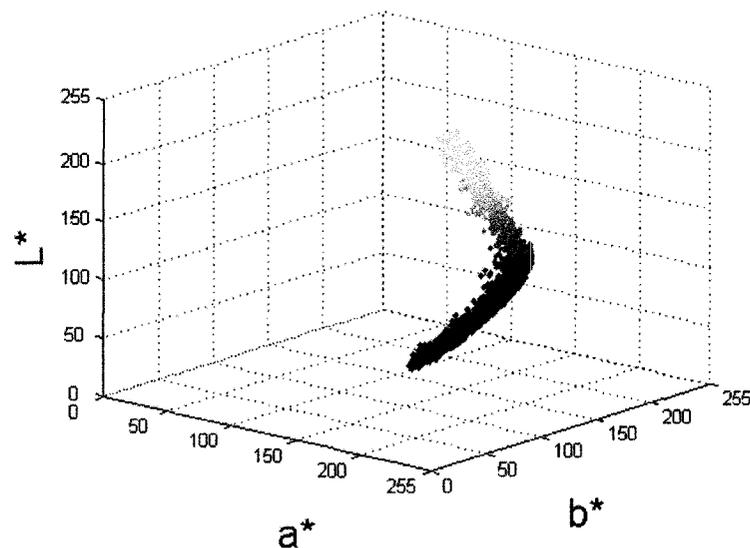


Figure 36 Couleur moyenne des fruits standard dans l'espace couleur ICCLAB

Les résultats de l'ANOVA à une entrée nous ont permis de regrouper les populations des 4 analyses d'images répétées comme une seule population, avec le risque alpha ( $\alpha = 0,001$  comme terme d'erreur) que ces populations ne puisse pas être statistiquement réunies. L'étude de la matrice de covariance de ces données couleur dans l'espace ICCLAB permet de mettre en évidence les variables les plus discriminantes (possédant la plus grande variance) ainsi que les corrélations existantes entre elles. La covariance se définit de la manière suivante, à partir de l'espérance des variables :

$$Cov(X, Y) = E[(X - \bar{X})(Y - \bar{Y})] \quad (4.1)$$

La covariance d'une variable avec elle-même se nomme la variance. La matrice de covariance de l'espace ICCLAB prend la forme suivante:

$$C = \begin{pmatrix} Var(L^*) & Cov(L^*, a^*) & Cov(L^*, b^*) \\ Cov(a^*, L^*) & Var(a^*) & Cov(a^*, b^*) \\ Cov(b^*, L^*) & Cov(b^*, a^*) & Var(b^*) \end{pmatrix} \quad (4.2)$$

Il est possible par la suite à partir de la matrice C d'extraire les coefficients de corrélation linéaire entre les variables :

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}} \in [-1, 1] \quad (4.3)$$

Voici les résultats obtenus sur les données en considérant l'ensemble de l'espace ICCLAB :

$$C = \begin{pmatrix} 2207 & 50 & 365 \\ 50 & 219 & 75 \\ 365 & 75 & 105 \end{pmatrix} \quad (4.4)$$

Tableau IX

Variance expliquée par les bandes du système ICCLAB sur tout l'espace

Bandes ICCLAB	Variance exprimée (Pourcentage)
L*	87 %
a*	9 %
b*	4 %

Tableau X

Coefficient de corrélation linéaire entre les bandes du système ICCLAB sur tout l'espace

Bandes ICCLAB	Coefficient de corrélation linéaire R [-1, 1]
L* - a*	0,07
L* - b*	0,76
a* - b*	0,50

On remarque que la bande L\* exprime la quasi-totalité de la variance de l'ensemble de la couleur canneberge (Tableau IX). On constate qu'il n'y a aucune corrélation linéaire entre L\* et a\* sur l'ensemble de l'espace (Tableau X).

Le nuage de point de notre base de données possède une forme de boomerang. Les couleurs rouges/ rouge foncées sont les couleurs pour lesquelles les fruits possèdent un Tacy élevé. Ces couleurs sont situées au-delà de la courbure du boomerang, et semblent former visuellement une droite dans l'espace ICCLAB. Une seconde étude sur les bandes du système ICCLAB se concentrera uniquement sur ces couleurs, en ne considérant la partie de l'espace ICCLAB délimitée par le plan  $L^* < 110$  (sommet de la courbure).

Tableau XI

Variance exprimée par les bandes du système ICCLAB pour  $L^* < 110$

Bandes ICCLAB	Variance exprimée (Pourcentage)
L*	66 %
a*	22 %
b*	12 %

Tableau XII  
 Coefficient de corrélation linéaire entre les bandes du système ICCLAB  
 pour  $L^* < 110$

Bandes ICCLAB	Coefficient de corrélation linéaire R [-1, 1]
$L^* - a^*$	0,97
$L^* - b^*$	0,96
$a^* - b^*$	0,99

On constate que sur cet espace restreint, la bande  $L^*$  possède une variance moindre que dans l'espace total, mais qui comptabilise tout de même les deux tiers de la variance du nuage de points (Tableau XI). Par contre, les coefficients de corrélation linéaire entre chacune des variables sont très élevés (Tableau XII), notamment entre  $a^*$  et  $b^*$ . Cela indique que dans cette partie de l'espace le nuage de points peut effectivement être modélisé par une droite. Cette zone du plan est particulièrement intéressante du fait de sa linéarité. Elle sera par la suite privilégiée pour la prédiction du Tacy.

#### 4.3.1 Analyse par composantes principales

Une analyse par composantes principales (ACP) [78] est ensuite appliquée sur l'ensemble de l'espace ICCLAB afin de nous fournir un nouvel espace dont les variables descriptives sont non corrélées entre elles.

L'ACP est un outil statistique servant à la réduction de la dimension et à l'interprétation des données. L'ACP est la recherche d'un nouvel espace orthogonal dans lequel les variables sont non corrélées entre elles. Pour une base de données possédant  $q$  observations de  $n$  caractéristiques chacune:  $x_i = [x_1, x_2, \dots, x_n]$  pour  $i = 1, 2, \dots, q$ , le vecteur moyen est d'abord calculé :

$$\bar{x} = \sum_{i=1}^q x_i \quad (4.5)$$

La matrice de covariance  $\mathbf{S}$  peut alors être calculée, avec  $T$  l'opérateur de transposée :

$$\mathbf{S} = \frac{1}{q-1} \sum_{i=1}^q (x_i - \bar{x})^T (x_i - \bar{x}) \quad (4.6)$$

Les coefficients de  $\mathbf{S}$  appartenant à la diagonale représentent les variances de la base de données suivant chacun des vecteurs de l'espace de représentation initiale. Tous les autres coefficients de la matrice  $s_{i,j}$ , avec  $i \neq j$ , mesurent la covariance entre la  $i^{\text{ème}}$  et la  $j^{\text{ème}}$  caractéristique des observations, indiquant leur niveau de corrélations. Plus la valeur absolue de l'élément  $s_{i,j}$  est élevée, plus la corrélation entre la  $i^{\text{ème}}$  et la  $j^{\text{ème}}$  caractéristique est élevée. A l'inverse, si la valeur absolue de l'élément  $s_{i,j}$  est faible, il y a peu de corrélations entre la  $i^{\text{ème}}$  et la  $j^{\text{ème}}$  caractéristique. La valeur 0 de  $s_{i,j}$  indique une absence totale de corrélation. La matrice  $\mathbf{S}$  est ainsi une matrice symétrique et carrée de dimension  $q \times q$ , qu'il est possible de la diagonaliser :

$$\mathbf{U}'\mathbf{S}\mathbf{U} = \mathbf{L} \quad (4.7)$$

$\mathbf{L}$  est une matrice diagonale et  $\mathbf{U}$  est une matrice orthonormale.  $\mathbf{L}$  représente la matrice de covariance dans l'espace des vecteurs propres. Le fait que la matrice  $\mathbf{L}$  soit diagonale prouve que les vecteurs propres sont bien non corrélés entre eux. Les termes diagonaux de  $\mathbf{L}$  sont les valeurs propres tandis que les vecteurs colonnes de  $\mathbf{U}$  sont les vecteurs propres. Les valeurs propres  $l_i$  sont trouvées en résolvant l'équation suivante :

$$\det(l_i\mathbf{I} - \mathbf{S}) = 0 \quad (4.8)$$

Où  $\mathbf{I}$  est la matrice identité de dimension  $q \times q$ . La résolution de l'équation précédente conduit à résoudre une équation de degré  $q$ , possédant  $q$  racines, les racines étant les valeurs propres. Les vecteurs propres sont ensuite déterminés à partir de leurs valeurs propres  $l_i$  :

$$(\mathbf{I} - \mathbf{S})\mathbf{A}_i = 0 \quad (4.9)$$

Puis :

$$\mathbf{u}_i = \frac{\mathbf{A}_i}{\sqrt{\mathbf{A}_i^T \mathbf{A}_i}} \quad (4.10)$$

Géométriquement, l'ACP effectue une translation du repère originale vers la valeur moyenne du nuage de points, suivi d'une rotation, dans laquelle l'orientation de chacun des vecteurs propres est fixée successivement, de manière à ce que le premier vecteur propre exprime la plus grande variabilité possible, le deuxième vecteur propre la plus grande variabilité suivant les directions restantes et ainsi de suite jusqu'au dernier vecteur propre. La réduction de la dimension s'effectue en ne conservant pas les derniers vecteurs propres, qui possèdent en général une faible variabilité qui correspond essentiellement à du bruit.

Dans le cas de la base de données de la couleur canneberges, les 3 composantes principales possèdent respectivement une variabilité de 90%, 9% et 1% (Tableau XIII). La première composante principale représente une très grande partie de la variabilité des données, tandis que la troisième composante principale n'est pas significative. L'espace de représentation des données sera alors réduit à deux dimensions, en ne conservant que les deux premières composantes principales. Les données possèdent dans ce nouveau repère la forme d'un boomerang (Figure 37).

Tableau XIII

Pourcentage de la variabilité totale exprimée par chaque composante principale

Composante principale	Variabilité exprimée (Pourcentage)
première composante	90%
deuxième composante	9%
troisième composante	1%

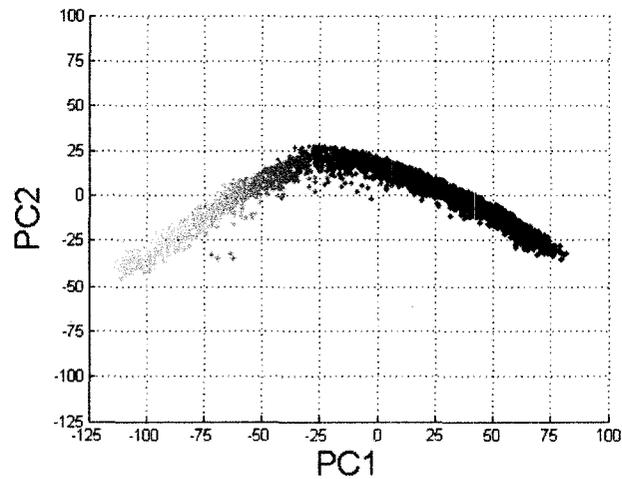


Figure 37 Couleur canneberge exprimée suivant les deux premiers vecteurs propres de l'ACP

#### 4.4 Comparaison des espaces ICCLAB et de l'ACP

L'ACP a permis de créer un nouvel espace dans lequel les données sont non corrélées, néanmoins, il reste intéressant d'étudier de plus près la relation qui existe entre les variables de l'espace ICCLAB et celles de l'ACP.

Le passage du repère ICCLAB vers le repère de l'ACP s'effectue ainsi par une translation du repère ICCLAB vers le centre du nuage de points, puis par une transformation isométrique  $W$  :

$$X_{ACP}^T = \left( X_{ICCLAB} - \bar{X}_{ICCLAB} \right)^T \cdot W^T \quad (4.11)$$

Nous étudierons par la suite la relation angulaire entre ces deux espaces de représentation grâce à la matrice  $W$  :

$$W = \begin{pmatrix} -0,98 & -0,03 & -0,16 \\ -0,08 & 0,94 & 0,30 \\ -0,14 & -0,31 & 0,93 \end{pmatrix} \quad (4.12)$$

W peut se réécrire plus formellement comme la composition d'une symétrie centrale à l'origine du repère de l'ACP, avec une rotation de l'espace ICCLAB :

$$W = -I_3 \cdot R \quad (4.13)$$

La symétrie centrale ne fait qu'inverser les signes de l'espace ICCLAB, donc la matrice R de rotation porte seule l'information sur les angles entre les directions des vecteurs des deux espaces. En se servant des angles d'Euler<sup>1</sup>, la matrice R peut se définir comme une composition d'une rotation de l'angle  $\gamma$  autour de l'axe L\*, suivi d'une rotation de l'angle  $\beta$  autour de l'axe a\*, suivi d'une rotation de l'angle  $\alpha$  autour de l'axe b\* :

$$R_{L^*a^*b^*}(\gamma, \beta, \alpha) = \begin{pmatrix} c\alpha \cdot c\beta & c\alpha \cdot s\beta \cdot s\gamma - s\alpha \cdot c\gamma & c\alpha \cdot s\beta \cdot c\gamma - s\alpha \cdot s\gamma \\ s\alpha \cdot c\beta & s\alpha \cdot s\beta \cdot s\gamma + c\alpha \cdot c\gamma & s\alpha \cdot s\beta \cdot c\gamma - c\alpha \cdot s\gamma \\ -s\beta & c\beta \cdot s\gamma & c\beta \cdot c\gamma \end{pmatrix} \quad (4.14)$$

$$R_{L^*a^*b^*}(\gamma, \beta, \alpha) = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} = \begin{pmatrix} 0,98 & 0,03 & 0,16 \\ 0,08 & -0,94 & -0,30 \\ 0,14 & 0,31 & -0,93 \end{pmatrix} \quad (4.15)$$

---

<sup>1</sup> La notation suivante est utilisée pour simplifier l'écriture de la matrice :  $c\alpha$  désigne  $\cos(\alpha)$  et  $s\alpha$  désigne  $\sin(\alpha)$

$$\begin{aligned}
\beta &= A \tan 2\left(-r_{31}, \sqrt{r_{11}^2 + r_{21}^2}\right) = -8,5^\circ \\
\alpha &= A \tan 2\left(r_{21}/c\beta, r_{11}/c\beta\right) = 5^\circ \\
\gamma &= A \tan 2\left(r_{32}/c\beta, r_{33}/c\beta\right) = 161^\circ
\end{aligned} \tag{4.16}$$

On constate que la rotation d'angle  $\gamma$  autour de  $L^*$  est de l'ordre de  $20^\circ$  (en terme de direction), et que les rotations autour des deux autres axes ( $\alpha$  et  $\beta$ ) sont pratiquement négligeable. La rotation n'affecte donc que les variables  $a^*$  et  $b^*$ . La matrice de rotation  $R$  exprime également la projection de l'espace ICCLAB sur celui de l'ACP :

$$R_{L^*a^*b^*}(\gamma, \beta, \alpha) = \begin{pmatrix} \cos(\widehat{L^*, PC1}) & \cos(\widehat{a^*, PC1}) & \cos(\widehat{b^*, PC1}) \\ \cos(\widehat{L^*, PC2}) & \cos(\widehat{a^*, PC2}) & \cos(\widehat{b^*, PC2}) \\ \cos(\widehat{L^*, PC3}) & \cos(\widehat{a^*, PC3}) & \cos(\widehat{b^*, PC3}) \end{pmatrix} \tag{4.17}$$

$$R_{L^*a^*b^*}(\gamma, \beta, \alpha) = \begin{pmatrix} \cos(10^\circ) & \cos(88^\circ) & \cos(80^\circ) \\ \cos(85^\circ) & \cos(161^\circ) & \cos(108^\circ) \\ \cos(81^\circ) & \cos(71^\circ) & \cos(159^\circ) \end{pmatrix} \tag{4.18}$$

L'angle entre  $L^*$  et  $PC1$  est très petit ( $10^\circ$ ), il apparaît donc que  $L^*$  est très similaire à  $PC1$  et qu'il contribue très peu à  $PC2$  et  $PC3$ . De plus, l'angle entre  $a^*$  et  $PC1$ , et  $b^*$  et  $PC1$  sont proches de  $90^\circ$ , et montrent que ces deux bandes contribuent très peu à  $PC1$ . Il est donc convenable de considérer que la rotation du repère ICCLAB vers celui de l'ACP se résume à une rotation autour de l'axe  $L^*$  uniquement (voir Figure 38 et Figure 39).

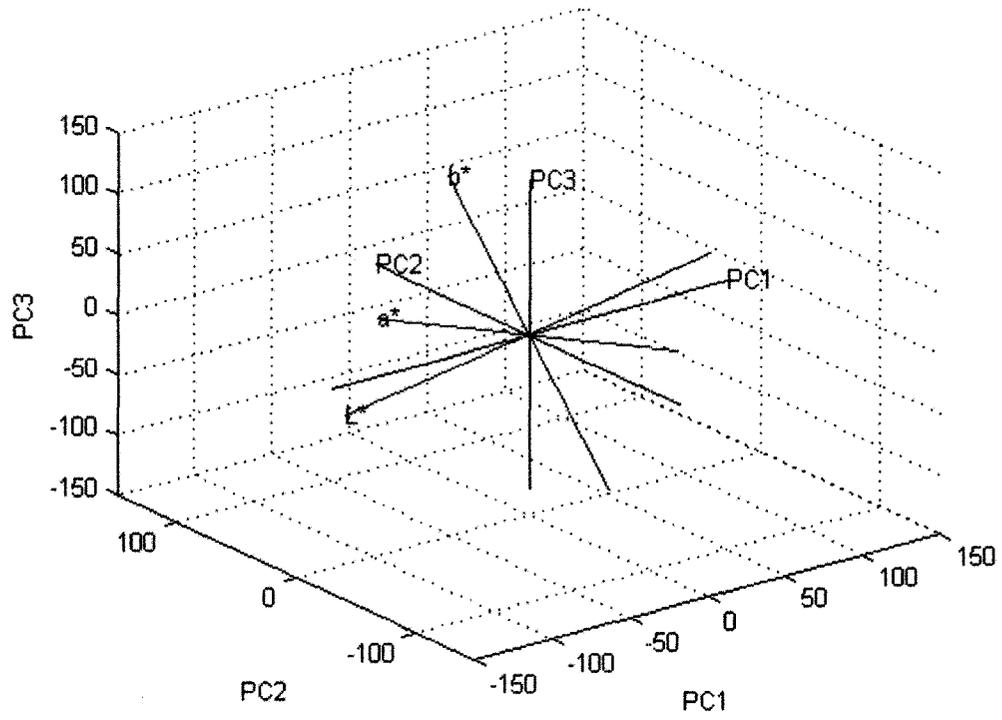


Figure 38 Diagramme de double projection sans les données

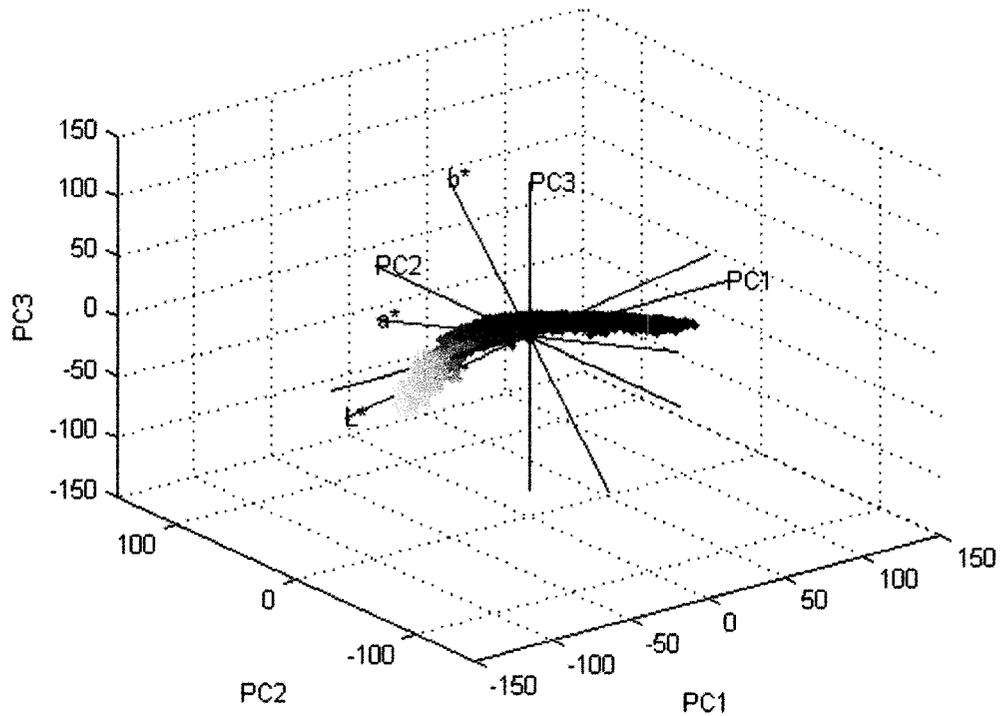


Figure 39 Diagramme de double projection avec les données

#### 4.5 Modélisation de la couleur canneberge

Dans le repère de l'ACP après projection sur la dernière composante principale, la couleur canneberge possède une allure de boomerang. Le nuage de points sera modélisé par une fonction, afin de faciliter son interprétation, au moyen d'une régression linéaire.

La régression linéaire permet de trouver le jeu de coefficients qui permet à un modèle de reproduire le plus fidèlement des données expérimentales. Les modèles linéaires correspondent aux fonctions de forme polynomiale. Les données  $y$  sont considérées comme des variables déterminées, mais auxquelles s'ajoute une erreur de mesure  $\epsilon$ , qui est supposée suivre une distribution normale de moyenne nulle. Pour obtenir l'estimation des coefficients d'un modèle linéaire donné, la méthode des moindres carrés

est utilisée afin de minimiser la somme au carré des résidus. Le résidu du  $i^{\text{ème}}$  point de donnée  $r_i$  est défini par la différence entre la valeur réelle  $y_i$  du point de donnée et sa valeur estimée par le modèle  $\hat{y}_i$ . Il s'apparente à l'erreur associée à la prédiction de la donnée :

$$r_i = y_i - \hat{y}_i \quad (4.19)$$

La somme des carrés des erreurs résiduels  $S$  est la suivante :

$$S = \sum_{i=1}^n r_i^2 \quad (4.20)$$

Où  $n$  est le nombre de points des données expérimentales. Afin de trouver les coefficients minimisant  $S$ , la méthode des moindres carrés dérive partiellement  $S$  en fonction de chaque coefficient, afin de trouver la valeur annulant la dérivée partielle.

Sous forme matricielle, le modèle prend la forme suivante :

$$y = X\beta + \varepsilon$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1^{m-1} & \cdots & x_1 & 1 \\ x_2^{m-1} & \cdots & x_2 & 1 \\ \vdots & \vdots & \vdots & 1 \\ x_n^{m-1} & \cdots & x_n & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} + \begin{bmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ \vdots \\ y_n - \hat{y}_n \end{bmatrix} \quad (4.21)$$

Où  $y$  est un vecteur colonne de dimension  $n$ ,  $\beta$  est un vecteur colonne de dimension  $m$  des coefficients du modèle,  $X$  est une matrice de dimension  $n \times m$  dont les colonnes représentent les données d'entrées du modèle, et  $\varepsilon$  un vecteur colonne de dimension  $n$  des erreurs résiduels. La solution de la méthode des moindres carrée est alors un vecteur  $b$ , représentant l'estimation des coefficients inconnus de  $\beta$ :

$$b = (X^T X)^{-1} X^T y \quad (4.22)$$

Les valeurs estimées du modèle peuvent être obtenues de la manière suivante :

$$\hat{y} = Xb = Hy \quad (4.23)$$

Avec :

$$H = X(X^T X)^{-1} X^T \quad (4.24)$$

Les résidus sont obtenus ainsi :

$$r = y - \hat{y} = (I - H)y \quad (4.25)$$

Néanmoins, la méthode des moindres carrés possède un inconvénient majeur, elle est très sensible aux observations aberrantes. En effet, la mise au carré des erreurs résiduelles amplifie l'effet des données aberrantes lors de la recherche des coefficients, conduisant souvent à un modèle moins juste. Une méthode de moindres carrés robuste permet d'améliorer les résultats, en attribuant un poids à chaque résidu. Un poids proche de 1 sera attribué aux données proches de la courbe de régression, tandis que les données aberrantes, qui sont éloignées de la courbe de régression, se verront attribuer un poids proche de 0.

$$S = \sum_{i=1}^n w_i r_i^2 \quad (4.26)$$

Les résidus sont supposés corrélés avec les données, ils seront normalisés par la loi de Student de  $(n - m - 1)$  degrés de liberté, afin qu'ils possèdent la même variance en tous les points de  $y$  :

$$r_{norm} = \frac{r_i}{\hat{\sigma}_i \sqrt{1 - h_i}} \quad (4.27)$$

$$\hat{\sigma}_i = \frac{|r|^2}{n - m - 1} - \frac{r_i^2}{(n - m - 1)(1 - h_i)} \quad (4.28)$$

où  $h_i$  est le  $i^{\text{ème}}$  terme de la diagonale de la matrice  $H$ , et  $\hat{\sigma}_i$  est l'estimé de la variance des résidus en excluant le  $i^{\text{ème}}$  résidu. Les résidus sont ensuite standardisés par une

constante  $K$  permettant de contrôler l'importance des données aberrantes dans le modèle :

$$u = \frac{r_{norm}}{K} \quad (4.29)$$

Les poids sont ensuite déterminés de la manière suivante :

$$w_i = \begin{cases} (1 - u_i^2)^2 & |u_i| < 1 \\ 0 & |u_i| \geq 1 \end{cases} \quad (4.30)$$

Le procédé est itératif, les poids sont recalculés et les coefficients du modèle réévalués successivement, jusqu'à obtenir la convergence des coefficients.

Pour la modélisation de la couleur canneberges, les modèles polynomiaux jusqu'au degré 7 ont été essayés (Tableau XIV), afin d'obtenir le coefficient de détermination le plus élevé. La méthode des moindres carrés et la méthode des moindres carrés robuste ont toutes les deux été testées.

Les valeurs de  $R^2$  pour tous les modèles testés avec une régression linéaire normale sont très bonnes ( $R^2 > 0,90$ ). Avec la régression linéaire robuste, les résultats sont encore légèrement améliorés pour tous les modèles. Les données aberrantes ont donc une influence assez faible sur le modèle. Finalement, le modèle polynomial du 7<sup>ème</sup> ordre obtenu avec une régression linéaire robuste sera conservé comme modèle de la couleur canneberge, car il possède le coefficient de détermination le plus élevé. L'étude de la bande de confiance à 95% de l'estimation des coefficients montre que ces derniers sont estimés avec une bonne précision (Tableau XV). Ce résultat est confirmé par la visualisation de la bande de confiance à 95% (Figure 41) et à 99% (Figure 42) de l'estimation des coefficients et des prédictions pour le modèle retenu.

Tableau XIV  
Coefficients de détermination des modèles

	qualité du modèle ( R <sup>2</sup> )	
	non robuste	robuste
quadratique	0,920	0,926
cubique	0,923	0,933
4ème ordre	0,939	0,947
5ème ordre	0,942	0,951
6ème ordre	0,943	0,952
7ème ordre	0,944	0,952

Tableau XV  
Coefficients du modèle polynomial du 7<sup>ème</sup> ordre  
(avec marge de confiance à 95%)

modèle polynomial du 7 <sup>ème</sup> ordre		
$f(x) = p_1 \cdot x^7 + p_2 \cdot x^6 + p_3 \cdot x^5 + p_4 \cdot x^4 + p_5 \cdot x^3 + p_6 \cdot x^2 + p_7 \cdot x + p_8$		
coefficients	non robuste	robuste
p <sub>1</sub>	5,79e-13 (3,68e-13; 7,91e-13)	5,89e-13 (3,95e-13; 7,84e-13)
p <sub>2</sub>	2,08e-11 (-3,99e-12; 4,56e-11)	2,30e-11 (2,19e-13; 4,58e-11)
p <sub>3</sub>	-1,23e-8 (-1,45e-8; -1,01e-8)	-1,25e-8 (-1,45e-8; -1,05e-8)
p <sub>4</sub>	-6,52e-8 (-2,79e-7; 1,49e-7)	-1,02e-7 (-2,99e-7; 9,4e-8)
p <sub>5</sub>	9,42e-5 (8,49e-5; 0,000103)	9,61e-5 (8,76e-5; 0,000104)
p <sub>6</sub>	-0,0068 (-0,00735; -0,00636)	-0,00669 (-0,00715; -0,00623)
p <sub>7</sub>	-0,354 (-0,367; -0,341)	-0,363 (-0,375; -0,351)
p <sub>8</sub>	15,86 (15,61; 16,11)	16,08 (15,85; 16,31)

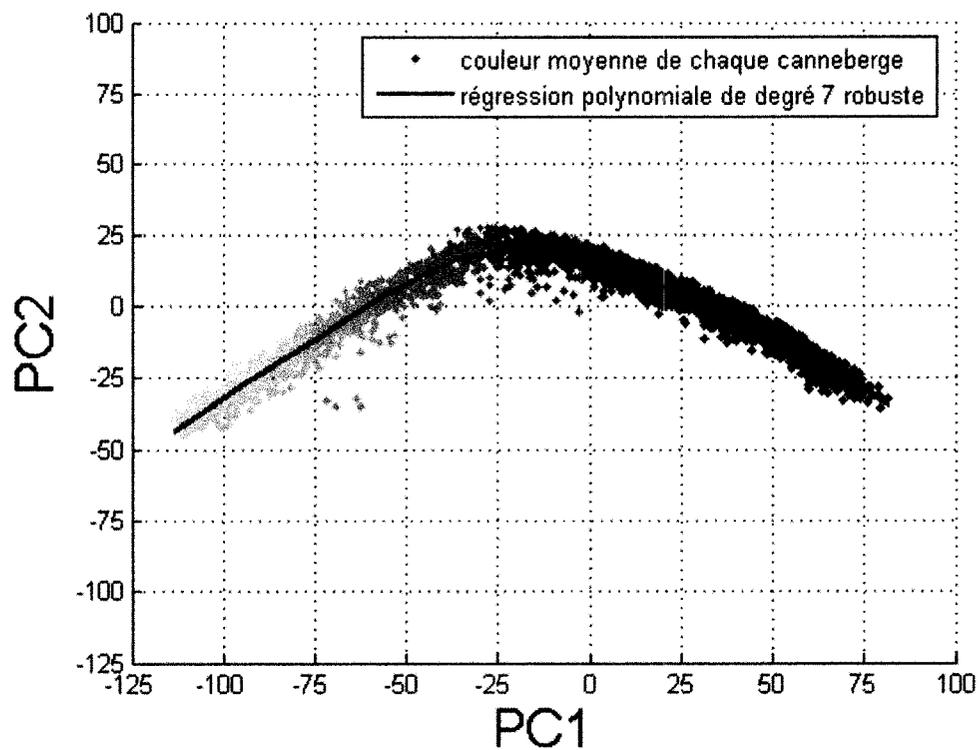


Figure 40 Régression de la couleur moyenne de chaque canneberge exprimée suivant les deux premières composantes principales de l'ACP ( $R^2 = 0,95$ )

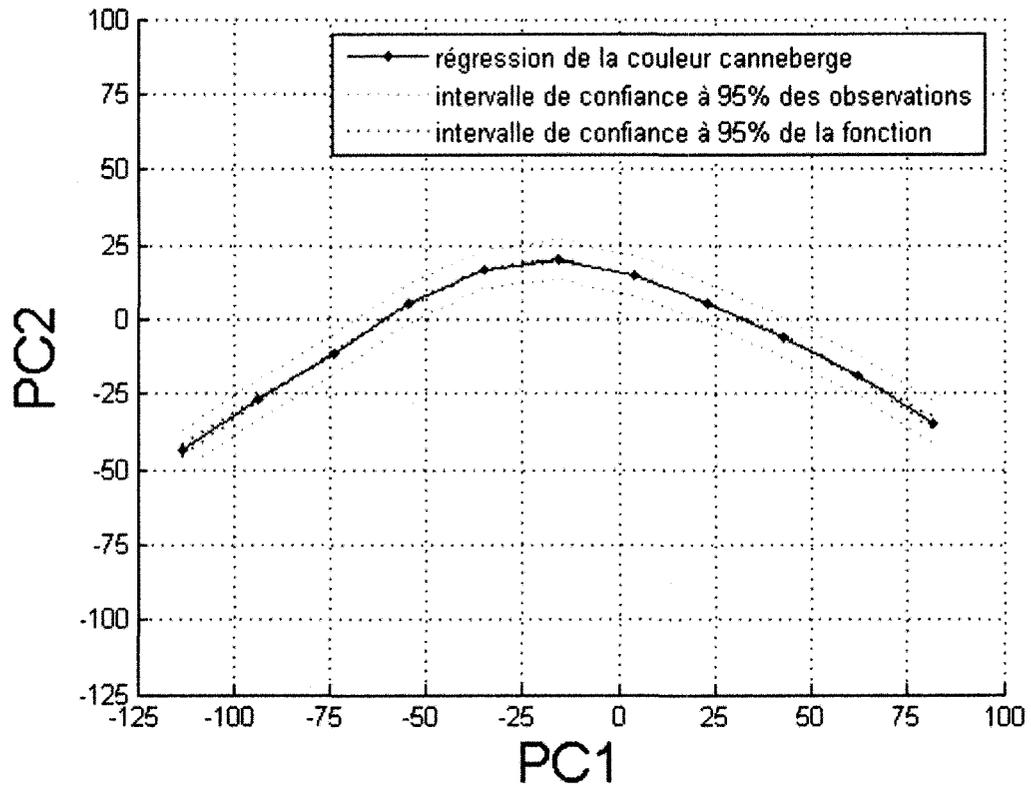


Figure 41 Intervalle de confiance à 95 % de la fonction et des observations de la régression de la couleur canneberge moyenne par fruit

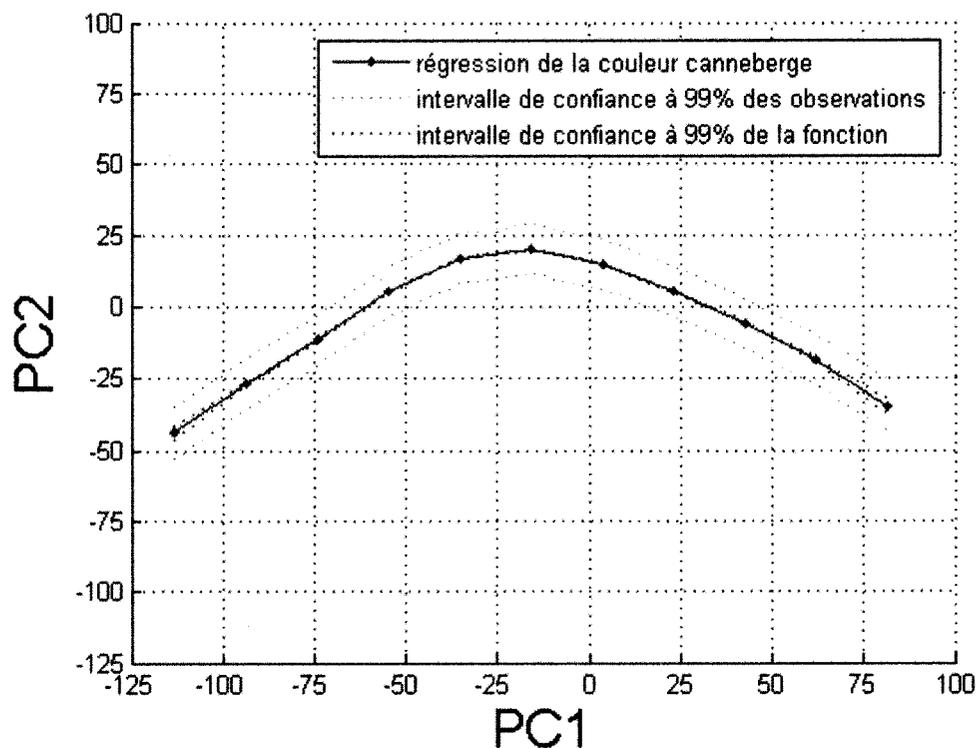


Figure 42 Intervalle de confiance à 99% de la fonction et des observations de la régression de la couleur canneberge moyenne par fruit

La même étude est effectuée cette fois-ci sur la couleur moyenne d'un lot standard par photo. Un coefficient de détermination très élevé de 0,99 est obtenu par une régression polynomiale quadratique. Ce modèle est donc retenu (Tableau XVI) dans le cas de la couleur moyenne d'un lot standard.

Tableau XVI  
Coefficients du modèle polynomial quadratique  
(avec marge de confiance à 95%)

modèle polynomial quadratique	
$f(x) = p_1 * x^2 + p_2 * x + p_3$	
coefficients	non robuste
$p_1$	-0,00632 (-0,00647; -0,00617)
$p_2$	-0,157 (-0,164; -0,150)
$p_3$	12,06 (11,64; 12,47)

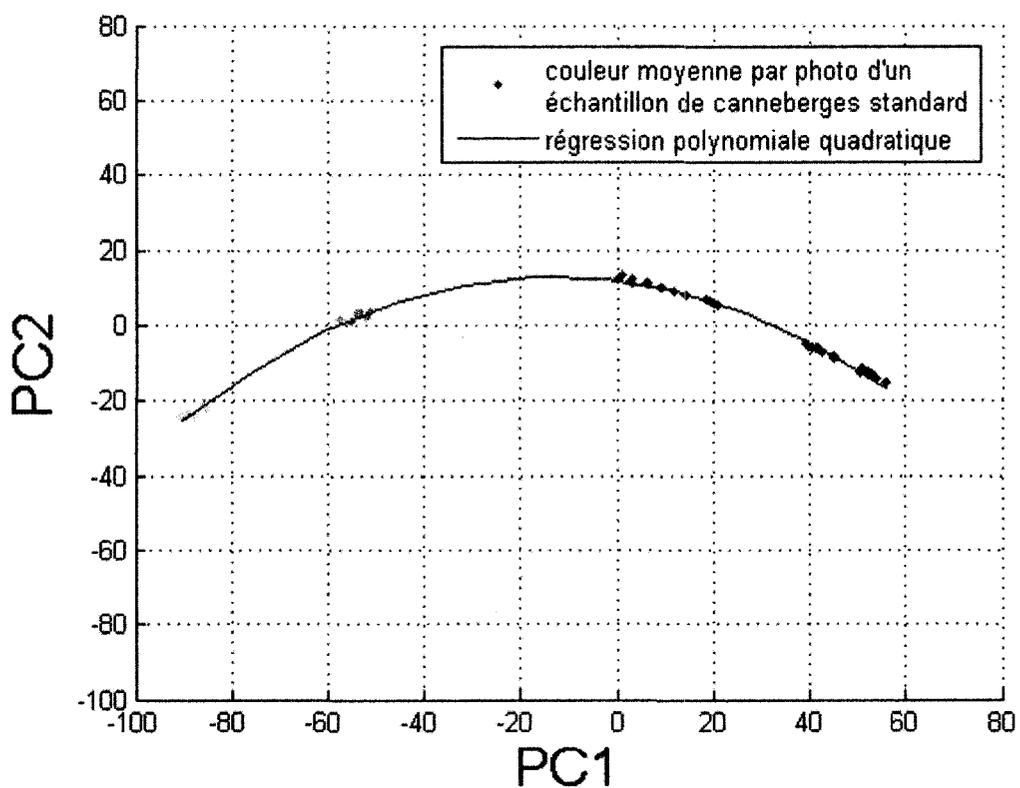


Figure 43 Régression de la couleur canneberge moyenne par photo exprimée suivant les deux premières composantes principales de l'ACP ( $R^2 = 0,99$ )

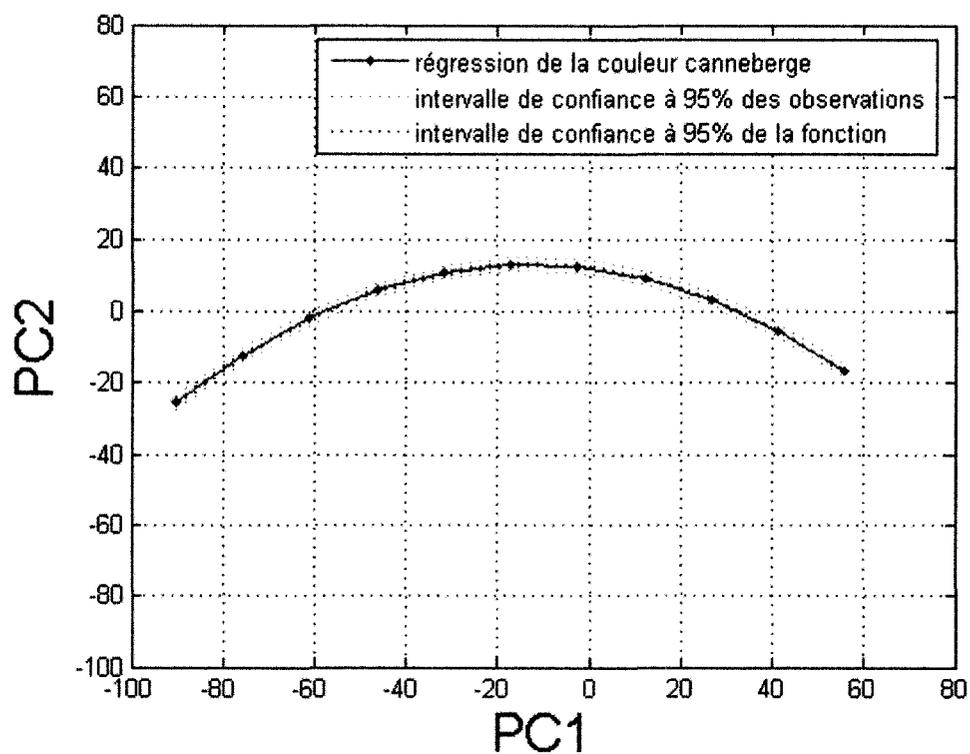


Figure 44 Intervalle de confiance à 95% de la fonction et des observations de la régression de la couleur canneberge moyenne par photo

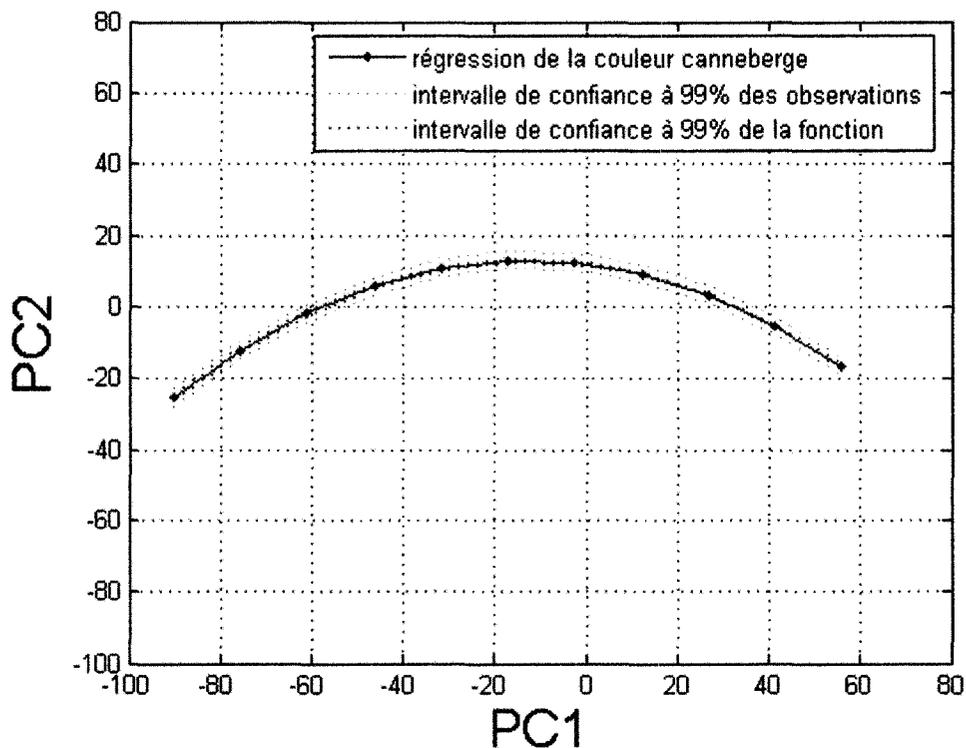


Figure 45 Intervalle de confiance à 99% de la fonction et des observations de la régression de la couleur canneberge moyenne par photo

La modélisation de la relation entre PC1 et PC2 permet de définir chaque coordonnée de la couleur canneberge par une seule variable, au lieu de 2. La modélisation a été effectuée sur les couleurs moyennes de chaque fruit standard (Figure 40) et aussi sur la couleur moyenne d'un lot standard par photo (Figure 43). Dans le dernier cas, la justesse de la modélisation est meilleure ( $R^2 = 0,99$ ), mais par contre la plage de modélisation est restreinte, car la moyenne de la couleur canneberge est effectuée sur une population de pixels plus importante, réduisant ainsi l'écart type entre les individus. Les intervalles de confiance à 95% (Figure 41 et Figure 44) et à 99% (Figure 42 et Figure 45) montrent que la modélisation de la couleur canneberge est obtenue avec une grande certitude, même jusqu'à 99%, car on contrôle très bien la variabilité.

#### 4.5.1 Échelle couleur canneberge

L'échelle couleur canneberge est établie sous sa forme continue à partir de l'axe PC1, en tenant compte de la relation entre PC1 et PC2 (fonction polynomiale du 7<sup>ème</sup> ordre sur la couleur moyenne des canneberges), puis en effectuant la transformation inverse de l'ACP (en supposant  $PC3 = 0$ ) pour retourner dans l'espace ICCLAB.



Figure 46 Échelle couleur continue

Néanmoins, sous sa forme continue, l'échelle couleur est difficile à manipuler par des experts humains, qui voudraient spécifier la couleur de fruit qu'ils désirent. Il est donc nécessaire de construire cette échelle avec un nombre de couleur limité, mais qui serait sélectionné par des experts humains. Cette nouvelle échelle sera reproduite sur support papier, et ne pourra donc être validée visuellement que sur support papier, à cause des problèmes liés à la reproduction couleur. Voici deux prototypes de l'échelle couleur "discontinue" :



Figure 47 Échelle couleur discontinue, écart couleur = 10, fond:  $L^* = 100$ ,  $a^* = 0$ ,  $b^* = 0$  (CIELAB)



Figure 48 Échelle couleur discontinue, écart couleur = 10, fond:  $L^* = 50, a^* = 0, b^* = 0$  (CIELAB)

Chacun des carrés de couleur composant ces échelles est composé d'une couleur unique. Ces couleurs ont été sélectionnées à partir de l'échelle continue, de façon à ce que la distance couleur (distance euclidienne) entre deux carrés successifs soit de 10 unités CIELAB. Chacune de ces échelles couvre la totalité de la plage de l'échelle continue. Les couleurs contenues dans ces deux échelles sont identiques, seule la couleur du fond les différencie. Dans la première, la couleur du fond entourant les couleurs est le blanc le plus lumineux ( $L^* = 100$  dans CIELAB), dans la deuxième, le fond est d'une couleur grise intermédiaire ( $L^* = 50$  dans CIELAB), qui est la couleur de fond recommandée par la CIE pour comparer des échantillons de couleurs entre eux. Il y a 14 couleurs dans chacune des échelles, et une réduction de ce nombre devra être effectuée en accord avec des experts humains, afin de ne conserver que les couleurs les plus représentatives (4 à 6 couleurs) de la canneberge.

#### **4.6 Conclusion**

La base de données de lots de canneberges standard créés lors de la saison de cueillette 2005 a servi à la modélisation de la couleur canneberge. La couleur canneberge a été dans un premier temps étudiée dans l'espace colorimétrique ICCLAB, il apparaît que dans cet espace à 3 dimensions la couleur canneberge est difficile à interpréter, sauf dans la zone de couleur rouge où les trois variables  $L^*$ ,  $a^*$  et  $b^*$  sont très fortement corrélées. L'ACP est ensuite utilisée pour réduire le nombre de dimensions de l'espace de représentation. Il apparaît alors que l'ACP peut se résumer à une rotation autour de l'axe  $L^*$ , et que les axes PC1 et  $L^*$  sont identiques. Il apparaît également que les deux premières composantes principales expriment à elles seules la quasi-totalité de la variabilité du nuage de points. La projection des données suivant les deux premières composantes principales donne l'allure d'un boomerang aux couleurs canneberges. La couleur canneberge a ensuite été modélisée par une régression linéaire avec un coefficient de détermination très élevé. Cette modélisation a ensuite servi à générer l'échelle couleur canneberge dans ses formes continues et discontinues. L'échelle

couleur canneberge est destinée à présenter sous format papier la gamme de couleur des canneberges réelles pour des experts humains. Ces experts devront donc décider de la forme de l'échelle convenant le mieux à leurs besoins. La compréhension du comportement de la couleur canneberge dans l'espace ICCLAB et celui de l'ACP facilitera la prédiction du Tacy en fonction de la couleur des fruits.

## CHAPITRE 5

### Prédiction du Tacy

#### 5.1 Introduction

Dans cette section, le modèle prédictif du Tacy à partir des photos des échantillons de canneberges sera établi. L'anthocyane est le pigment donnant la couleur rouge à la canneberge, et il se situe directement sous l'épiderme du fruit. Le modèle prédictif tiendra ainsi compte de la couleur des fruits et de leurs dimensions. Dans un premier temps, un modèle permettant d'estimer la surface volumique de la canneberge à partir des photos sera établi. Ensuite différents modèles prédictifs du Tacy seront créés et validés à partir de la base de données de la saison 2005 afin de sélectionner le plus performant. Le modèle prédictif retenu pour la saison 2005 sera de nouveau validé sur la base de données de la saison 2006.

#### 5.2 Modélisation de la surface de la canneberge

L'anthocyane, le pigment caractéristique de la canneberge, se situe essentiellement sous l'épiderme du fruit. Il est donc important d'évaluer les dimensions de la surface externe des fruits afin de connaître la quantité d'anthocyane qu'ils renferment. Pour cela, un modèle volumique de la canneberge sera établi, en comparant les modèles sphérique (1), ellipsoïdal (2) et sphéroïdal (3).

##### 5.2.1 Modèle sphérique (1)

Une sphère peut être définie comme la surface formée par la rotation d'un cercle autour de son diamètre ( $2R$ ). L'aire d'une sphère est donnée par la formule suivante :

$$Aire_{sphère} = 4\Pi R^2 \quad (4.31)$$

L'aire d'un disque issue de la projection de la sphère précédente est le suivant:

$$Aire_{disque} = \Pi R^2 \quad (4.32)$$

Ainsi, l'aire d'une sphère de rayon R est proportionnelle à l'aire d'un disque de rayon R.

### 5.2.2 Modèle ellipsoïdal (2)

L'ellipsoïde est une surface du second degré de l'espace euclidien à trois dimensions, elle admet un centre et au moins trois plans de symétrie. L'intersection d'une ellipse avec un plan est une ellipse. L'équation d'un ellipsoïde est de la forme suivante :

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \quad (4.33)$$

avec a, b et c des paramètres strictement positifs et égaux aux longueurs des demi-axes de l'ellipsoïde. La projection d'un ellipsoïde parallèlement à l'axe z forme une ellipse ayant pour aire :

$$Aire_{ellipse} = \Pi ab \quad (4.34)$$

Dans le cas particulier où a=b=c, la surface d'une sphère de rayon a est retrouvée. Dans le cas où seulement deux de ces paramètres sont égaux, on parle d'ellipsoïde de révolution ou bien encore de sphéroïde.

### 5.2.3 Modèle sphéroïdal (3)

Un sphéroïde est un cas particulier de l'ellipsoïde dans lequel deux des trois axes de l'ellipsoïde sont égaux entre eux. Un sphéroïde peut être construit par la rotation d'une ellipse autour de l'un de ses axes et est défini par l'équation suivante :

$$\frac{x^2 + y^2}{a^2} + \frac{z^2}{c^2} = 1 \quad (4.35)$$

Si  $a > c$ , alors on parle d'un sphéroïde oblate (aplatis tel un galet), si  $a < c$ , alors on parle d'un sphéroïde prolata (allongé tel un ballon de rugby). Les équations de l'aire de ces deux sphéroïdes sont les suivantes :

$$Aire_{sphéroïde\_prolate} = 2\pi a \left( a + c \frac{\arcsin(e)}{e} \right) \quad (4.36)$$

$$e_{prolate} = \sqrt{1 - \frac{a^2}{c^2}} \quad (4.37)$$

$$Aire_{sphéroïde\_oblate} = 2\pi a^2 + \pi \frac{c^2}{e} \ln \left( \frac{1+e}{1-e} \right) \quad (4.38)$$

$$e_{oblate} = \sqrt{1 - \frac{c^2}{a^2}} \quad (4.39)$$

### 5.2.4 Caractérisation de la canneberge dans les photos par ses informations morphologiques

Dans les photos, l'information disponible pour chaque fruit est la projection de leurs volumes dans le plan focal de la caméra. La surface résultante (projetée) est caractérisée par son nombre de pixels (aire de la surface) et peut être modélisée par certaines figures géométriques. Dans le cas des canneberges, la surface projetée peut être approximée par :

a. Soit un disque de rayon R (disque de Waddel):

$$R_{disque} = \sqrt{\frac{Aire_{surface}}{\Pi}} \quad (4.40)$$

b. Soit une ellipse ayant une inertie équivalente, de demi-grand axe a et de demi-petit axe b calculé à partir de ses moments d'inertie [79]:

$$a = \sqrt{\frac{2 \left[ \mu_{20} + \mu_{02} + \sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2} \right]}{\mu_{00}}} \quad (4.41)$$

$$b = \sqrt{\frac{2 \left[ \mu_{20} + \mu_{02} - \sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2} \right]}{\mu_{00}}} \quad (4.42)$$

Ces deux figures géométriques seront considérées comme les projections du volume des canneberges et serviront de base aux différents modèles 3D.

Le modèle 3D ellipsoïdal (2) n'est pas applicable, car pratiquement aucune information sur la hauteur du fruit n'est disponible dans les photos. Les deux modèles se rapprochant le plus de la forme de la canneberge sont les modèles sphérique (1) et sphéroïde prolata (3). Afin de caractériser la différence entre ces deux modèles, l'aire d'un sphéroïde prolata (3) sera comparée avec celle d'une sphère (1) ayant la même aire projetée, en fonction de l'élongation.

L'élongation est une mesure permettant de caractériser la forme d'une ellipse, prenant des valeurs égales ou supérieures à 1. Ainsi, un cercle possède une élongation de 1 (modèle sphérique), tandis qu'une ellipse très allongée possédera une valeur plus élevée (modèle sphéroïdal prolata).

$$élongation = \frac{axe_{majeur}}{axe_{mineur}} \quad (4.43)$$

La figure suivante montre l'erreur (en pourcentage) de l'approximation de l'aire d'un sphéroïde prolata par l'aire d'une sphère possédant la même aire projetée (nombre de pixels dans l'image), en fonction de l'élongation :

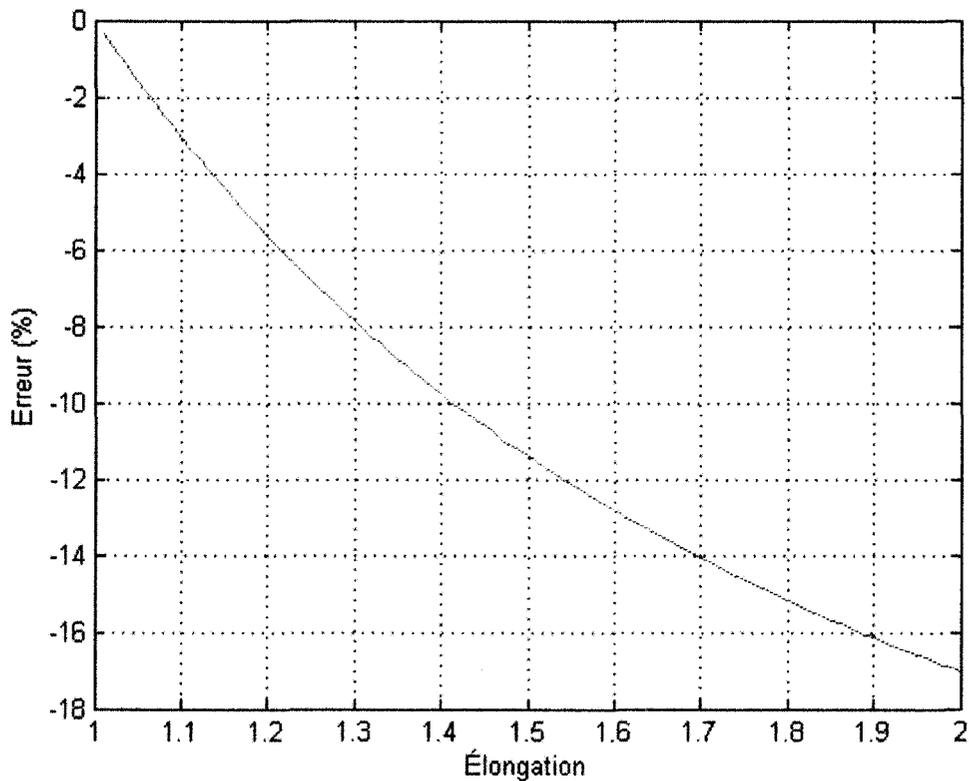


Figure 49 Erreur entre le modèle sphérique (1) et sphéroïdal (3) en fonction de l'élongation

$$erreur = \frac{aire\_sphéroïde - aire\_sphère}{aire\_sphéroïde} * 100 \quad (4.44)$$

On remarque que l'aire d'une sphère est plus grande que l'aire d'un sphéroïde prolata, quelle que soit l'élongation. Plus l'élongation augmente, plus l'erreur augmente. Il reste à estimer l'élongation des canneberges photographiées.

L'histogramme suivant montre la distribution de l'ensemble des 12000 fruits composant la base de données 2005 (standard + test) en fonction de leur élongation :

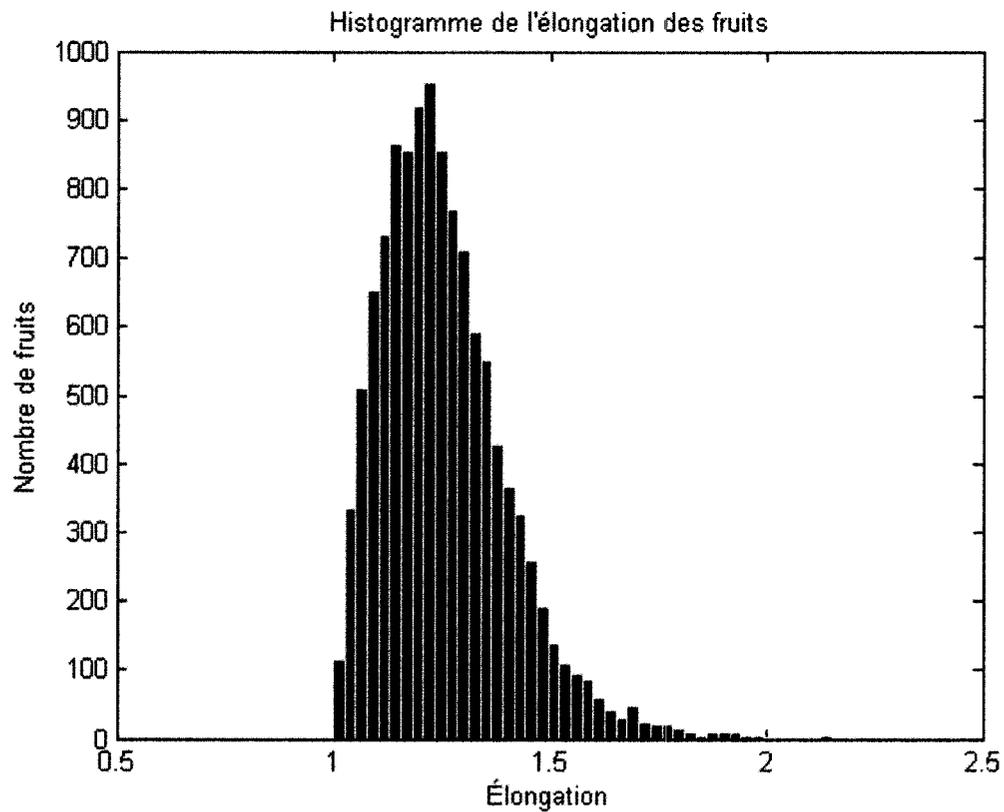


Figure 50 Histogramme de l'élongation des canneberges

Élongation :

- a. Valeur moyenne : 1,26;
- b. Déviation standard : 0,16.

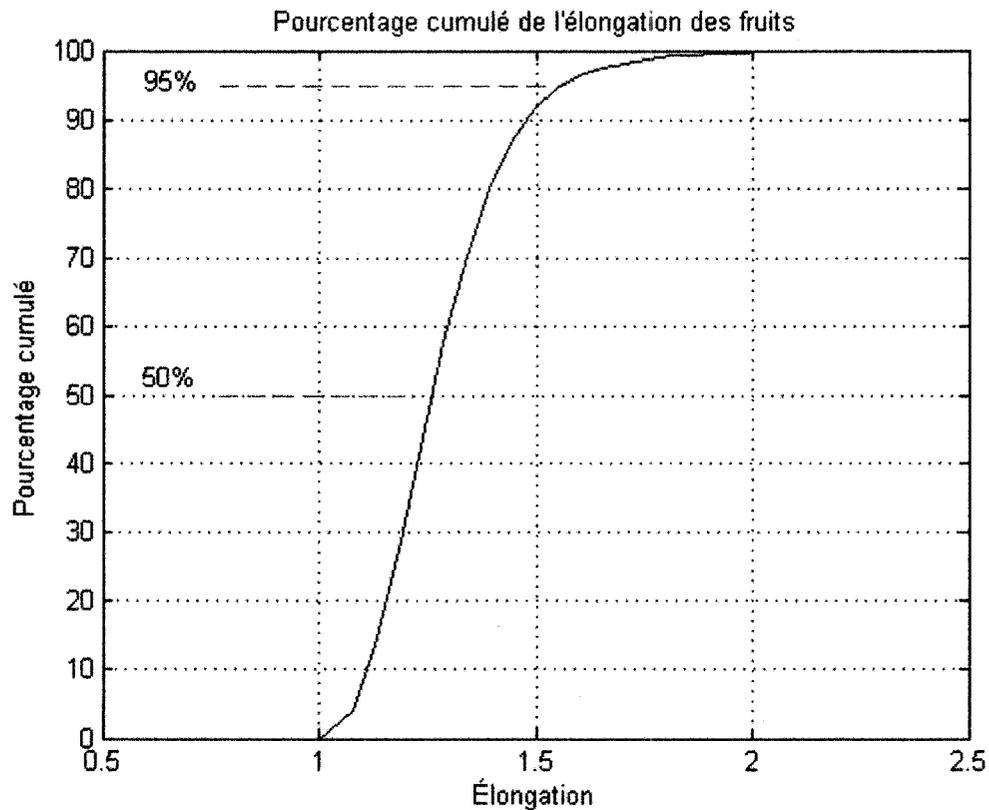


Figure 51 Pourcentage cumulé de l'élongation des fruits

En observant l'histogramme (Figure 50) ainsi que le pourcentage cumulé (Figure 51) de l'élongation des canneberges, il est possible de remarquer que 95 % des fruits possèdent une élongation inférieure à 1,6. Les canneberges possèdent ainsi une surface projetée elliptique, ayant une faible élongation. L'erreur d'approximation de la surface du modèle sphéroïdal prolate par la surface du modèle sphérique peut être considérée comme inférieure à 12 % pour 95 % des fruits (Figure 49). Les deux modèles sont relativement similaires bien que le sphéroïdal prolate paraisse meilleur comparativement au modèle sphérique pour estimer l'aire des fruits de canneberge.

Des mesures directes sur quelques échantillons de fruits indiquent que la forme d'une canneberge se rapproche de celle d'un sphéroïde prolate. Néanmoins, la détermination

de l'aire du modèle sphéroïdal nécessite quelques calculs, tandis que l'aire du modèle sphérique s'estime directement à partir du nombre de pixels de la canneberge. De plus, l'erreur entre les deux modèles est faible (inférieure à 12 %), et l'estimation du rayon du disque du fruit dans l'image est plus robuste aux erreurs de segmentation que l'estimation des axes de l'ellipse. Le modèle sphérique sera donc utilisé comme approximation de l'aire de la surface de la canneberge.

### **5.3 Prédiction de l'anthocyane**

Dans cette section, le modèle permettant de prédire au mieux le Tacy sera déterminé. Les modèles prédictifs seront du type régression linéaire multiple. Les lots contenant des fruits pourris, désormais laissés dans les échantillons par le laboratoire, ne sont pas pris en compte dans cette étude (voir ANNEXE 5). Les bases de données utilisées comportent toutes 4 photos par échantillons de lots de canneberges, ainsi que leur Tacy mesuré par la méthode chimique. Dans un premier temps, les échantillons standards 2005 seront utilisés pour établir les différents modèles de régressions. Par la suite, ces modèles de régression seront testés avec les données des lots photographiés lors de la cueillette de 2005, en excluant les échantillons standard. Puis le cas des lots présentant simultanément des fruits jaunes et des fruits très foncés (mélanges extrêmes) sera étudié. Enfin, le modèle prédictif retenu pour la saison 2005 sera validé une nouvelle fois, sur la base de données de la saison 2006.

#### **5.3.1 Modèles de régression linéaire**

Dans cette section, une série de régression linéaire multiple sera effectuée afin de prédire au mieux le Tacy d'un échantillon à partir de sa couleur moyenne (moyenne de la couleur de tous les pixels appartenant aux canneberges d'une image), ainsi que de la dimension des fruits. Les échantillons standards 2005 seront utilisés pour établir les différents modèles de régression, à partir de la moyenne des informations des quatre

images d'un échantillon afin d'augmenter la justesse du modèle. Les caractéristiques suivantes seront étudiées comme couleur moyenne d'un échantillon :

- a. L\*;
- b. a\*;
- c. PC1;
- d. PC2.

Soit M le nombre de pixels appartenant à des canneberges dans une image, et N le nombre de canneberge dans une image.

La couleur moyenne ( $X_1$ ) d'un lot de canneberges est définie de la manière suivante :

$$X_1 = \frac{\sum_{k=1}^M \text{Couleur\_pixel}_k}{M} \quad (4.45)$$

Néanmoins, dans le repère L\*-a\* (ainsi que PC1-PC2), il est possible de remarquer une symétrie des valeurs de a\* par rapport au point de coordonnées L\* = 120 (sommet de la courbe). Or pour une même valeur de a\*, si L\* > 120 les fruits ne contiendront que très peu d'anthocyane, alors que si L\* < 120, les fruits en contiendront beaucoup. Ainsi, les variables a\* et PC2 ne pourront servir à prévoir la teneur en anthocyane que pour des valeurs de L\* < 120.

La dimension sera également prise en compte, à partir des paramètres suivants :

- a. nombre moyen de pixels d'une canneberge par photo (surface moyenne des fruits) :

$$\text{surface\_moyenne} = \frac{M}{N} \quad (4.46)$$

- b. nombre de pixels appartenant aux canneberges par photo (surface totale de tous les fruits) :

$$surface\_totale = M \quad (4.47)$$

Voici la plage [min : max] des diverses variables pour les échantillons standard, d'abord dans le cas où toute la plage couleur est considérée :

- a.  $L^*$  : [44 : 183];
- b. PC1 : [-88 : 53];
- c. Surface moyenne : [1017 : 2684];
- d. Surface totale : [101336 : 136765];
- e. Tacy : [6 : 59].

Voici la plage [min : max] des diverses variables pour les échantillons standards, en enlevant les échantillons possédant une couleur rosée et jaune-vert c'est-à-dire les échantillons ayant un Tacy inférieur à 10 :

- a.  $L^*$  : [44 : 92];
- b.  $a^*$  : [148 : 173];
- c. PC1 : [1,7 : 53];
- d. PC2 : [-13 : 12];
- e. Surface moyenne : [1017 : 2684];
- f. Surface totale : [101336 : 136765];
- g. Tacy : [26 : 59].

Les régressions linéaires ont été effectuées grâce à la boîte à outils *mbcmodel* (*Model-Based Calibration Toolbox*) disponible dans MATLAB. Plusieurs modèles de régression linéaire multiple ont été créés à partir des échantillons standard 2005, avec une variable de couleur et une de taille comme entrées du modèle. Le critère utilisé pour évaluer les modèles est la racine carrée de l'erreur quadratique moyenne (RMSE) définie de la manière suivante :

$$\sqrt{\frac{\sum_{k=1}^P (Tacy\_chimique_k - Tacy\_prédit_k)^2}{(P - m)}} \quad (4.48)$$

avec :

- $P$  : Nombre d'échantillons dans la base de modélisation;
- $Tacy\_chimique_k$  : Tacy chimique (réel) de l'échantillon  $k$ ;
- $Tacy\_prédit_k$  : Tacy prédit par le modèle pour l'échantillon  $k$ ;
- $m$  : nombre de coefficients du modèle estimés à partir des données de la base d'apprentissage.

Plus l'erreur RMSE est faible et proche de 0, meilleur est le modèle.

Le meilleur modèle sans filtrage des données à couleurs trop claires est le suivant (voir Tableau XVII et Figure 52):

- Variable couleur :  $X_1 = L^*$ ;
- Variable surfacique :  $X_2 =$  surface totale;
- RMSE = 1,98;
- $R^2 = 0,99$ .

Tableau XVII  
Modèle et coefficients

Modèle cubique sur la couleur et linéaire sur la taille :		
Tacy prédit = $a + b \cdot X_1 + c \cdot X_2 + d \cdot X_1^2 + e \cdot X_1^3$		
Coefficient	Valeur	Erreur standard
a	97,11	4,9
b	-1,98	0,15
c	2,6e-4	2,9e-5
d	0,011	1,5e-3
e	- 2,35e-5	4,6e-6

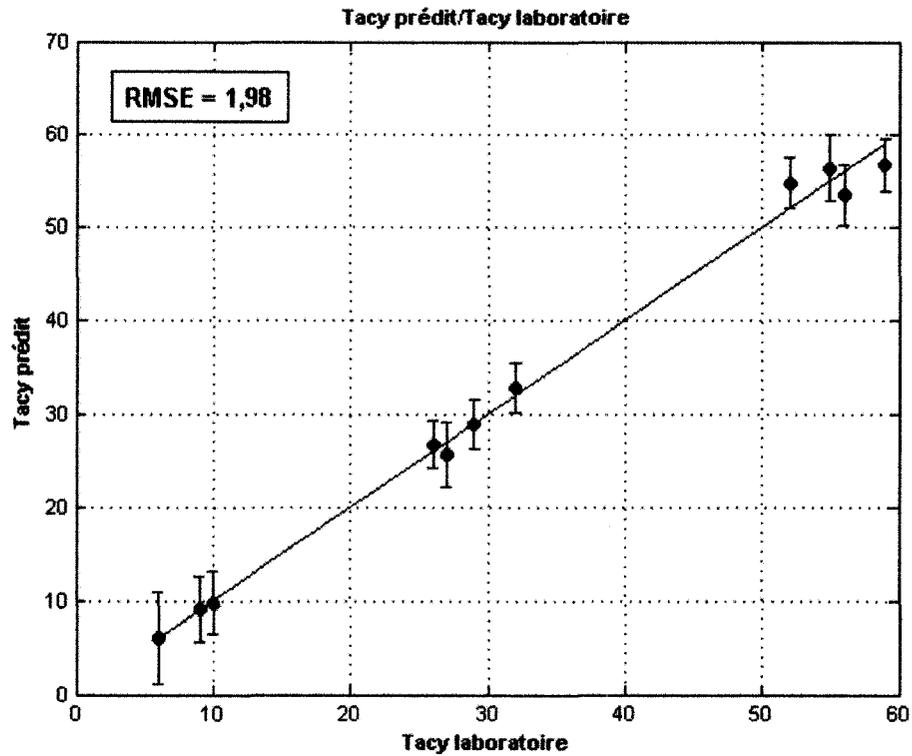


Figure 52 Tacy prédit VS Tacy laboratoire (meilleur modèle sur la totalité de l'espace)

Tous les intervalles de confiance des données dans cette section sont définis à 95%.

Le meilleur modèle avec filtrage des données à couleurs jaune-pâle est le suivant (voir Tableau XVIII et Figure 53) :

- a. Variable couleur :  $X1 = a^*$ ;
- b. Variable surfacique :  $X2 = \text{surface totale}$ ;
- c.  $RMSE = 2,13$ ;
- d.  $R^2 = 0,98$ .

Tableau XVIII  
Modèle et coefficients

Modèle linéaire sur la couleur et sur la taille :		
Tacy prédit = a + b*X <sub>1</sub> + c*X <sub>2</sub>		
Coefficient	Valeur	Erreur standard
a	240,24	6,27
b	- 1,39	0,036
c	2,3e-4	2,90e-5

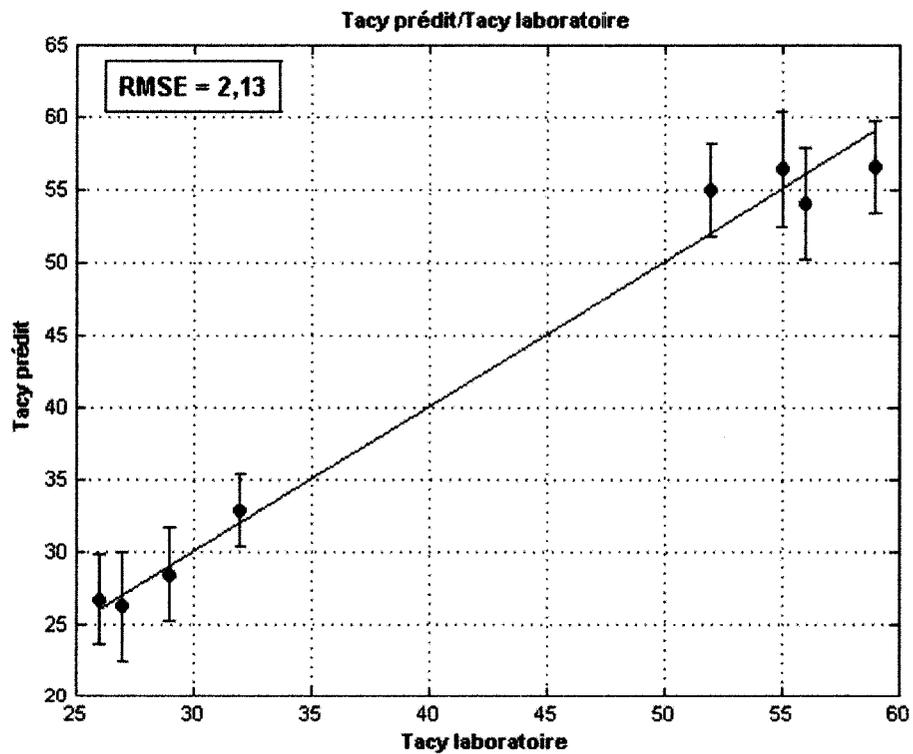


Figure 53 Tacy prédit VS Tacy laboratoire (meilleur modèle sur l'espace restreint)

D'autres modèles de régression sont aussi étudiés, ceux ne tenant pas compte de l'information sur la dimension des fruits présentent une erreur (RMSE) bien plus importante, aux environs de 4,0 (résultats non inclus). Ce fait confirme les conclusions de Francis concernant l'effet de la taille des fruits sur la teneur en anthocyane [9]. Néanmoins, pour nos meilleurs modèles (Figure 52 et Figure 53), l'erreur de régression (RMSE) varie de 1,98 à 2,13.

### 5.3.2 Validation des modèles de régression

Tous les modèles établis précédemment ont été ensuite testés à partir des données des lots photographiés<sup>1</sup> lors de la cueillette de 2005, en excluant les lots standard. Il est à noter que les échantillons spéciaux possédant 25 % de fruits jaunes, 25 % de fruits rosés, 25 % de fruits rouges et 25% de fruits rouge foncé (mélanges extrêmes) sont également exclus de la base de test, car ce type de lots soulève le problème de l'effet boomerang de la couleur canneberge, ce type de lots sera étudié ultérieurement. Pour établir un modèle linéaire entre les couleurs ambiguës ( $a^*$ , PC2) et l'anthocyane, un filtrage exclut les échantillons dont la valeur de  $L^*$  est supérieure à 120 (base de données réduite). Le critère de qualité du modèle lors du test est toujours l'erreur RMSE qui doit être le plus petit possible pour l'évaluation du Tacy moyen. Dans le cas de la validation des modèles, la définition de l'erreur RMSE est la suivante :

$$\sqrt{\frac{\sum_{k=1}^Q (Tacy\_chimique_k - Tacy\_prédit_k)^2}{Q}} \quad (4.49)$$

---

<sup>1</sup> Voir en ANNEXE 5 les photos de 4 lots commerciaux typique de la saison de cueillette 2005

avec :

- a.  $Q$  : Nombre d'échantillons dans la base de test;
- b.  $Tacy\_chimique_k$  : Tacy chimique (réel) de l'échantillon  $k$ ;
- c.  $Tacy\_prédit_k$  : Tacy prédit par le modèle pour l'échantillon  $k$ .

Voici la plage des données de test complète par photo (sans moyenne des 4 photos, sans les mélanges extrêmes) :

- a.  $L^*$  : [40,59 : 137,20];
- b.  $PC1^*$  : [-43,21 : 56,90];
- c. Surface moyenne : [1151 : 2308];
- d. Surface totale : [99097 : 132239];
- e. Tacy : [8 : 64].

Voici la plage des données de test par photo de la base réduite, c'est-à-dire avec filtrage des données dont  $L^* > 120$  (sans moyenne des 4 photos – sans les mélanges extrêmes) :

- a.  $L^*$  : [40,59 : 76,09];
- b.  $A^*$  : [147,61 : 163,68];
- c.  $PC1^*$  : [20,03 : 56,90];
- d.  $PC2^*$  : [-14,16 : 2,36];
- e. Surface moyenne : [1151 : 2308];
- f. Surface totale : [104414 : 132239];
- g. Tacy : [40 : 64].

Tableau XIX  
Validation des modèles polynomiaux de prédiction du Tacy  
(incluant 2 séries de données aberrantes, et sans les mélanges extrêmes)

numéro modèle	variable couleur	degré de liberté couleur	variable surfacique	degré de liberté de la surface	base de données	RMSE (4 photos séparées)	RMSE (4 photos groupées)
1	L*	2	surface moyenne	1	complète	5,3	5,2
2	L*	3	surface moyenne	1	complète	5,5	5,4
3	L*	2	surface totale	1	complète	6,0	6,0
4	L*	3	surface totale	1	complète	6,4	6,3
5	PC1	2	surface moyenne	1	complète	5,3	5,2
6	PC1	3	surface moyenne	1	complète	5,4	5,3
7	PC1	2	surface totale	1	complète	6,0	5,9
8	PC1	3	surface totale	1	complète	6,4	6,3
9	L*	1	surface moyenne	1	réduite	4,9	4,8
10	L*	2	surface moyenne	1	réduite	5,1	5,0
11	L*	3	surface moyenne	1	réduite	5,0	4,9
12	L*	1	-	-	réduite	5,1	5,0
13	L*	2	-	-	réduite	5,3	5,2
14	PC1	1	surface moyenne	1	réduite	4,8	4,7
15	PC1	2	surface moyenne	1	réduite	5,0	4,9
16	L*	1	surface totale	1	réduite	5,5	5,4
17	L*	2	surface totale	1	réduite	6,3	6,2
18	PC1	1	surface totale	1	réduite	5,5	5,4
19	PC1	2	surface totale	1	réduite	6,2	6,1
20	a*	1	-	-	réduite	5,1	5,0
21	a*	2	-	-	réduite	4,7	4,6
22	a*	1	surface moyenne	1	réduite	4,7	4,5
23	a*	2	surface moyenne	1	réduite	4,4	4,3
24	a*	1	surface totale	1	réduite	5,3	5,2
25	a*	2	surface totale	1	réduite	5,3	5,2
26	PC2	1	surface moyenne	1	réduite	4,7	4,5
27	PC2	2	surface moyenne	1	réduite	4,4	4,3
28	PC2	1	surface totale	1	réduite	5,3	5,2
29	PC2	2	surface totale	1	réduite	5,2	5,0

### 5.3.3 Choix du modèle prédictif

L'erreur RMSE varie entre des valeurs allant de 4 à 5 unités de Tacy. Les erreurs RMSE obtenues par une analyse de 4 photos indépendante ou par une analyse des moyennes des 4 photos (permise grâce à l'ANOVA) présentent des termes d'erreurs quasi équivalents (4,4 et 4,3) pour le modèle prédictif utilisant les variables  $a^*$  et la taille moyenne avec un terme quadratique pour le  $a^*$ . Les meilleurs résultats sont obtenus sur la base de données filtrée ( $L^* < 120$ ), avec les couleurs  $a^*$  et PC2, et la taille moyenne des fruits.

Du fait de sa linéarité et de son erreur RMSE relativement faible, le modèle suivant sera étudié plus en détail :

- a. Modèle polynomial de degré 1 pour la couleur et la surface;
- b. variable couleur :  $X_1 = a^*$  ;
- c. variable surfacique :  $X_2 =$  surface moyenne;
- d.  $RMSE = 2,24$  et  $R^2 = 0,98$  (sur données de modélisation);
- e.  $RMSE = 4,7$  (sur données de test).

Tableau XX  
Modèle et coefficients

Modèle linéaire sur la couleur et sur la taille :		
Tacy prédit = $a + b \cdot X_1 + c \cdot X_2$		
Coefficient	Valeur	Erreur standard
a	280,88	6,40
b	- 1,42	0,037
c	- 4,33e-3	5,57e-4

Parmi les lots examinés par le laboratoire avec trois répétitions, l'estimation du Tacy du lot complet de marchandise s'effectue dans 80% des cas avec un terme d'erreur type inférieur à  $\pm 2$  unités de Tacy et jusqu'à 95% des cas avec un terme d'erreur type inférieur à  $\pm 3$  unités de Tacy (résultats non présentés). Des erreurs types potentielles de  $\pm 2$  et de  $\pm 3$  unités peuvent donc être considérées pour toutes les estimations du Tacy du lot complet effectuée le laboratoire.

La Figure 54 montre que l'erreur type de la moyenne de prédiction du Tacy entre les 4 photos d'un même échantillon de la base de test est très variable, mais qu'elle reste néanmoins inférieure à l'erreur type du laboratoire. De plus, la déviation standard ne dépend pas du Tacy de l'échantillon. La Figure 55 montre qu'il existe bien une corrélation entre le Tacy prédit et celui annoncé par les données chimique du laboratoire. Nous comparons ensuite le pourcentage de succès de prédiction pour des écarts potentiels de  $\pm 2$  et de  $\pm 3$ , similaire à ceux du laboratoire, en regroupant les moyennes des 4 photos entre elles (Figure 57 et Figure 58), mais aussi en considérant les photos individuellement (Figure 59 et Figure 60). Il apparaît que le pourcentage de prédictions réussies varie entre 70 % et 90 % selon les cas, en considérant une erreur type similaire à celle du laboratoire (voir tableaux suivants).

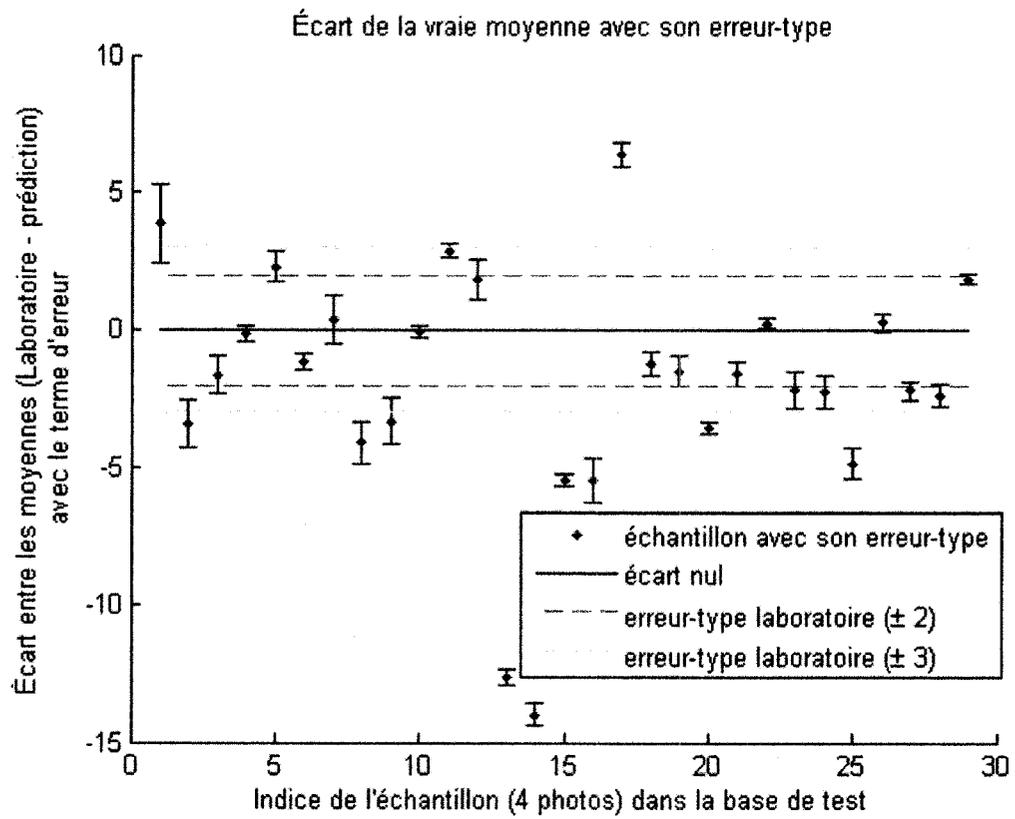


Figure 54 Écart de la vraie moyenne avec son erreur type

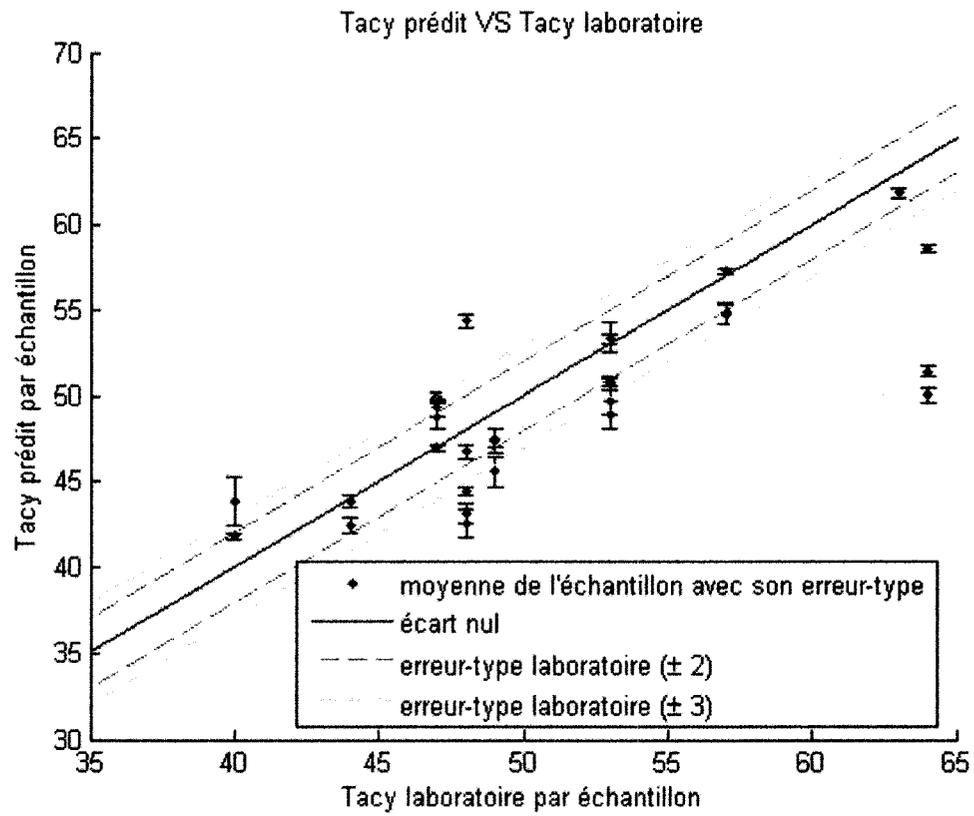


Figure 55 Corrélation entre le Tacy du laboratoire et le Tacy prédit

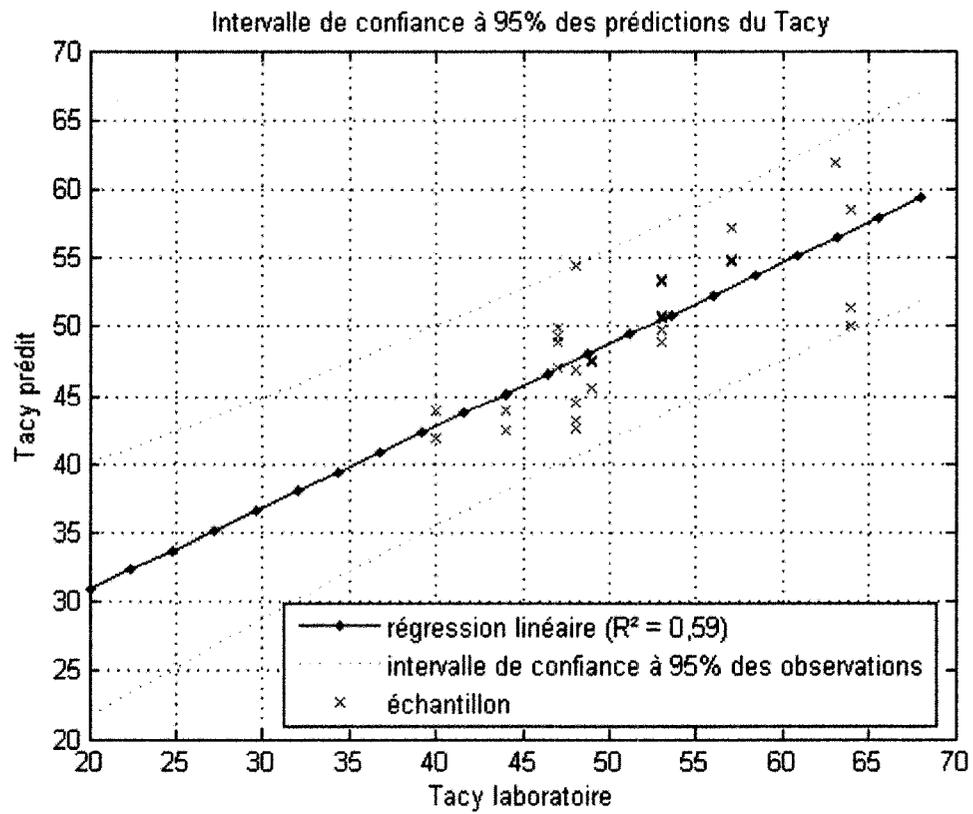


Figure 56 Intervalle de confiance à 95% des prédictions du Tacy après regroupement des 4 photos

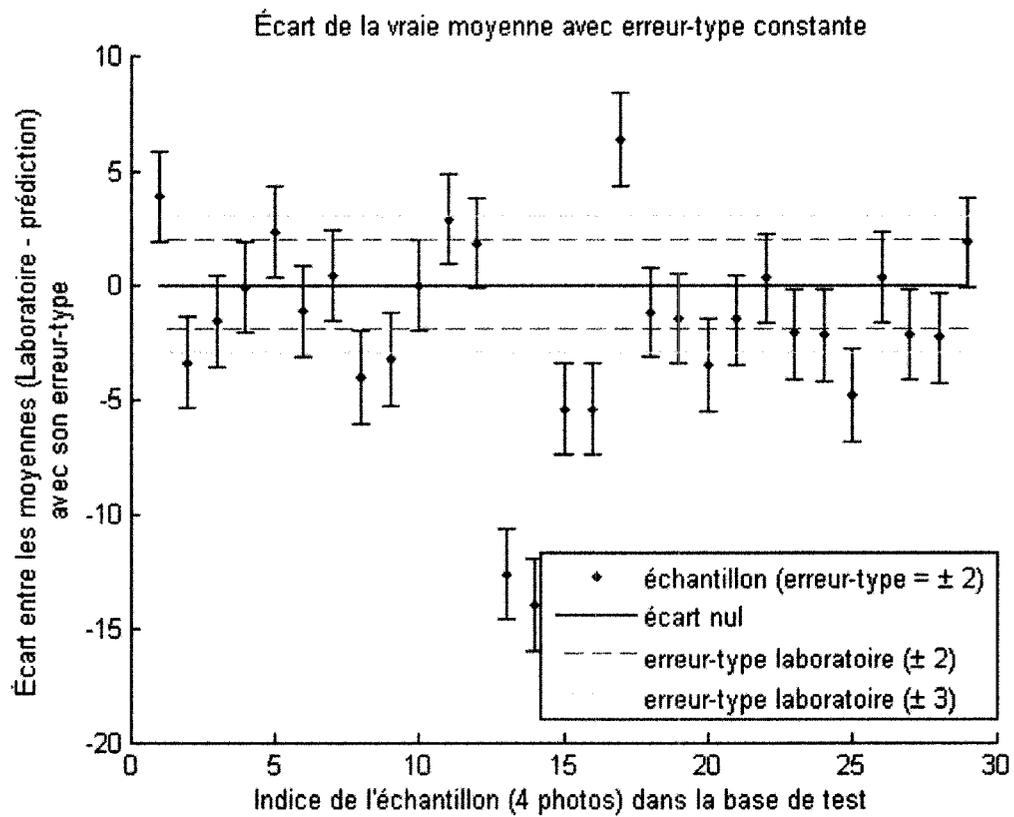
Figure 57 Modèle d'erreur type de  $\pm 2$  après regroupement des 4 photos

Tableau XXI

Statistique de succès de prédiction avec une erreur type de  $\pm 2$ 

	erreur type laboratoire de $\pm 2$	erreur type laboratoire de $\pm 3$
nombre d'échantillons	29	29
nombre de succès	22	24
nombre d'échecs	7	5
pourcentage de succès	76 %	83 %

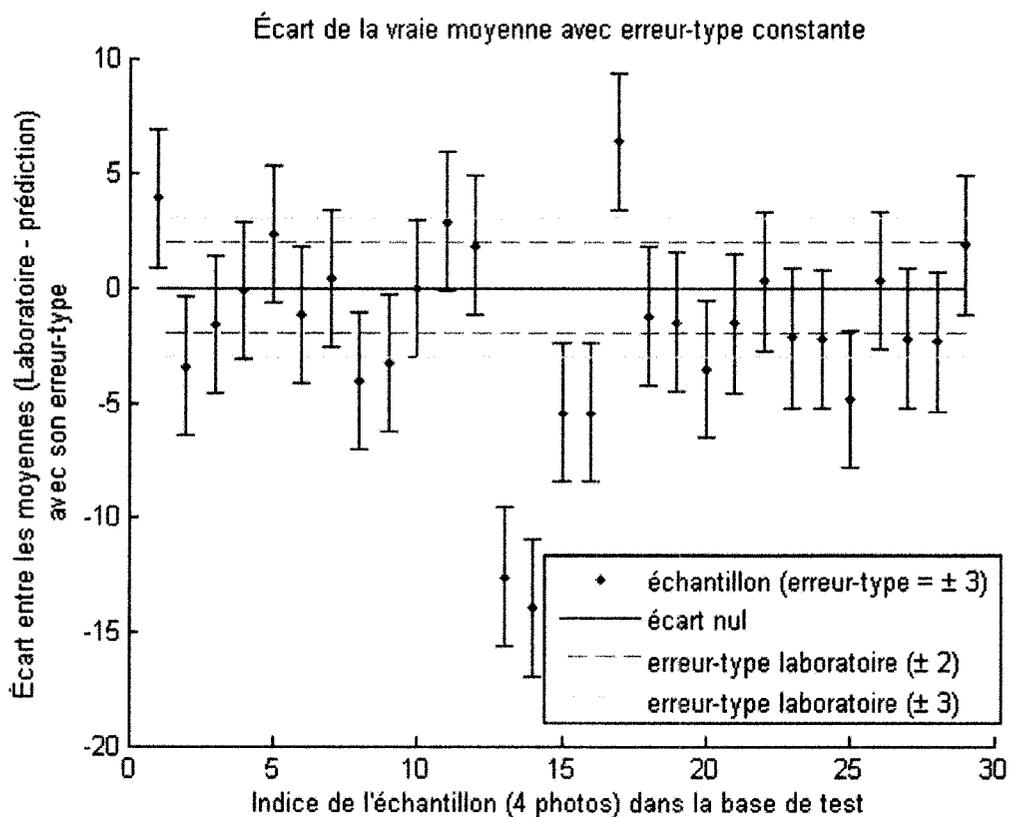


Figure 58 Modèle d'erreur type de  $\pm 3$  après regroupement des 4 photos

Tableau XXII

Statistique de succès de prédiction avec une erreur type de  $\pm 3$

	erreur type laboratoire de $\pm 2$	erreur type laboratoire de $\pm 3$
nombre d'échantillons	29	29
nombre de succès	24	26
nombre d'échecs	5	3
pourcentage de succès	83 %	90 %

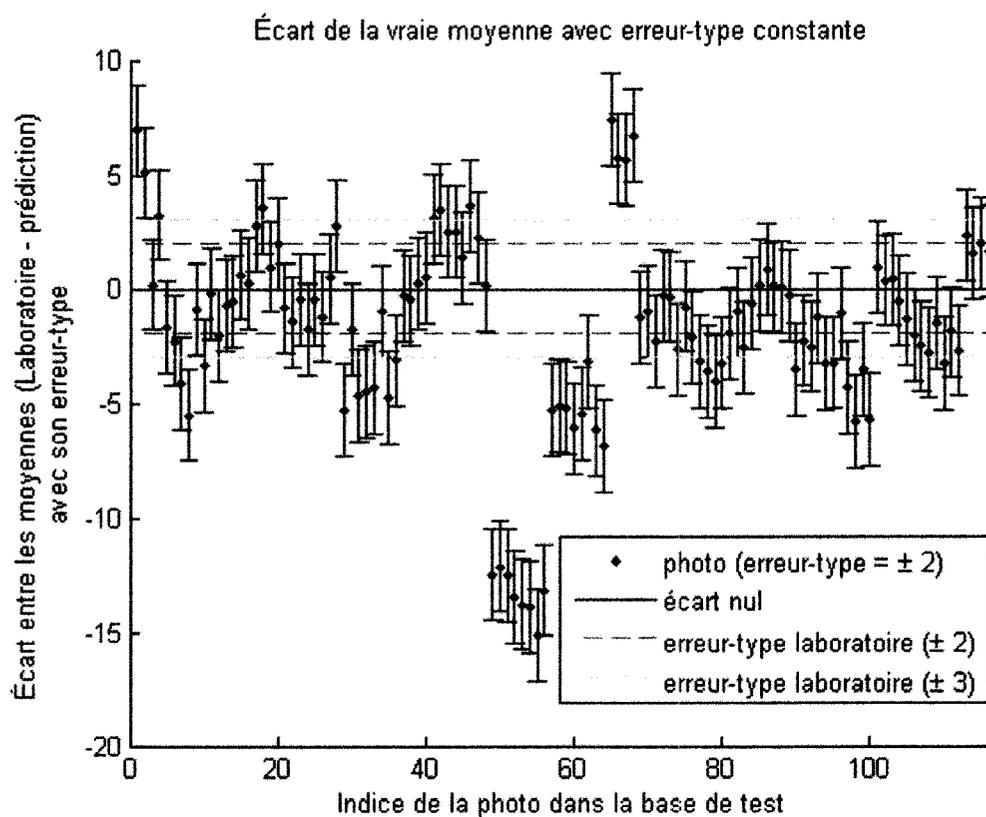
Figure 59 Modèle d'erreur type de  $\pm 2$  avec photos séparées

Tableau XXIII

Statistique de succès de prédiction avec une erreur type de  $\pm 2$ 

	erreur type laboratoire de $\pm 2$	erreur type laboratoire de $\pm 3$
nombre de photos	116	116
nombre de succès	85	91
nombre d'échecs	31	25
pourcentage de succès	73 %	78 %

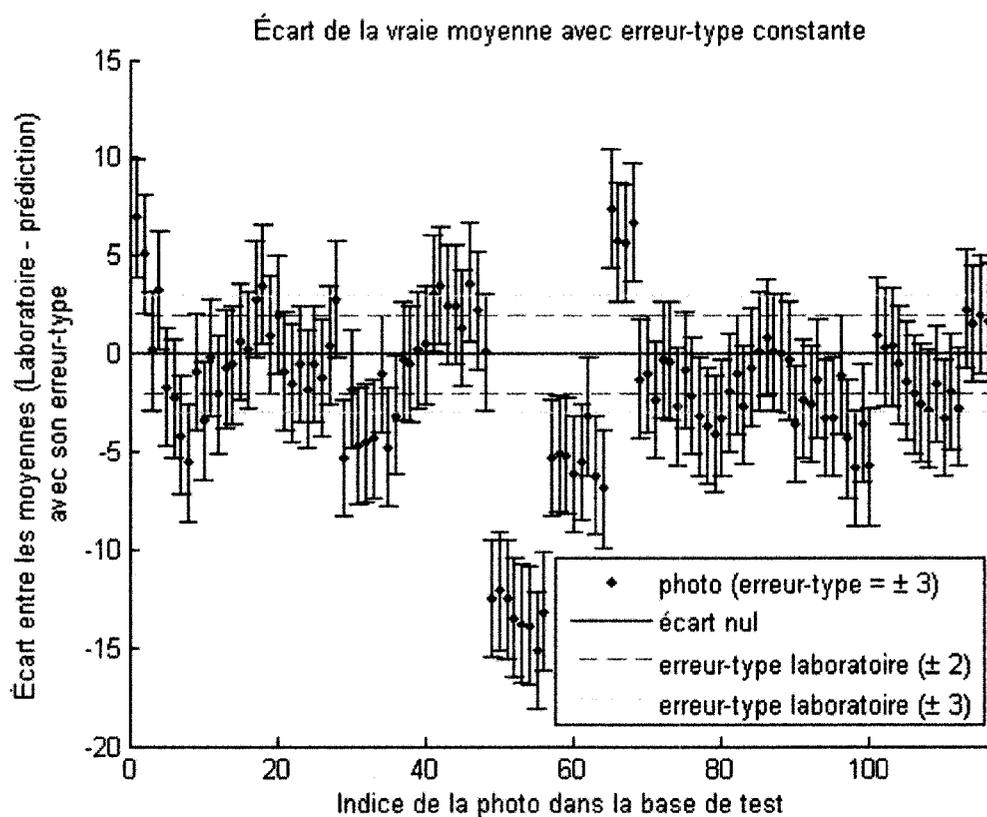


Figure 60 Modèle d'erreur type de  $\pm 3$  avec photos séparées

Tableau XXIV

Statistique de succès de prédiction avec une erreur type de  $\pm 3$

	erreur type laboratoire de $\pm 2$	erreur type laboratoire de $\pm 3$
nombre de photos	116	116
nombre de succès	91	102
nombre d'échecs	25	14
pourcentage de succès	78 %	88 %

L'analyse des écarts de la moyenne montre que la plus grande source d'erreur provient du lot numéro 442, dont le Tacy est de 64. Pour les trois échantillons de ce lot, la valeur de 64 leur est attribuée (plus ou moins arbitrairement). Pour chacun des 3 échantillons, l'erreur (Tacy\_prédit – Tacy\_chimique) est la suivante :

- a. -12;
- b. -13;
- c. -5.

Ceci, ajouté à la vérification visuelle des photos, montre une grande différence de couleur entre les deux premiers échantillons et le troisième. L'erreur pourrait également être interprétée comme provenant d'une erreur d'extrapolation du modèle, or un lot dont un échantillon unique est disponible et dont le Tacy est de 63 possède un écart de 1 unité avec la prédiction du modèle. Le lot numéro 443 dont seule la valeur 48 de Tacy est fournie pour les 3 échantillons est l'autre source d'erreur importante. Toutes les photos issues de ces deux lots seront par la suite considérées comme des données aberrantes et seront enlevées de la base de test par la suite (voir ANNEXE 5).

Il apparaît donc qu'en enlevant les données aberrantes, le nombre de prédictions erronées est considérablement diminué, même en prenant l'hypothèse d'une erreur type de  $\pm 2$  pour les prédictions du laboratoire (Tableau XXV). En effectuant la moyenne des prédictions du Tacy des 4 photos d'un même échantillon et en excluant les données aberrantes, et en prenant l'hypothèse d'une erreur type de  $\pm 2$  pour les prédictions du laboratoire et une erreur type de  $\pm 3$  pour nos prédictions, il n'y a plus de prédiction erronée (Tableau XXVI).

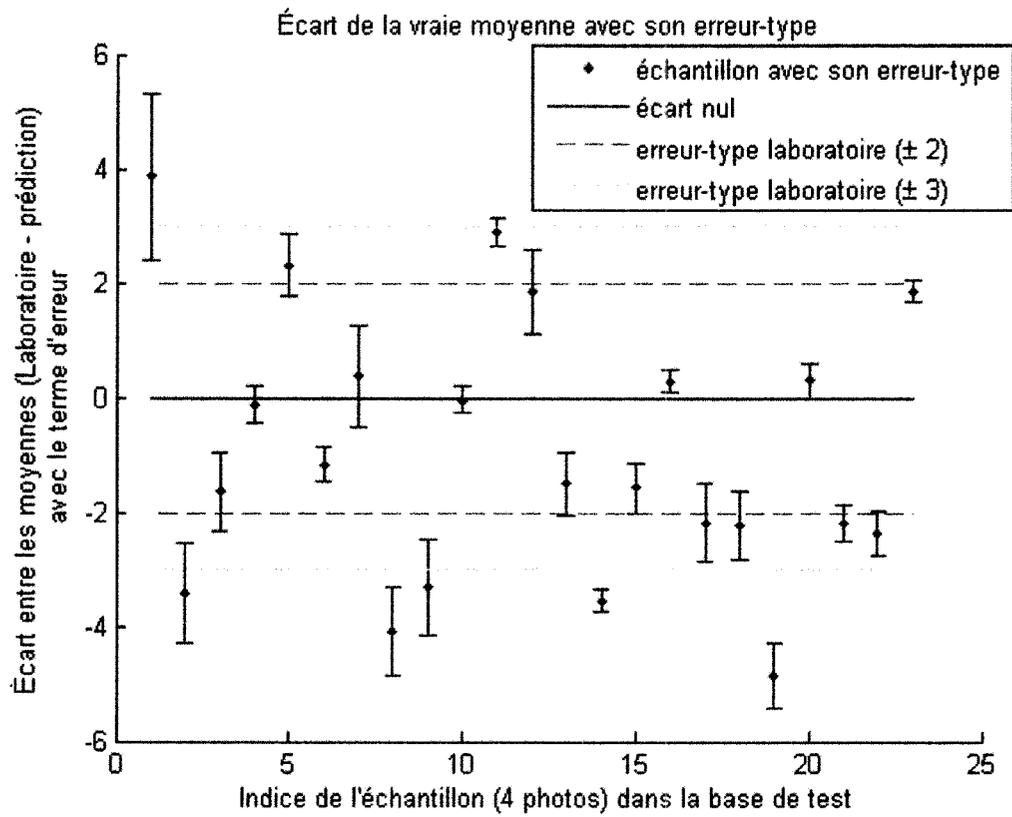


Figure 61 Écart de la vraie moyenne avec son erreur type (sans données aberrantes)

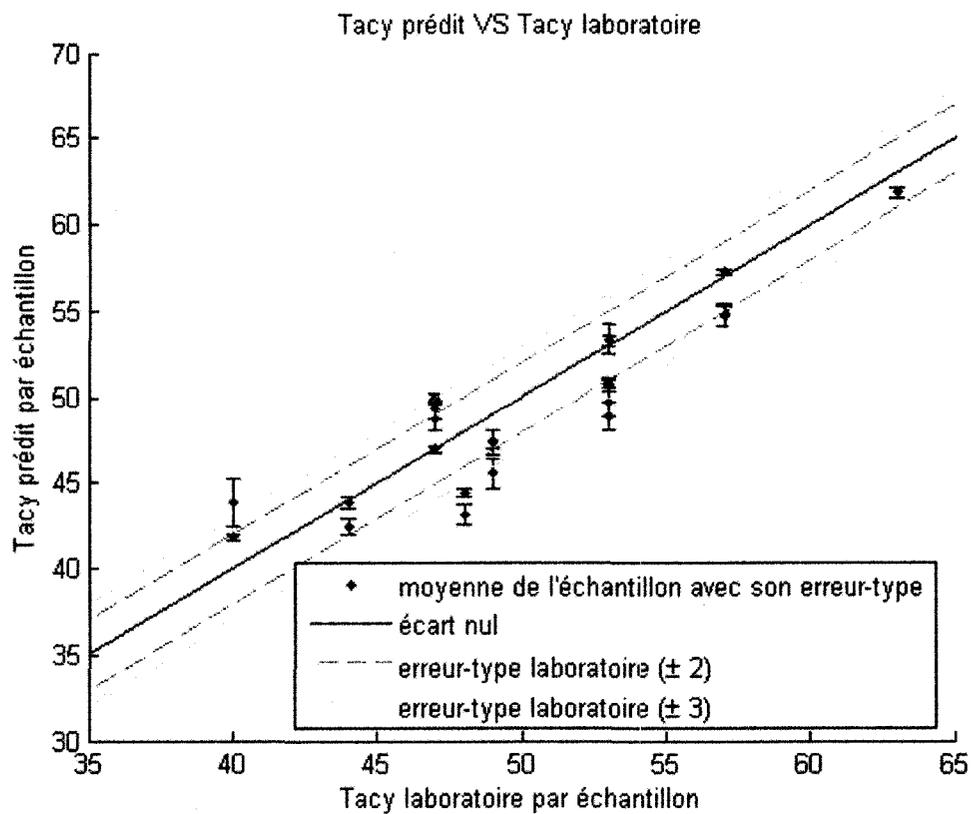


Figure 62 Corrélation entre le Tacy du laboratoire et le Tacy prédit (sans données aberrantes)

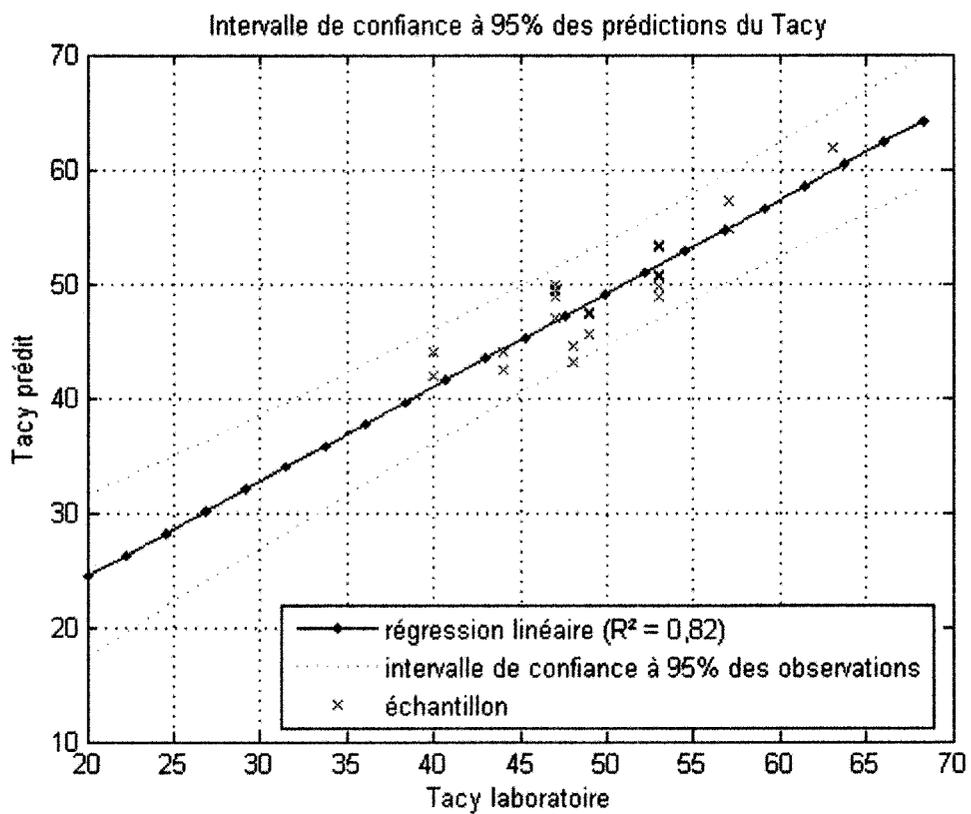


Figure 63 Intervalle de confiance à 95% des prédictions du Tacy après regroupement des 4 photos (sans données aberrantes)

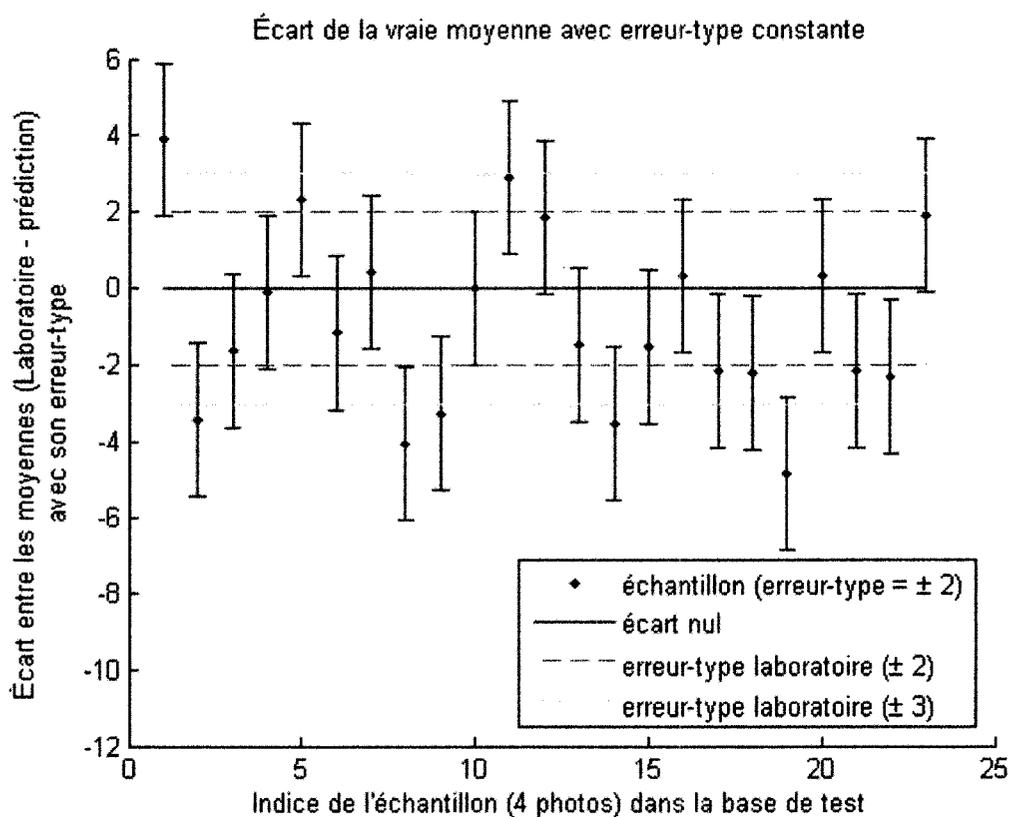


Figure 64 Modèle d'erreur type de  $\pm 2$  après regroupement des 4 photos  
(sans données aberrantes)

Tableau XXV

Statistique de succès de prédiction avec une erreur type de  $\pm 2$

	erreur type laboratoire de $\pm 2$	erreur type laboratoire de $\pm 3$
nombre d'échantillons	23	23
nombre de succès	21	23
nombre d'échecs	2	0
pourcentage de succès	91 %	100 %

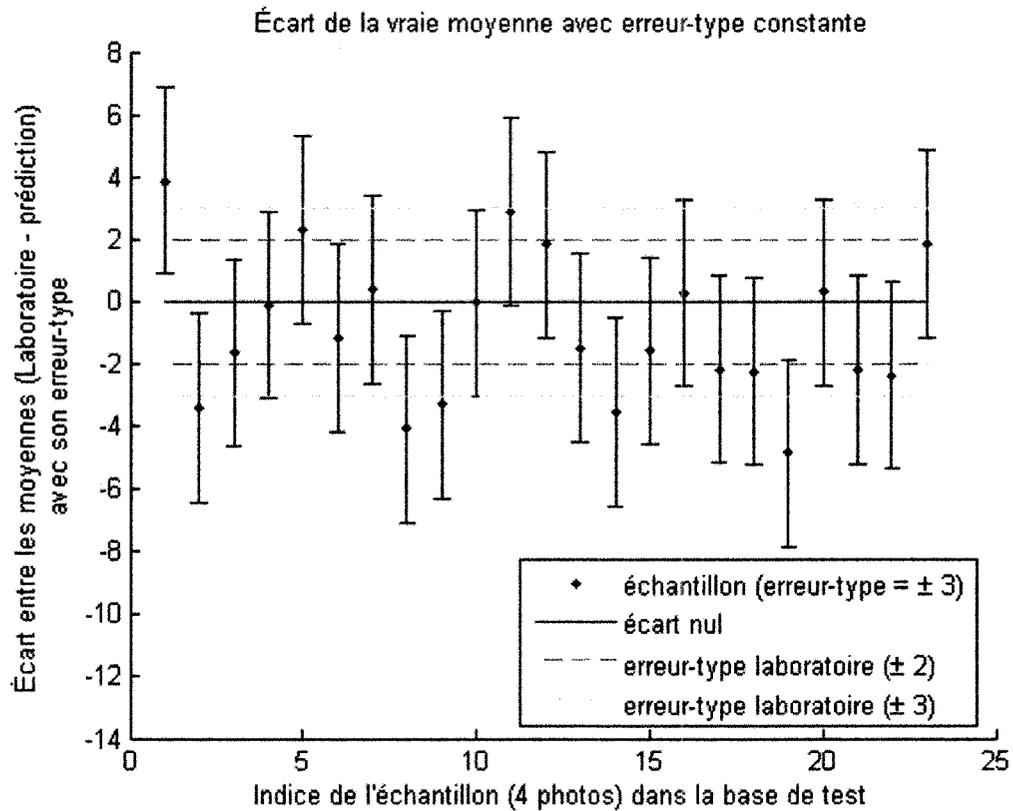


Figure 65 Modèle d'erreur type de  $\pm 3$  après regroupement des 4 photos  
(sans données aberrantes)

Tableau XXVI

Statistique de succès de prédiction avec une erreur type de  $\pm 3$

	erreur type laboratoire de $\pm 2$	erreur type laboratoire de $\pm 3$
nombre d'échantillons	23	23
nombre de succès	23	23
nombre d'échecs	0	0
pourcentage de succès	100 %	100 %

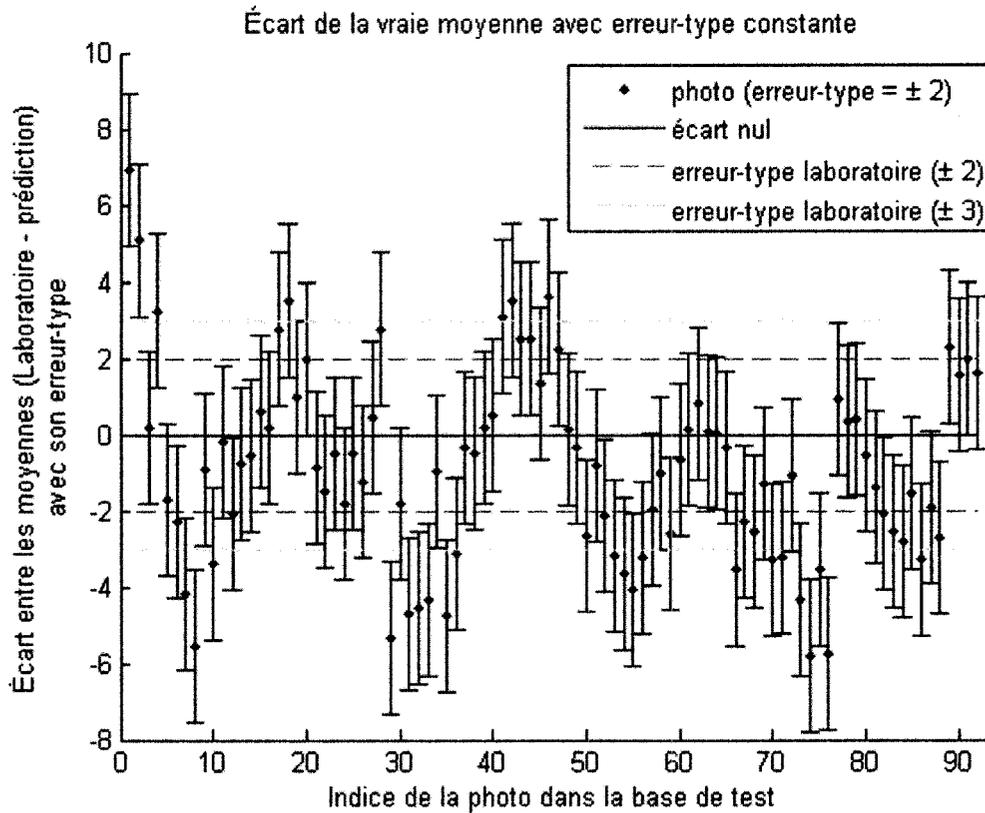


Figure 66 Modèle d'erreur type de  $\pm 2$  avec photos séparées (sans données aberrantes)

Tableau XXVII

Statistique de succès de prédiction avec une erreur type de  $\pm 2$

	erreur type laboratoire de $\pm 2$	erreur type laboratoire de $\pm 3$
nombre de photos	92	92
nombre de succès	80	86
nombre d'échecs	12	6
pourcentage de succès	87 %	93 %

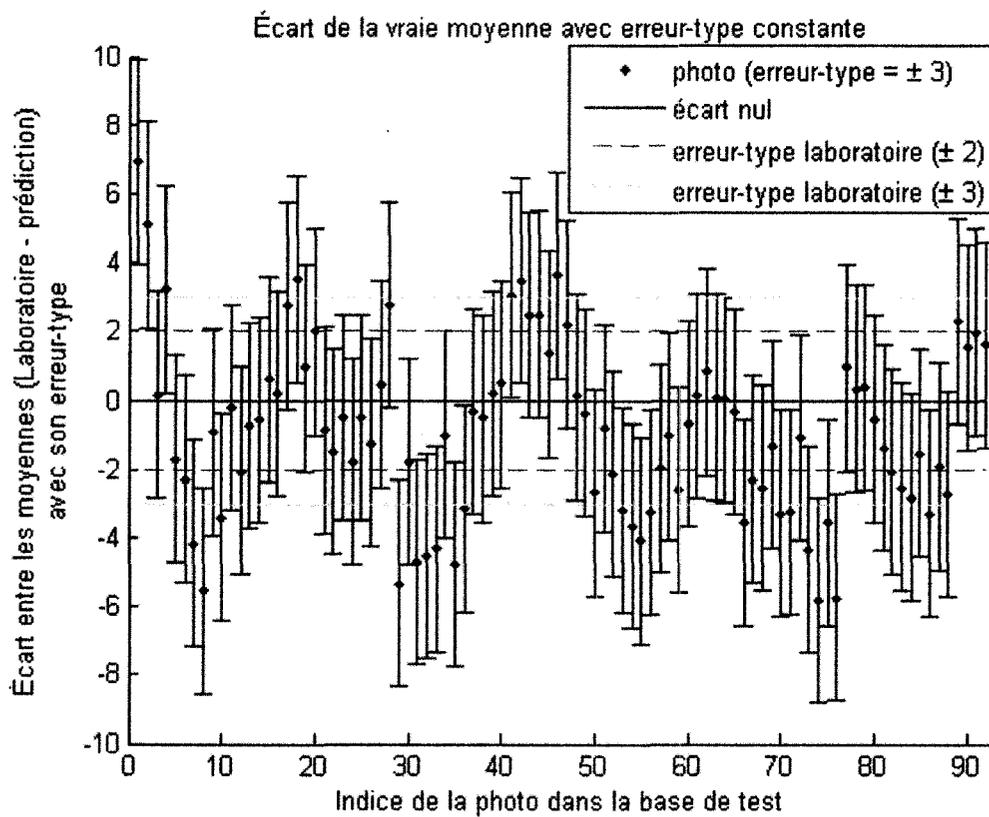


Figure 67 Modèle d'erreur type de  $\pm 3$  avec photos séparées (sans données aberrantes)

Tableau XXVIII

Statistique de succès de prédiction avec une erreur type de  $\pm 3$

	erreur type laboratoire de $\pm 2$	erreur type laboratoire de $\pm 3$
nombre de photos	92	92
nombre de succès	86	91
nombre d'échecs	6	1
pourcentage de succès	93 %	99 %

Les quatre figures suivantes montrent les différentes vues de la surface produite par la modélisation du Tacy à partir de la bande a\* (X1) et de la taille moyenne (X2) :

$$Tacy\_prédit = 280,88 - 1,42X_1 - 0,00433X_2 \quad (4.50)$$

Les points bleus représentent les données ayant servi à construire le modèle prédictif. Les points rouges représentent la valeur de Tacy fournie par le laboratoire pour chaque échantillon (4 photos), en ayant pour coordonnées la moyenne de la couleur a\* et de la taille moyenne de chaque fruit, effectuée sur les 4 photos. Les données aberrantes ne sont pas représentées.

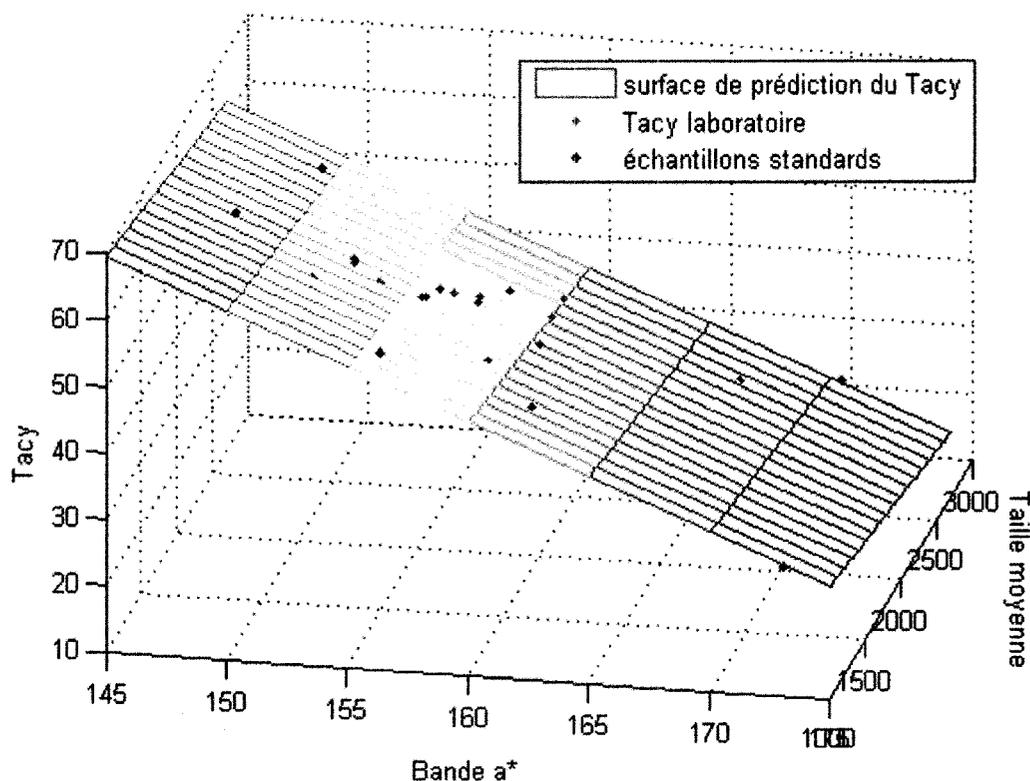


Figure 68 Modèle prédictif du Tacy (vue 1)



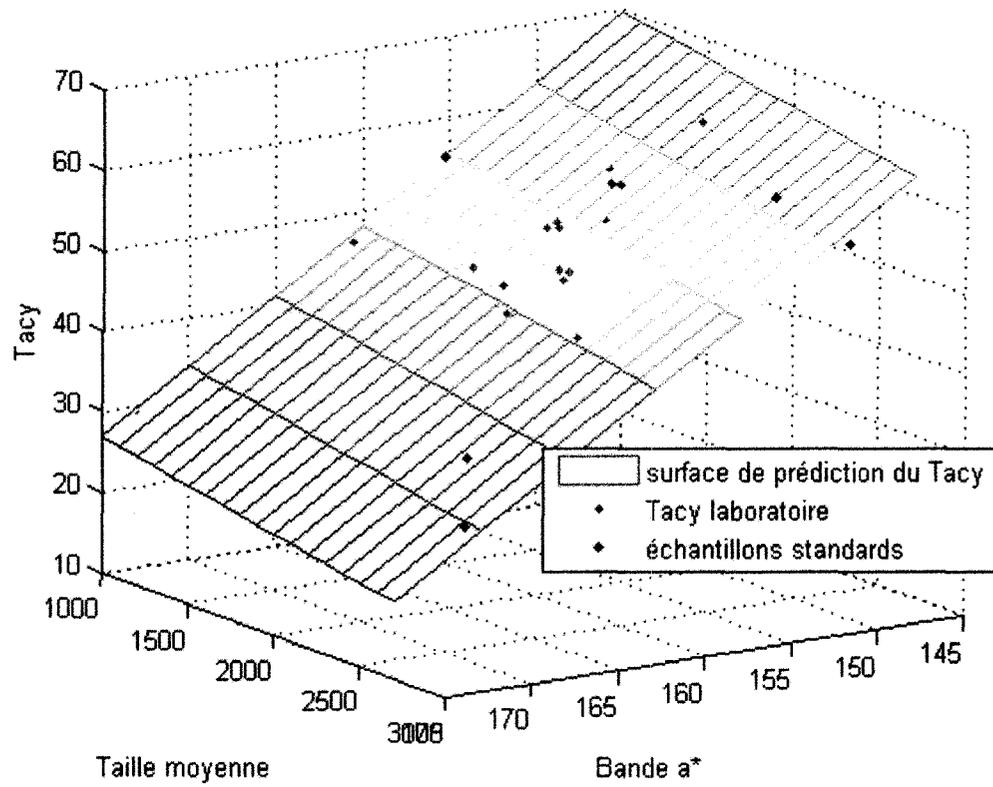


Figure 70 Modèle prédictif du Tacy (vue 3)

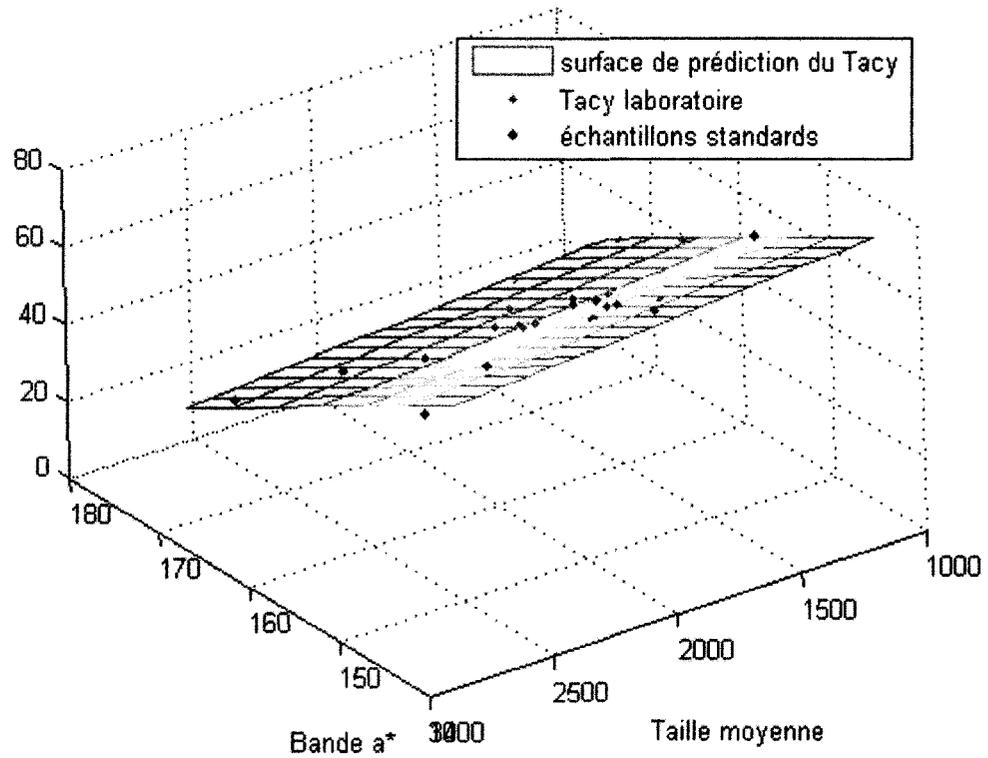


Figure 71 Modèle prédictif du Tacy (vue 4)

Après avoir enlevé les données aberrantes mises en évidence dans l'analyse précédente, l'erreur RMSE sur le reste de la base de test est de nouveau calculée.

Tableau XXIX

Validation des modèles polynomiaux de prédiction du Tacy  
(sans les 2 lots de mélanges extrêmes et ni les 2 séries de données aberrantes)

numéro modèle	variable couleur	degré de liberté couleur	variable surfacique	degré de liberté de la surface	base de données	RMSE (4 photos séparées)	RMSE (4 photos groupées)
1	L*	2	surface moyenne	1	complète	3,6	3,5
2	L*	3	surface moyenne	1	complète	3,7	3,6
3	L*	2	surface totale	1	complète	4,3	4,1
4	L*	3	surface totale	1	complète	4,6	4,4
5	PC1	2	surface moyenne	1	complète	3,5	3,4
6	PC1	3	surface moyenne	1	complète	3,6	3,4
7	PC1	2	surface totale	1	complète	4,2	4,0
8	PC1	3	surface totale	1	complète	4,5	4,4
9	L*	1	surface moyenne	1	réduite	2,9	2,8
10	L*	2	surface moyenne	1	réduite	3,1	2,9
11	L*	3	surface moyenne	1	réduite	3,1	2,9
12	L*	1	-	-	réduite	3,0	2,9
13	L*	2	-	-	réduite	3,2	3,0
14	PC1	1	surface moyenne	1	réduite	2,9	2,7
15	PC1	2	surface moyenne	1	réduite	3,0	2,8
16	L*	1	surface totale	1	réduite	3,5	3,4
17	L*	2	surface totale	1	réduite	4,3	4,1
18	PC1	1	surface totale	1	réduite	3,5	3,3
19	PC1	2	surface totale	1	réduite	4,2	4,0
20	a*	1	-	-	réduite	3,2	3,0
21	a*	2	-	-	réduite	2,9	2,7
22	a*	1	surface moyenne	1	réduite	2,7	2,5
23	a*	2	surface moyenne	1	réduite	2,6	2,4
24	a*	1	surface totale	1	réduite	3,3	3,1
25	a*	2	surface totale	1	réduite	3,4	3,1
26	PC2	1	surface moyenne	1	réduite	2,8	2,5
27	PC2	2	surface moyenne	1	réduite	2,6	2,4
28	PC2	1	surface totale	1	réduite	3,4	3,1
29	PC2	2	surface totale	1	réduite	3,2	3,0

Il est maintenant clair et démontré que nous pouvons obtenir des termes d'erreur sur la prédiction avec le modèle (EPM) très similaires aux termes d'erreurs obtenus par la méthode du laboratoire (EML), c'est-à-dire entre  $\pm 2$  à  $\pm 3$  unités sur la moyenne du Tacy en rapport avec les données du laboratoire. Pour validation voir le Tableau XXX.

Tableau XXX

Prédiction du Tacy pour les échantillons de la base de test  
(sans données aberrantes, ni mélanges extrêmes)

Indice de l'échantillon #	Tacy chimique laboratoire	Tacy prédit (regroupement des 4 photos)
1	40	44
2	49	46
3	49	47
4	44	44
5	47	49
6	63	62
7	53	53
8	53	49
9	53	50
10	47	47
11	47	50
12	47	49
13	49	48
14	48	44
15	44	42
16	57	57
17	57	55
18	57	55
19	48	43
20	53	53
21	53	51
22	53	51
23	40	42

### 5.3.4 Discussion

Il apparaît que sans les données aberrantes, les meilleurs modèles sont ceux utilisant les variables couleur a\* et taille moyenne des 4 photos combinées d'un échantillon de fruits. Le fait d'effectuer la moyenne des quatre photos nous permet d'avoir une prédiction quasi parfaite (100%) comparativement à prendre des photos unitaires. Les modèles tenant compte de la couleur PC2 et de la taille moyenne donnent une erreur légèrement plus élevée, mais tout de même similaire. Il est donc inutile de privilégier la voie de la couleur PC2, car non seulement elle n'améliore pas les performances, mais en plus l'industrie agro-alimentaire est peu familière à son utilisation. Au contraire, la notion d'espace couleur CIELAB est largement employée. La couleur L\*, qui offre la plus grande variabilité, mais elle ne permet pas une prédiction aussi précise du Tacy que l'utilisation du a\*. L'information couleur a\* sera donc retenue pour la prédiction du Tacy car elle offre les meilleures performances. En plus, cette variable n'est pas sensible à l'intensité lumineuse, il est donc possible de la conserver pour faire des mesures colorimétriques en plein champ, là où la lumière naturelle change constamment. L'utilisation de la taille moyenne et de la couleur moyenne a\* de tous les fruits d'un échantillon revient à dire que les 100gr de fruits de l'échantillon sont équivalent à 100gr du fruit moyen du même échantillon. Cela indique que le nombre de fruits qui est présenté au modèle est une donnée implicite, vu que l'on utilise seulement la moyenne des fruits. Ce procédé est ainsi directement applicable pour des mesures dans le champ, où il sera difficile de réunir 100gr de fruits, à condition d'avoir suffisamment de fruits pour garantir un bon échantillonnage. Néanmoins, l'application dans le champ nécessitera une calibration de la lumière, afin de fixer un seuil sur la luminance pour éviter l'ambiguïté sur la symétrie de l'échelle couleur par rapport à la variable a\*, et également une calibration de la dimension des fruits. Il reste à tester le modèle pour une plage de Tacy plus étendue afin de mieux connaître sa qualité, et aussi résoudre le problème des lots de type mélange extrême.

### 5.3.5 Cas des mélanges extrêmes

Certains des lots de test réalisés pour les besoins expérimentaux contiennent toute la gamme de couleur canneberge, allant du jaune jusqu'au rouge très foncé. À cause de l'effet boomerang de la couleur canneberge suivant la variable  $a^*$  (également suivant PC2), les canneberges jaunes, qui ne contiennent que très peu de Tacy, possèdent des valeurs de couleur suivant  $a^*$  similaire à celle des canneberges rouge foncé qui elles renferment beaucoup d'anthocyane. Notre modèle prédictif, qui prend en compte la couleur  $a^*$  moyenne de tous les fruits de l'image, n'est donc pas efficace pour traiter les lots de types mélanges extrêmes. En effet, la couleur moyenne selon  $a^*$  d'un échantillon contenant par exemple 50% de fruits jaunes et 50% de fruits rouge foncé sera similaire à celle d'un lot contenant 100% de fruits rouge foncé. Le modèle prédictif donnerait alors des estimations de Tacy similaires pour les deux lots précédents, comme si ces deux lots contenaient 100% de fruits rouge foncé. Bien entendu, les lots de type mélange extrême contiennent moins de Tacy que les lots rouge foncé uniforme. Notre modèle prédictif a donc tendance à surestimer le Tacy des mélanges extrêmes, car il prend comme données d'entrée la couleur moyenne de l'ensemble des canneberges composant le lot. Il est donc nécessaire de modifier la présentation des données au modèle prédictif, afin de résoudre ce problème.

Le modèle prédictif prend comme données d'entrée la couleur moyenne du lot de canneberge, ainsi que la dimension moyenne des fruits du lot. En somme, les données d'entrée du modèle sont les caractéristiques moyennes des fruits composant le lot, le modèle évalue en fait le Tacy à partir du fruit moyen du lot, qui est représentatif de l'ensemble des fruits du lot. Les problèmes dus à l'effet boomerang peuvent être résolus en soumettant chacun des fruits de l'image individuellement au modèle prédictif. Le modèle prédictif devra premièrement établir si le fruit est plutôt jaune ou rouge, en évaluant de quel côté la couleur du fruit se situe par rapport au sommet du boomerang. Pour cela, la bande  $L^*$  de la couleur du fruit sera évaluée, car elle permet de définir la

couleur canneberge de manière non ambiguë. Si la valeur de la bande  $L^*$  du fruit est supérieure à une valeur seuil  $L_{seuil}^*$ , définissant l'abscisse du sommet du boomerang, alors le fruit sera considéré comme jaune, il sera ensuite soumis à un modèle prédictif spécifique pour fruit jaune utilisant la couleur  $a^*$  du fruit et sa surface. Si par contre la valeur de la bande  $L^*$  du fruit est inférieure ou égale à  $L_{seuil}^*$ , le fruit sera considéré comme rouge, il sera ensuite soumis à un autre modèle prédictif utilisant la couleur  $a^*$  du fruit et sa surface. Enfin, une fois que le Tacy de chaque fruit sera établi, le Tacy du lot complet sera déduit en effectuant la moyenne des Tacy individuels des fruits pondérés par leur surface. Ainsi, si un fruit possède une plus grande surface que les autres fruits du lot, sa contribution au Tacy du lot sera plus importante. L'algorithme suivant explique les étapes vues précédemment :

Soit un lot de canneberge comportant  $N$  fruits. Pour chacun des fruits  $i$ ,  $i$  allant de 1 à  $N$  :

si  $L_i^* > L_{seuil}^*$  (fruit jaune) :

$$Tacy_{fruit\_i} = -12,43 + 0,13 \cdot a_i^* \quad (4.51)$$

sinon (fruit rouge):

$$Tacy_{fruit\_i} = 280,88 - 1,42 \cdot a_i^* - 4,33 \cdot 10^{-3} \cdot surface_{fruit\_i}$$

Puis le Tacy du lot est donné par la formule suivante :

$$Tacy_{lot} = \frac{\sum_{i=1}^N Tacy_{fruit\_i} \cdot surface_{fruit\_i}}{\sum_{i=1}^N surface_{fruit\_i}} \quad (4.52)$$

La valeur de  $L_{seuil}^*$  déterminée à partir de la base de données 2005 est égale à 110.

Le modèle utilisé pour évaluer le Tacy des fruits de couleur jaune ou pâle, est un modèle prédictif établi à partir des échantillons standard de Tacy inférieur à 10, et présente une

erreur RMSE de 0,79 sur ces dernières. Il utilise uniquement la variable  $a^*$  car pour les fruits jaune ou pâle l'influence de la surface sur le Tacy est négligeable.

Le modèle hybride peut également être formulé sous une autre forme. Les modèles prédictifs du Tacy peuvent être écrits sous la forme générale :

$$Tacy_{fruit\_i} = \alpha(L_i^*) \cdot a_i^* + \beta(L_i^*) \cdot surface_{fruit\_i} + \gamma(L_i^*) \quad (4.53)$$

Avec  $\alpha(L_i^*)$ ,  $\beta(L_i^*)$  et  $\gamma(L_i^*)$  des fonctions définissant les coefficients des modèles prédictifs. Si  $L_i^* > L_{seuil}$ , alors les coefficients appartiennent au modèle prédictif pour les fruits jaunes, sinon au modèle prédictif pour les fruits rouges.

Par la suite, afin d'alléger l'écriture, le terme  $(L_i^*)$  des fonctions ne sera plus répété.

Après recombinaison des termes, le Tacy du lot se reformule alors :

$$Tacy_{lot} = \alpha \cdot \left( \frac{\sum_{i=1}^N a_i^* \cdot surface_{fruit\_i}}{\sum_{i=1}^N surface_{fruit\_i}} \right) + \beta \cdot \left( \frac{\sum_{i=1}^N surface_{fruit\_i}^2}{\sum_{i=1}^N surface_{fruit\_i}} \right) + \gamma \quad (4.54)$$

Ce modèle hybride est linéaire suivant la couleur. Cela veut dire que la contribution de la couleur au modèle prédictif est la même si l'on considère la couleur moyenne d'un lot (à condition qu'il soit homogène) ou bien la moyenne des couleurs individuelles des fruits avec pondération par la surface des fruits :

$$couleur\_moyenne_{lot} = \frac{\sum_{i=1}^N couleur_{fruit\_i} \cdot surface_{fruit\_i}}{\sum_{i=1}^N surface_{fruit\_i}} \quad (4.55)$$

Par contre, le modèle hybride n'est pas linéaire en fonction de la surface des fruits. Cela veut dire que la contribution de la surface au modèle prédictif n'est la même si l'on considère la surface moyenne des fruits d'un lot ou bien la moyenne des surfaces individuelles des fruits avec pondération par la surface des fruits :

$$\frac{\sum_{i=1}^N surface_{fruit\_i}}{N} \neq \frac{\sum_{i=1}^N surface_{fruit\_i}^2}{\sum_{i=1}^N surface_{fruit\_i}} \quad (4.56)$$

Cela entraîne une erreur sur le terme de la surface moyenne lorsque l'on utilise le modèle hybride au lieu du modèle classique. Néanmoins, l'inégalité ci-dessus devient une égalité lorsque toutes les surfaces des fruits ( $surface_{fruit}$ ) du lot sont égales entre elles :

$$\frac{N \cdot surface_{fruit}}{N} = \frac{N \cdot surface_{fruit}^2}{N \cdot surface_{fruit}} \quad (4.57)$$

Ainsi plus les surfaces des fruits dans un lot sont similaires, plus la non-linéarité (et l'erreur) tend à disparaître. Cette erreur est due à la normalisation en fonction de la surface des fruits, mais elle est inévitable si l'on considère les modèles prédictifs tenant compte de la surface des fruits. D'autres types de normalisation, sur le nombre de fruits par exemple, entraîneraient également ce type d'erreur.

Le Tableau XXXI montre les résultats obtenus sur la base de test comportant des mélanges extrêmes (indices d'échantillon 16 et 17) ainsi que des lots formés de fruits jaune ou pâle (indices d'échantillon 4 et 19). L'erreur RMSE sur cette base de test est de

2,82. De plus, tous les échantillons testés respectent nos limites préétablies d'erreur type de  $\pm 3$  unités (résultats détaillés non présentés). Ce modèle prédictif hybride permet ainsi de traiter toutes les couleurs de canneberges, allant du jaune au rouge foncé, mais également les lots de type mélange extrême, tout en ayant une bonne précision sur les hautes valeurs de Tacy grâce au modèle en a\*. En contrepartie, des erreurs sur le terme de la surface sont introduites à cause de la non-linéarité de la surface lors de la normalisation.

Tableau XXXI

Prédiction du Tacy avec modèle hybride pour les échantillons de la base de test  
(sans données aberrantes, avec mélanges extrêmes)

Indice de l'échantillon #	Tacy chimique laboratoire	Tacy prédit (regroupement des 4 photos)
1	40	41
2	49	45
3	49	47
4	20	19
5	44	43
6	47	48
7	63	61
8	53	52
9	53	48
10	53	49
11	47	46
12	47	49
13	47	48
14	49	47
15	48	44
16	27	27
17	33	28
18	44	41
19	8	10
20	57	56
21	57	54
22	57	54
23	48	42
24	53	52
25	53	50
26	53	49
27	40	40

### 5.3.6 Validation avec les données de la récolte 2006

#### 5.3.6.1 Base de données 2006

La base de données recueillie lors de la saison 2006 contient plus de 1700 photos de lots commerciaux de canneberges. Cette base de données impressionnante a pu être collectée grâce à l'aide des techniciennes du laboratoire d'analyse chimique. Un ensemble de 111 lots qui a été segmenté constituera notre base de données afin de valider la méthode établie lors de la saison 2005. Les paramètres d'acquisition ont grandement été modifiés entre la saison 2005 et 2006, notamment la configuration de la scène est surtout le modèle de la caméra numérique. Il est donc nécessaire de régénérer les coefficients du modèle prédictif afin qu'ils soient adaptés aux données de 2006. Une partie de la base de données de 2006 servira à estimer les nouvelles valeurs des coefficients de l'équation du modèle prédictif, tandis que la totalité de la base de données sera utilisée pour la validation du modèle. À cause de l'effet boomerang de la couleur canneberge, seuls les lots de couleurs homogènes peuvent être utilisés pour le calcul des coefficients. Néanmoins, aucun lot homogène de couleur standard n'a été créé expérimentalement durant la récolte 2006. Pour pallier cet inconvénient, des lots homogènes pouvant être assimilés à des lots standard ont été recherchés dans la base de données. Un ensemble de 10 lots a finalement été sélectionné pour former la base d'apprentissage. La couleur de ces lots va du rouge au rouge très foncé ; aucun lot homogène de couleur jaune ou pâle n'a pu être trouvé. Ces lots ne sont toutefois pas parfaitement homogènes, ils contiennent soit quelques fruits jaunes, soit des fruits de plusieurs intensités de rouge différentes.

Voici la plage [min ; max] des variables pour les échantillons 2006 considérés comme standard :

a.  $a^* : [151 : 165]$ ;

- b. surface moyenne : [1430 :1775];
- c. Tacy :[38 : 63].

Les lots de couleur très foncée, ayant des valeurs de  $a^*$  inférieur à 150, n'ont pas été retenu pour le calcul des coefficients, car la variable  $a^*$  n'est plus proportionnelle au Tacy si cette gamme de couleurs est également considérée.

### 5.3.6.2 Modèle prédictif

Le modèle prédictif hybride est sélectionné pour la prédiction du Tacy des données de 2006. Les coefficients de ce modèle doivent être réévalués à cause du changement de caméra et de disposition de la scène entre 2005 et 2006. Comme il a été mentionné précédemment, un ensemble de 10 lots homogènes de couleur rouge présents dans la base de données servira à recalculer le modèle prédictif pour les fruits rouges du modèle hybride. Le modèle prédictif pour les fruits jaunes ne pourra pas être recalculé à cause de l'absence de données, le modèle de 2005 sera ainsi conservé comme approximation. La valeur de  $L_{seuil}^*$  est déterminée égale à 100.

Le modèle prédictif pour les fruits rouges est le suivant :

$$Tacy\_prédit = 279 - 1,39 \cdot a^* - 0,00542 \cdot surface\_moyenne \quad (4.58)$$

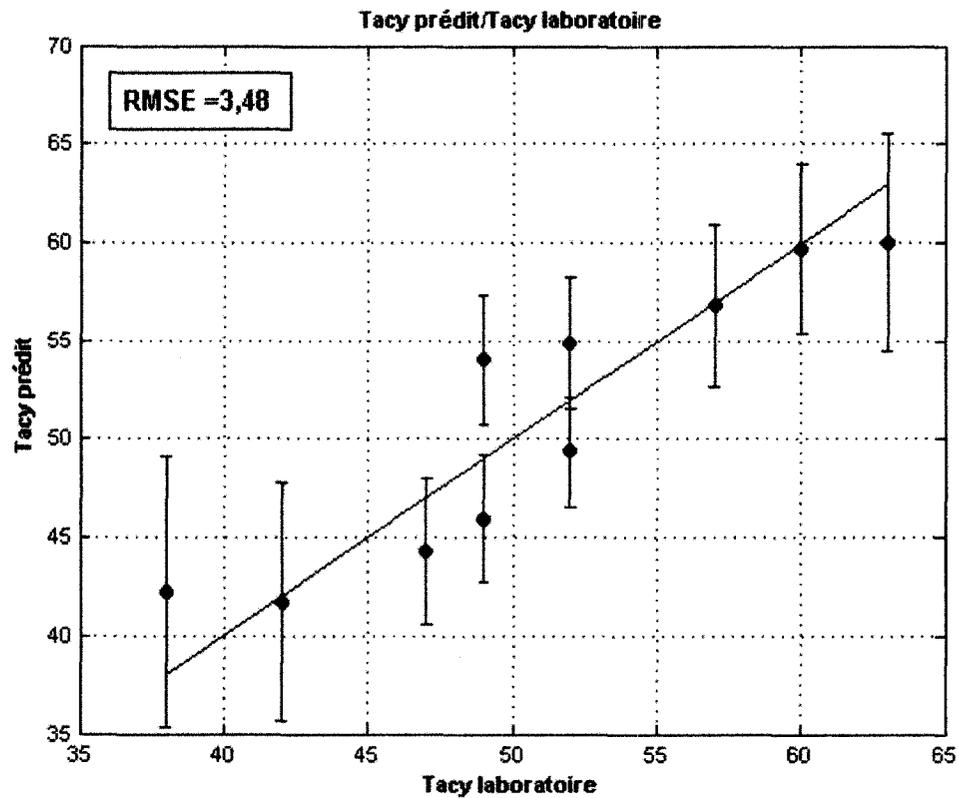


Figure 72 Modélisation du Tacy laboratoire sur les données de 2006

Avec :

- a. RMSE (sur données de modélisation) : 3,48 ;
- b.  $R^2 = 0,84$ .

Il est possible de remarquer que la qualité de la modélisation est inférieure à celle obtenue en 2005. Ceci peut être attribué au fait que les lots utilisés n'ont pas été standardisés comme lors de la saison précédente. L'autre explication provient du fait que cette saison, les photos ont été prises par les techniciennes du laboratoire, chacune d'elle possède ses propres habitudes de placement vis-à-vis du montage et influence ainsi l'éclairage de la scène par leur ombre portée. Tous ces lots n'ont pas été photographiés le même jour, donc plusieurs techniciennes ont participé à la prise de ces photos. De

plus, durant toute la période de la photographie de ces lots, le montage a aussi été déplacé, influençant également l'éclairage de la scène.

Il est également possible de voir que les coefficients des modèles de prédiction des fruits rouges des saisons 2005 et 2006 sont très similaires :

Modèle prédictif des fruits rouges 2005 :

$$Tacy\_prédit = 280,88 - 1,42 \cdot a^* - 4,33 \cdot 10^{-3} \cdot surface\_moyenne \quad (4.59)$$

Modèle prédictif des fruits rouges 2006 :

$$Tacy\_prédit = 279,00 - 1,39 \cdot a^* - 5,42 \cdot 10^{-3} \cdot surface\_moyenne$$

Ce résultat semble a priori surprenant, car les caméras A70 et A620 utilisées respectivement en 2005 et 2006 sont différentes, donc leur comportement colorimétrique devrait différer. Néanmoins, les deux caméras appartiennent à la même famille de produits du constructeur Canon, dont la caméra A620 est une version plus performante de la caméra A70, capable de restituer le gamut sRGB avec plus de précision. La similarité des deux modèles prédictifs montre que le comportement colorimétrique des deux caméras utilisées est le même sur les couleurs du fruit de la canneberge. La légère différence entre les coefficients des modèles peut provenir du changement de caméra comme il a été stipulé, ou bien également de la qualité inférieure de la base de données utilisé en 2006 pour la modélisation.

Le modèle hybride utilisé pour les données de 2006 est donc le suivant pour un lot de canneberge comportant N fruits:

. Pour chacun des fruits  $i$ ,  $i$  allant de 1 à N :

si  $L_i^* > L_{seuil}^*$  (fruit jaune) :

$$Tacy_{fruit\_i} = -12,43 + 0,13 \cdot a_i^* \quad (4.60)$$

sinon (fruit rouge):

$$Tacy_{fruit\_i} = 279 - 1,39 \cdot a_i^* - 5,42 \cdot 10^{-3} \cdot surface_{fruit\_i}$$

Puis le Tacy du lot est donné par la formule suivante :

$$Tacy_{lot} = \frac{\sum_{i=1}^N Tacy_{fruit\_i} \cdot surface_{fruit\_i}}{\sum_{i=1}^N surface_{fruit\_i}} \quad (4.61)$$

### 5.3.6.3 Validation du modèle hybride 2006

Le modèle hybride 2006 est validé sur la totalité de la base de données de 2006. La base de validation comprend ainsi les 10 lots ayant servi à la modélisation, car ces derniers n'étant pas de vrais lots standard, leur comportement dans le modèle hybride peut être différent de celui dans le modèle qu'ils ont généré. Les valeurs de Tacy de la base de données vont de 23 jusqu'à 94.

L'erreur RMSE du modèle hybride sur la base de validation est de 5,40 (voir ANNEXE 7). Le coefficient de détermination  $R^2$  entre Tacy mesuré par le laboratoire et le Tacy prédit par le modèle est élevé ( $R^2 = 0,82$  voir Figure 73). Cela indique que le Tacy prédit suit la tendance du Tacy laboratoire. Des hypothèses d'erreur type de  $\pm 2$  et  $\pm 3$  sur la valeur du Tacy du laboratoire ont été considérées. L'hypothèse d'une erreur type de  $\pm 3$  sur la valeur du Tacy prédit est effectuée. Il apparaît qu'en considérant une erreur type de  $\pm 2$ , le pourcentage de prédiction réussie est de 70%, tandis qu'en considérant une erreur type de  $\pm 3$  le pourcentage de prédiction réussie est de 77% (voir Tableau XXXII).

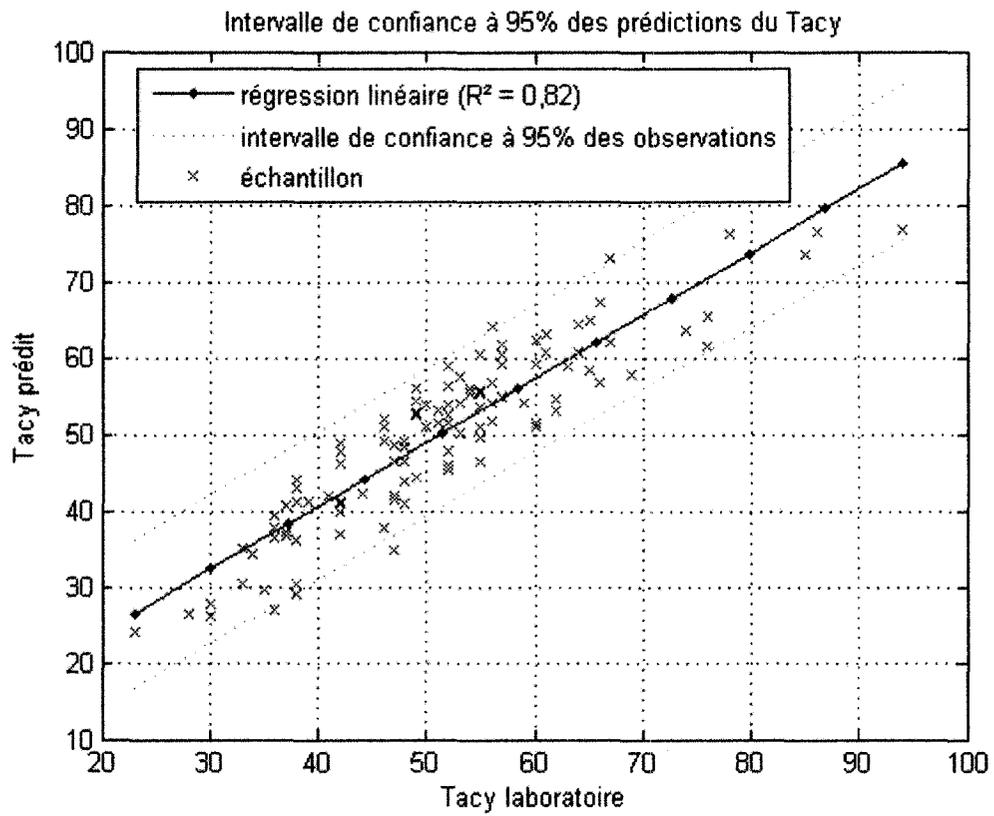


Figure 73 Intervalle de confiance à 95% des prédictions du Tacy  
avec regroupement des 4 photos

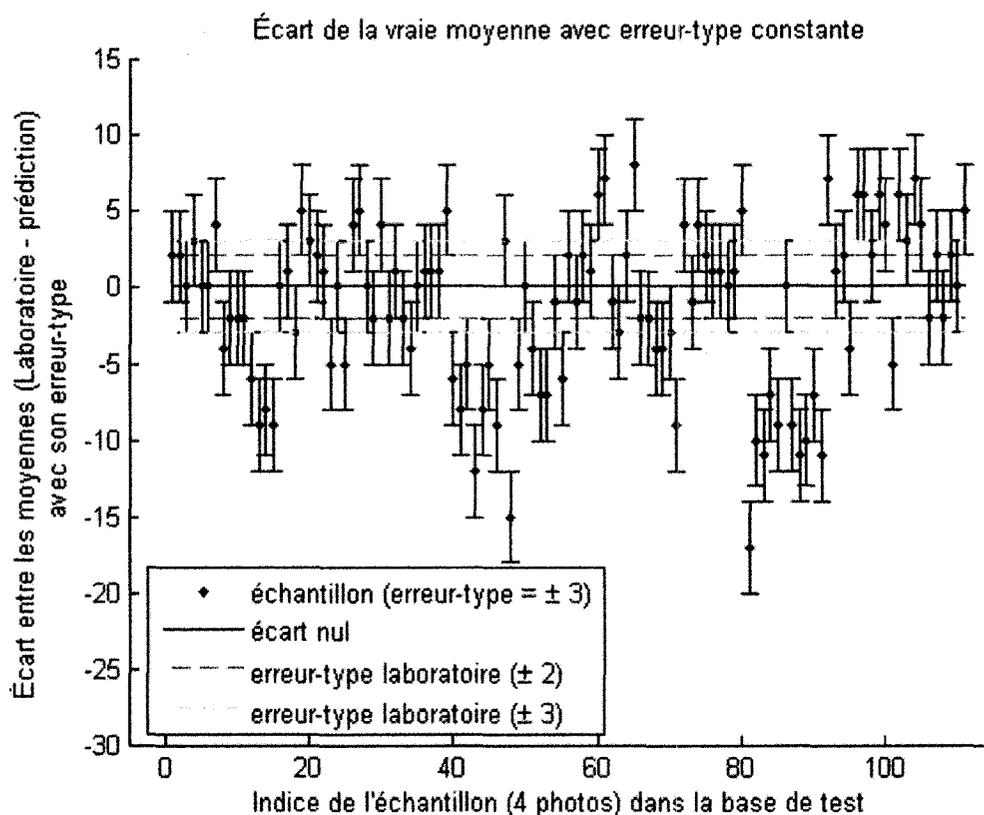
Figure 74 Modèle d'erreur type de  $\pm 3$  avec regroupement des 4 photos

Tableau XXXII

Statistique de succès de prédiction avec une erreur type de  $\pm 3$ 

	erreur type laboratoire de $\pm 2$	erreur type laboratoire de $\pm 3$
nombre d'échantillons	111	111
nombre de succès	78	86
nombre d'échecs	33	25
pourcentage de succès	70%	77%

La valeur marchande d'un lot de canneberge augmente en fonction de la quantité de Tacy présent dans les fruits, mais à partir d'un Tacy de 40 mg par 100 g la valeur marchande atteint un prix plafond. Un lot dont le Tacy sera plus élevé que 40 aura la

même valeur marchande qu'un lot de ayant un Tacy de 40. Il est donc intéressant de regarder les pourcentages de succès des prédictions sur deux plages distinctes de variation de Tacy. La première comprend les valeurs faibles de Tacy (Tacy laboratoire  $\leq 50$ ), zone où la valeur marchande du lot dépend du Tacy. La seconde comprend les valeurs élevées de Tacy (Tacy laboratoire  $> 50$ ), zone où la valeur marchande du lot est indépendante du Tacy. La valeur intermédiaire de 50 est choisie au lieu 40 (plafond de paiement) pour tenir compte des erreurs types et ainsi étudier le comportement du modèle autour de la valeur critique de 40. Les résultats montre qu'en considérant une erreur type de  $\pm 3$ , sur la plage des Tacy laboratoire  $\leq 50$  le pourcentage de prédictions réussies est de 85%, tandis que sur la plage des Tacy laboratoire  $> 50$  ce pourcentage n'est que de 71% (voir Tableau XXXIII et Tableau XXXIV). Le modèle prédictif est ainsi plus précis sur la plage Tacy laboratoire  $\leq 50$ . Ceci est normal, car les lots de Tacy supérieur à 60 ont été exclus des données ayant servi à déterminer les nouveaux coefficients du modèle, car ils entraînent une non-linéarité dans la relation. Le modèle offre donc un plus grand pourcentage de succès dans la zone où la valeur marchande du lot dépend du Tacy. Cependant, le pourcentage de succès plus faible dans la plage des Tacy élevés n'a pas d'influence sur la valeur marchande des lots. En effet, pour toutes les valeurs de Tacy prédites par le modèle supérieures à 50, les Tacy laboratoire correspondant sont tous supérieurs ou égaux à 46, ce qui est bien au-dessus de la valeur critique de 40.

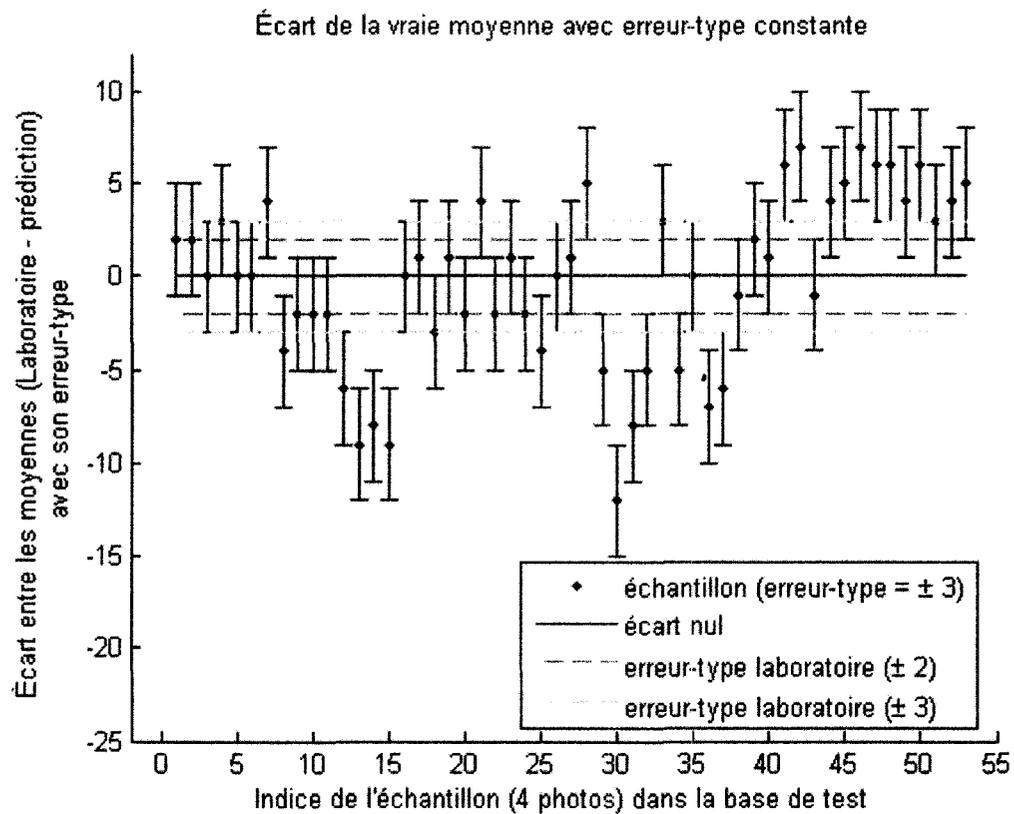


Figure 75 Modèle d'erreur type de  $\pm 3$  avec regroupement des 4 photos  
pour Tacy laboratoire  $\leq 50$

Tableau XXXIII

Statistique de succès de prédiction avec une erreur type de  $\pm 3$  pour  
Tacy laboratoire  $\leq 50$

	erreur type laboratoire de $\pm 2$	erreur type laboratoire de $\pm 3$
nombre d'échantillons	53	53
nombre de succès	39	45
nombre d'échecs	14	8
pourcentage de succès	74%	85%

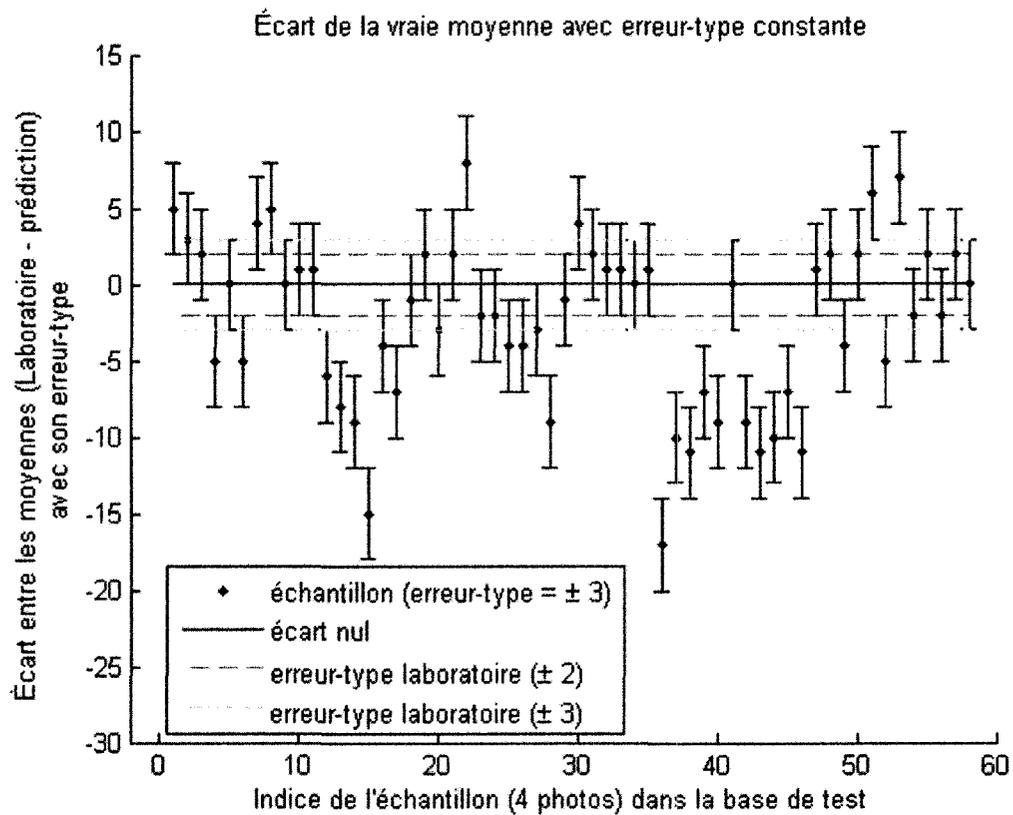


Figure 76 Modèle d'erreur type de  $\pm 3$  avec regroupement des 4 photos  
pour Tacy laboratoire  $> 50$

Tableau XXXIV

Statistique de succès de prédiction avec une erreur type de  $\pm 3$  pour  
Tacy laboratoire  $> 50$

	erreur type laboratoire de $\pm 2$	erreur type laboratoire de $\pm 3$
nombre d'échantillons	58	58
nombre de succès	39	41
nombre d'échecs	19	17
pourcentage de succès	67%	71%

## 5.4 Tacy numérique

Voici le résumé de la démarche permettant de prédire le Tacy. Dans un premier temps des lots ayant des canneberges homogènes en taille et en couleur doivent être formés, ces échantillons dit standards permettront d'établir les coefficients du modèle prédictif. Ce dernier est constitué de deux parties distinctes, chacune modélisant une branche du boomerang de la couleur canneberge. Les échantillons standards de couleur rouge à rouge foncé serviront à former le modèle pour les Tacy élevés, à partir de leur couleur moyenne  $a^*$  et de leur taille moyenne. Les échantillons standards de couleur jaune à rose serviront à former le modèle pour les Tacy faibles, uniquement à partir de leur couleur moyenne  $a^*$ .

Une fois les coefficients du modèle prédictif déduit, il est possible de prédire le Tacy pour chacune des quatre photos (répétitions) d'un lot.

Soit une des quatre photos d'un lot de canneberge comportant  $N$  fruits. Pour chacun des fruits  $i$ ,  $i$  allant de 1 à  $N$  (coefficient du modèle 2005) :

$$\begin{aligned} &\text{si } L_i^* > L_{seuil}^* \text{ (fruit jaune) :} \\ &\quad Tacy_{fruit\_i} = -12,43 + 0,13 \cdot a_i^* \end{aligned} \tag{4.62}$$

sinon (fruit rouge):

$$Tacy_{fruit\_i} = 280,88 - 1,42 \cdot a_i^* - 4,33 \cdot 10^{-3} \cdot surface_{fruit\_i}$$

Puis le Tacy de la photo du lot est donné par la formule suivante :

$$Tacy_{lot} = \frac{\sum_{i=1}^N Tacy_{fruit\_i} \cdot surface_{fruit\_i}}{\sum_{i=1}^N surface_{fruit\_i}} \tag{4.63}$$

La moyenne des quatre Tacy obtenus à partir des 4 répétitions est considérée comme le Tacy prédit par le modèle. Les résultats de ce chapitre montrent que cette approche permet d'obtenir des taux de succès de 100% sur les données de 2005 et de 77% sur les données de 2006 en considérant un terme d'erreur type de  $\pm 3$  unités.

## 5.5 Conclusion

Dans cette partie, le modèle prédictif du Tacy à partir des photographies des lots de canneberges a été établi en considérant un terme d'erreur type de  $\pm 3$  unités. Ce modèle utilise la variable colorimétrique  $a^*$  de l'espace ICCLAB et la surface des fruits de canneberges. La surface des fruits a été modélisée par le nombre de pixels du fruit dans l'image, en prenant l'hypothèse d'un modèle sphérique pour la forme des canneberges. Les résultats des régressions des modèles montrent que la dimension (surface) des fruits influence le Tacy ; plus les fruits d'un lot sont petits, plus le Tacy du lot est élevé. Sur les données de la saison 2005, il apparaît que le pourcentage de succès des prédictions est plus élevé lorsque la moyenne des Tacy prédits par les quatre photos est considérée, plutôt que le Tacy prédit par une photo individuelle. Le problème de l'effet boomerang de la variable  $a^*$  sur la couleur canneberge est résolu en considérant un modèle prédictif hybride, dans lequel un modèle prédictif différent est appliqué au fruit selon son niveau de luminance. Ce modèle hybride permet d'obtenir 100% de prédictions réussies sur les données de 2005 en enlevant les données aberrantes, et 77% sur les données de 2006.

## CONCLUSION

L'objectif de ce travail est d'établir une méthode non destructive, s'appuyant sur la vision par ordinateur pour prédire le taux d'anthocyane des fruits de canneberge. Cette méthode doit à long terme remplacer la méthode d'extraction chimique de l'anthocyane, car cette dernière entraîne le rejet de produits chimiques dans l'environnement et nécessite beaucoup de temps. La méthode non destructive doit posséder des termes d'écart-types (SD) et d'erreurs types (SEM) équivalents à ceux de la méthode chimique.

La littérature montre qu'il existe une relation entre la couleur externe des fruits, leur calibre et le Tacy. Le matériel utilisé a permis d'acquérir les photos des échantillons de canneberge afin d'en extraire les couleurs et leur surface. Ces photos ont été segmentées à partir de l'algorithme de partage des eaux, néanmoins la segmentation nécessite une correction par un expert humain. L'ANOVA à une entrée a démontré qu'on ne prend qu'un risque  $\alpha = 0,001$  en considérant les quatre photos avec brassage aléatoire d'un même échantillon comme identique. Il est ainsi statistiquement correct de regrouper les informations extraites des quatre photos pour une plus grande précision.

L'étude de la couleur canneberge sur les échantillons de couleur standard dans l'espace colorimétrique ICCLAB a montré que la variable couleur possédant la plus grande variabilité est la bande  $L^*$ . De plus, la couleur canneberge possède la forme d'un boomerang, où les branches du boomerang sont similaires à deux segments de droite. L'analyse par composantes principales a permis de réduire à deux dimensions l'espace de représentation de la couleur canneberge, puis une régression linéaire a permis de modéliser la couleur canneberge, sous forme d'une échelle couleur à une dimension.

Enfin, le modèle prédictif du Tacy à partir des photographies des lots de canneberges a été établi en considérant un terme d'erreur type de  $\pm 3$  unités. Divers modèles de

régression linéaire multiple ont été testés. Le modèle le plus performant utilise la variable colorimétrique  $a^*$  de l'espace ICCLAB et la surface des fruits de canneberges. La surface des fruits a été modélisée par le nombre de pixels du fruit dans l'image, en prenant l'hypothèse d'un modèle sphérique pour la forme des canneberges. Les résultats des régressions des modèles confirment que la dimension (surface) des fruits influence le Tacy. Sur les données de la saison 2005, il apparaît que le pourcentage de succès des prédictions est plus élevé lorsque la moyenne des Tacy prédits par les quatre photos est considérée. Le problème de l'effet boomerang de la variable  $a^*$  sur la couleur canneberge est résolu en considérant un modèle prédictif hybride, dans lequel un modèle prédictif différent est appliqué au fruit selon son niveau de luminance. Ce modèle hybride permet d'obtenir 100% de prédictions réussies sur les données de 2005 en enlevant les données aberrantes, et 77% sur les données de 2006.

## RECOMMANDATIONS

Les résultats montrent que les prédictions de Tacy sur les données de 2006 sont moins fiables que sur les données de la saison 2005. La raison la plus probable à cette baisse de performance est l'absence de normalisation de l'environnement du laboratoire. En effet, le laboratoire étant de dimension réduite, le montage peut être déplacé ou des objets peuvent être placés à ses côtés, altérant la qualité de l'éclairage de la scène. Aussi, plusieurs techniciennes du laboratoire chimique se sont portées volontaire pour prendre les clichés des lots de canneberges en notre absence en 2006. Chacune d'entre elles possède sa propre manière de procéder, influençant d'une certaine manière l'éclairage de la scène. Les clichés de 2006 n'ont donc pas été tous pris dans les mêmes conditions, contrairement à la saison 2005 où un unique manipulateur formé prenait les photos. Ceci prouve la nécessité de travailler dans un environnement standardisé. L'utilisation d'une enceinte normalisée permettrait d'assurer la constance des mesures colorimétrique en contrôlant l'environnement et l'illuminant de la scène. Il serait possible d'utiliser le système de manière identique quelque soit l'environnement externe ou le manipulateur. Enfin, l'utilisation d'un éclairage arrière de la scène combinée avec l'enceinte normalisée permettrait de résoudre les difficultés de segmentation de l'image en obtenant directement le masque binaire des fruits.

Les travaux futurs devront explorer le volet analytique de la relation entre le Tacy, la couleur et la surface des fruits. La couleur devra être analysée directement au niveau des pixels des fruits, sans effectuer de moyenne qui engendre une perte d'information. Ainsi les pixels de couleur jaune qui ne contribuent pas au Tacy devront être directement seuillés. Aussi, la mesure de la surface des fruits devra être modifiée, en utilisant cette fois-ci le modèle sphéroïdal qui semble être plus précis. Enfin lorsque le Tacy d'un échantillon est directement connu, la méthode devra prédire le Tacy chimique sans tenir compte de l'erreur type, qui s'applique uniquement à l'incertitude sur le Tacy du lot complet.

## **ANNEXE 1**

### **Spécifications des illuminants fluorescents**

Les spécifications des différentes familles des illuminants fluorescents sont résumées dans les tableaux et figures suivantes.

Tableau XXXV  
Spécificités des illuminants fluorescents standard

Illuminant CIE	Coordonnée x	Coordonnée y	Température couleur (K)	IRC	dénomination usuelle
F1	0,3131	0,3371	6430	76	<i>Daylight</i>
F2	0,3721	0,3751	4230	64	<i>Cool White</i>
F3	0,4091	0,3941	3450	57	<i>White</i>
F4	0,4402	0,4031	2940	51	<i>Warm White</i>
F5	0,3138	0,3452	6350	72	<i>Daylight</i>
F6	0,3779	0,3882	4150	59	<i>Lite White</i>

Tableau XXXVI  
Spécificités des illuminants fluorescents à large bande

Illuminant CIE	Coordonnée x	Coordonnée y	Température couleur (K)	IRC	dénomination usuelle
F7	0,3129	0,3292	6500	90	D65 <i>Daylight</i>
F8	0,3458	0,3586	5000	95	D50 Sylvania F40 <i>design 50</i>
F9	0,3741	0,3727	4150	90	<i>Cool White</i> <i>Deluxe</i>

Tableau XXXVII  
Spécificités des illuminants fluorescents à bande étroite

Illuminant CIE	Coordonnée x	Coordonnée y	Température couleur (K)	IRC	dénomination usuelle
F10	0,3458	0,3588	5000	81	TL85 Ultralume 50
F11	0,3805	0,3769	4000	83	TL84 SP41 Ultralume 40
F12	0,4370	0,4042	3000	83	TL83 Ultralume 30

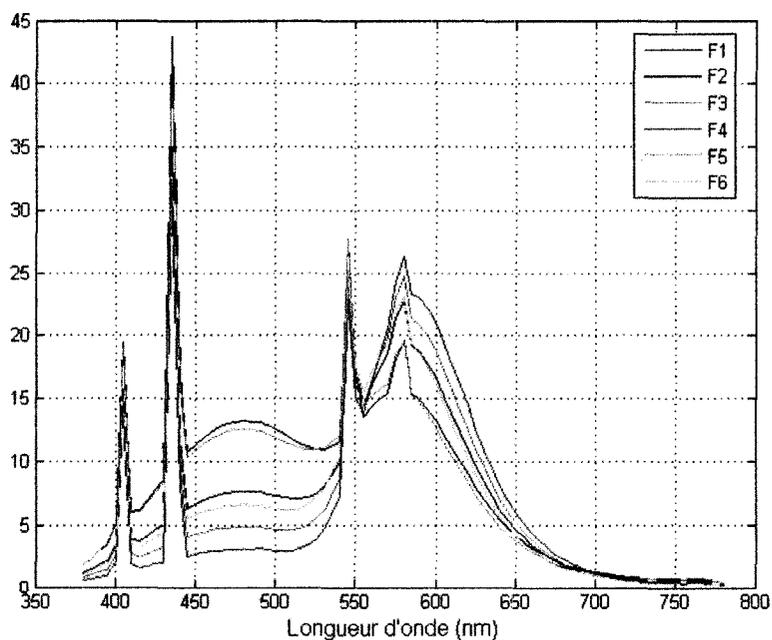


Figure 77 Spectres des illuminants fluorescents standards

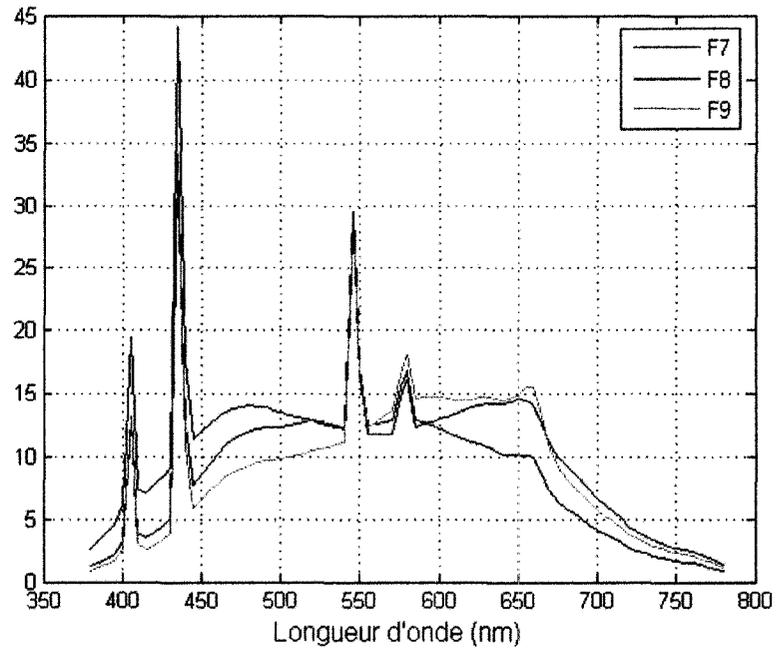


Figure 78 Spectres des illuminants fluorescents à bande large

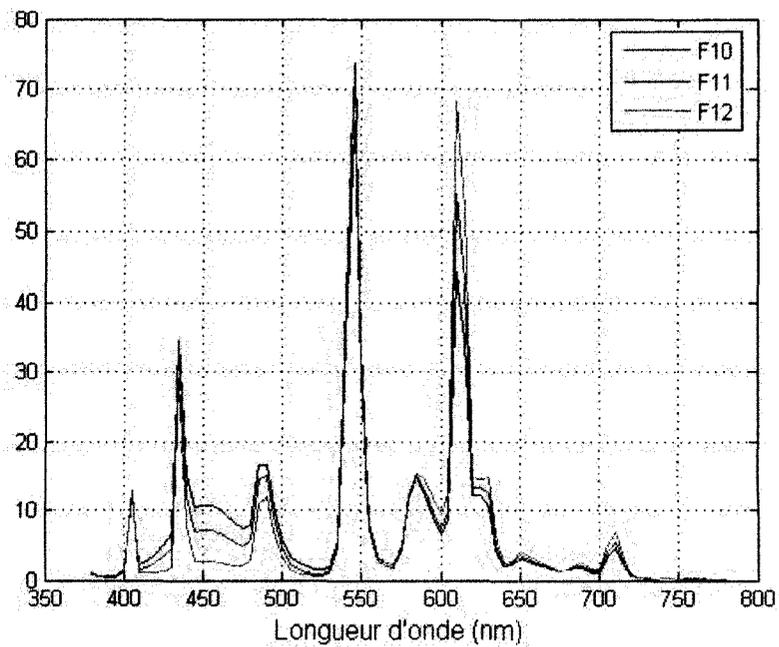


Figure 79 Spectres des illuminants fluorescents à bande étroite

## **ANNEXE 2**

**Mire couleur GretagMacbeth 24 cases**

Les deux figures suivantes identifient la mire couleur 24 cases de GretagMacbeth et indiquent les valeurs de chacune des cases dans différents espaces colorimétriques.

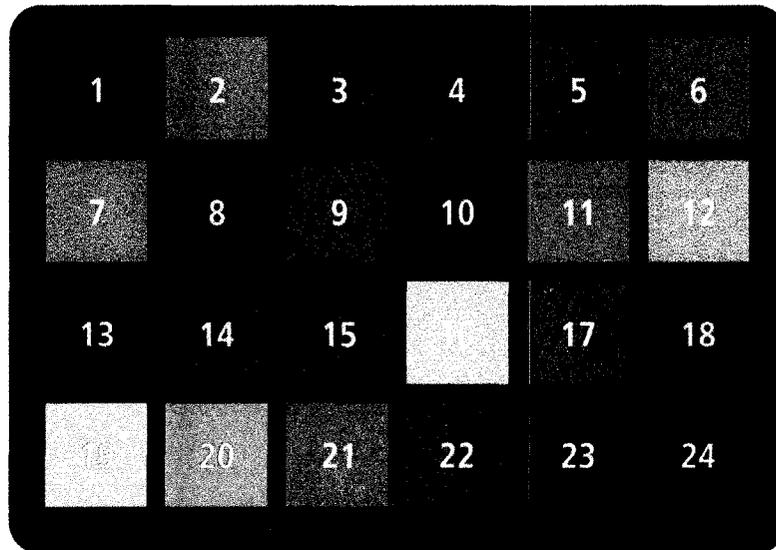


Figure 80 Modèle de la mire couleur 24 cases de GretagMacbeth

*(Tiré de [80])*

		sRGB								
		R	G	B						
1.	dark skin	115	82	68	37.986	13.555	14.059	3 YR	3.7 / 3.2	
2.	light skin	194	150	130	65.711	18.13	17.81	2.2 YR	6.47 / 4.1	
3.	blue sky	98	122	157	49.927	-4.88	-21.925	4.3 PB	4.95 / 5.5	
4.	foliage	87	108	67	43.139	-13.095	21.905	6.7 GY	4.2 / 4.1	
5.	blue flower	133	128	177	55.112	8.844	-25.399	9.7 PB	5.47 / 6.7	
6.	bluish green	103	189	170	70.719	-33.397	-0.199	2.5 BG	7 / 6	
7.	orange	214	126	44	62.661	36.067	57.096	5 YR	6 / 11	
8.	purplish blue	80	91	166	40.02	10.41	-45.964	7.5 PB	4 / 10.7	
9.	moderate red	193	90	99	51.124	48.239	16.248	2.5 R	5 / 10	
10.	purple	94	60	108	30.325	22.976	-21.587	5 P	3 / 7	
11.	yellow green	157	188	64	72.532	-23.709	57.255	5 GY	7.1 / 9.1	
12.	orange yellow	224	163	46	71.941	19.363	67.857	10 YR	7 / 10.5	
13.	blue	56	61	150	28.778	14.179	-50.297	7.5 PB	2.9 / 12.7	
14.	green	70	148	73	55.261	-38.342	31.37	0.25 G	5.4 / 8.65	
15.	red	175	54	60	42.101	53.378	28.19	5 R	4 / 12	
16.	yellow	231	199	31	81.733	4.039	79.819	5 Y	8 / 11.1	
17.	magenta	187	86	149	51.935	49.986	-14.574	2.5 RP	5 / 12	
18.	cyan	8	133	161	51.038	-28.631	-28.638	5 B	5 / 8	
19.	white (.05*)	243	243	242	96.539	-0.425	1.186	N	9.5 /	
20.	neutral 8 (.23*)	200	200	200	81.257	-0.638	-0.335	N	8 /	
21.	neutral 6.5 (.44*)	160	160	160	66.766	-0.734	-0.504	N	6.5 /	
22.	neutral 5 (.70*)	122	122	121	50.867	-0.153	-0.27	N	5 /	
23.	neutral 3.5 (.105*)	85	85	85	35.656	-0.421	-1.231	N	3.5 /	
24.	black (1.50*)	52	52	52	20.461	-0.079	-0.973	N	2 /	

Cie L\*a\*b\* values use Illuminant D50 2 degree observer sRGB values for Illuminate D65.

Figure 81 Valeurs des cases de la mire couleur 24 cases GretagMacbeth  
(Tiré de [80])

## **ANNEXE 3**

**Analyse de variabilité typique entre les 4 photos d'un même échantillon ou lot**

Tableau XXXVIII  
Analyse de variabilité de l'échantillon standard A2

<b>L*</b>				
	REP 1	REP 2	REP 3	REP 4
Moyenne	56,0	53,3	54,2	56,5
SD	18,1	16,8	18,8	17,5
SEM	1,9	1,8	2,0	1,9
CV%	<b>32,2</b>	<b>31,6</b>	<b>34,7</b>	<b>31,0</b>
Max	100,6	99,5	107,3	114,2
Min	24,3	23,9	27,1	26,4
Étendue	76,3	75,6	80,2	87,8
N = S =				
<b>a*</b>				
	REP 1	REP 2	REP 3	REP 4
Moyenne	155,7	154,5	154,6	156,4
SD	11,1	10,4	11,8	10,5
SEM	1,2	1,1	1,3	1,1
CV%	<b>7,2</b>	<b>6,7</b>	<b>7,7</b>	<b>6,7</b>
Max	181,7	176,4	181,3	181,0
Min	132,9	129,3	132,4	130,8
Étendue	48,7	47,1	49,0	50,2
N = S =				
<b>b*</b>				
	REP 1	REP 2	REP 3	REP 4
Moyenne	147,1	146,3	146,5	147,5
SD	8,4	7,8	8,9	8,0
SEM	0,9	0,8	0,9	0,9
CV%	<b>5,7</b>	<b>5,4</b>	<b>6,1</b>	<b>5,4</b>
Max	167,5	164,6	165,6	167,2
Min	131,7	130,0	130,9	129,8
Étendue	35,8	34,6	34,7	37,5
N = S =				
<b>Taille</b>				
	REP 1	REP 2	REP 3	REP 4
Moyenne	1443,3	1486,2	1436,1	1467,7
SD	276,5	306,9	311,1	316,7
SEM	29,3	32,5	33,0	33,6
CV%	<b>19,2</b>	<b>20,6</b>	<b>21,7</b>	<b>21,6</b>
Max	2098,0	2339,0	2430,0	2249,0
Min	428,0	709,0	553,0	574,0
Étendue	1670,0	1630,0	1877,0	1675,0
N = S =				

Tableau XXXIX

Grand résumé de l'analyse de variabilité de l'échantillon standard A2

	L*	a*	b*	Taille	Nb objets
Moyenne					
SD	1,5	0,9	0,6	23,0	-
SEM	0,7	0,4	0,3	11,5	-
CV%	2,7	0,6	0,4	1,6	-
Max	56,5	156,4	147,5	1486,2	-
Min	53,3	154,5	146,3	1436,1	-
Etendue	3,2	1,9	1,2	50,2	-
N = S =					

Les coefficients de variation basés sur 4 moyennes par variable sont inférieurs à 3%, ce qui est bien inférieur au 10% généralement acceptable par les normes de l'industrie.

Tableau XL  
ANOVA à une entrée de l'échantillon standard A2

L*					
Source	degrés de liberté	Somme des carrés	Moyenne des carrés		
	3	586,0	195,3	0,61	NS <sup>1</sup>
	352	111936,0	318,0		
	355	112522,0			
a*					
Source	degrés de liberté	Somme des carrés	Moyenne des carrés		
	3	208,9	69,6	0,58	NS <sup>1</sup>
	352	42548,2	120,9		
	355	42757,2			
b*					
Source	degrés de liberté	Somme des carrés	Moyenne des carrés		
	3	86,9	29,0	0,42	NS <sup>1</sup>
	352	24249,5	68,9		
	355	24336,4			
Taille					
Source	degrés de liberté	Somme des carrés	Moyenne des carrés		
	3	141169,4	47056,5	0,51	NS <sup>1</sup>
	352	32361003,1	91934,7		
	355	32502172,5			

À un risque alpha ( $\alpha$  de 0,001), il n'y a pas de différence significative entre la moyenne des populations, et ce pour toutes les variables (L\*, a\*, b\*, taille). Il est donc possible de regrouper les objets des 4 répétitions.

<sup>1</sup> NS<0,001

Tableau XLI  
Analyse de variabilité de l'échantillon standard B2

<b>L*</b>				
	REP 1	REP 2	REP 3	REP 4
Moyenne	93,0	96,4	94,4	95,3
SD	19,6	21,7	22,0	20,1
SEM	2,1	2,3	2,3	2,1
CV%	21,1	22,5	23,3	21,1
Max	155,4	158,7	144,7	142,5
Min	50,7	56,8	45,1	54,6
Étendue	104,7	101,9	99,6	87,9
N = S =				
<b>a*</b>				
	REP 1	REP 2	REP 3	REP 4
Moyenne	173,4	174,7	174,3	175,6
SD	7,5	7,0	8,1	7,3
SEM	0,8	0,7	0,9	0,8
CV%	4,3	4,0	4,6	4,2
Max	187,5	188,7	189,6	188,1
Min	156,6	157,4	150,4	156,0
Étendue	30,9	31,3	39,2	32,1
N = S =				
<b>b*</b>				
	REP 1	REP 2	REP 3	REP 4
Moyenne	159,8	160,5	160,4	161,0
SD	5,3	4,7	5,5	4,7
SEM	0,6	0,5	0,6	0,5
CV%	3,3	2,9	3,4	2,9
Max	169,0	170,1	169,9	168,9
Min	146,9	148,6	141,6	146,3
Étendue	22,1	21,5	28,2	22,7
N = S =				
<b>Taille</b>				
	REP 1	REP 2	REP 3	REP 4
Moyenne	1459,6	1412,1	1415,9	1390,1
SD	293,5	270,3	308,1	278,5
SEM	31,3	28,8	32,8	29,7
CV%	20,1	19,1	21,8	20,0
Max	2178,0	2048,0	2082,0	2294,0
Min	764,0	731,0	652,0	784,0
Étendue	1414,0	1317,0	1430,0	1510,0
N = S =				

Tableau XLII

Grand résumé de l'analyse de variabilité de l'échantillon standard B2

	L*	a*	b*	Taille	Nb objets
Moyenne					
SD	1,4	0,9	0,5	29,1	-
SEM	0,7	0,4	0,2	14,5	-
CV%	1,5	0,5	0,3	2,0	-
Max	96,4	175,6	161,0	1459,6	-
Min	93,0	173,4	159,8	1390,1	-
Etendue	3,3	2,1	1,2	69,4	-
N = S =					

Les coefficients de variation basés sur 4 moyennes par variable sont inférieurs à 3%, ce qui est bien inférieur au 10% généralement acceptable par les normes de l'industrie.

Tableau XLIII  
ANOVA à une entrée de l'échantillon standard B2

<b>L*</b>					
Source	degrés de liberté	Somme des carrés	Moyenne des carrés		
	3	527,7	175,9	0,40	NS <sup>1</sup>
	348	151894,6	436,5		
	351	152422,3			
<b>a*</b>					
Source	degrés de liberté	Somme des carrés	Moyenne des carrés		
	3	206,8	68,9	1,23	NS <sup>1</sup>
	348	19454,7	55,9		
	351	19661,5			
<b>b*</b>					
Source	degrés de liberté	Somme des carrés	Moyenne des carrés		
	3	61,7	20,6	0,80	NS <sup>1</sup>
	348	8894,7	25,6		
	351	8956,4			
<b>Taille</b>					
Source	degrés de liberté	Somme des carrés	Moyenne des carrés		
	3	223279,3	74426,4	0,90	NS <sup>1</sup>
	348	28863156,3	82940,1		
	351	29086435,6			

À un risque alpha ( $\alpha$  de 0,001), il n'y a pas de différence significative entre la moyenne des populations, et ce pour toutes les variables (L\*, a\*, b\*, taille). Il est donc possible de regrouper les objets des 4 répétitions.

<sup>1</sup> NS<0,001

Tableau XLIV  
Analyse de variabilité de l'échantillon standard C2

L*				
	REP 1	REP 2	REP 3	REP 4
Moyenne	148,8	151,5	150,2	150,2
SD	20,9	22,3	20,0	22,1
SEM	2,5	2,7	2,4	2,6
CV%	14,0	14,7	13,3	14,7
Max	185,1	201,8	185,1	197,2
Min	101,2	108,8	98,4	105,4
Étendue	83,9	93,1	86,6	91,9
N = S =				
a*				
	REP 1	REP 2	REP 3	REP 4
Moyenne	169,3	167,9	166,5	166,9
SD	11,0	13,0	10,8	11,9
SEM	1,3	1,6	1,3	1,4
CV%	6,5	7,8	6,5	7,1
Max	187,6	189,6	189,3	186,8
Min	140,6	137,7	146,1	138,4
Étendue	47,0	51,8	43,3	48,4
N = S =				
b*				
	REP 1	REP 2	REP 3	REP 4
Moyenne	161,4	160,9	161,2	160,7
SD	4,3	4,2	4,2	4,0
SEM	0,5	0,5	0,5	0,5
CV%	2,7	2,6	2,6	2,5
Max	171,3	169,6	172,5	171,6
Min	153,1	152,9	152,8	149,7
Étendue	18,2	16,7	19,7	22,0
N = S =				
Taille				
	REP 1	REP 2	REP 3	REP 4
Moyenne	1556,0	1589,1	1553,4	1563,5
SD	284,5	326,9	268,3	276,2
SEM	33,8	39,1	31,8	32,8
CV%	18,3	20,6	17,3	17,7
Max	2120,0	3199,0	2104,0	2045,0
Min	919,0	1018,0	834,0	814,0
Étendue	1201,0	2181,0	1270,0	1231,0
N = S =				

Tableau XLV

Grand résumé de l'analyse de variabilité de l'échantillon standard C2

	L*	a*	b*	Taille	Nb objets
Moyenne					
SD	1,1	1,2	0,3	16,3	-
SEM	0,5	0,6	0,2	8,2	-
CV%	0,7	0,7	0,2	1,0	-
Max	151,5	169,3	161,4	1589,1	-
Min	148,8	166,5	160,7	1553,4	-
Etendue	2,6	2,8	0,7	35,7	-
N = S =					

Les coefficients de variation basés sur 4 moyennes par variable sont inférieurs à 3%, ce qui est bien inférieur au 10% généralement acceptable par les normes de l'industrie.

Tableau XLVI  
ANOVA à une entrée de l'échantillon standard C2

<b>L*</b>					
Source	degrés de liberté	Somme des carrés	Moyenne des carrés		
	3	245,4	81,8	0,18	NS <sup>1</sup>
	279	126993,8	455,2		
	282	127239,1			
<b>a*</b>					
Source	degrés de liberté	Somme des carrés	Moyenne des carrés		
	3	330,7	110,2	0,81	NS <sup>1</sup>
	279	38155,3	136,8		
	282	38486,0			
<b>b*</b>					
Source	degrés de liberté	Somme des carrés	Moyenne des carrés		
	3	19,7	6,6	0,38	NS <sup>1</sup>
	279	4864,1	17,4		
	282	4883,8			
<b>Taille</b>					
Source	degrés de liberté	Somme des carrés	Moyenne des carrés		
	3	56176,0	18725,3	0,22	NS <sup>1</sup>
	279	23418951,3	83938,9		
	282	23475127,3			

À un risque alpha ( $\alpha$  de 0,001), il n'y a pas de différence significative entre la moyenne des populations, et ce pour toutes les variables (L\*, a\*, b\*, taille). Il est donc possible de regrouper les objets des 4 répétitions.

<sup>1</sup> NS<0,001

## **ANNEXE 4**

**Couleur moyenne des fruits composant les échantillons standard de la saison de récolte 2005 représentée dans l'espace ICCLAB**

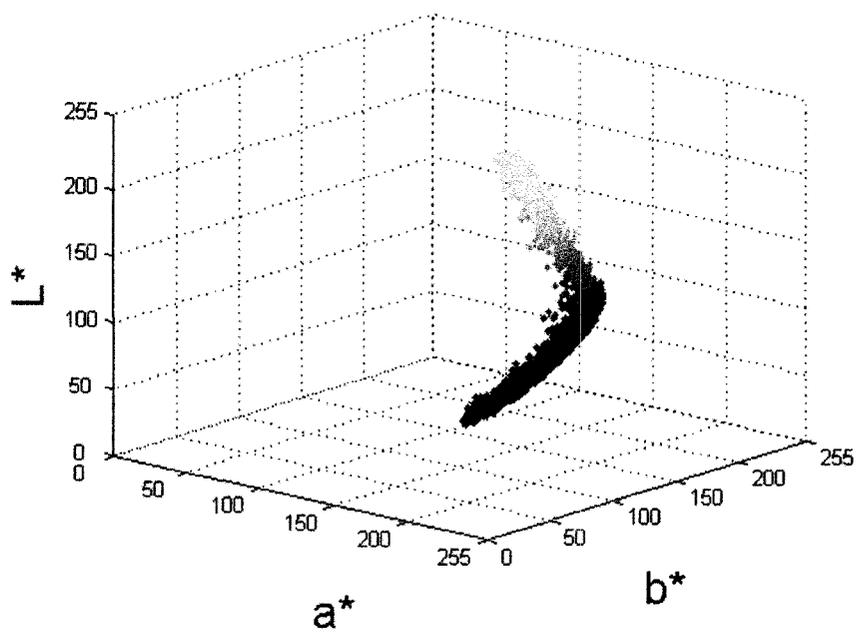


Figure 82 Couleur moyenne des fruits standard de canneberge dans l'espace ICCLAB

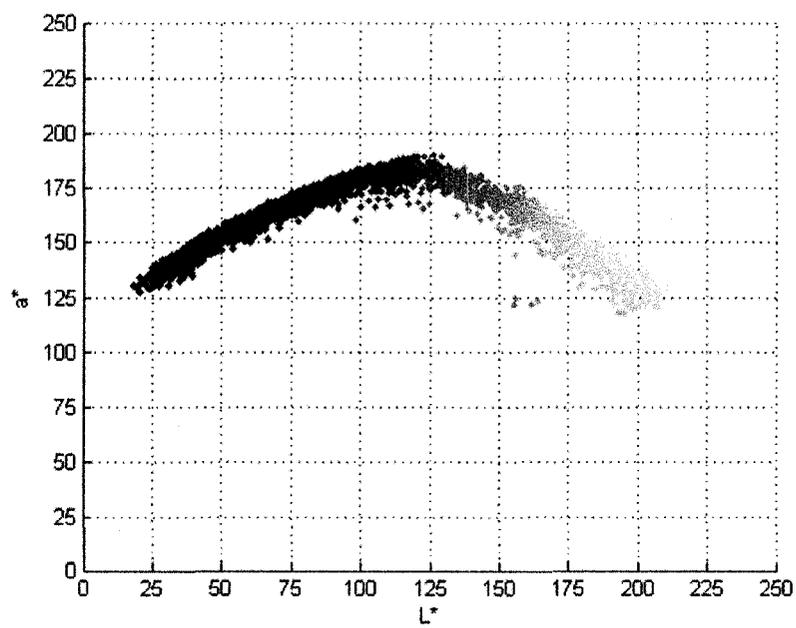


Figure 83 Couleur moyenne des fruits standard de canneberge suivant  $L^*$  et  $a^*$

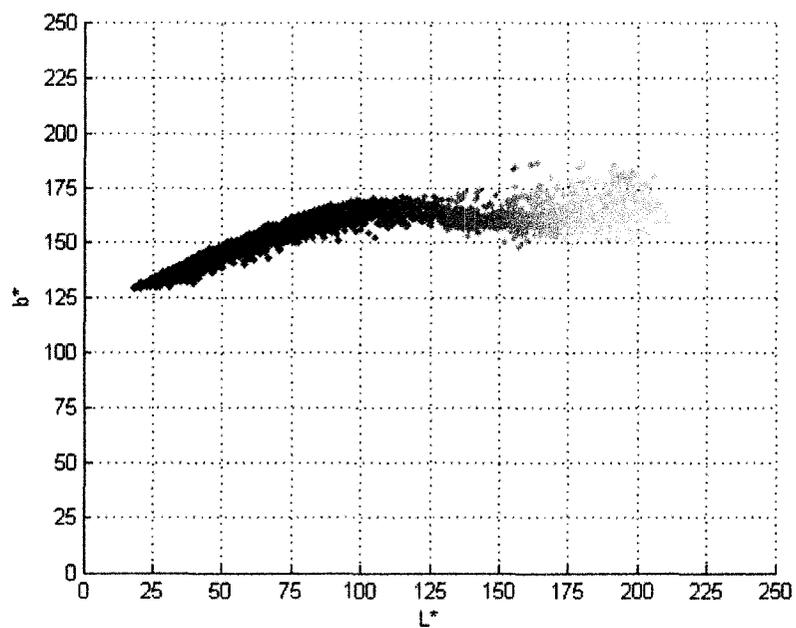


Figure 84 Couleur moyenne des fruits standard de canneberge suivant L\* et b\*

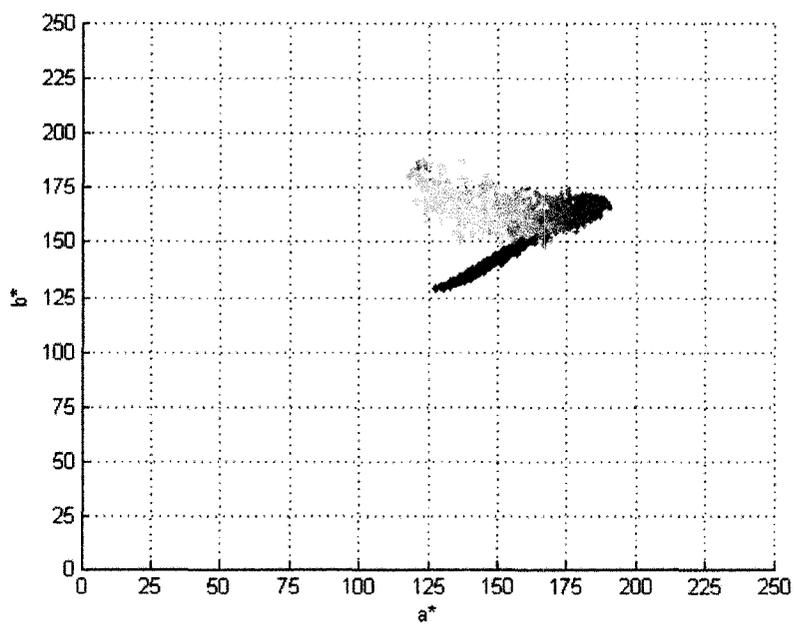


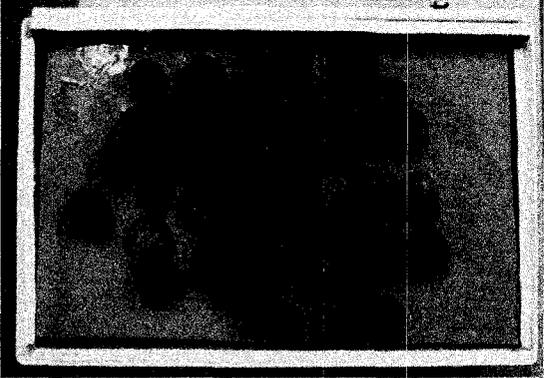
Figure 85 Couleur moyenne des fruits standard de canneberge suivant a\* et b\*

## **ANNEXE 5**

**Photos de lots typiques de la saison de récolte 2005**

Tableau XLVII

Photos des trois échantillons du lot n° 446 (29-oct-05) présentant 12% de fruits pourris

photo n° 1 de l'échantillon n° 1	
	
photo n° 1 de l'échantillon n° 2	
	
photo n° 1 de l'échantillon n° 3	
	

Les photos des échantillons présentant un trop grand nombre de fruits pourris ne sont pas segmentées correctement lors de la partie de traitement de l'image. Il n'est donc pas possible de prédire le Tacy de ces échantillons par la méthode de vision artificielle.

Tableau XLVIII

Échantillon type n°1 de la base de test (lot n°450, 27-oct-05)

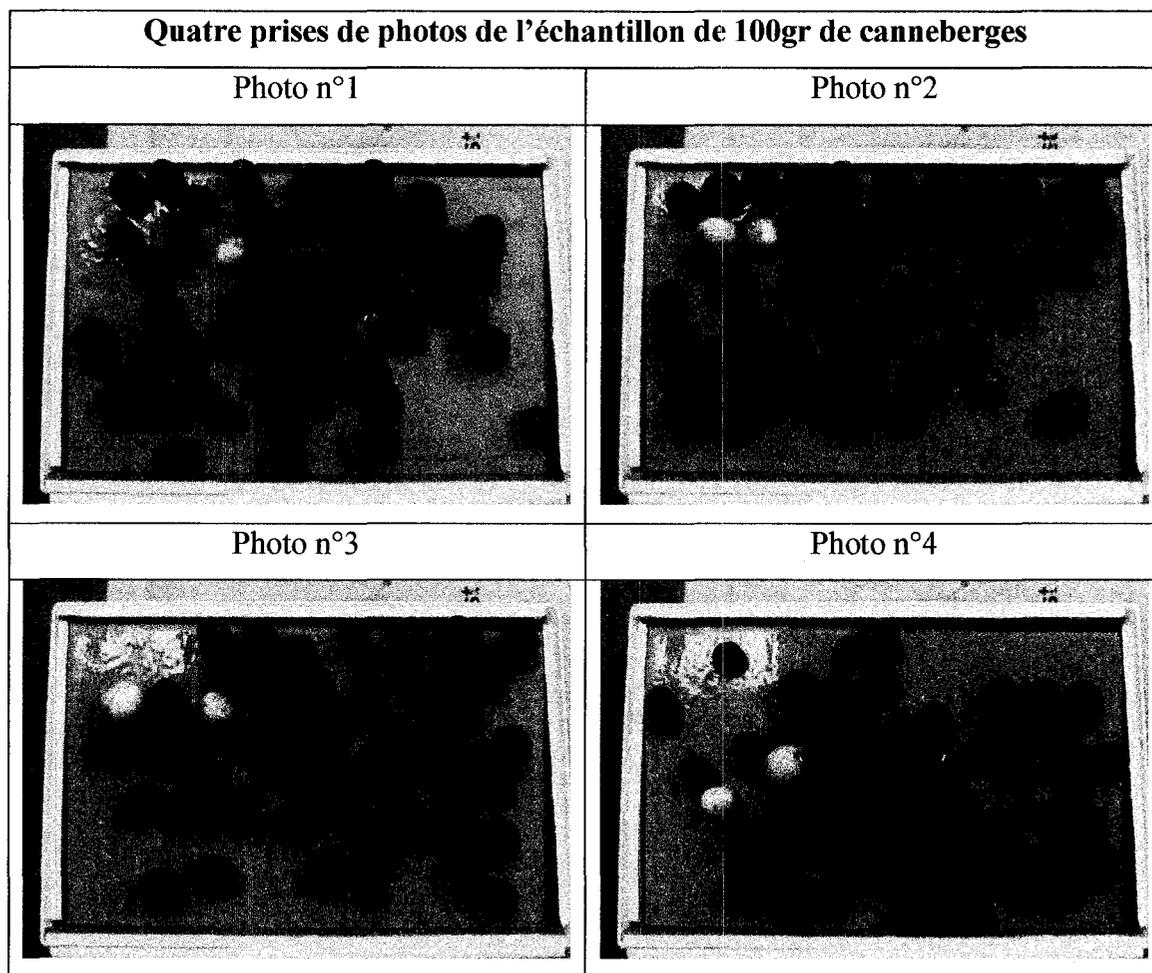


Tableau XLIX

Tacy chimique et Tacy prédit (lot n°450, 27-oct-05)

<b>Tacy moyen 'chimique laboratoire'</b>	<b>Modèle prédictif Tacy moyen</b>
40	42

Tableau L

Échantillon type n°2 de la base de test (lot n°445, 28-oct-05)

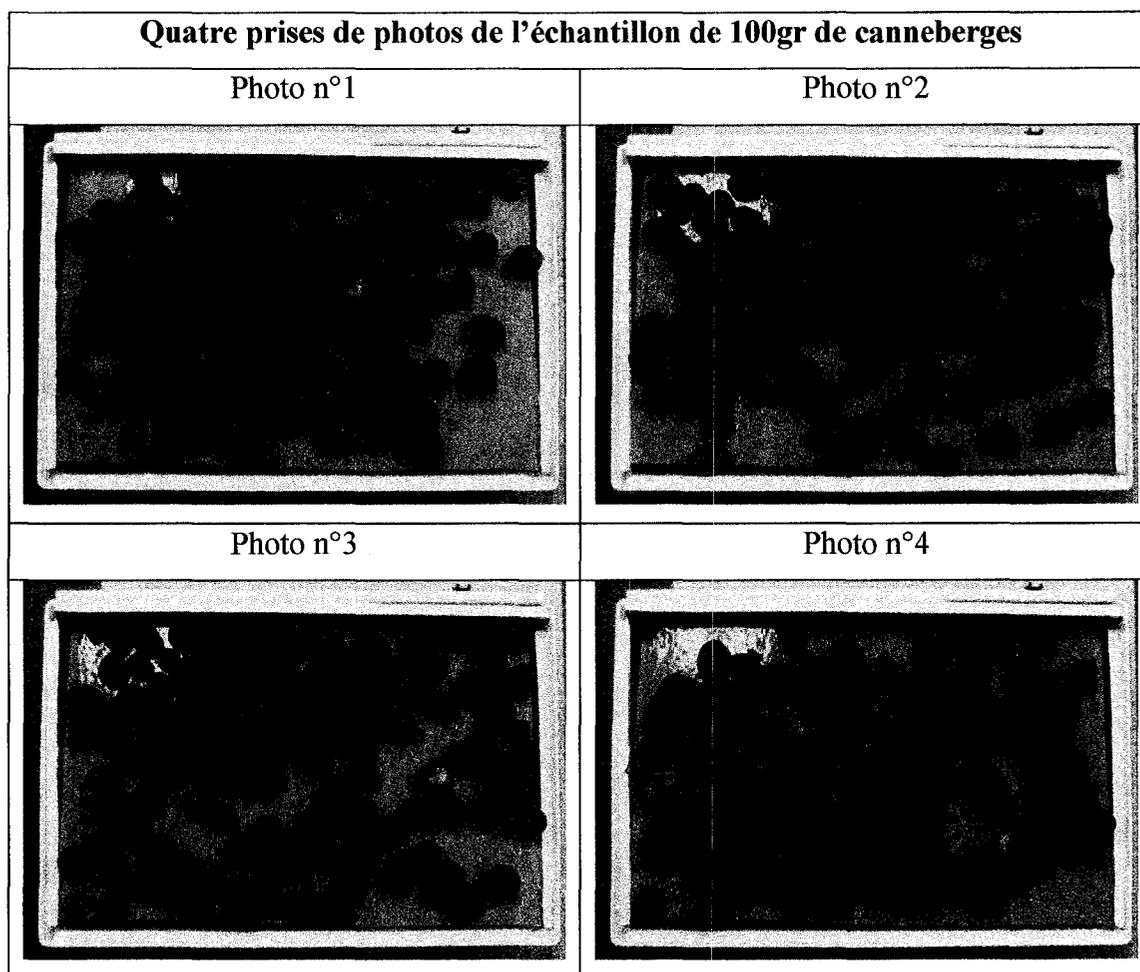


Tableau LI

Tacy chimique et Tacy prédit (lot n°445, 28-oct-05)

<b>Tacy moyen 'chimique laboratoire'</b>	<b>Modèle prédictif Tacy moyen</b>
48	44

Tableau LII

Échantillon type n°3 de la base de test (lot n°449, 29-oct-05)

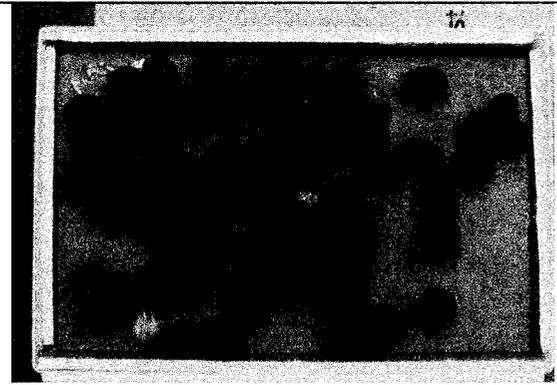
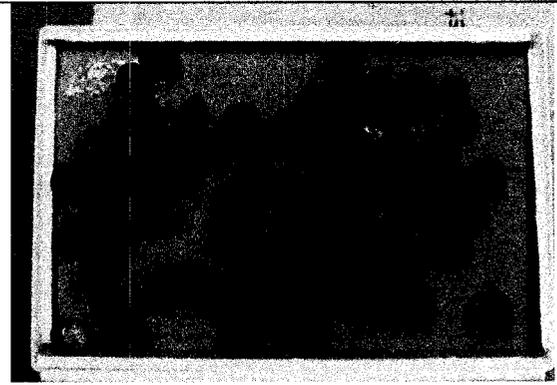
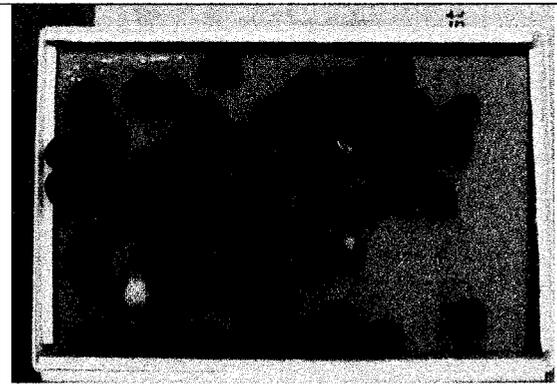
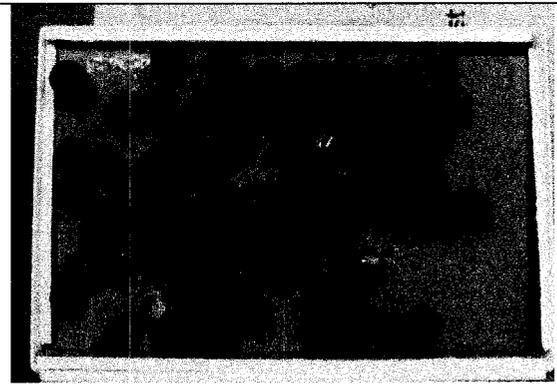
<b>Quatre prises de photos de l'échantillon de 100gr de canneberges</b>	
<b>Photo n°1</b>	<b>Photo n°2</b>
	
<b>Photo n°3</b>	<b>Photo n°4</b>
	

Tableau LIII

Tacy chimique et Tacy prédit (lot n°449, 29-oct-05)

<b>Tacy moyen 'chimique laboratoire'</b>	<b>Modèle prédictif Tacy moyen</b>
53	53

Tableau LIV  
Échantillon type n°4 de la base de test (lot n°431, 26-oct-05)

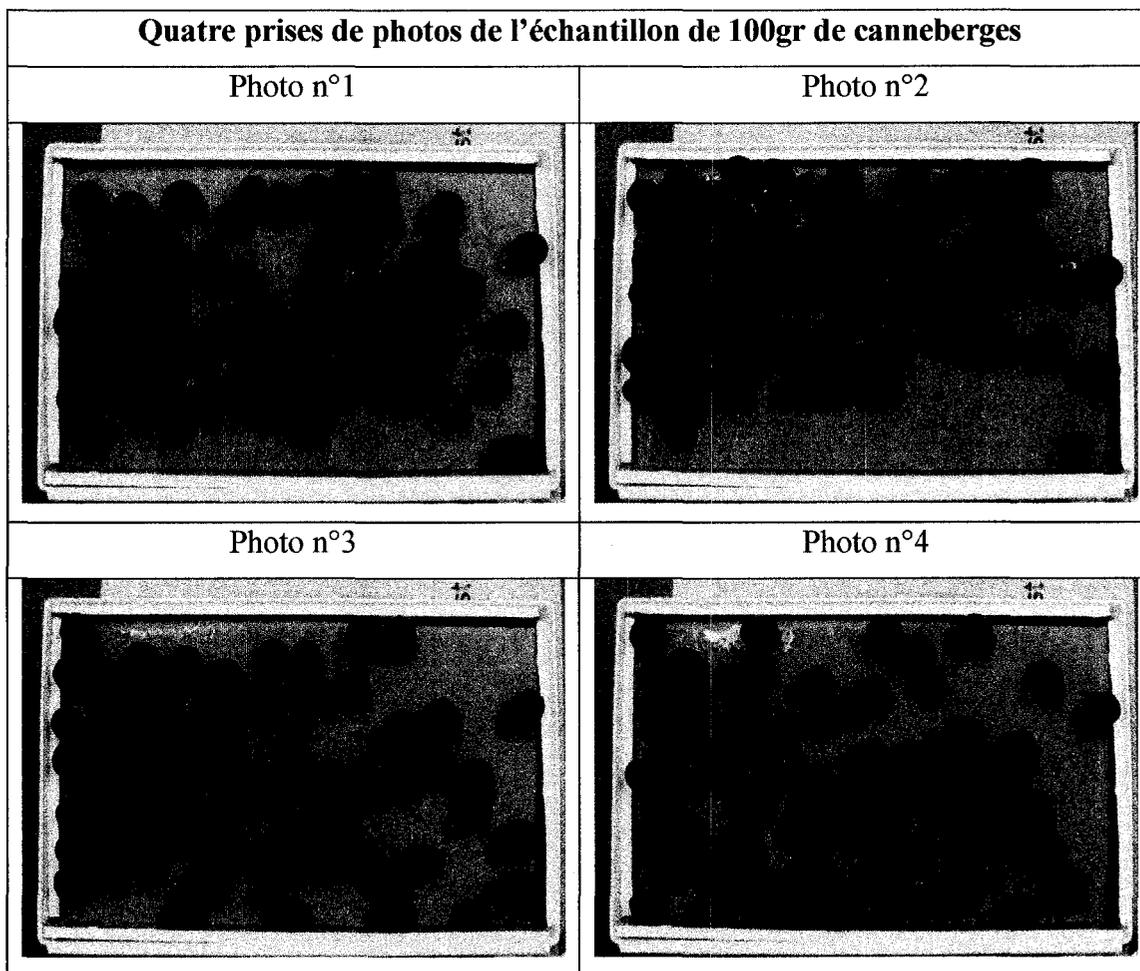


Tableau LV  
Tacy chimique et Tacy prédit (lot n°431, 26-oct-05)

<b>Tacy moyen 'chimique laboratoire'</b>	<b>Modèle prédictif Tacy moyen</b>
63	62

## Tableau LVI

Photos des trois échantillons du lot n° 442 (29-oct-05) considérés comme des données aberrantes

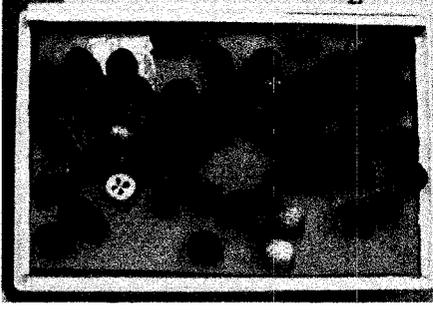
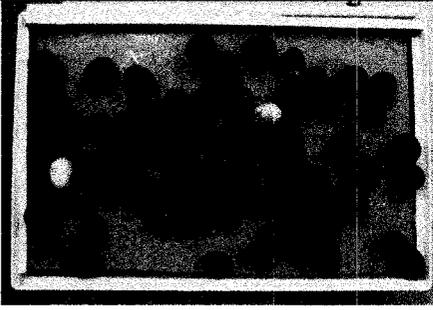
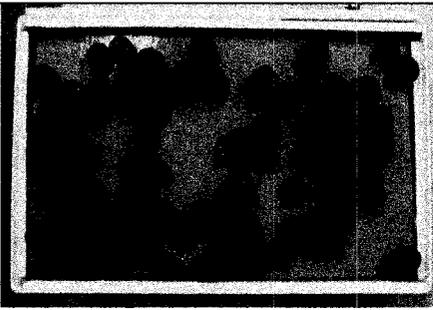
photo n° 1 de l'échantillon n° 1		
	 A black and white micrograph showing a dark field with several bright, irregularly shaped spots. One prominent spot in the lower-left quadrant has a distinct four-lobed pattern, resembling a cross or a star.	
photo n° 1 de l'échantillon n° 2		
	 A black and white micrograph showing a dark field with several bright, irregularly shaped spots. A prominent bright spot is visible in the upper-right quadrant, and another is in the lower-left quadrant.	
photo n° 1 de l'échantillon n° 3		
	 A black and white micrograph showing a dark field with several bright, irregularly shaped spots. The spots are more numerous and less distinct than in the previous two samples.	

Tableau LVII

Tacy chimique et Tacy prédit des 3 échantillons du lot n°442 (29-oct-05)

	Tacy moyen 'chimique laboratoire'	Modèle prédictif Tacy moyen
échantillon n° 1	64	50
échantillon n° 2		48
échantillon n° 3		58

## Tableau LVIII

Photos des trois échantillons du lot n° 443 (27-oct-05) considérés comme des données aberrantes

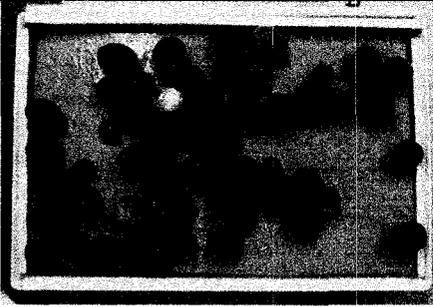
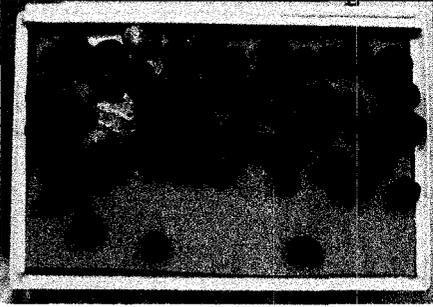
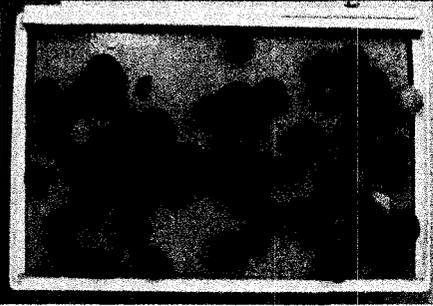
	photo n° 1 de l'échantillon n° 1	
		
	photo n° 1 de l'échantillon n° 2	
		
	photo n° 1 de l'échantillon n° 3	
		

Tableau LIX

Tacy chimique et Tacy prédit des 3 échantillons du lot n°443 (27-oct-05)

	Tacy moyen 'chimique laboratoire'	Modèle prédictif Tacy moyen
échantillon n° 1	48	41
échantillon n° 2		53
échantillon n° 3		45

Les données de Tacy des lots n° 442 et 443 sont considérées comme des données aberrantes, car il est possible de constater visuellement pour ces deux lots que les couleurs des fruits sont notablement différentes pour au moins un échantillon parmi les trois (échantillon n° 3 pour le lot n° 442 et échantillon n° 2 pour le lot n° 443). Les valeurs de Tacy obtenues par le modèle prédictif pour chacun des échantillons confirment l'inspection visuelle des photos. Toutefois, pour le lot n° 443, il est possible de remarquer que la moyenne des Tacy des 3 échantillons obtenus par le modèle prédictif est égale à 46, valeur très proche du Tacy moyen 'chimique laboratoire'. Le lot n° 443 serait donc un lot dont l'écart type entre les 3 Tacy 'chimique laboratoire' est très élevé.

Tableau LX  
Échantillon du mélange extrême n° 1 (échantillon non commercial)

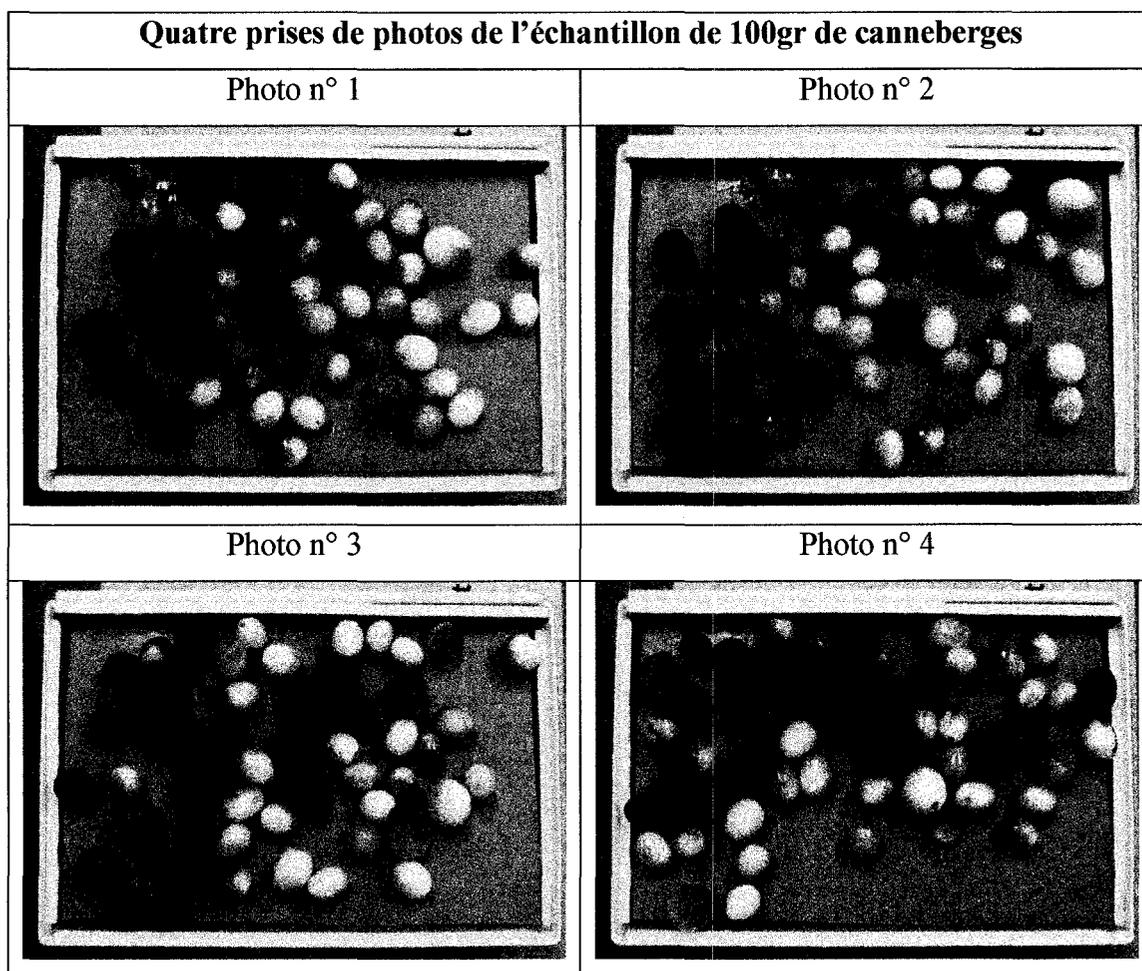


Tableau LXI  
Tacy chimique et Tacy prédit du mélange extrême n° 1 (échantillon non commercial)

Tacy moyen 'chimique laboratoire'	Modèle prédictif hybride
27	27

Tableau LXII

Échantillon du mélange extrême n° 2 (échantillon non commercial)

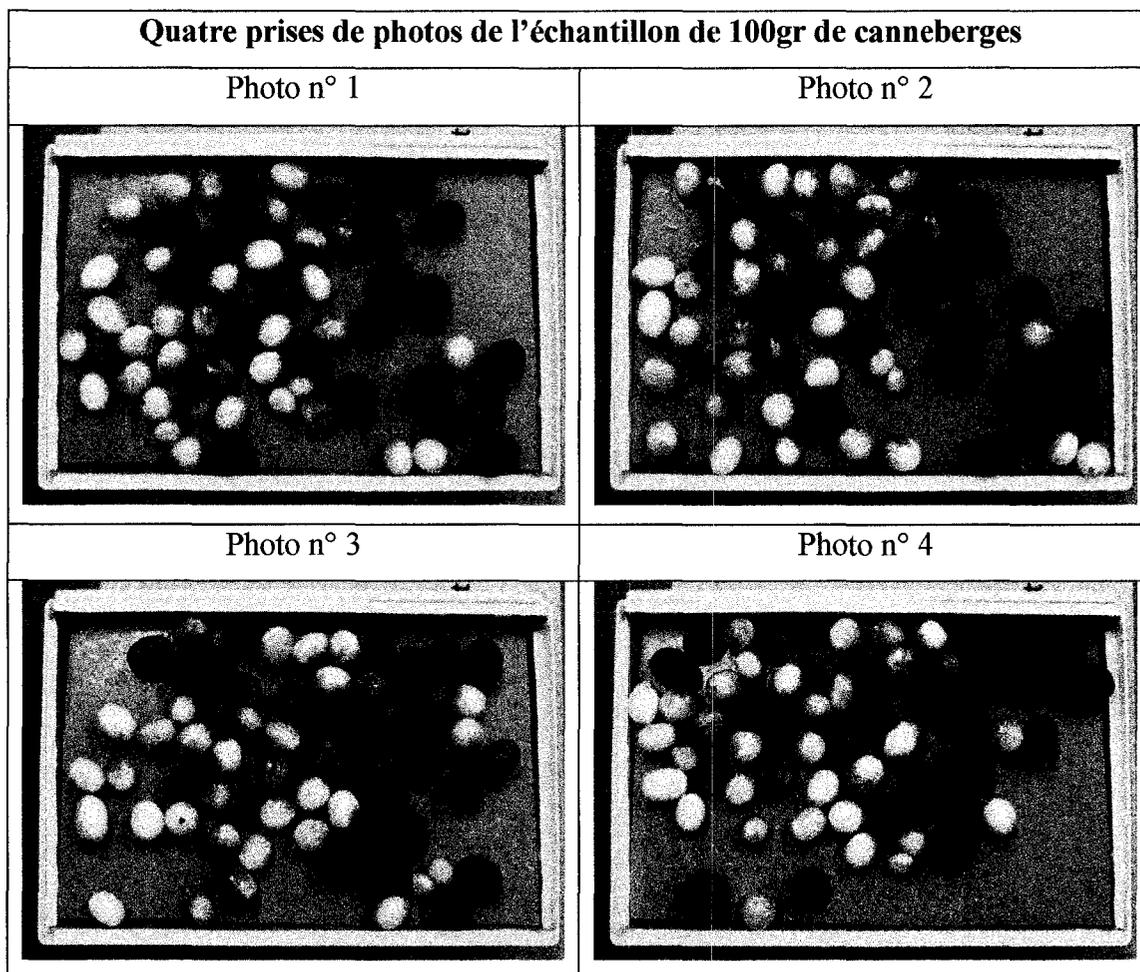


Tableau LXIII

Tacy chimique et Tacy prédit du mélange extrême n° 2 (échantillon non commercial)

Tacy moyen 'chimique laboratoire'	Modèle prédictif hybride
33	28

Les mélanges extrêmes sont des échantillons possédant 25 % de fruits jaunes / verdâtres, 25 % de fruits rosés, 25 % de fruits rouges et 25 % de fruits rouge foncé. Il est possible de constater à partir des 2 échantillons présentés ci-dessus, que dans le cas de mélanges extrêmes, le modèle prédictif hybride permet une estimation du Tacy dans les tolérances d'erreurs types de  $\pm 3$ .

## **ANNEXE 6**

### **Photos de lots typiques de la saison de récolte 2006**

Tableau LXIV  
Échantillon typique de la saison 2006 n° 1

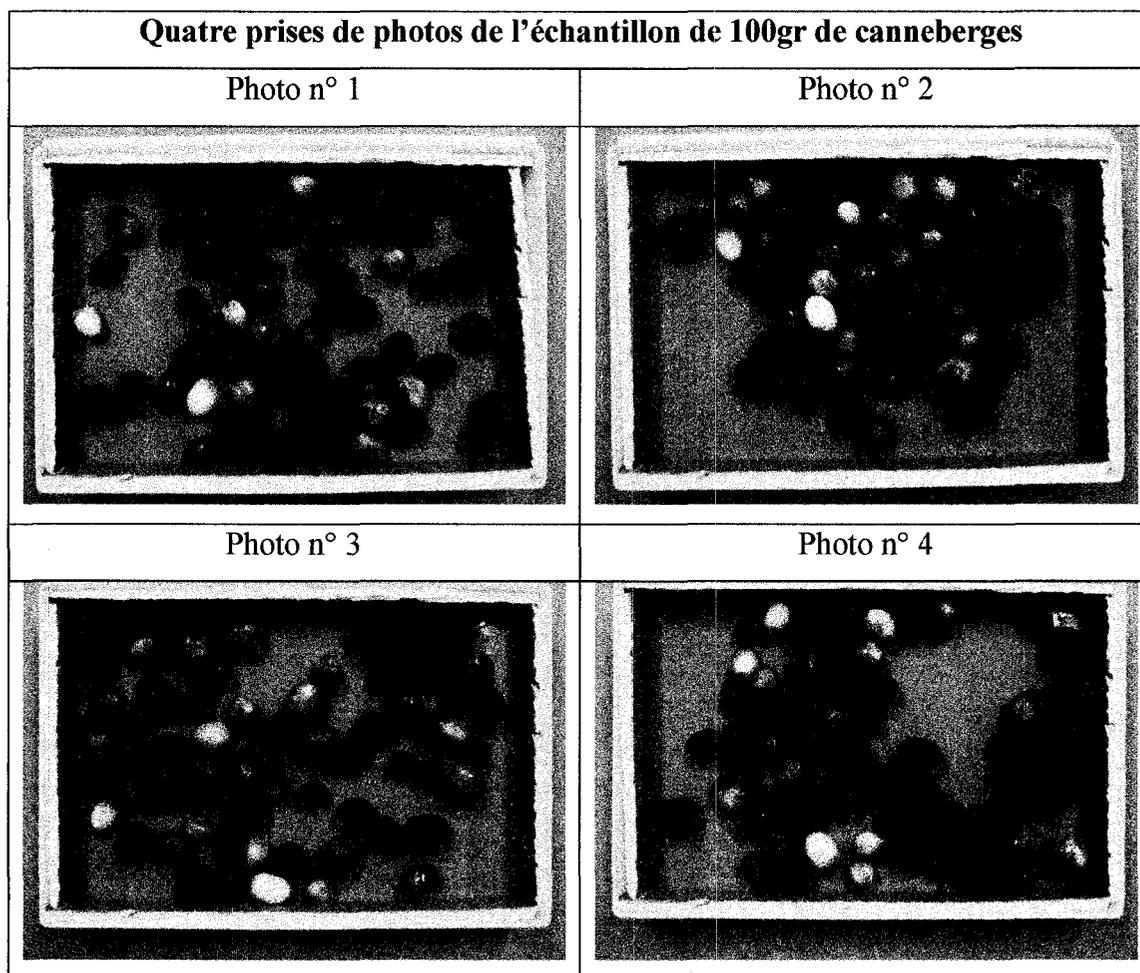


Tableau LXV  
Tacy chimique et Tacy prédit de l'échantillon typique de la saison 2006 n° 1

<b>Tacy moyen 'chimique laboratoire'</b>	<b>Modèle prédictif hybride</b>
28	26

Tableau LXVI  
Échantillon typique de la saison 2006 n° 2

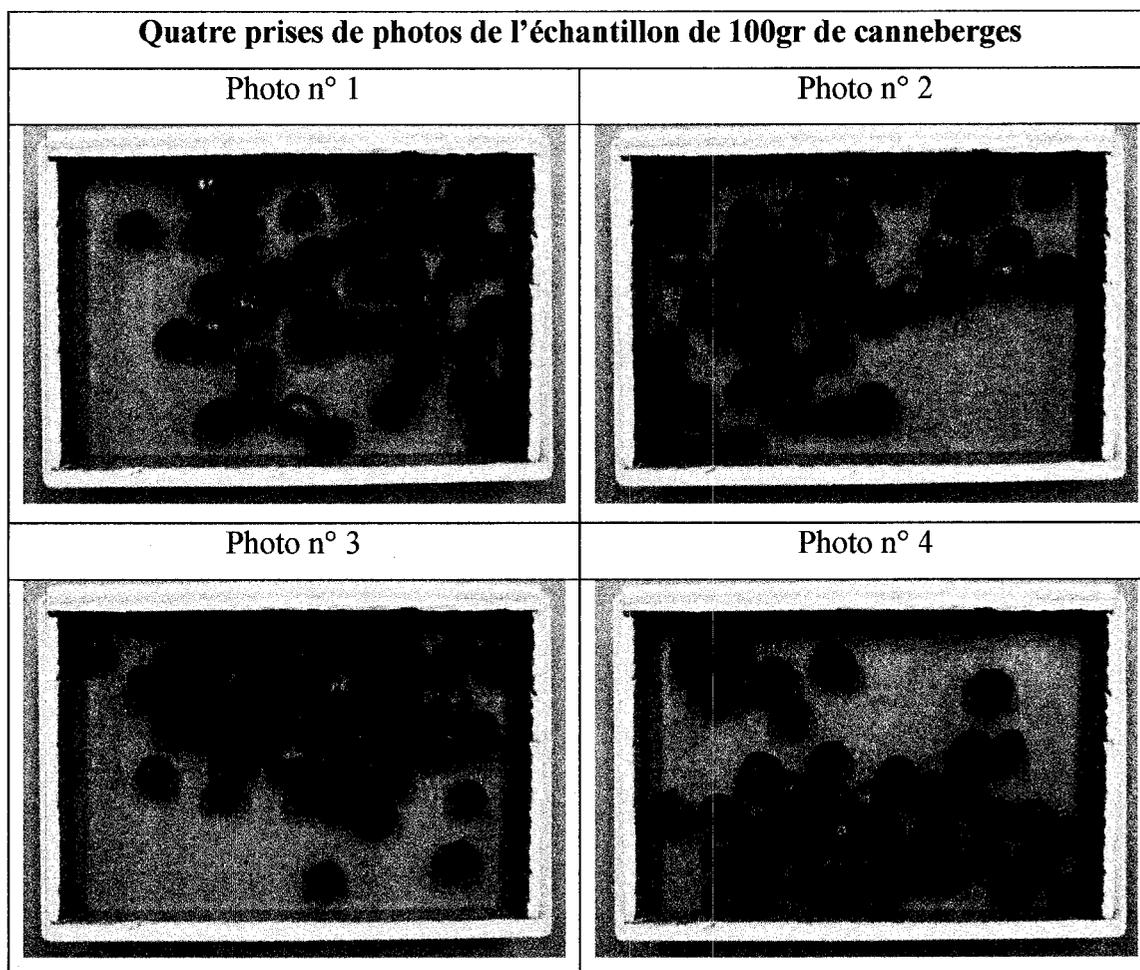


Tableau LXVII  
Tacy chimique et Tacy prédit de l'échantillon typique de la saison 2006 n° 2

<b>Tacy moyen 'chimique laboratoire'</b>	<b>Modèle prédictif hybride</b>
37	41

Tableau LXVIII  
Échantillon typique de la saison 2006 n° 3

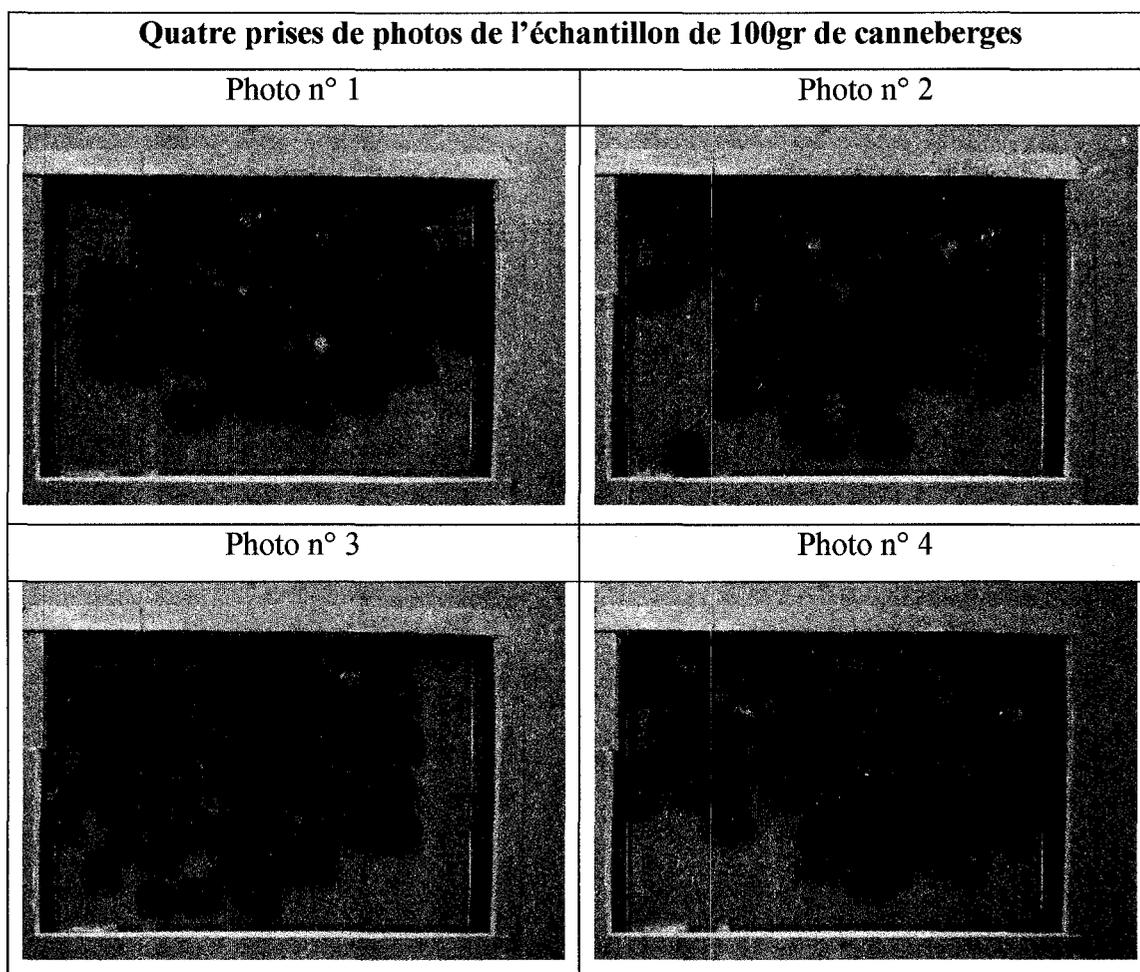


Tableau LXIX  
Tacy chimique et Tacy prédit de l'échantillon typique de la saison 2006 n° 3

<b>Tacy moyen 'chimique laboratoire'</b>	<b>Modèle prédictif hybride</b>
49	44

Tableau LXX  
Échantillon typique de la saison 2006 n° 4

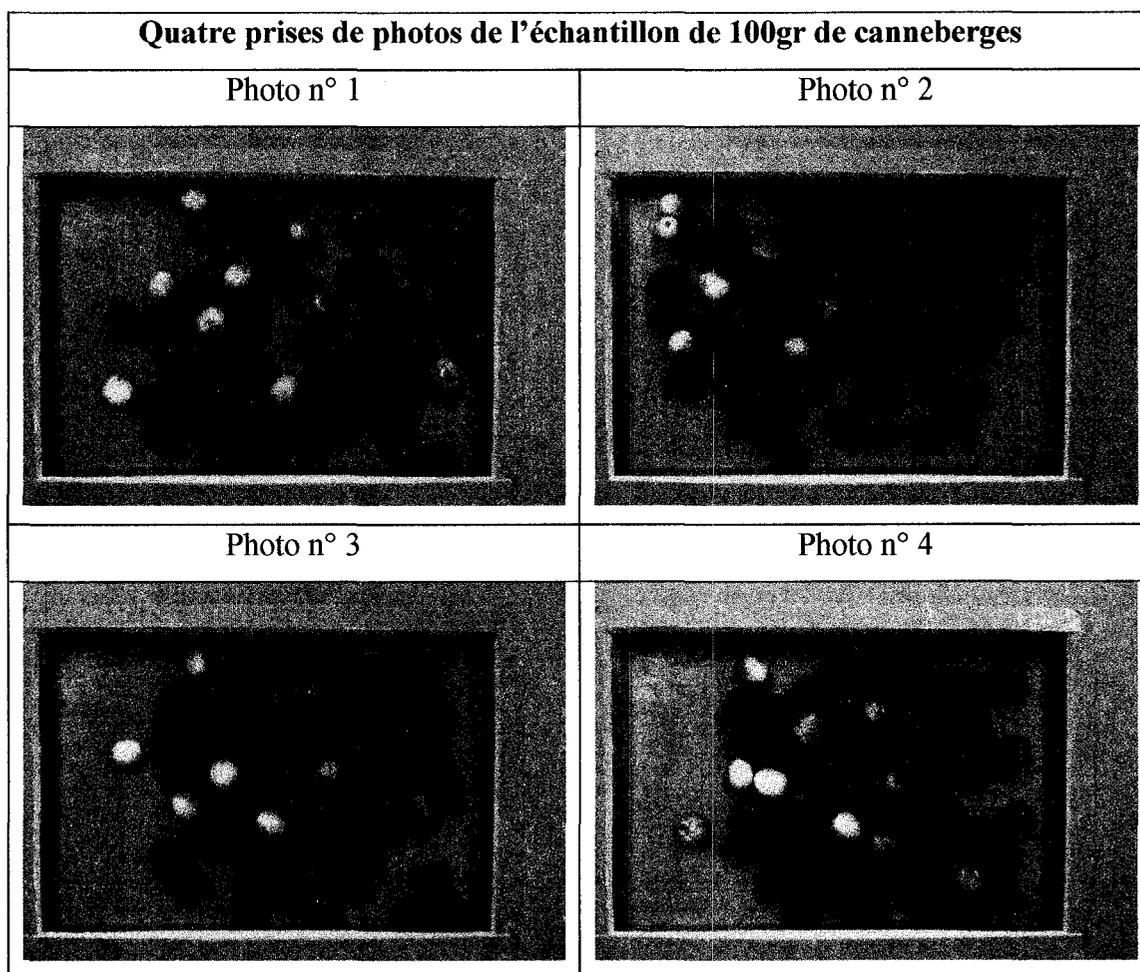


Tableau LXXI  
Tacy chimique et Tacy prédit de l'échantillon typique de la saison 2006 n° 4

<b>Tacy moyen 'chimique laboratoire'</b>	<b>Modèle prédictif hybride</b>
50	51

## **ANNEXE 7**

### **Prédiction du Tacy pour les échantillons de la base de données 2006**

Tableau LXXII

Prédiction du Tacy pour les échantillons de la base de données 2006

Indice de l'échantillon #	Tacy chimique laboratoire	Tacy prédit (regroupement des 4 photos)
1	36	38
2	33	35
3	36	36
4	36	39
5	37	37
6	37	37
7	37	41
8	30	26
9	28	26
10	30	28
11	38	36
12	35	29
13	38	29
14	38	30
15	36	27
16	34	34
17	23	24
18	33	30
19	57	62
20	57	60
21	57	59
22	48	49
23	59	54
24	64	64
25	67	62
26	53	57
27	55	60
28	61	61

## Prédiction du Tacy pour les échantillons de la base de données 2006 (suite)

Indice de l'échantillon #	Tacy chimique laboratoire	Tacy prédit (regroupement des 4 photos)
29	42	40
30	42	46
31	44	42
32	47	48
33	48	46
34	48	44
35	48	48
36	50	51
37	52	53
38	55	56
39	49	54
40	52	46
41	55	47
42	42	37
43	47	35
44	46	38
45	47	42
46	60	51
47	38	41
48	76	61
49	49	44
50	47	47
51	63	59
52	52	45
53	48	41
54	60	59
55	47	41
56	52	54

## Prédiction du Tacy pour les échantillons de la base de données 2006 (suite)

Indice de l'échantillon #	Tacy chimique laboratoire	Tacy prédit (regroupement des 4 photos)
57	42	41
58	39	41
59	41	42
60	38	44
61	42	49
62	42	41
63	53	50
64	54	56
65	56	64
66	57	55
67	56	54
68	52	48
69	55	51
70	64	61
71	60	51
72	50	54
73	52	51
74	52	56
75	51	53
76	54	55
77	53	54
78	55	55
79	56	57
80	38	43
81	94	77
82	86	76
83	85	74
84	65	58

## Prédiction du Tacy pour les échantillons de la base de données 2006 (suite)

Indice de l'échantillon #	Tacy chimique laboratoire	Tacy prédit (regroupement des 4 photos)
85	62	53
86	51	51
87	66	57
88	69	58
89	74	64
90	62	55
91	76	65
92	49	56
93	66	67
94	61	63
95	56	52
96	46	52
97	42	48
98	52	54
99	67	73
100	49	53
101	55	50
102	42	48
103	46	49
104	52	59
105	49	53
106	78	76
107	60	62
108	55	53
109	60	62
110	65	65
111	46	51

## BIBLIOGRAPHIE

- [1] Blank, F. (1958). Anthocyanins, flavones, xanthones. In W. Ruhland (Ed.), *Handbuch der Pflanzenphysiologie* (Vol. 10, pp. 300-353). Berlin, Germany: Springer Verlag.
- [2] Ribéreau-Gayon, P. (1959). *Recherches sur les anthocyanes des végétaux. Application au genre Vitis*. Unpublished Doctoral dissertation, University of Bordeaux, Librairie Générale de l'Enseignement, Paris, France.
- [3] Sakamura, S., & Francis, F. J. (1961). The anthocyanins of the American cranberry. *Journal of Food Science*, 26, 318-321.
- [4] Zapalis, C., & Francis, F. J. (1965). Cranberry anthocyanins. *Journal of Food Science*, 30, 396-399.
- [5] Fuleki, T., & Francis, F. J. (1968). Quantitative methods for anthocyanins I. Extraction and determination of total anthocyanin in cranberries. *Journal of Food Science*, 33, 72-77.
- [6] Lepage, R., Lussier, J.-S., & Doyon, G. (2005, 26 septembre au 1 octobre). *Estimation de la teneur en anthocyane au moyen de l'analyse par composantes principales (ACP) de la couleur des canneberges*. Paper presented at the 4e ateliers sur Traitement et Analyse de l'Information : Méthodes et Applications (TAIMA' 2005), Hammamet (Tunisie).
- [7] Johnson, L. E., Clydesdale, F. M., & Francis, F. J. (1976). Use of expanded color scales to predict chemical and visual changes in solutions. *Journal of Food Science*, 41(1), 74-77.
- [8] Hulme, A. C. (1970). Antocyanidins and anthocyanins. In Academic Press (Ed.), *The biochemistry of fruits and their products* (Vol. 1, pp. 276-280). London. UK.
- [9] Francis, F. J. (1957). Color and pigment measurement in fresh cranberries. *American society for horticultural science*, 69, 296-301.
- [10] Francis, F. J., & Clydesdale, F. M. (1970). Colour measurement of foods. XVIII. Cranberry products. *Food Product Development*, 4(2), 54, 56, 60-62, 83 & 86.
- [11] Du, C.-J., & Sun, D.-W. (2006). Learning techniques used in computer vision for food quality evaluation: a review. *Journal of Food Engineering*, 72(1), 39-55.

- [12] Thomas, P., Kannan, A., Degwekar, V. H., & Ramamurthy, M. S. (1995). Non-destructive detection of seed weevil-infested mango fruits by X-ray imaging. *Postharvest Biology and Technology*, 5(1-2), 161-165.
- [13] Lammertyn, J., Dresselaers, T., Van Hecke, P., Jancsok, P., Wevers, M., & Nicolai, B. M. (2003). Analysis of the time course of core breakdown in 'Conference' pears by means of MRI and X-ray CT. *Postharvest Biology and Technology*, 29(1), 19-28.
- [14] Clark, C. J., Hockings, P. D., Joyce, D. C., & Mazucco, R. A. (1997). Application of magnetic resonance imaging to pre- and post-harvest studies of fruits and vegetables. *Postharvest Biology and Technology*, 11(1), 1-21.
- [15] Sonogo, L., Ben-Arie, R., Raynal, J., & Pech, J. C. (1995). Biochemical and physical evaluation of textural characteristics of nectarines exhibiting woolly breakdown: NMR imaging, X-ray computed tomography and pectin composition. *Postharvest Biology and Technology*, 5(3), 187-198.
- [16] Lammertyn, J., Dresselaers, T., Van Hecke, P., Jancsok, P., Wevers, M., & Nicolai, B. M. (2003). MRI and x-ray CT study of spatial distribution of core breakdown in 'Conference' pears. *Magnetic Resonance Imaging*, 21(7), 805-815.
- [17] Grassmann, H. G. (1853) Zur theorie der farbenmischung. *Annalen der Physik und Chemie*, 89, 69. English translation "Theory of Compound Colors" published in *Philosophical Magazine*, vol 4(7), pp. 254-264 (1854). Reprinted in MacAdam (1993), pp 10-13.
- [18] Diaz, R., Gil, L., Serrano, C., Blasco, M., Molto, E., & Blasco, J. (2004). Comparison of three algorithms in the classification of table olives by means of computer vision. *Journal of Food Engineering Applications of computer vision in the food industry*, 61(1), 101-107.
- [19] Leemans, V., Magein, H., & Destain, M.-F. (1999). Defect segmentation on 'Jonagold' apples using colour vision and a Bayesian classification method. *Computers and Electronics in Agriculture*, 23(1), 43-53.
- [20] Bulanon, D. M., Kataoka, T., Ota, Y., & Hiroma, T. (2002). AE--Automation and Emerging Technologies: A Segmentation Algorithm for the Automatic Recognition of Fuji Apples at Harvest. *Biosystems Engineering*, 83(4), 405-412.
- [21] Aleixos, N., Blasco, J., Navarron, F., & Molto, E. (2002). Multispectral inspection of citrus in real-time using machine vision and digital signal processors. *Computers and Electronics in Agriculture*, 33(2), 121-137.

- [22] Lu, R. (2004). Multispectral imaging for predicting firmness and soluble solids content of apple fruit. *Postharvest Biology and Technology*, 31(2), 147-157.
- [23] Carlomagno, G., Capozzo, L., Attolico, G., & Distante, A. (2004). Non-destructive grading of peaches by near-infrared spectrometry. *Infrared Physics & Technology. Workshop on Advanced Infrared Technology and Application*, 46(1-2), 23-29.
- [24] Kondo, N. (2003). *Fruit grading robot*. Paper presented at the Advanced Intelligent Mechatronics, 2003. AIM 2003. Proceedings. 2003 IEEE/ASME International Conference on.
- [25] Calpe, J., Pla, F., Monfort, J., Diaz, P., & Boada, J. C. (1996). *Robust low-cost vision system for fruit grading*. Paper presented at the Electrotechnical Conference, 1996. MELECON '96., 8th Mediterranean.
- [26] Blasco, J., Aleixos, N., & Molto, E. (2003). Machine Vision System for Automatic Quality Grading of Fruit. *Biosystems Engineering*, 85(4), 415-423.
- [27] Mendoza, F., Dejmek, P., & Aguilera, J. M. (2006). Calibrated color measurements of agricultural foods using image analysis. *Postharvest Biology and Technology*, 41(3), 285-295.
- [28] B. Hernández, C. S., C. Alberdi, S. Alfonso, M. Berroguí, J. M. Diñeiro, (2004). Design and performance of a color chart based in digitally processed images for sensory evaluation of piquillo peppers (*Capsicum annuum*). *Color Research & Application*, 29(4), 305-311.
- [29] Sarkar, N. R. (1991). Machine vision for quality control in the food industry. In M. Dekker (Ed.), *Instrumental Methods for Quality Assurance in Foods* (pp. 167-187). New York.
- [30] Brosnan, T., & Sun, D.-W. (2004). Improving quality inspection of food products by computer vision--a review. *Journal of Food Engineering*, 61(1), 3-16.
- [31] Francis, F. J., & Clydesdale, F. M. (1975). *Food Colorimetry: Theory and Applications*. Westport, CT: AVI Publishing Co.
- [32] Ding, L., & Goshtasby, A. (2001). On the Canny edge detector. *Pattern Recognition*, 34(3), 721-725.
- [33] Adams, R., & Bischof, L. (1994). Seeded region growing. *Transactions on Pattern Analysis and Machine Intelligence, IEEE*, 16(6), 641-647.

- [34] Lorenzo-Navarro, J., & Hernandez-Tejera, M. (1994). *Image segmentation using a modified split and merge technique*. Paper presented at the Third International Workshop on Computer Aided Systems Theory (EUROCAST '93), 21-26 Feb. 1993 Cybernetics and Systems, Las Palmas, Spain.
- [35] Feng, G., & Qixin, C. (2004). *Study on color image processing based intelligent fruit sorting system*. Paper presented at the Fifth World Congress on Intelligent Control and Automation, 2004. WCICA 2004.
- [36] Kondo, N., Ahmad, U., Monta, M., & Murase, H. (2000). Machine vision based quality evaluation of Iyokan orange fruit using neural networks. *Computers and Electronics in Agriculture*, 29(1-2), 135-147.
- [37] Jain, A. K., Duin, R. P. W., & Mao, J. (2000). Statistical pattern recognition: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1), 4-37.
- [38] Abdullah, M. Z., Mohamad-Saleh, J., Fathinul-Syahir, A. S., & Mohd-Azemi, B. M. N. (2006). Discrimination and classification of fresh-cut starfruits (*Averrhoa carambola* L.) using automated machine vision system. *Journal of Food Engineering*, 76(4), 506-523.
- [39] Kim, J., Mowat, A., Poole, P., & Kasabov, N. (2000). Linear and non-linear pattern recognition models for classification of fruit from visible-near infrared spectra. *Chemometrics and Intelligent Laboratory Systems*, 51(2), 201-216.
- [40] Ben-Hanan, U., Peleg, K., & Gutman, P.-O. (1992). Classification of fruits by a boltzmann perceptron neural network. *Automatica*, 28(5), 961-968.
- [41] Unay, D., & Gosselin, B. (2005). *Artificial Neural Network-Based Segmentation and Apple Grading by Machine Vision*. Paper presented at the IEEE International Conference on Image Processing, 2005. ICIP 2005.
- [42] Encyclopédie informatique Comment Ça Marche. (2006). *Vidéo et imagerie numérique - La lumière (licence Creative Commons)*. Retrieved 19-04, 2006, from <http://www.commentcamarche.net/video/lumiere.php3>
- [43] Colour and Vision Research Laboratories (Institute of Ophthalmology UCL). (2006). *Colour and Vision database - Cone Fundamentals*. Retrieved 31-01, 2007, from <http://cvision.ucsd.edu/>
- [44] Fairchild, M. D. (1998). *Color Appearance Models*. Reading, MA: Addison-Wesley.

- [45] Niemetzky, G. (2004). *Gestion de la couleur : calibration & profils ICC pour l'image numérique et la chaîne graphique* (2e éd. ed.). Paris: Eyrolles.
- [46] Guild, J. (1926). A critical survey of modern developments in the theory and technique of colorimetry and allied sciences. *Proc. Opt Convention*, 61-146.
- [47] Wright, W. D. (1981). *50 Years of the 1931 CIE Standard Observer for Colorimetry* (1st ed.): Farbe.
- [48] Grassman, H. (1854). *On the theory of compound colours*. Leipzig.
- [49] Courtellemont, P. (2006). *Traitement des images couleurs - Partie 2 :Espaces de representation couleur*. Retrieved 31-01, 2007, from <http://perso.univ-lr.fr/pcourtel/html/cours/dea/part2.ppt>
- [50] Munsell Color Science Laboratory. (2006). *CIE Standard Illuminant Data*. Retrieved 31-01, 2007, from <http://www.cis.rit.edu/mcsl/online/cie.php>
- [51] Commission Internationale de l'Eclairage. (1986). *Colorimetry* (2nd ed.). Bureau Central de la CIE Vienne.
- [52] Wikipedia The Free Encyclopedia. (2006). *CIE 1931 color space*. Retrieved 31-01, 2007, from [http://en.wikipedia.org/wiki/CIE\\_1931\\_color\\_space](http://en.wikipedia.org/wiki/CIE_1931_color_space)
- [53] *Chromaticity diagram*. (2006). Retrieved 01-02, 2007, from <http://www.rpi.edu/~schubert/Light-Emitting-Diodes-dot-org/chap10/F10-07-R.jpg>
- [54] Brown, W. R. J., & MacAdam, D. L. (1949). Visual sensitivities to combined chromaticity and luminance differences. *Journal of the Optical Society of America*, 39(10), 808-834.
- [55] Wikipedia The Free Encyclopedia. (2006). *MacAdam ellipse (sous licence GNU FDL)*. Retrieved 01-02, 2007, from [http://en.wikipedia.org/wiki/MacAdam\\_ellipse](http://en.wikipedia.org/wiki/MacAdam_ellipse)
- [56] ENST - TSI. (2006). *Les espaces de couleur RVB et Lab*. Retrieved 01-02, 2007, from [http://www.tsi.enst.fr/tsi/enseignement/ressources/mti/RVB\\_ou\\_LAB/html/color\\_space.html](http://www.tsi.enst.fr/tsi/enseignement/ressources/mti/RVB_ou_LAB/html/color_space.html)
- [57] Luo, M. R., & Rigg, B. (1986). Chromaticity-discrimination ellipses for surface colours. *Color Research & Application*, 11, 25-42.

- [58] Hayden, S. L., & Oulton, D. P. (1994). The quantification of small visual colour differences. *Journal of the Society of Dyers and Colourists*, 110, 104-111.
- [59] Commission internationale de l'éclairage. (1995). *Industrial colour-difference evaluation*. Bureau Central CIE Vienne.
- [60] Commission internationale de l'éclairage. (2001). Improvement to industrial colour-difference evaluation.
- [61] Nathan Moroney. (2003). A hypothesis regarding the poor blue constancy of CIELAB. *Color Research & Application*, 28(5), 371-378.
- [62] M. R. Luo, G. C., B. Rigg,. (2001). The development of the CIE 2000 colour-difference formula: CIEDE2000. *Color Research & Application*, 26(5), 340-350.
- [63] Gaurav Sharma, W. W., Edul N. Dalal,. (2005). The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Research & Application*, 30(1), 21-30.
- [64] H. Mangine, K. J., C. Noel,. (2005). A preliminary comparison of CIE color differences to textile color acceptability using average observers. *Color Research & Application*, 30(4), 288-294.
- [65] M. Melgosa. (2006). Improvement of CMC upon CIEDE2000 for a new experimental dataset. *Color Research & Application*, 31(3), 239-241.
- [66] Ákos Borbély, Á. S., János Schanda,. (2001). The concept of correlated colour temperature revisited. *Color Research & Application*, 26(6), 450-457.
- [67] Steven D. Hordley. (2006). Scene illuminant estimation: Past, present, and future. *Color Research & Application*, 31(4), 303-314.
- [68] Xiao, F., Farell, J. E., Dicarlo, J. M., & Wandell, B. A. (2003, May). *Preferred Color Spaces for White Balancing*. Paper presented at the Proceedings of SPIE -- Volume 5017 -- Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications IV, Santa Clara, CA.
- [69] Viggiano, J. A. (2004, June). *Comparison of the accuracy of different white-balancing options as quantified by their color constancy*. Paper presented at the Proceedings of SPIE -- Volume 5301 -- Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications V.

- [70] Finlayson, G. D., Drew, M. S., & Funt, B. V. (1994). Color constancy: generalized diagonal transforms suffice. *Journal of the Optical Society of America*, 11, 3011-3019.
- [71] Hurvich, L. M., & Jameson, D. (1966). *The perception of brightness and darkness* (1 ed.): Allyn and Bacon.
- [72] Alain Trémeau - Laboratoire LIGIV. (2006). *Analyse et traitement des images couleur*. Retrieved 01-02, 2007, from <http://www.lirmm.fr/ModuleImage/ImagesCouleur-juin2006-partie2.pdf>
- [73] Jordibares. (2006). *Color 101 - color tutorial : introduction*. Retrieved 01-02, 2007, from [www.jordibares.co.uk/tutorials/color/color.html](http://www.jordibares.co.uk/tutorials/color/color.html).
- [74] *Gamut sRGB VS Adobe RGB*. Retrieved 11-03, 2007, from [http://arnaud.klinger.free.fr/photo/gestion\\_couleurs/adb98\\_srgb\\_gamut.jpg](http://arnaud.klinger.free.fr/photo/gestion_couleurs/adb98_srgb_gamut.jpg)
- [75] Osma-Ruiz, V., Godino-Llorente, J. I., Saenz-Lechon, N., & Gomez-Vilda, P. (2007). An improved watershed algorithm based on efficient computation of shortest paths. *Pattern Recognition*, 40(3), 1078-1090.
- [76] Frucci, M., Ramella, G., & Sanniti di Baja, G. Using resolution pyramids for watershed image segmentation. *Image and Vision Computing, In Press, Corrected Proof*.
- [77] Mickey, R. M., Dunn, O. J., & Clark, V. (2004). *Applied statistics : analysis of variance and regression* (3rd ed. / Ruth M. Mickey, Olive Jean Dunn, Virginia A. Clark. ed.). N.J.: Hoboken : Wiley-Interscience.
- [78] Tzeng, D.-Y., & Berns, R. S. (2005). A review of principal component analysis and its applications to color technology. *Color Research & Application*, 30(2), 84-98.
- [79] Rocha, L., Velho, L., & Carvalho, P. C. P. (2004). *Motion reconstruction using moments analysis*. Paper presented at the 17th Brazilian Symposium on Computer Graphics and Image Processing, 2004.
- [80] GretagMacbeth.*Colorchecker*. Retrieved 01-05, 2006, from <http://www.gretagmacbeth.com/>