

THÈSE DE DOCTORAT

UNIVERSITÉ DE LILLE

Pour obtenir le grade de
DOCTEUR

Spécialité : automatique, productique

Présentée et soutenue par : Balkiss HAMAD

Contribution à l'optimisation de systèmes de tailles de vêtements par une approche morphologique orientée clients

École Doctorale Lille Sciences Pour l'Ingénieur (EDSPI)
Laboratoire de recherche : GEMTEX – Génie des Matériaux Textiles

Thèse codirigée par Pascal Bruniaux et Sébastien Thomassey

Soutenue le 05 juin 2020, devant le jury composé de :

Président	M. Denis HAMAD	Professeur à l'université du Littoral
Rapporteur	M. Dominique ADOLPHE	Professeur à l'université de Haute-Alsace
Rapporteur	M. Imed KACEM	Professeur à l'université de Lorraine
Examinatrice	Mme Agnieszka CICHOCKA	Maître assistant à l'université de Łódź
Directeur de thèse	M. Pascal BRUNIAUX	Professeur à l'ENSAIT
Co-Directeur de thèse	M. Sébastien THOMASSEY	Maître de Conférences HDR à l'ENSAIT

Remerciements

Tout d'abord, j'adresse mes remerciements à la Région Hauts-de-France pour le financement du projet DIGitalisation de la conception Textile (DIGTEX_2) dont ce travail est issu.

Mes remerciements vont également à Monsieur Denis HAMAD pour m'avoir fait l'honneur d'accepter de présider mon jury de thèse ; Messieurs Dominique Adolphe et Imed Kacem d'avoir accepté d'être les rapporteurs de cette thèse, de leur intérêt pour mes travaux, de leurs commentaires constructifs sur ce manuscrit ; Madame Agnieszka Cichocka d'avoir accepté d'être l'examinatrice de ce travail.

Mes vifs remerciements vont à mon directeur de thèse, Pascal Bruniaux et à mon co-directeur Sébastien Thomassey. Je tiens à leur témoigner toute ma reconnaissance pour ces trois années qui n'auraient pas pu être possibles sans leurs précieux conseils, leur disponibilité et pour l'enrichissement intellectuel que cela m'apporté.

Tous mes remerciements vont également à tous les membres du laboratoire GEMTEX avec qui j'ai partagé mon quotidien et passé des moments agréables.

Je remercie ma sœur, mon frère et sa femme, sans oublier mes petits-neveux Amine et Aziz, ma nièce Yakout ainsi que tous mes amis pour leurs encouragements.

Enfin, je tiens à exprimer ma profonde gratitude à mon père et ma mère pour tout ce qu'ils ont fait pour moi. Je ne pourrais vous remercier assez... que Dieu vous protège.

Résumé

Dans un contexte de mondialisation qui facilite la mise en concurrence des entreprises, il est important que l'industrie de l'habillement se focalise sur le développement de nouvelles approches visant à satisfaire le client final tout en minimisant les coûts. Ces industriels sont confrontés à d'énormes et multiples défis pour répondre aux enjeux actuels qui ont pour ambitions de réduire les impacts environnementaux et la surproduction, et d'augmenter la satisfaction client. Dans ce contexte, l'analyse des morphologies de la population ciblée et la définition de système de tailles des vêtements adaptés sont des leviers importants pour répondre au plus juste aux attentes des clients. Par conséquent, cette thèse a pour objectif de définir une méthodologie de construction des systèmes de tailles optimaux pour une population donnée, puis pour chacune des classes morphologiques issues d'une classification non-supervisée à partir de mensurations sélectionnées et normalisées. La problématique majeure pour la construction de ce système de tailles est de trouver le meilleur compromis entre le bien-être des futurs vêtements, le taux de couverture de la population et le nombre de tailles proposées. Le défi est alors de traduire ce compromis à l'aide d'algorithmes d'optimisation multi-objectif et d'outils d'aide à la décision. Trois modèles de construction de systèmes de tailles sont développés, chacun proposant des degrés d'intégration du client et de facilité de mise en œuvre industrielle différents.

Mots clés : morphologies, classification non-supervisée, mensurations, système de tailles, bien-être des vêtements, optimisation multi-objectif.

Abstract

In a context of increasing globalization that facilitates competition among retailers worldwide, it is important for the apparel industry to focus on developing new approaches to satisfy the end customer while minimizing costs. As they face enormous and multiple challenges, they are forced to urgently adopt adjustment and optimization plans. It is therefore particularly useful to analyze the morphologies of the target population, to revisit the method of construction of the garment sizing system and to identify customer expectations in terms of fit. This thesis aims to define a methodology for building an optimal sizing system for a given population and then on each of the morphological classes resulting from a non-supervised classification based on selected and standardized measurements. The major issue for the construction of a sizing system is to find the best compromise between the suitability of future clothing, the coverage rate of the population and the number of sizes offered. The challenge is to convert this compromise into objective functions to be solved using multi-objective optimization algorithms and decision support tools. Three models for building sizing systems are developed, each offering different degrees of customer integration and ease for industrial implementation.

Keywords : morphologies, unsupervised classification, measurements, sizing system, clothing fit, multi-objective optimization.

Table des matières

Remerciements.....	i
Résumé.....	iii
Abstract.....	iv
LISTE DES FIGURES	vii
LISTE DES TABLEAUX.....	x
LISTE DES ABBREVIATIONS	xi
Introduction générale	1
Chapitre 1 Etat de l'art.....	5
1.1 Introduction	5
1.2 Analyse de la forme du corps humain.....	5
1.2.1 Campagne de mensuration française	6
1.2.2 Projet CAESAR	8
1.2.3 Techniques d'analyse des mesures anthropométriques	9
1.3 Système des tailles de vêtements	17
1.3.1 Système de tailles français	17
1.3.2 Normes de tailles des vêtements	18
1.3.3 Construction de système de tailles	21
1.4 Modélisation du corps humain	26
1.5 Conclusion.....	30
Chapitre 2 Outils et méthodes utilisés	33
2.1 Introduction	33
2.2 Classification des morphologies.....	33
2.2.1 Sélection de variables	34
2.2.2 Normalisation des variables	37
2.2.3 Classification.....	41
2.3 Optimisation	60
2.3.1 Formulation d'un problème d'optimisation	60
2.3.2 Notion de dominance Pareto	61
2.3.3 Les méthodes de résolution.....	62
2.3.4 Discussion	68
2.4 Conclusion.....	68
Chapitre 3 Contributions méthodologiques	71
3.1 Introduction	71
3.2 Classification des morphologies.....	71

3.2.1	Sélection des mensurations	72
3.2.2	Normalisation des mensurations	73
3.2.3	Algorithme de classification non-supervisée	75
3.3	Optimisation des systèmes de tailles par morphologies	75
3.3.1	Modèle GBSS: Grid Based Sizing System	76
3.3.2	Modèle BBSS: Box Based Sizing System	81
3.3.3	Modèle FFSS: Free Form Sizing System.....	90
3.3.4	Algorithme de construction du BBSS et FFSS	92
3.4	Conclusion.....	96
Chapitre 4 Mise en œuvre et résultats		99
4.1	Introduction	99
4.2	Classification non supervisée de morphologies	99
4.2.1	Description de la base de données	99
4.2.2	Prétraitement de données	99
4.2.3	Sélection des variables	100
4.2.4	Normalisation des variables	105
4.2.5	Classification non supervisée.....	107
4.3	Optimisation système de tailles	110
4.3.1	Modèle GBSS	110
4.3.2	Modèles orientées clients (BBSS et FFSS).....	120
4.3.3	Comparaison des modèles GBSS, BBSS et FFSS	135
4.4	Conclusion.....	137
Conclusion générale.....		141
Liste des publications.....		146
BIBLIOGRAPHIE		147
Annexe 1 : Description mesures anthropométriques		155
Annexe 2 : Exemples de questions posés lors de l'enquête clients		156
Annexe 3 : Exemple de l'arbre de décision généré.....		157
Annexe 4 : Systèmes de tailles pour les mensurations Poitrine/Taille sur la population totale		158
Annexe 5 : Systèmes de tailles pour les mensurations Stature/Poitrine par mophotype		162
Annexe 6 : Systèmes de tailles pour les mensurations Poitrine/Hanche par mophotype		169

LISTE DES FIGURES

Figure 1. Illustration des conformations chez la femme [5]	7
Figure 2. Les trois postures du projet CAESAR [11]	8
Figure 3. Principales techniques d'analyse	10
Figure 4. Exemple des mesures anthropométriques [19].....	10
Figure 5. Représentation graphique d'une distribution de loi normale [22]	11
Figure 6. Processus de l'ACP adopté par Gupta[28]	13
Figure 7. Sept formes du corps représentatives des femmes [34].....	15
Figure 8. Classification en deux étapes [40]	16
Figure 9. Répartition des femmes françaises selon leur taille de vêtements[47].....	18
Figure 10. Méthode théorique de définition d'un système de tailles	21
Figure 11. Exemple d'un pictogramme femme[56].....	22
Figure 12. Principe de construction d'un système de tailles à partir d'un SOM proposé par Doustaneh et <i>al.</i> [60].....	25
Figure 13. Exemple d'alignement d'un modèle avec un scan 3D par Hirshberg et <i>al.</i> [64].....	27
Figure 14. Deux phases pour la modélisation du corps par Seo et <i>al.</i> [66]	27
Figure 15. Fonction du Correctif de Gregory [69].....	28
Figure 16. Modélisation de mannequin 3D par Hsiao[70]	29
Figure 17. Système de modélisation interactif de Zeng et <i>al.</i> [71].....	29
Figure 18. Méthodologie proposée	31
Figure 19. Processus ECD adapté à la classification des morphologies	34
Figure 20. Phénomène de sur-apprentissage[76]	35
Figure 21. Exemple de deux composantes principales	36
Figure 22. Exemple de diagramme d'éboulis	37
Figure 23. Processus de normalisation par la sphère minimale englobante [63].....	39
Figure 24. Sphère minimale englobante d'un modèle 3D humain [84].....	39
Figure 25. Illustration de deux types de classification.....	42
Figure 26. Méthodes de calcul de l'erreur de prédiction [97].....	44
Figure 27. Procédure de la validation croisée.....	44
Figure 28. Taxonomie des approches de clustering [102].....	47
Figure 29. Dendrogramme obtenu après une classification ascendante hiérarchique avec quatre classes [63].....	48
Figure 30. Illustration de différents critères d'agrégation [107].....	50
Figure 31. Illustration de déroulement de l'algorithme K-moyennes[37].	52
Figure 32. Architecture de la carte auto-organisatrice 2D.....	55
Figure 33. Illustration de différentes topologies de la carte auto-organisatrice.....	56
Figure 34. Illustration de front de Pareto pour un problème de minimisation bi-objectif [122]	62
Figure 35. Classification des méthodes de résolution [124]	63
Figure 36. Critères contradictoires des métaheuristiques [124].....	64
Figure 37. Fonctionnement général d'un AG [128].....	65
Figure 38. Exemple de types de croisement [130].....	67
Figure 39. Exemple d'une mutation binaire[131].....	68
Figure 40. Processus de classification non-supervisée de morphologies.	72
Figure 41. Exemple de différents types de morphologies [132].....	72

Figure 42. Exemple de deux morphologies similaires normalisées par le cercle minimum englobant.....	73
Figure 43. Illustration d'algorithme de cercle minimal englobant.	74
Figure 44. Illustration de grille de tailles.	76
Figure 45. Catégories pour l'évaluation du système de tailles.	77
Figure 46. Processus de construction d'un système de tailles optimisé du modèle GBSS.....	79
Figure 47. Illustration d'un front de Pareto obtenu avec un modèle GBSS	80
Figure 48. Processus de prédiction des mensurations.....	84
Figure 49. Illustration du calcul de deux écarts-types pour taille de vêtement.....	85
Figure 50. Illustration de la régression linéaire utilisée pour estimer les écarts types pour différentes tailles.....	86
Figure 51. Illustration d'un système de tailles générées par le modèle BBSS.	87
Figure 52. Illustration de la loi normale.....	90
Figure 53. Exemple de codage d'un chromosome.....	93
Figure 54. Problème de l'opérateur de croisement classique	94
Figure 55. Principe de l'opérateur de croisement proposé	95
Figure 56. Illustration de la variation de Fitness en fonction nombre de tailles.....	95
Figure 57. Comparaison entre les modèles.	97
Figure 58. Principe général de la méthodologie proposée	98
Figure 59. Scree plot des composantes principales de l'ACP.	101
Figure 60. Biplot de l'ACP.	102
Figure 61. Contribution des variables sur la deuxième composante.	103
Figure 62. Contribution des variables sur la première composante.....	103
Figure 63. Score Laplacien	104
Figure 64. Descripteur de forme 2D du torse par ces points de mensurations.	105
Figure 65. Cercle minimum englobant.	106
Figure 66. Cercle minimum englobant de rayon $r=1$	106
Figure 67. Biplot de l'ACP après la normalisation.....	106
Figure 68. Indices Davies-Bouldin et Silhouette en fonction du nombre de classes.....	107
Figure 69. Les deux morphotypes représentant les deux classes morphologiques.....	109
Figure 70. Systèmes de tailles générés par le modèle GBSS sur la population totale pour les mensurations stature/tour de poitrine.....	112
Figure 71. Résultats obtenus pour l'ensemble des combinaisons de pas (a) et le front de Pareto correspondant (b) pour le système de tailles poitrine/hanche.....	112
Figure 72. Pas de la grille pour les systèmes de tailles du front de Pareto	113
Figure 73. Front de Pareto obtenu sur la population totale et les deux classes morphologiques pour les systèmes stature/poitrine	114
Figure 74. Comparaison des performances des systèmes de tailles générés pour la population totale et pour les deux morphologies	115
Figure 75. Comparaison des performances d'un système à 12 tailles pour la population totale et des deux systèmes à 6 tailles pour chacune des morphologies	115
Figure 76. Systèmes obtenus avec 12 tailles pour la population totale (a), 6 tailles pour la morphologie 1 (b) et 6 tailles pour la morphologie 2 (c).....	116
Figure 77. Systèmes de tailles générés par le modèle GBSS sur la population totale pour les tour de hanche/tour de poitrine	117
Figure 78. Front de Pareto obtenu sur la population totale et les deux classes morphologiques pour les systèmes hanche/poitrine	118
Figure 79. Comparaison des performances des systèmes de tailles générés pour la population totale et pour les deux morphologies	118

Figure 80. Comparaison des performances d'un système à 12 tailles pour la population totale et des deux systèmes à 6 tailles pour chacune des morphologies	119
Figure 81. Distributions des mesures du tour de poitrine des clients parmi les tailles de l'enseigne T_i ($i = 1, \dots, 4$)	122
Figure 82. Lois semi-normales obtenues à partir des tours de poitrine des clients et de moyenne T_i ($i = 1, \dots, 4$)	123
Figure 83. Prédiction des écarts types pour chacune des tailles par régression linéaire.....	124
Figure 84. Exemple de SOM pour l'initialisation de la population de l'algorithme génétique sur un système stature/poitrine.....	125
Figure 85. Convergence de la fonction fitness du modèle BBSS (a) et le modèle FFSS (b)	126
Figure 86. Systèmes de tailles générés par le modèle BBSS sur la population totale pour les mensurations stature/tour de poitrine.....	127
Figure 87. Evaluation par la distance de Manhattan (a) et la fonction objectif P (b) des systèmes des tailles obtenus avec le modèle BBSS pour les mensurations stature/poitrine..	128
Figure 88. Comparaison des performances des systèmes de tailles générés pour la population totale et pour les deux morphologies avec les critères Manhattan (a) et P (b)	128
Figure 89. Exemples d'optimums locaux sur les morphologies 1 et 2 avec les systèmes de tailles générés sur la population totale	129
Figure 90. Systèmes de tailles générés par le modèle BBSS sur la population totale pour les mensurations tour de hanche/tour de poitrine	130
Figure 91. Évaluation par la distance de Manhattan (a) et la fonction objectif P (b) des systèmes des tailles obtenus avec le modèle BBSS pour les mensurations hanche/poitrine .	130
Figure 92. Comparaison des performances des systèmes de tailles générés pour la population totale et pour les deux morphologies avec les critères Manahattan (a) et P (b).....	131
Figure 93. Systèmes de tailles générés par le modèle FFSS sur la population totale pour les mensurations stature/tour de poitrine.....	132
Figure 94. Évaluation par la distance de Manhattan (a) et la fonction objectif Q (b) des systèmes des tailles obtenus avec le modèle FFSS pour les mensurations stature/poitrine...	132
Figure 95. Comparaison des performances des systèmes de tailles générés pour la population totale et pour les deux morphologies avec les critères Manahattan (a) et Q (b).....	133
Figure 96. Systèmes de tailles générés par le modèle FFSS sur la population totale pour les mensurations tour de hanche/tour de poitrine	134
Figure 97. Evaluation par la distance de Manhattan (a) et la fonction objectif Q (b) des systèmes des tailles obtenus avec le modèle FFSS pour les mensurations hanche/poitrine ..	134
Figure 98. Comparaison des performances des systèmes de tailles générés pour la population totale et pour les deux morphologies avec les critères Manahattan (a) et Q (b).....	134
Figure 99. Comparaison des trois modèles avec le critère Manhattan sur la morphologie 2 et les mensurations hanche/poitrine	136
Figure 100. Comparaison des trois modèles avec le critère P sur la population totale et les mensurations stature/poitrine	136
Figure 101. Comparaison des trois modèles avec le critère Q sur la morphologie 1 et les mensurations stature/poitrine	137
Figure 102. Illustration mensurations CAESAR.	155
Figure 103. Questionnaire de satisfaction clients	156
Figure 104. Exemple d'une partie d'arbre CART généré pour le tour de poitrine.	157

LISTE DES TABLEAUX

Tableau 1. Exemple de campagnes de mensurations.....	6
Tableau 2. Classification de la population en termes de tailles [24]	12
Tableau 3. Normes de tailles des vêtements.	20
Tableau 4. Exemple de suppression des tailles de vêtements [17]	24
Tableau 5. Classification Supervisée Vs Non-supervisée.....	42
Tableau 6. Mesures de distances classiques	46
Tableau 7. Coefficients de corrélation.	100
Tableau 8. Les variables ordonnées selon le score Laplacien	104
Tableau 9. Nombre d'individus pour les deux classes.....	108
Tableau 10. Mensurations des deux centroïdes	108
Tableau 11. Mensurations clés des quatre tailles de l'enseigne.....	121
Tableau 12. SMAPE sur les mensurations prédites.	121
Tableau 13. Paramètres de l'algorithme génétique.....	125

LISTE DES ABBREVIATIONS

ACP *Analyse en Composante Principales*

AG *Algorithme Génétique*

BBSS *Box Based Sizing System*

CAESAR *Civilian American and European Surface Anthropometry Resource*

CAO *Conception Assistée par Ordinateur*

ECD *Extraction des Connaissances à partir de Données*

FFSS *Free Form Sizing System*

FL *Fit Level*

GBSS *Grid Based Sizing System*

IFTH *Institut Français de la mode et de l'habillement*

SMAPE *Symmetric Mean Absolute Percent Error*

SOM *Self Organizing Map*

Introduction générale

Dans le contexte actuel du marché de l'habillement, fortement concerné par les enjeux environnementaux qui visent à réduire la surproduction et influencé par les exigences croissantes des consommateurs, les enseignes de la distribution textile se doivent d'innover continuellement. A l'ère du "big data", l'accès aux données sensibles des consommateurs est devenu une opportunité pour ces entreprises afin de concevoir, produire et distribuer des produits en adéquation avec la demande de leur clientèle. Parmi l'ensemble des informations disponibles, la connaissance des données morphologiques des clients s'avère très utile pour ajuster les tailles standards des vêtements. En effet, l'inadéquation des tailles conduit à des résultats néfastes sur l'activité de l'entreprise : augmentation des invendus, des ventes perdues et retours de produits, couplé à plus long terme d'une baisse de la satisfaction du client. Les tailles doivent donc être adaptées aux attentes de la clientèle et notamment aux mensurations de celle-ci. Ces dernières années, l'usage de scanner corporel 3D a permis d'optimiser le processus de prises de mesures à la fois en précision et en rapidité. Cet outil de mesure permet d'acquérir des bases de données conséquentes et représentatives d'une population donnée. Cette technologie avancée ouvre également de nouvelles perspectives d'analyse de la forme du corps humain conduisant à la création de mannequins morphotypes (avatars 3D) qui peuvent être utilisés pour des cabines d'essayage virtuelles ou du prototypage rapide. Néanmoins, ces avatars sont généralement fabriqués pour des tailles standards en faisant abstraction des évolutions morphologiques pour les autres tailles. La solution proposée pour pallier ce problème est de redéfinir des tailles propres à chaque morphologie.

Ces travaux de thèse ont pour objectif de contribuer à la création de nouveaux systèmes de tailles plus adaptés aux morphologies d'une population donnée. Pour ce faire, une classification non-supervisée a été réalisée pour identifier les différentes classes morphologiques d'une population à partir d'une campagne de mensurations, et extraire ensuite les morphologies représentatives de chaque classe. Par la suite, différents modèles de constructions de système de tailles optimal en adéquation avec les résultats de la classification ont été proposés. Le premier modèle GBSS (Grid Based Sizing System), inspiré de la littérature, est fondé sur la création d'une grille de tailles entre deux mensurations du corps. Nous avons tenté de l'améliorer en introduisant un nouveau critère de performance, appelé FL (Fit Level), mesurant le bien-être comme la distance entre la mensuration réelle du corps humain et la taille assignée. L'optimisation du système de taille consiste alors à le construire par une incrémentation du nombre de tailles en ajoutant le voisin

qui minimise ce critère. Ensuite, nous avons exploré les diverses possibilités de combinaisons d'intervalles entre chaque taille. Les résultats définissent un front de Pareto qui constitue un outil d'aide à la décision intéressant afin de définir le meilleur compromis entre le nombre de tailles n et le critère FL. Cependant, ce modèle est limité par sa sensibilité à la convergence vers des optimums locaux. Au niveau pratique, ce modèle nécessite également de fixer subjectivement des tolérances qui influent sur les résultats obtenus. Les solutions retenues consistent à :

- D'une part, impliquer les clients dans la construction du système de tailles par l'intégration de leurs exigences en termes de tailles et de bien aller. Ces préférences sont extraites d'une enquête de satisfaction menée par une enseigne.
- D'autre part, mettre en œuvre un algorithme de construction du système de tailles basé sur des techniques méta-heuristique. Ainsi, un algorithme génétique a été développé spécifiquement pour notre problème afin d'optimiser les tailles aux mensurations de la population ciblée.

Ces deux évolutions sont implémentées dans des modèles nommés BBSS (Box Based Sizing System) et FFSS (Free Form Sizing System). Le principe général de ces modèles consiste à utiliser deux bases de données pour l'apprentissage. Une première base de données, contient les données de mensurations obtenues par une campagne de mensurations sur une population. Cette base est utilisée pour l'optimisation du processus de contraction du système de tailles. La seconde base de données, de taille plus modeste que la première, est constituée à partir de l'enquête de satisfaction des clients et est utilisée pour définir les tolérances préférentielles des clients. Ces tolérances clients, extraites de la seconde base, deviennent des paramètres de la fonction objectif utilisée pour l'optimisation du système de tailles sur la première de base de données. Ainsi, cette approche semi-supervisée permet d'apprendre les préférences des clients sur une base de données de petite taille et de les généraliser sur une base de données de mensurations plus conséquente et représentative d'une population plus large.

Le modèle BBSS représente chaque taille par une boîte délimitée par les tolérances estimées afin de maintenir une couverture de la population. La fonction objectif utilisée consiste à minimiser la distance entre les clients potentiels et leur taille la plus proche et à minimiser le nombre de clients non couverts par le système. Ces tolérances sont aussi utilisées comme un moyen de pondération des mensurations lors du calcul de la distance car toutes les mensurations n'ont pas la même importance pour le client.

Le modèle FFSS s'appuie sur une fonction objectif qui maximise la probabilité que les tailles du système soient adaptées aux clients selon les préférences qu'ils ont déclarés dans l'enquête de satisfaction. Ce modèle propose une approche très orientée sur les données et le client.

Ce manuscrit s'articule autour de quatre chapitres :

Le premier chapitre résume les travaux de recherche qui portent sur l'acquisition et le traitement des données anthropométriques dans un premier temps, et dans un deuxième temps sur la construction des systèmes de tailles de vêtements et la modélisation 3D du corps humain. Ce chapitre expose également la méthodologie générale que nous proposons de suivre tout au long de ces travaux.

Le deuxième chapitre est dédié aux méthodes et outils mathématiques permettant de mettre en œuvre cette méthodologie, notamment la classification et l'optimisation.

Dans le troisième chapitre, nous détaillons la méthodologie proposée. Ce chapitre permet également de répondre aux questions de recherche qui émergent lors du développement de nos modèles :

- Comment choisir les mensurations du corps pertinentes pour caractériser les morphologies ?
- Comment normaliser ces mesures en préservant les proportions du corps humain ?
- Comment identifier les morphologies dans une population à partir des mensurations ?
- Comment construire un système de tailles optimal en intégrant les tolérances clients ?
- Comment définir les tolérances clients en termes de tailles ?

Dans le quatrième chapitre, nous présentons les résultats de l'application de ces différentes méthodes sur la base de données CAESAR et à partir de l'enquête de satisfaction menée par une enseigne de prêt à porter féminin.

Enfin, la conclusion expose les principaux apports de notre travail, définit les limites des modèles proposées et présentent les évolutions possibles qui pourront faire l'objet de travaux futurs.

Chapitre 1 Etat de l'art

1.1 Introduction

Les informations concernant la forme du corps humain sont fondamentales pour le domaine de l'industrie du vêtement. Les résultats de l'analyse de ces informations permettent de reconnaître les changements révolus dans la population et d'avoir une approximation au plus précise des mesures du corps, ce qui influence directement la conception des vêtements. Ainsi, une analyse précise de la forme du corps est une garantie importante pour les enseignes de répondre aux exigences du consommateur en matière d'ajustement du vêtement. En prenant en compte ces analyses, les enseignes de prêt-à-porter peuvent offrir à leurs clients de nouveaux systèmes de tailles plus adaptés à leurs morphologies. Dans ce chapitre, nous présentons différents exemples de campagnes de mensuration qui montrent l'importance internationale du besoin de catégoriser le corps humain. Ensuite, nous passons en revue les principales techniques d'analyse de ce type de données issues de ces campagnes. Les travaux de recherche qui portent sur la construction des systèmes de tailles et la modélisation du corps humain en 3D sont également résumés. Enfin, à partir de ces états de l'art, nous exposons la méthodologie qui sera adoptée tout au long de ces travaux.

1.2 Analyse de la forme du corps humain

Dans la dernière décennie, des progrès remarquables ont été effectués dans le cadre du développement de systèmes de numérisation complète du corps humain comme l'usage des scanners corporels 3D. Ils permettent de numériser en 3D la forme exacte du corps, d'obtenir des données plus précises telles que les mensurations et d'offrir un moyen pratique d'exporter un modèle 3D dans les logiciels de CAO (Conception assistée par ordinateur) [1]. Selon D'Apuzzo, cette technique numérique de mesure du corps humain est rapide, précise et très facile à utiliser [2]. Les chercheurs du domaine de l'habillement reconnaissent depuis longtemps la nécessité d'établir une analyse ciblée du corps humain au sein d'une population [3]. En effet, une population est caractérisée par une large variabilité morphologique qui est très significatif d'un pays à l'autre et même au sein d'un même pays. Cette variabilité est liée aux différences d'âges, à la culture et au mode de vie. Ainsi, nombreux sont les pays ayant déjà lancé de vastes campagnes de mensurations en utilisant ces scanners 3D, afin de

connaître véritablement l'évolution de la morphologie de leur population (**Tableau 1**). Les bases de données issues de ces campagnes sont utilisées pour divers types de projets dans différents domaines industriels tels que le textile, l'automobile et l'aéronautique. Selon le besoin de l'étude et le type de scanner corporel 3D utilisé, le nombre de mesures collectées varie.

Pays	Date de l'enquête	Âges des participants	Nombre de participants	Scanner utilisé	Nombre des mesures
États-Unis et l'Europe : CAESAR [4]	1998–2002	18–65	4400, H & F	WB4, Cyberware	57(+40 manuels)
France: IFTH [5]	2003–2005	5–70	11 562, H & F	Vitronic	55(+10 manuels)
Allemagne: SizeGerman[6]	2007–2008	6–87	13 400, H & F	Vitronic	43
Royaume-Uni: SizeUK[7]	1999–2002	16–90	11 000, H & F	TC2	130(+10 manuels)
Canada et aux États-Unis: SizeNorthAmerica[8]	2017-2018	6-75	17 820, H & F	VITUS Smart XXL	140

Tableau 1. Exemple de campagnes de mensurations.

Dans ce qui suit, nous détaillons la campagne de mensuration française de l'IFTH et le projet CAESAR qui est la source de données de nos travaux.

1.2.1 Campagne de mensuration française

La campagne nationale de mensuration réalisée par l'IFTH (Institut Français de la Mode et de l'Habillement) a permis la publication d'une synthèse des nouvelles mensurations de la population française adulte. Cette campagne a été menée du 8 avril 2003 au 23 avril 2005 sur un échantillon de 11 562 personnes, âgées de 5 à 70 ans. Deux cabines de mesures 3D ont été déplacées sur 37 sites dans toutes les régions de France pour atteindre le nombre d'échantillon nécessaire à l'analyse de la population Française suivant différents critères, en particulier le critère morphologique fortement associé à la dimension et au sexe.

Les principaux résultats font apparaître que la taille moyenne de l'homme français est de 175,6 cm et celle de la femme 162,5 cm. En un siècle, cela représente une augmentation de 11 cm pour les hommes, et 8 cm pour les femmes. Une femme sur cinq porte un vêtement de taille 40 et plus d'un tiers des femmes s'habille en 40 ou 42 [5].

Afin de modéliser plus précisément la morphologie des femmes, une partition plus fine de la population féminine est faite en considérant la notion de conformation (ou encore DROP). Cette notion est une demie-différence entre le tour de poitrine et le tour de bassin [5].

$$DROP = \frac{1}{2} Tourdebassin - \frac{1}{2} Tourdepoitrine$$

Trois groupes de conformations apparaissent chez les femmes (**Figure 1**):

- Une conformation mince, avec un tour de bassin inférieur ou égal au tour de poitrine, associée au DROP 0.
- Une conformation normale, avec un tour de bassin supérieur au tour de poitrine de 0 à 8 cm, faisant référence au DROP]0-4].
- Une conformation forte, avec un tour de bassin supérieur d'au moins 8 cm au tour de poitrine, associée au DROP > 4.

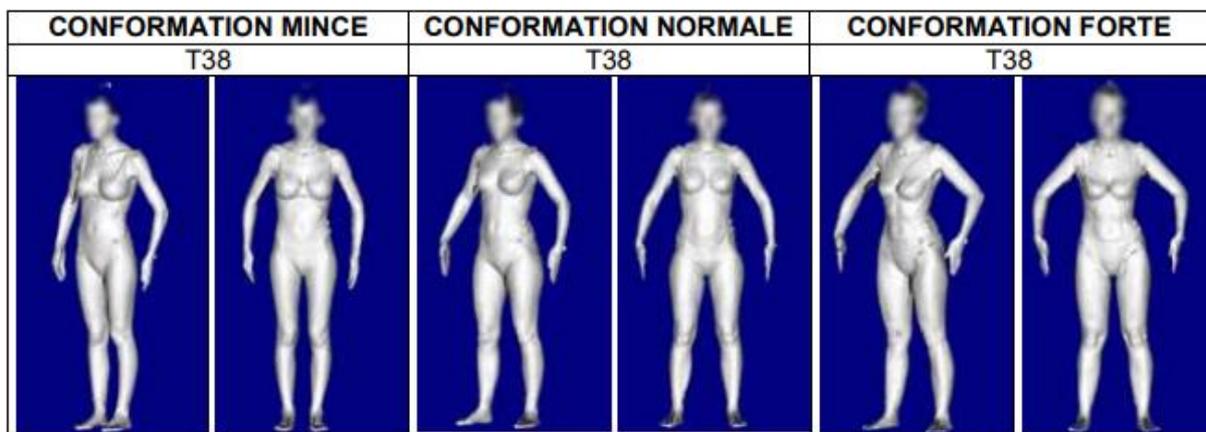


Figure 1. Illustration des conformations chez la femme [5]

L'objectif principal de la campagne est de réactualiser les tailles et patrons des vêtements pour l'industrie d'habillement en fonction des morphologies. Les résultats de la campagne de mensuration révèlent les différences de morphologie de la population française. Cette initiative a été effectuée pour apporter les réponses nécessaires aux industriels de la mode. Néanmoins, ces statistiques basées sur des mesures unidimensionnelles ne nous reflètent que partiellement la véritable variété des formes du corps humain. En effet, avec une information bidimensionnelle ou tridimensionnelle, nous pouvons améliorer la pratique existante et ajouter de nouvelles perspectives à la création des systèmes de tailles et à l'amélioration de l'ajustement du vêtement à la morphologie du corps [9]. En outre, le scanner 3D fournit un

nombre important de mesures corporelles. Certaines d'entre elles ne seront pas significatives pour la partition de la population. Cela nécessite donc un choix des mesures pertinentes pour la conception des vêtements.

1.2.2 Projet CAESAR

Le projet CAESAR (Civilian American and European Surface Anthropometry Resource) est une étude anthropométrique réalisée entre 1998 et 2001 sur près de 4400 civils américains et européens âgés de 18 à 65 ans qui possèdent des morphologies différentes [4]. Les scans 3D ont été obtenus en utilisant le scanner Cyberware WB4[10]. Cette étude combine les données démographiques, les mesures anthropométriques traditionnelles effectuées par un ergonome, les modèles 3D représentant trois postures distinctes et les coordonnées des marqueurs anthropométriques placés sur le sujet (**Figure 2**) [11]. Le choix des postures a été effectué pour minimiser les surfaces cachées lors de l'utilisation du scanner [9]. Cependant, le coût pour l'obtention des données du projet CAESAR est de 27 000 dollars.

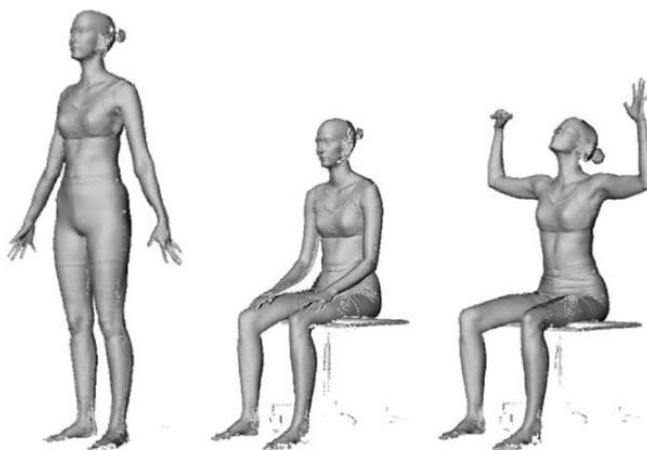


Figure 2. Les trois postures du projet CAESAR [11]

Cette enquête fait partie des premières études anthropométriques à grande échelle qui ont fourni une description détaillée et précise de la forme du corps et fait l'objet de nombreux travaux de recherche. Certains travaux ont été menés sur la localisation des points de repère anatomique et l'extraction automatique des mesures qui nécessitent un temps de prétraitement des scans [10,12]. Ces mesures correspondent à des distances entre des points de repère anatomiques ou à des circonférences définies à des endroits spécifiques.

Robinette [13] a investigué la précision des mesures extraites du scanner et a montré que cette technique de prise de mesures donne des résultats fiables. En revanche, elle ne permet

pas d'identifier avec rigueur les points non marqués. Par ailleurs, la mise en place de marqueurs avant la numérisation est une procédure qui nécessite une intervention humaine fastidieuse. D'autres travaux portent sur la définition des descripteurs tridimensionnels de la forme du corps humain en 3D [14–16]. L'idée est de transformer un nuage de points de surface numérisés en une représentation compacte qui contient les informations représentatives de la forme du corps humain. Un des descripteurs développé au sein du laboratoire GEMTEX est basée sur le calcul des distances géodésiques par rapport un ensemble de points des références [15]. Cette distance géodésique entre deux points correspond à la longueur du chemin le plus court qui les sépare [15]. Elle est invariante par rapport aux transformations (rotation, translation,...) mais elle reste sensible à la position et au nombre des points à considérer.

1.2.3 Techniques d'analyse des mesures anthropométriques

La numérisation en 3D du corps humain n'est pas en soi suffisante pour extraire l'information sur la forme du corps humain. En effet, l'exploitation optimale des données 3D générées est dépendante du choix des outils d'analyse et de traitement. Avec le progrès des technologies et l'essor des logiciels de calculs avancés, la tendance des techniques appliquées dans ce domaine est passée de simples calculs statistiques à des techniques d'exploration de données telles que l'analyse en composantes principales (ACP) et la classification.

La revue de la littérature montre que les travaux réalisés dans ce cadre se classifient en quatre catégories principales : analyse descriptive univariée, analyse descriptive bivariée, analyse descriptive multivariée et classification [17, 18] (**Figure 3**).

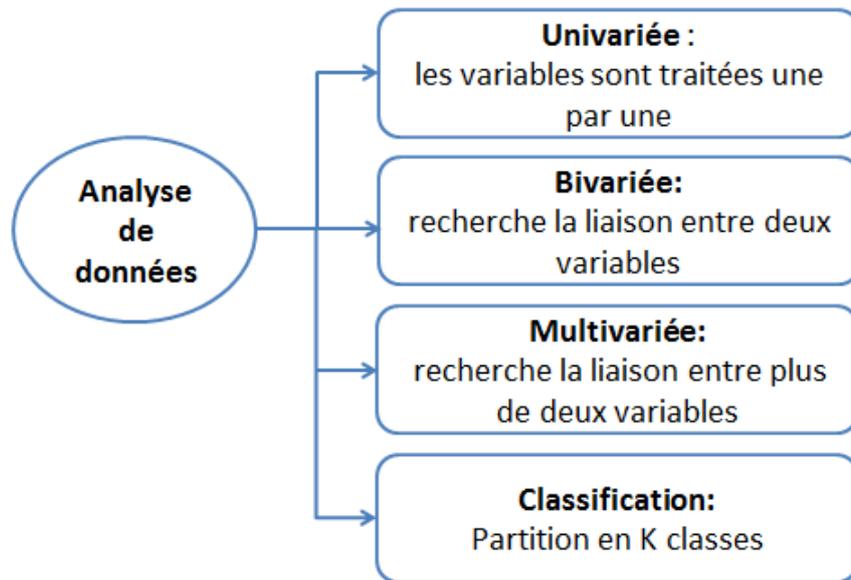


Figure 3. Principales techniques d'analyse

1.2.3.1 Analyse descriptive univariée

Les mesures anthropométriques sont utilisées pour l'étude des mesures du corps humain à des fins de classification et de comparaison morphologique. Elles comprennent des mesures prises horizontalement (tour de poitrine, tour de bassin, tour de taille...) et verticalement (stature, hauteur latérale du tour de taille,...) (**Figure 4**) [19].

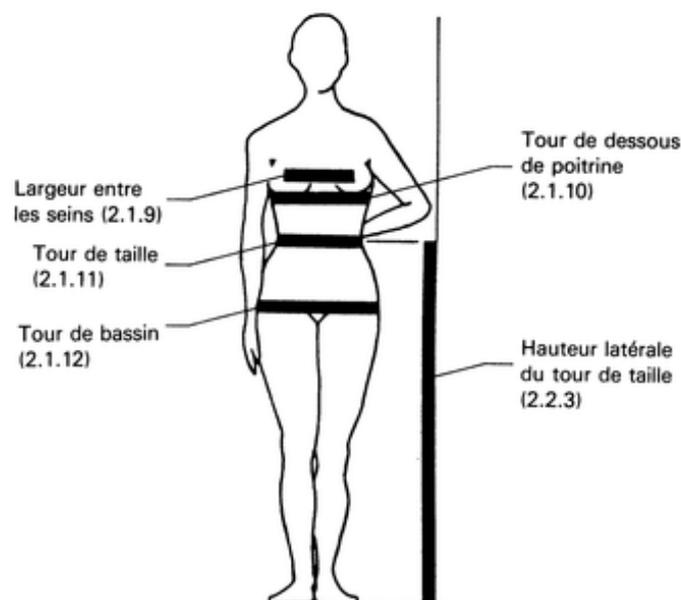


Figure 4. Exemple des mesures anthropométriques [19]

Plusieurs chercheurs ont essayé d'établir des analyses statistiques des mensurations des corps d'une population donnée et de les représenter sous une forme graphique. Ces représentations statistiques servent à synthétiser les données étudiées au moyen d'un petit nombre de valeurs. Nous distinguons essentiellement des caractéristiques de tendance centrale (moyenne arithmétique, médiane,...) et des caractéristiques de dispersion (l'écart type) qui servent à préciser la variabilité de la population et à résumer l'éloignement de l'ensemble des observations par rapport à leur tendance centrale [20]. La représentation graphique de ces caractéristiques au sein d'une population tend généralement à suivre une distribution gaussienne et se rapproche d'une courbe en forme de cloche de la loi normale [21]. La courbe est définie par la donnée de la valeur moyenne μ et de l'écart type σ comme l'illustre la **Figure 5**.

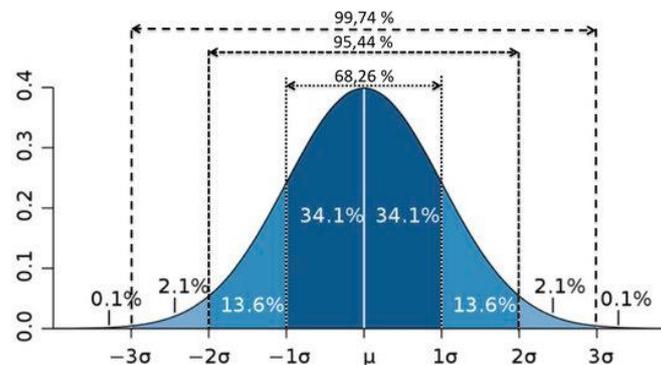


Figure 5. Représentation graphique d'une distribution de loi normale [22]

La technique d'analyse basée sur l'écart-type consiste alors à additionner et soustraire l'écart type de la moyenne de l'ensemble des données pour former les différentes catégories de la population. En appliquant cette technique, Gupta et *al.* [23], ont mené une étude anthropométrique sur 500 étudiants Indiens, âgés de 18 à 26 ans. La première classification des données de la population a été effectuée sur la base de la taille des sujets à l'aide de l'écart-type. La population a été divisée en trois catégories principales de tailles, à savoir : Petite = $< \mu - \sigma$, Moyenne = $\mu \pm \sigma$, et grande = $\mu + \sigma$. Mpampa et *al.* [24], ont choisi un intervalle plus grand entre $[\mu - 2\sigma, \mu + 2\sigma]$, afin de prendre en compte la majorité de la population. Six catégories de tailles sont définies (**Tableau 2**). L'élargissement de l'intervalle influe forcément sur le nombre final de tailles de vêtements à produire.

Body type	Height (cm)
Very short < Mean - 2SD	< 164
Mean - 2SD < Short < Mean - SD	164-171
Mean - SD < Normal < Mean	171-178
Mean < Tall < Mean + SD	178-185
Mean + SD < Very tall < Mean + 2SD	185-192
Too much tall > Mean + 2SD	192 >

Tableau 2. Classification de la population en termes de tailles [24]

L'application de l'écart-type est particulièrement adaptée pour être utilisée avec des données respectant une distribution normale. Cependant, elle est affectée par la détermination de la taille de l'intervalle et l'échantillon. De plus, cette technique est limitée aux ensembles de données pour une seule mesure du corps. Pour approfondir l'analyse et étudier la liaison entre deux mesures, il est nécessaire de procéder à une analyse bivariée.

1.2.3.2 Analyse descriptive bivariée

La détermination du coefficient de corrélation semble être la méthode la plus couramment utilisée par les chercheurs pour identifier les mesures clés nécessaires à l'élaboration des systèmes de tailles[17]. Selon Petrova[25], la première étude sur les mesures corporelles a été menée en 1941 par O'Brien et Shelton sur 10042 femmes aux États-Unis. Les auteurs ont utilisé une technique de distribution bivariée. Les résultats montrent que les meilleurs prédicteurs de la forme du corps étaient la stature pour les dimensions verticales et le poids pour les dimensions horizontales. Le poids était considéré comme problématique d'un point de vue pratique car il était prévisible que les femmes s'opposent à donner leur poids mais pas leur tour de poitrine ou leur tour de taille. Ainsi, O'Brien et Shelton ont pensé qu'il était nécessaire de considérer les combinaisons de la stature avec d'autres mesures[26]. Ils ont choisi alors le tour de poitrine au lieu du poids pour le haut du corps, le tour de hanche pour le bas du corps. De même, Otieno [27] présente une analyse statistique sur 20 mesures corporelles recueillies auprès de 150 femmes âgées de 19 à 35 ans au Royaume-Uni. Les données ont été analysées à l'aide de statistiques descriptives à partir des valeurs moyennes, des écart-types et des percentiles. Les dimensions clés ont été analysées à l'aide des coefficients de corrélation. Les cinq mesures (stature, buste, taille, hanche, entrejambe) qui ont le coefficient de corrélation le plus élevé avec d'autres variables sont choisies comme des dimensions clés.

L'analyse bivariée se focalise sur la relation entre deux variables et néglige le fait que cette relation peut être influencée par d'autres variables. En outre, si le nombre de variables est important, il est difficile de prendre en compte l'ensemble des combinaisons possibles. L'analyse statistique multivariée est donc recommandée lorsqu'il s'agit de traiter plusieurs variables.

1.2.3.3 Analyse descriptive multivariée

Dans le cadre du développement de systèmes de tailles à partir de données anthropométriques, plusieurs auteurs se sont intéressés à l'application de l'Analyse en Composantes principales (ACP) pour extraire les dimensions clés. Par la suite, ces dimensions clés peuvent être utilisées pour diviser la population en sous-groupes homogènes. Gupta [28] propose une sélection des dimensions clés basée sur l'ACP illustré en **Figure 6**. Deux composantes principales sont retenues : composante pour les circonférences et composante pour les longueurs du corps. De chaque composante, une dimension clé a été sélectionnée représentant la valeur propre la plus élevée. Le fait que cette valeur soit forte signifie que la variable est étroitement associée à la composante. Dans l'ensemble, les variables de circonférences choisies sont le tour de poitrine pour le haut du corps et le tour de hanche pour le bas du corps. Le choix de ces variables comme dimensions clés confirme les résultats de l'analyse des corrélations pour le haut du corps, tandis que pour le bas du corps le tour de hanche a été choisi au lieu de tour de taille.

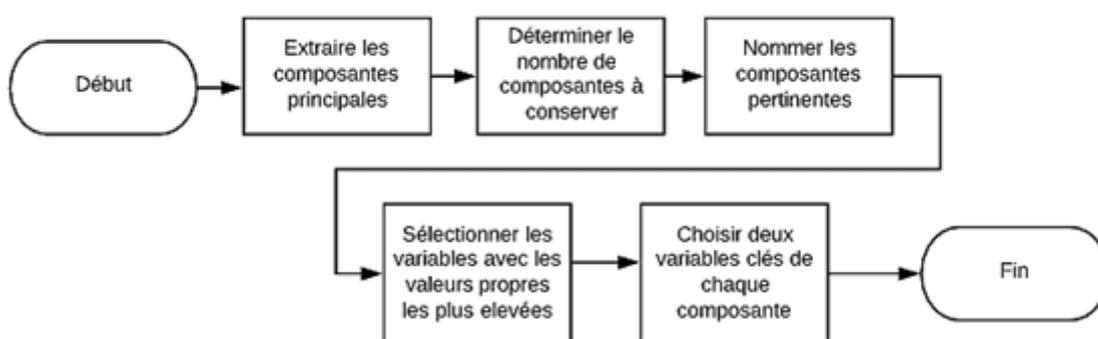


Figure 6. Processus de l'ACP adopté par Gupta[28]

Le même principe a été adopté par Vithanage et *al.* [29]. Ils proposent le développement d'un guide de taille de pantalon pour les femmes Sri-lankaises. Les deux dimensions sélectionnées sont le tour de taille pour la première composante et la longueur d'entrejambe pour la seconde. Hsu a mené une enquête anthropométrique sur 956 femmes taïwanaises. L'ACP a

été appliquée sur 16 mesures corporelles. Deux composantes principales ont été identifiées comme étant la composante des circonférences et la composante des longueurs. Les deux composantes ont été examinées avec la méthode de Ward pour une partition de la population et cinq morphologies de corps ont été identifiées[30]. Cette méthode est particulièrement adaptée lorsque la classification est effectuée après une ACP, les individus à classer étant repérés par leurs coordonnées sur les premières composantes factorielles.

L'intérêt de l'ACP réside dans la réduction de la dimension de représentation des données mais la qualité de résultats dépend du nombre de composantes à retenir. La limite de cette technique vient du fait que c'est une méthode de projection, et que la perte d'information induite par la projection peut entraîner des interprétations erronées[31]. Il existe d'autres méthodes d'analyse multivariées comme l'Analyse Factorielle des Correspondances (AFC) et l'Analyse des Correspondances Multiples (ACM) permettant le traitement des variables qualitatives [32].

1.2.3.4 Classification

L'objectif de la classification est de construire des classes d'individus telles que les individus dans une même classe se ressemblent le plus et telles que les individus de classes distinctes se démarquent le plus [33]. Il existe une très large famille de méthodes détaillées dans la section 2.2.3.2.2. Nous présentons dans ce qui suit ses applications en anthropométrie afin de classer une population donnée en sous-groupes homogènes ayant des caractéristiques similaires.

Robinet et Carrier [34] décrivent la procédure utilisée pour la classification des mesures corporelles à partir d'un échantillon statistique de 11 500 personnes (hommes, femmes et enfants) de 5 à 79 ans issue de l'IFTH. Une méthode hiérarchique ascendante de classification non supervisée a été utilisée pour identifier sept formes corporelles féminines et six formes masculines. Pour les sept formes féminines, ils ont cherché à trouver des similitudes par rapport aux neuf formes de corps identifiées par l'étude américaine menée par [TC]² au début des années 2000, en utilisant les mesures moyennes ainsi qu'une analyse visuelle des images du scanner corporel comme illustré en **Figure 7**. Par la suite, ils ont effectué une comparaison entre la population américaine et française par rapport à l'appartenance aux classes des différentes formes représentatives des deux populations. Le calcul du pourcentage d'appartenance montre que les modèles de représentation morphologique ne peuvent pas être transférés d'un pays à l'autre. De ce fait, les enseignes de vêtements doivent s'adapter à une population cible.



Figure 7. Sept formes du corps représentatives des femmes [34]

Viktor et *al.* [35], ont appliqué l'un des algorithmes de partitionnement le plus répandu k -moyennes (K -means en anglais) pour classer les scans corporels 3D des hommes obtenus à partir de l'enquête CAESAR. Le nombre de classes k a été fixé dès le début à cinq représentant les tailles universelles des vêtements: S, M, L, XL, XXL. Ils ont représenté chaque groupe par leur centroïde. Pour Jeyasingh et Appavoo [36], les données de 620 hommes ont été classées en six catégories codées : XS, S, M, L, XL et XXL. Zakaria [37] a également fait recours à la même méthode de partitionnement pour segmenter l'échantillon de la population en trois groupes: S, M et L. Le fait de devoir choisir a priori le paramètre k contient une part de subjectivité. Une façon de surmonter cet inconvénient consiste à utiliser des critères de validation du meilleur nombre de classe comme l'indice de Davies-Bouldin (DB), l'indice de Calinski ou encore l'indice de silhouette [38].

Ces indices sont généralement fondés sur l'évaluation de la dispersion intra-classes et inter-classes. Une autre façon est d'appliquer une classification non-supervisée en deux étapes « TwoStep Cluster ». Cette technique a l'avantage de détecter automatiquement le nombre de classes optimal pour les données d'apprentissage [39].

Shahrabi et *al.*[40], ont effectué une analyse en deux étapes en combinant une méthode hiérarchique et une méthode non-hiérarchique (**Figure 8**).

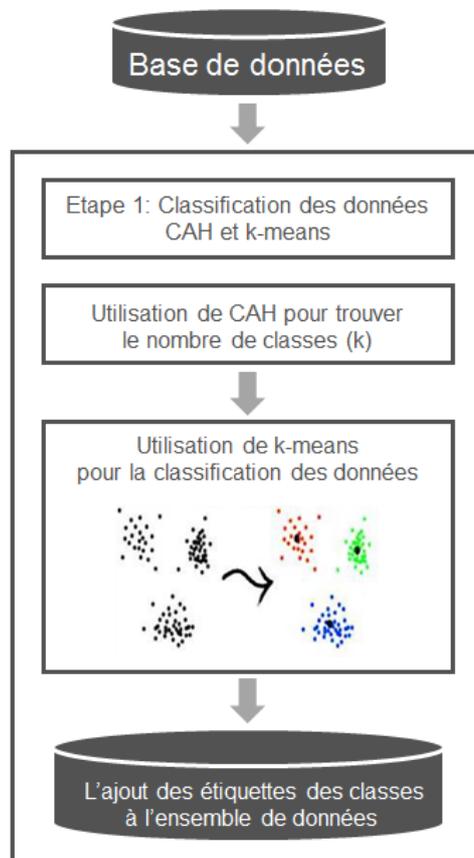


Figure 8. Classification en deux étapes [40]

La méthode de classification ascendante hiérarchique (CAH) a été d'abord appliquée pour identifier le nombre de classes. Par la suite, ils ont opté pour la méthode non-hiérarchique, K-means, afin de détecter les cinq classes représentant les différentes formes de corps chez 600 hommes. Le même principe est utilisé par Hsu et *al.* [30] à partir de 16 variables anthropométriques de bas de corps de 956 femmes. Les résultats de cette étude ont permis de déterminer aussi cinq morphologies féminines. Chung et *al.*[41], ont aussi appliqué la même classification en deux étapes sur les élèves Taïwanais. Trois morphologies ont été identifiées dénommées petite, moyenne et grande.

De nombreuses méthodes ont ainsi été développées pour analyser la morphologie humaine ouvrant des voies à la mise en place de système de tailles de vêtements reflétant au mieux la réalité du corps de la population étudiée. Il s'avère qu'un ajustement plus précis des vêtements et de meilleures expériences clients se concrétiseraient par une hausse de revenus pour l'enseigne de prêt-à-porter.

1.3 Système des tailles de vêtements

Un système de tailles peut être défini comme un tableau de chiffres qui représente la valeur des mesures clés utilisées pour classer les différentes morphologies dans la population cible en sous-groupes homogènes [42]. Les personnes d'un même sous-groupe sont censées avoir les mêmes caractéristiques de forme corporelle et partager la même taille de vêtement.

L'objectif principal d'un système de tailles est de fournir le meilleur ajustement pour la plupart des clients avec un nombre optimal de tailles qui profite à la fois aux consommateurs et aux fabricants. Un trop grand nombre de tailles à choisir entraînera de la confusion dans la recherche de la bonne taille et sera économiquement pénalisant. A l'inverse, trop peu de tailles entraîneront de l'insatisfaction en raison de la perte de l'ajustement du vêtement. Fralix [43] affirme que les systèmes de tailles devraient permettre d'obtenir un processus de fabrication rentable et abordable et devraient également pouvoir satisfaire la majorité de la population pour laquelle ils sont conçus. Par conséquent, nous allons cerner les limites des systèmes de tailles actuels et présenter les différentes normes et méthodes qui ont été proposées afin de pallier à ces limites.

1.3.1 Système de tailles français

En France, le tour de poitrine et le tour de taille sont utilisés pour la désignation en taille commerciale chez la femme[44]. Il est important de prendre la mesure au plus près du corps.

Elle est calculée comme suit et le résultat est arrondi à l'entier pair le plus proche.

$$taillecommerciale = \frac{\frac{1}{2}tourdetaille + \frac{1}{2}tourdepoitrine}{2}$$

Cependant, trouver la bonne taille de vêtement est souvent difficile. Les tailles varient aussi selon la marque commerciale, voire même au sein d'une même marque selon les catégories de produits. Chaque marque dispose de son propre guide de tailles réalisé en fonction de sa stratégie marketing et de sa cible commerciale. Une même taille peut varier de quelques centimètres entre deux marques. Cela provoque que le client peut se perdre dans sa recherche. Cette variation de tailles est justifiée par le Vanity Size (ou la « taille flatteuse » qui consiste à réduire les tailles sur les étiquettes afin de flatter le client [45], [46].

Les marques font croire à la cliente qu'elle rentre par exemple dans un 38 sauf qu'il s'agit d'un 40. Il n'existe aucune législation sur l'étiquetage des tailles. Un autre problème est que ces marques sont trop peu nombreuses à tenir compte des différences morphologiques dans une

population. Les tailles les plus convoitées ne sont pas toujours disponibles et les gammes de tailles sont parfois très limitées. Les tailles commerciales féminines chez bon nombre de marques s'arrêtent au 44, voire au 42 et ne voient pas la nécessité d'aller au-delà.

ClicknDress[47] a réalisé en juin 2016 une étude auprès de plus de 55.000 femmes françaises âgées de 17 à 65 ans. L'étude montre qu'environ 40% des femmes font une taille 44 et plus contre 6% des françaises font une taille 36 (**Figure 9**).

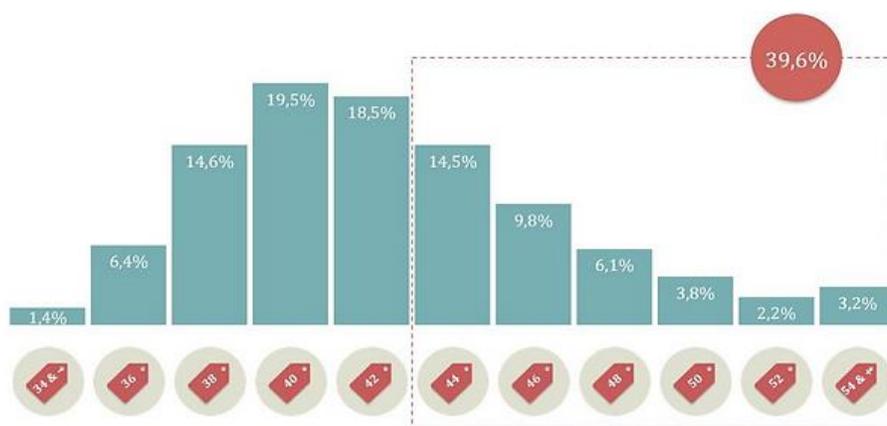


Figure 9. Répartition des femmes françaises selon leur taille de vêtements[47]

Sur près de 8.000 références produits étudiées dans l'étude, il apparaît que les femmes faisant une taille 36 ont 3 fois plus de choix que celles faisant une taille 44 alors qu'elles sont 2,3 fois moins nombreuses. Une initiative par les marques de mode est de développer leur propre ligne «grande tailles» en élargissant leur gamme de tailles de vêtements.

Les marques peuvent aussi jouer sur les tailles pour exprimer un style spécifique. American Vintage par exemple propose un style un peu ample. Donc, les vêtements sont coupés pour être portés très large. À l'inverse, The Kooples qui cherche un look très ajusté par rapport au corps, il faudra prendre la taille au-dessus pour être à l'aise.

Néanmoins, reste un vrai décalage entre la réalité des morphologies et la proposition des marques de prêt à porter.

1.3.2 Normes de tailles des vêtements

Dans une enquête anthropométrique, chaque mesure du corps est identifiée par un code, un nom, et une description précisant la nature de la mesure. Il est implicitement supposé que les mesures de diverses enquêtes sont différentes lorsque leurs noms sont distincts, mais ce n'est

pas toujours le cas. A titre d'exemple, certaines enquêtes utilisent les termes tour de dos pour dessous poitrine, tour de hanche pour tour de bassin, stature pour taille. Cela peut indéniablement être source de confusion. Pour éviter une telle confusion inutile, les normes sont utilisées comme références.

Il existe alors des tentatives de normalisation, au niveau international avec les normes ISO 7250, ISO 8559 et au niveau européen avec la norme EN 13402. Ces normes sont présentées dans le **Tableau 3**.

Ces normes fournissent un panel de recommandations sur les mesures corporelles élaborées par consensus, aux fabricants de textile pour pouvoir ensuite établir un système des tailles de vêtements. Mais il n'existe aucune réglementation relative à la concordance des tailles. Généralement, le barème de mensuration du corps comporte des mesures fondées uniquement sur une valeur de stature. Par conséquent, différents barèmes doivent être déterminés pour différentes statures. De plus, la norme NF EN 13402 sur laquelle se sont basées de nombreuses enseignes de prêt-à-porter date de 1970. Or au fil du temps, les personnes grandissent et les morphologies évoluent. Face à cette évolution des morphologies, il est utile de changer les barèmes de tailles et de s'adapter à la population étudiée plutôt qu'à des moyennes désuètes.

Normes	Descriptions
<p>ISO 7250-1:2017[48] Définitions des mesures de base du corps humain pour la conception technologique -- Partie 1: Définitions des mesures du corps et repères</p> <p>ISO 7250-2:2010[49] Définitions des mesures de base du corps humain pour la conception technologique -- Partie 2: Résumés statistiques des mesurages du corps de populations nationales</p> <p>ISO 7250-3:2015 [50] Définitions des mesures de base du corps humain pour la conception technologique -- Partie 3: Gammes de conception régionales et mondiales pour utilisation dans les normes de produits</p>	<p>Cette norme en trois parties fournit une description de mesures anthropométriques qui peuvent être utilisées pour la création de bases de données anthropométriques ainsi que des résumés statistiques des mesurages du corps.</p>
<p>ISO 8559-1:2017[51] Désignation des tailles de vêtements -- Partie 1: Définitions anthropométriques pour la mesure du corps</p> <p>ISO 8559-2:2017[52] Désignation des tailles de vêtements -- Partie 2: Indicateurs des dimensions principales et secondaires</p> <p>ISO 8559-3:2018[53] Désignation des tailles des vêtements -- Partie 3: Méthodologie de création de barèmes de mensuration du corps et des intervalles</p>	<p>Cette norme définit un repère et des indications de mesure du corps humain. Elle peut être utile pour déterminer les mesures clés par exemple la mesure primaire des vestes pour femmes est le tour de poitrine, et les mesures secondaires sont la taille et le tour de hanches. Elle définit aussi les intervalles de mesures qui doivent être utilisés pour chaque mensuration.</p>
<p>NF EN 13402-1 Juin 2001[54] Désignations des tailles de vêtements -- Partie 1 : termes, définitions et procédés de mesurage du corps</p> <p>NF EN 13402-2 Février 2003[55] Désignation des tailles de vêtements -- Partie 2 : mesures primaires et secondaires</p> <p>NF EN 13402-3 Septembre 2017[56] Désignation des tailles de vêtements -- Partie 3 : mesures et intervalles</p>	<p>Cette norme donne la définition des différentes mesures corporelles à prendre en considération pour désigner la taille des vêtements et prescrit ensuite les mesures primaires et secondaires pour des types spécifiés de vêtements. Elle préconise à la fin l'utilisation d'un pictogramme sur lequel figurent les mensurations principales du corps humain pour standardiser les étiquettes de tailles.</p>

Tableau 3. Normes de tailles des vêtements.

1.3.3 Construction de système de tailles

L'expérience de l'essayage en cabine nous rappelle que les tailles fixées aujourd'hui par les enseignes de prêt-à-porter sont bien souvent en conflit avec la diversité des corps. Bon nombre de vêtements sont mal mesurés et plusieurs clients éprouvent des inconforts directement liés aux décisions des concepteurs en matière de tailles. De plus, à l'heure de l'essor du e-commerce, les clients ont tendance à se faire rembourser leurs produits avec le motif « mauvaise taille » qui se traduit par une perte directe pour la marque. A cet égard, la reconnaissance de la morphologie du corps humain sert de base à la construction d'un système de tailles plus performant. Pour Gupta deux éléments déterminent la précision et la performance de ce système, c'est à dire: les données anthropométriques utilisées et la méthode choisie pour l'analyse de ces données[28].

Selon McCulloch et *al.*[57] , un système de tailles performant et rentable doit répondre aux critères suivants :

1. Le guide de tailles doit être conforme aux mesures corporelles pour assurer le meilleur ajustement.
2. La couverture de la population cible devrait être au maximum.
3. La gamme des tailles proposées devrait être au minimum.

La conception théorique d'un système de taille a fait l'objet de nombreux travaux exposés dans la littérature[28], illustré en **Figure 10**.

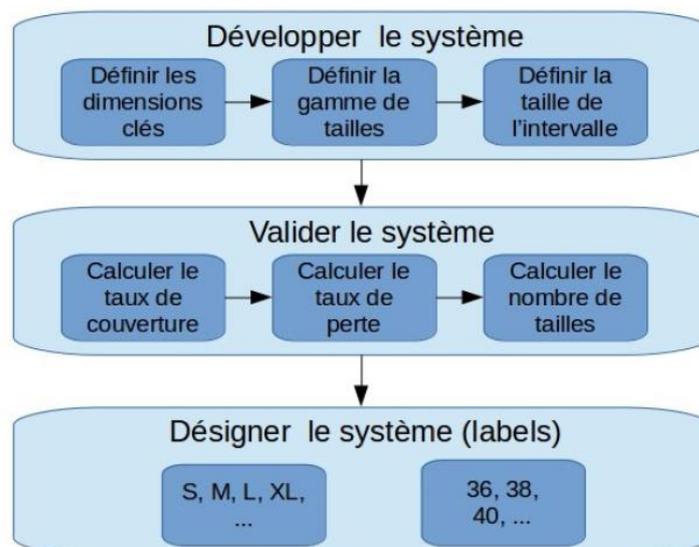


Figure 10. Méthode théorique de définition d'un système de tailles

Le principe général est composé de trois étapes :

1. *Développer le système*: cette étape consiste à définir les dimensions clés qui doivent être mesurées sur la personne. Après, il faut décider de la gamme de taille et fixer la taille de l'intervalle pour chaque dimension. Par exemple, si la stature d'un groupe d'échantillons est de 145-185 cm, cela fait une gamme de 40 cm. Si l'intervalle de taille est fixé à 8 cm, le nombre de tailles est donc estimé à cinq tailles. Par conséquent, le développement d'un bon système de calibrage dépend fortement du choix des mesures clés, de la gamme de tailles et de l'intervalle choisi.
2. *Valider le système*: cette étape se base sur le calcul du taux de couverture, la perte globale (quantifiant le bien-être et l'ajustement des vêtements), et le nombre de tailles. Un système de taille dont le taux de couverture se situe entre 65 % et 80 % est considéré comme satisfaisant selon Petrova [25]. La perte globale est introduite par McCulloch et al. [57] comme la distance entre la mesure réelle du corps humain et les tailles assignées. La distance euclidienne est l'un des moyens fondamentaux et le plus populaires utilisés pour déterminer cette distance. La perte globale devrait être aussi faible que possible pour un meilleur ajustement.
3. *Désigner le système*: cette étape porte sur la désignation de la taille faisant référence à la façon dont chaque taille est identifiée sur le vêtement ou sur l'étiquette. L'étiquetage de taille peut être numérique, alphabétique ou graphique. L'une des méthodes est l'utilisation de pictogramme introduit dans la norme NF EN 13402-3 avec les mesures clés en cm (**Figure 11**).

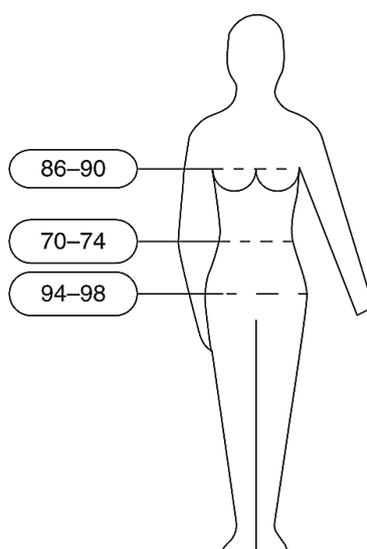


Figure 11. Exemple d'un pictogramme femme[56]

Salusso et *al.*[58] ont souligné qu'un système de tailles bien conçu doit être basé sur une classification des données couvrant la plus grande partie de la population avec un nombre minimum de tailles. Dans cette perspective, certains chercheurs commencent par une segmentation de la population en groupes. Après, ils établissent un système de tailles[17], [18], [41].

Hrz̃enjak et *al.*[59] ont développé un système de tailles pour le bas du corps pour les filles croates âgées de 13 à 20 ans. Un échantillon de 4071 filles a été segmenté en trois groupes homogènes: Type 1, Type 2 et Type 3 au moyen d'une classification en deux étapes. Ils ont appliqué une ACP pour extraire les dimensions clés : stature et tour de taille. La taille de l'intervalle est fixée respectivement à 6 cm et 2 cm. Le système de taille a permis d'établir 56 catégories de tailles pour les trois groupes avec un taux couverture de 77,7 %. Un exemple de désignation de taille de vêtement est T3/166/76 qui représente le type 3, 166 pour la stature et 76 pour le tour de taille.

Cependant, les systèmes obtenus par ces méthodes génèrent généralement un trop grand nombre de tailles. Conformément à la pratique adoptée par Petrova[25] et Zakaria[17], les tailles qui couvrent moins de 2 % de la population ont été exclues du système de tailles afin de réduire le nombre de tailles. La lacune de cette méthode réside dans le fait de supprimer de nombreuses tailles avec un taux de couverture inférieur à 2%, ce qui engendre la diminution du pourcentage total de la population couverte. De plus, elle conduira à un système non linéaire (trous dans le système) suite à la suppression aléatoire de tailles comme le montre le **Tableau 4** (les tailles 5, 8, 13, 14, 19 et 20 sont supprimées).

Size roll	Body types	Key dimensions (cm)		Accommodation rate		Aggregate loss (cm)
		Height	Bust girth	n	%	
1	Small	114–121.9	54–59.9	54	10.3	2.6
2			60–65.9	14	2.7	2.5
3	Medium	122–129.9	54–59.9	60	11.4	2.6
4			60–65.9	54	10.3	2.5
5			114–121.9	66–71.9	8	1.5*
6		122–129.9	66–71.9	29	5.5	2.5
7		72–77.9	11	2.1	2.7	
8		78–83.9	4	0.8*	2.5	
9	Large	130–137.9	54–59.9	13	2.5	2.7
10			60–65.9	46	8.8	2.7
11		66–71.9	31	5.9	2.6	
12		72–77.9	23	4.4	2.8	
13		78–83.9	7	1.3*	2.6	
14		84–89.9	3	0.6*	3.0	
15		138–145.9	60–65.9	15	2.9	2.5
16		66–71.9	36	6.9	2.5	
17		72–77.9	16	3.0	2.6	
18		78–83.9	17	3.2	2.5	
19	146–153.9	84–89.9	6	1.1*	2.7	
20		60–65.9	3	0.6*	2.6	
21		66–71.9	17	3.2	2.6	
22		72–77.9	18	3.4	2.7	
23		78–83.9	22	4.2	2.5	
24		84–89.9	14	2.7	2.8	
TOTAL				521	99.3	2.6

Tableau 4. Exemple de suppression des tailles de vêtements [17]

La plupart des systèmes de tailles linéaires sont basés sur un incrément constant à travers toutes les mesures. Cela présente l'avantage de faciliter le classement, l'étiquetage et la gradation de la taille. Mais, il présente l'inconvénient que les restrictions de construction du système peuvent conduire à un mauvais ajustement, comme le fait de se limiter à un nombre réduit de tailles.

Une alternative est les systèmes de tailles non linéaires qui peuvent être établis à l'aide d'une technique d'apprentissage non supervisé basée sur le réseau de neurones artificiels: la « carte auto-organisatrice » ou « carte de Kohonen », souvent désignée par le terme anglais « Self Organizing Maps (SOM) » permettant le regroupement des données en classes distinctes.

Doustaneh et al.[60] utilisent cette technique en trois niveaux pour créer un système de tailles pour 670 hommes Iraniens (**Figure 12**).

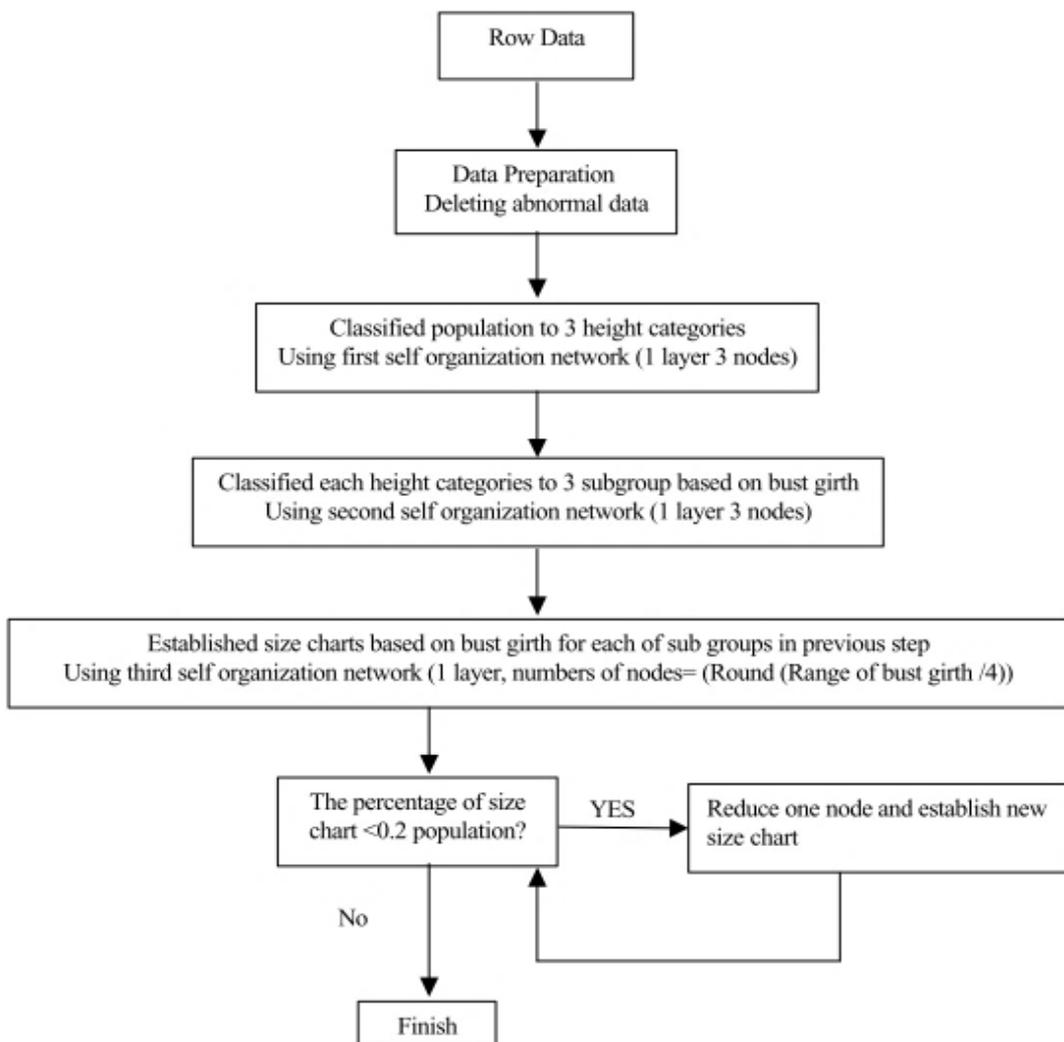


Figure 12. Principe de construction d'un système de tailles à partir d'un SOM proposé par Doustaneh et *al.*[60]

Dans un premier temps, la population a été divisée en trois catégories de statures en utilisant une seule couche de 3 neurones. Dans un deuxième temps, chaque catégorie de stature a été subdivisée en trois groupes en fonction de tour de poitrine. Dans un troisième et dernier temps, ils ont choisi un intervalle de 4 cm pour diviser la gamme de tailles de chaque sous-groupe de poitrines. Cette division rend le système linéaire. Si le taux de couverture était inférieur à 2%, le nombre de neurones dans le réseau était réduit par un. Au total, 26 tailles ont été créées par cette méthode. Cette étude suggère que l'utilisation de SOM permet d'améliorer le taux de couverture et le taux de perte par comparaison avec la méthode statistique.

Vadood et *al.*[61] présentent un guide de tailles pour les costumes des hommes Iraniens. Dix mesures de corps clés sont choisies pour la création de costumes. Puis, la population est

divisée en 5 et 7 classes en utilisant un SOM. Le taux de perte a été calculé en fonction de trois mesures principales : tour de poitrine, tour de taille et tour de hanches. L'utilisation de 5 classes ayant une répartition équitable du nombre de leurs membres et un taux de perte acceptable apparaît comme le choix le plus adaptée à la population étudiée. Les différentes méthodes décrites ci-dessus présentent l'avantage de classer des populations cibles en différentes morphologies à partir des mesures anthropométriques et de définir un système de tailles. Des méthodes plus récentes basées sur des SOM permettent de réduire considérablement le nombre de tailles et de résoudre en grande partie les problèmes d'ajustement des vêtements. En revanche, choisir le bon nombre de couches et le bon nombre de neurones par couche n'est pas évident et demande une certaine expérience.

C'est à partir des morphologies extraites et le tableau de tailles que des modèles de corps 3D adaptés aux différentes morphologies de la population seront paramétrés et intégrés dans des outils de Conception Assistée par Ordinateur (CAO) pour l'essayage virtuel ou le prototypage rapide de vêtements. La modélisation 3D du corps humain représente un pont important entre la numérisation 3D et la conception assistée par ordinateur.

1.4 Modélisation du corps humain

Actuellement, les modèles 3D du corps humain présentent un grand intérêt dans une grande variété de domaines d'application tels que : l'industrie automobile, l'architecture, le film d'animation, et le textile. La construction du modèle du corps humain est l'étape fondamentale du design de mode interactif. De nos jours, les outils disponibles permettent la conception complète du modèle ou la modification d'un modèle existant. Cette conception nécessite des données 3D précises sur le corps humain. Ces données peuvent être collectées à partir du scanner 3D. Il existe deux principales méthodes pour reconstruire un modèle 3D [62][63]. La première permet de reconstruire le corps humain en faisant correspondre le modèle provenant du scanner 3D à un modèle référence(préexistant)[64], [65]. La deuxième méthode permet d'établir le modèle du corps humain à partir de courbes caractéristiques « Feature Curves».

La **Figure 13** illustre un exemple de déformation de modèle référence de corps humain. L'alignement est réalisé à travers le calcul de l'Iterative Closest Point (ICP). C'est la distance moyenne minimale entre le modèle référence et le nuage des points du scan 3D. Cette méthode présente l'avantage de ne pas s'appliquer à des points références spécifiques mais

sur l'ensemble des points. L'inconvénient est qu'elle nécessite un temps de traitement important et est dépendante de la qualité de maillage de modèle exemple.

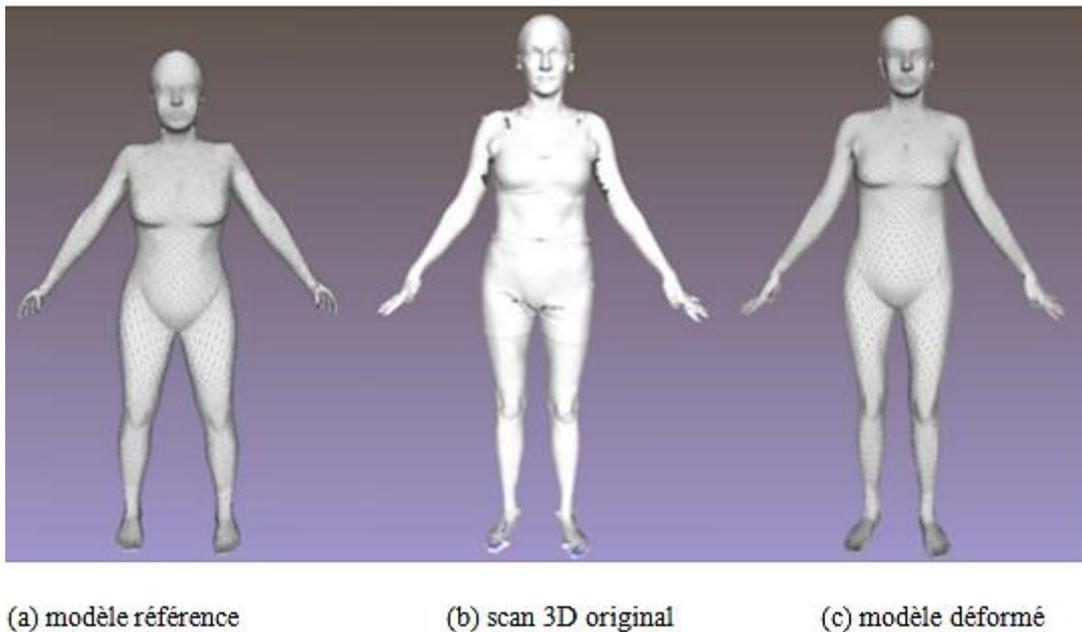


Figure 13. Exemple d'alignement d'un modèle avec un scan 3D par Hirshberg et *al.* [64]. Seo et *al.*[66] utilisent la même technique de reconstruction basée sur modèle référence. L'algorithme comporte deux phases principales (**Figure 14**): l'ajustement du squelette et le raffinement. La phase d'ajustement cherche à trouver l'approximation linéaire du modèle scanné au modèle de référence par déformation du squelette. La phase de raffinement consiste à améliorer la précision de l'ajustement en minimisant la différence de forme entre le modèle de référence et le modèle scanné.

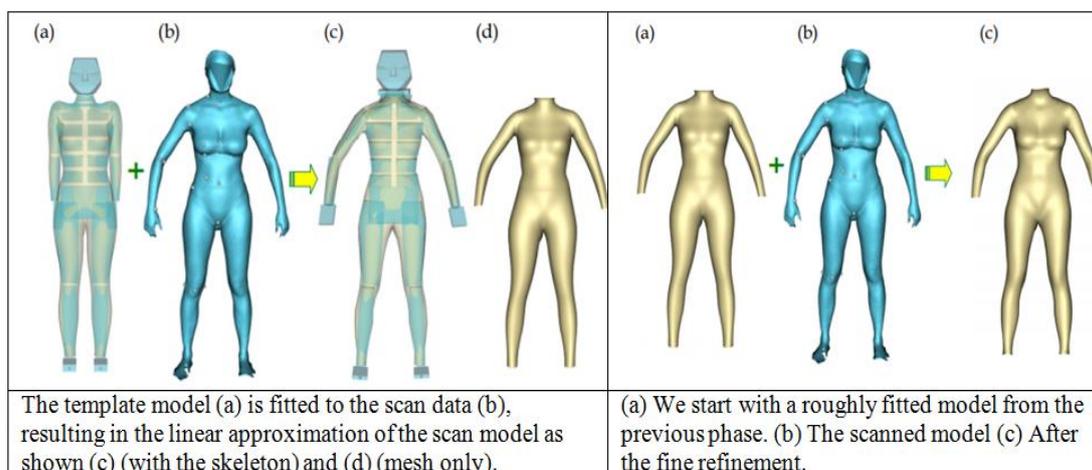


Figure 14. Deux phases pour la modélisation du corps par Seo et *al.* [66]

Cette méthode de correspondance entre les deux modèles donne un rendu visuel proche. Mais, elle ne reproduit pas fidèlement le modèle de référence d'un point de vue dimensionnel [67].

Wang[68] a adopté la deuxième méthode (**Figure 15**). Il repère les points anatomiques sur le nuage de points du corps humain scanné. En effet, il est possible de repérer au préalable les positions de points anatomiques à l'aide de marqueurs et de les localiser à la surface du corps au moment du balayage corporel. Les courbes caractéristiques ont été déterminées à partir de ces points représentant les circonférences du corps. Sur ces courbes, différentes facettes (patches) sont superposées. Ces facettes sont déformées par un lissage de la surface en utilisant les correctifs de l'interpolation de Gregory afin de générer des surfaces continues maillées. C'est la méthode de modélisation qui donne le plus de précision vue la génération minimale des erreurs d'approximation[69]. C'est la raison pour laquelle elle est massivement utilisée dans la plupart de logiciels CAO.

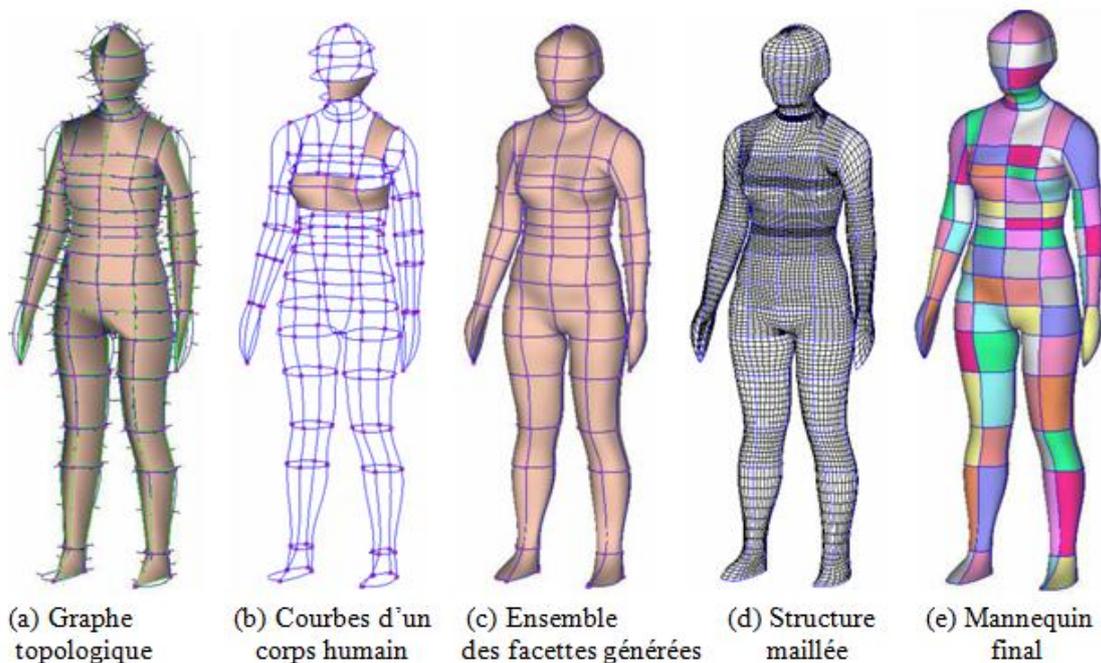


Figure 15. Fonction du Correctif de Gregory [69]

Hsiao[70] a utilisé aussi la modélisation basée sur les courbes caractéristiques pour la modélisation de mannequin (**Figure 16**). L'ingénierie inverse permet de produire un modèle géométrique du mannequin. Les courbes ont été extraites des points de références de la surface de l'objet. Ensuite, le modèle de mannequin 3D a été construit en connectant les surfaces B-spline.

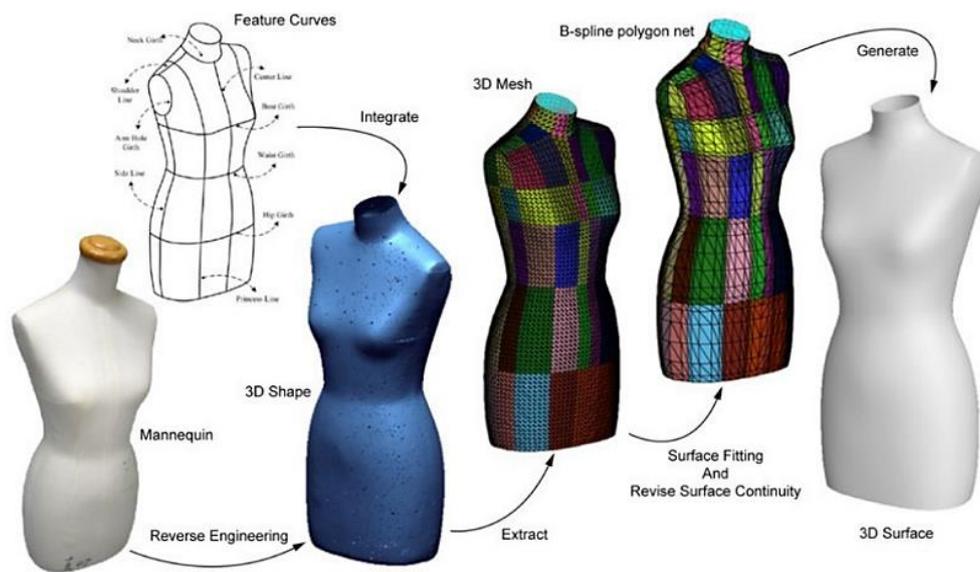


Figure 16. Modélisation de mannequin 3D par Hsiao[70]

Zeng et al.[71] ont conçu un système interactif pour la modélisation du corps humain 3D en déformant automatiquement les paramètres anthropométriques pour chaque facette du maillage 3D du corps (**Figure 17**). Cette déformation est acquise par une régression linéaire à partir de paramètres sélectionnés.

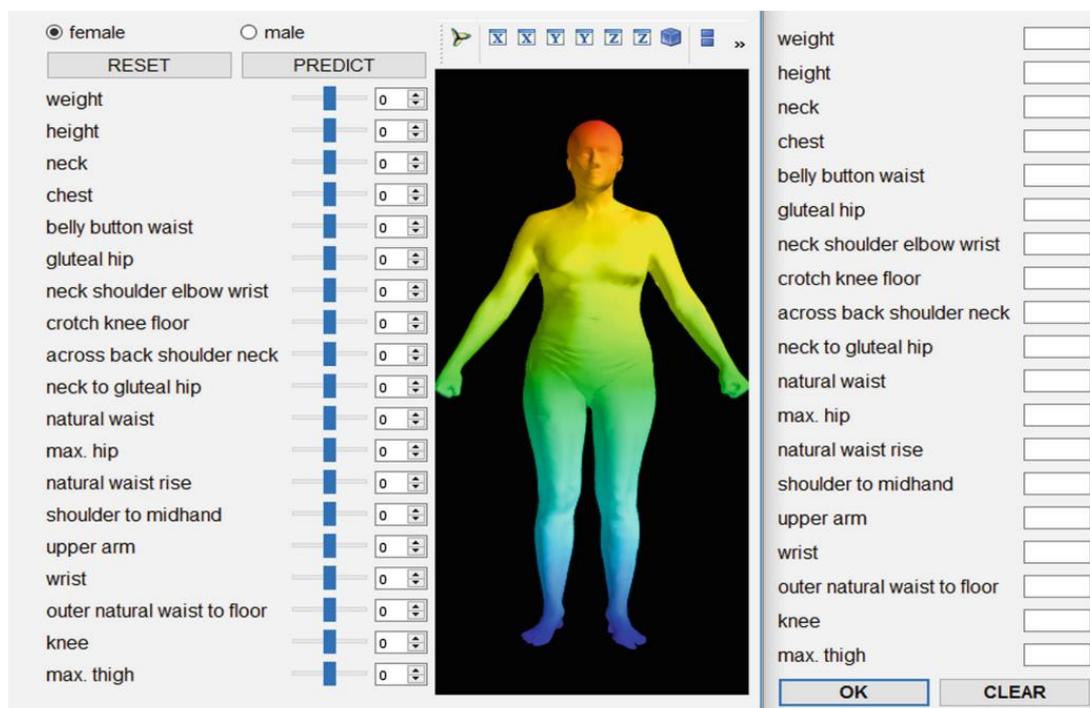


Figure 17. Système de modélisation interactif de Zeng et al.[71]

L'apport de ce système est dans la facilité de contrôler le mannequin avec certaines mensurations. Le problème réside au choix du mannequin initial qui représente une forme idéale du corps humain et à l'évolution de certaines mensurations qui peut conduire à des aberrations morphologiques.

1.5 Conclusion

Dans ce chapitre, nous avons passé en revue l'analyse des données anthropométriques pour mieux caractériser la morphologie humaine d'une population donnée et les différentes méthodes de construction des systèmes de tailles de vêtements. Ces travaux existants reposent généralement sur une gradation linéaire des tailles standardisées avec des paramètres fixes facilitant la mise en œuvre industrielle. Ils cherchent aussi à améliorer le bien-être du vêtement, sans tenir compte de nombre de tailles proposé ou de la diversité des morphologies cibles. Cette uniformité ne convient pas à tous les clients ce qui diminue le pourcentage de couverture. De ce fait, le système doit prendre en compte les différentes morphologies des clients et avoir un compromis entre trois critères (bien-être, nombre de tailles et couverture de la population) en faisant recours à l'optimisation. Une analyse morphologique de la population, l'optimisation des systèmes de tailles de vêtements en symbiose avec cette précédente analyse constituent les deux axes de recherche de cette thèse. Le premier axe concerne la définition de classes morphologiques caractérisant des formes humaines représentatives d'une population donnée afin d'en extraire les morphologies les plus significatives par une classification. Le deuxième axe est la mise en place des systèmes de tailles optimaux. Finalement, la création de mannequins 3D de corps humain représentant les morphotypes de chaque classe, paramétrés par les systèmes des tailles est envisagée.

La **Figure 18** illustre les différentes étapes de la méthodologie qui sera mise en place pour répondre à ces objectifs et sera détaillée dans le chapitre 3. Le chapitre 2 propose un état de l'art sur les méthodes et outils mathématiques permettant de satisfaire ces objectifs.

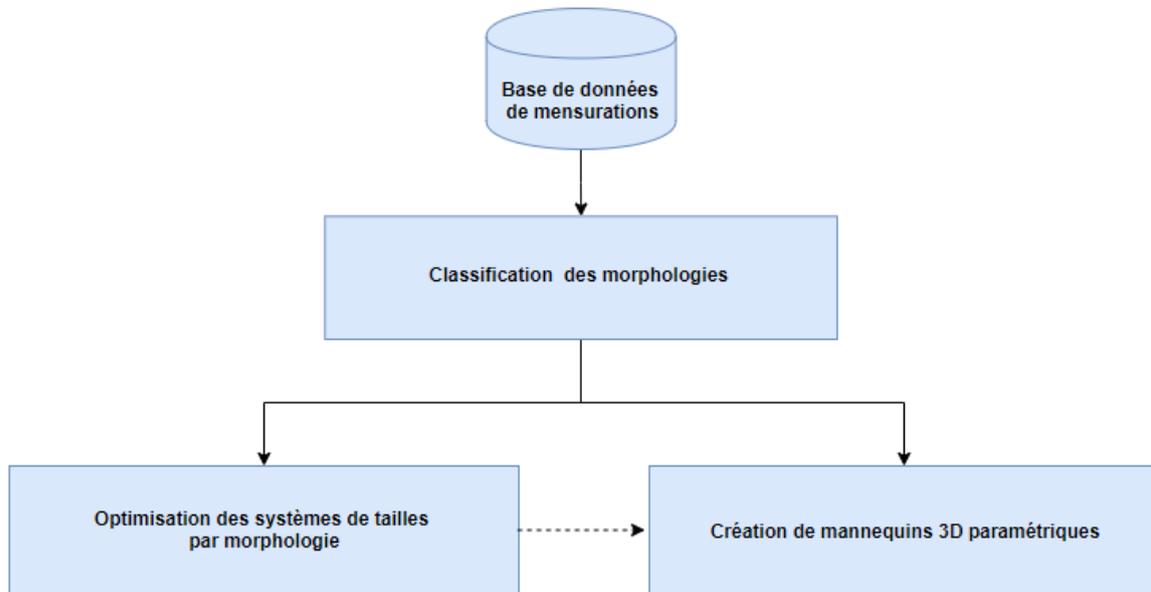


Figure 18. Méthodologie proposée

Chapitre 2 Outils et méthodes utilisés

2.1 Introduction

Dans ce chapitre, nous présenterons les fondements théoriques des outils et méthodes qui sont concernés par notre méthodologie : la classification des morphologies et l'optimisation des systèmes de tailles par classe morphologique. La classification consiste à regrouper les individus ayant des morphologies similaires. Elle comporte une phase de sélection pour choisir les mesures clés à partir d'une base de données de mensurations, une phase de normalisation pour rendre comparables ces mesures et le choix d'un algorithme de classification pour regrouper les mesures similaires. L'optimisation est un outil d'aide à la décision qui permet de trouver une solution optimale au problème à résoudre. Nous nous focaliserons sur l'optimisation multi-objectif car plusieurs critères sont à considérer simultanément lors de création d'un système de tailles tels que le nombre de taille, la pourcentage de couverture et le bien aller de vêtements. La principale différence avec l'optimisation mono-objectif vient alors de la notion de dominance de Pareto qui sera décrite avec un panorama des méthodes de résolution. Le but de ce chapitre est de détailler le principe général de chaque concept, de présenter les diverses méthodes existantes que nous pourrions utiliser afin d'évaluer leurs apports et leurs limites. Grâce à cette évaluation, nous allons pouvoir choisir les algorithmes qui nous semblent les plus appropriés à la mise en œuvre de notre méthodologie développée dans le chapitre 3.

2.2 Classification des morphologies

Pour la mise en œuvre de cette classification, nous avons adapté à notre problématique un processus ECD (Extraction des Connaissances à partir de bases de Données) [72] divisé en trois étapes comme l'illustre la **Figure 19** :

- 1) Sélection de variables: c'est une étape qui permet l'identification des variables pertinentes à considérer dans la classification.
- 2) Normalisation: c'est un processus de transformation des variables choisies afin de les rendre comparables entre-elles.

- 3) Classification: c'est le choix de l'algorithme de classification pour construire des classes homogènes d'individus.

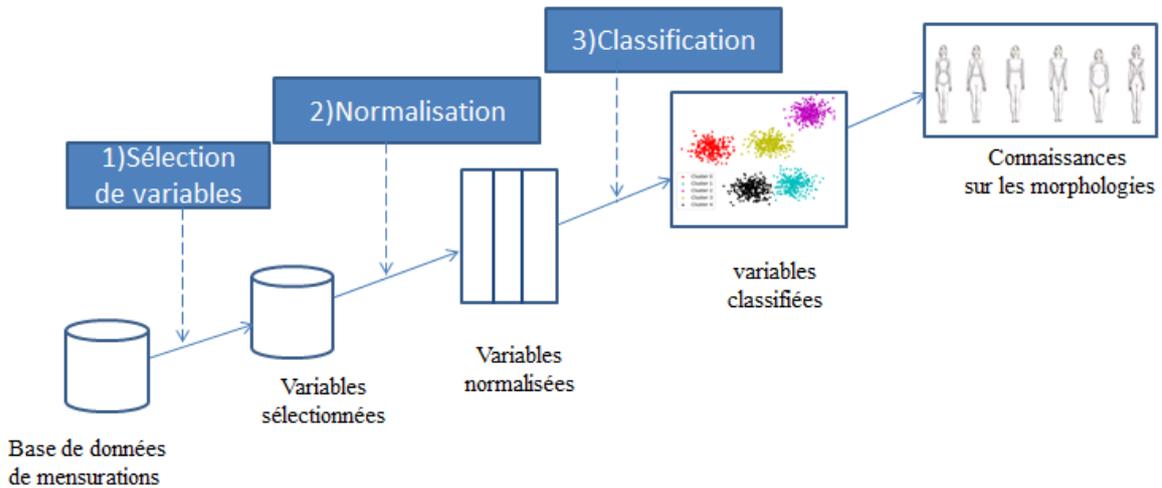


Figure 19. Processus ECD adapté à la classification des morphologies

Pour chacune des trois étapes, plusieurs méthodes sont envisageables. Les sections suivantes présentent un aperçu des techniques utilisées pour ces tâches.

2.2.1 Sélection de variables

L'étape de sélection de variables est particulièrement importante dans le processus de la classification. Elle consiste à choisir parmi un ensemble de variables, un sous-ensemble de variables pertinentes (en anglais *feature subset selection*) permettant de réduire la dimension du problème étudié sans nuire aux performances de la classification. Certaines variables peuvent être inutiles, comme par exemple une variable répartie uniformément, elle ne met en évidence aucune structure. Cette sélection est considérée comme un problème difficile à résoudre dans un contexte non-supervisé en raison de l'absence des labels des classes qui guideraient leur choix [73].

2.2.1.1 Méthodes de sélection de variables

Deux grandes familles de méthodes de sélection existent [74, 75] :

- *Enveloppes (Wrappers)*: Ces méthodes recourent à l'algorithme de classification par enveloppe pour évaluer le degré d'utilité des variables dans le processus de la classification. Le principe de base est de générer des sous-ensembles de variables et de les évaluer. Par la suite, l'algorithme sélectionne le meilleur sous-ensemble de

variables générées. Elles sont essentiellement basées sur les machines à vecteurs de supports (en anglais support vector machine, SVM). Les méthodes enveloppes sont très sensibles, voire susceptibles au phénomène de sur-apprentissage (en anglais « overfitting ») illustré dans la **Figure 20**. Ce phénomène se caractérise par une erreur d'apprentissage très faible alors que l'erreur sur les données de test est relativement élevée. De plus, elles souffrent d'une grande complexité de calcul.

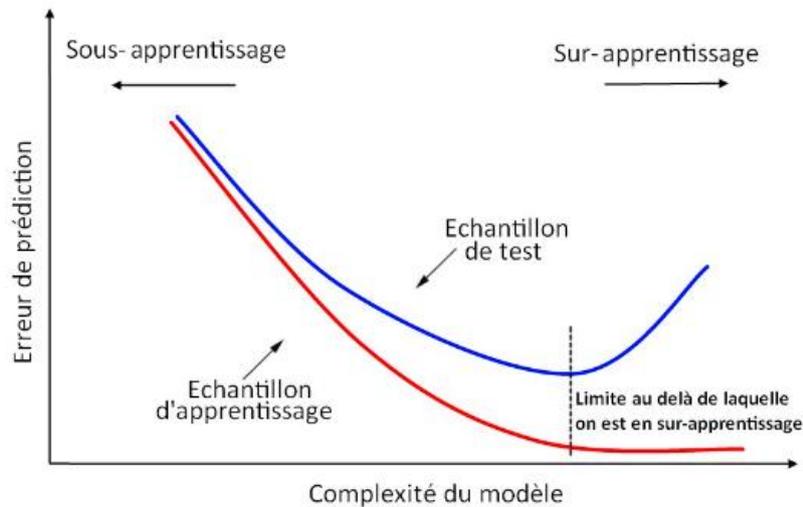


Figure 20. Phénomène de sur-apprentissage[76]

- *Filtres*: Contrairement aux méthodes précédentes, ces méthodes sélectionnent les variables indépendamment de l'algorithme de classification. Elles présélectionnent un sous-ensemble de variables, puis elles appliquent ces variables dans la phase de classification. Les filtres se basent seulement sur les caractéristiques des données et reposent sur le calcul d'un score qui permet de calculer la pertinence de chaque variable. Cette indépendance rend les méthodes filtres plus robustes face au problème de sur-apprentissage avec une complexité et un coût de calcul plus faible. Par contre, on leur reproche souvent l'ignorance de leur impact sur les performances de l'algorithme d'apprentissage suite à ce choix des sous-ensembles.

2.2.1.2 Méthode d'extraction de variables

L'Analyse en Composantes Principales (ACP) est une méthode d'analyse de données, couramment utilisée dans la littérature (**section 1.2.3.3**) comme une méthode de sélection pour réduire la dimension des données en quelques variables contenant les informations les plus pertinentes dans les données.

L'ACP proposée par Karl Pearson dès 1901 est l'une des méthodes descriptives multidimensionnelles les plus utilisées. L'objectif de l'ACP est de réduire la dimension de l'espace des données à un nombre inférieur de variables appelées « composantes principales » ou encore « axes factoriels » [77].

C'est une technique basée sur la projection orthogonale. Elle consiste à projeter les observations multidimensionnelles dans un sous-espace de dimension inférieure en maximisant la variance des projections [78]. En pratique, cette réduction repose sur le calcul de la matrice des variances-covariances. Ainsi, la première composante (PC1) correspond à celle dont la projection orthogonale des observations présente la dispersion (ou la variance) la plus grande. La deuxième composante (PC2) est calculée sur le résidu non expliqué par la première tout en étant orthogonal (**Figure 21**). Cette orthogonalité entre les deux composantes traduit l'absence de corrélation linéaire. Le calcul des autres composantes se fait de la même manière. Les premiers axes concentrent généralement l'essentiel de l'information et fournissent la majeure partie du pourcentage de la variance expliquée. À noter que la variance expliquée par les premiers axes devrait atteindre au moins 50 % de la variance afin de prouver l'utilité de l'ACP [79].

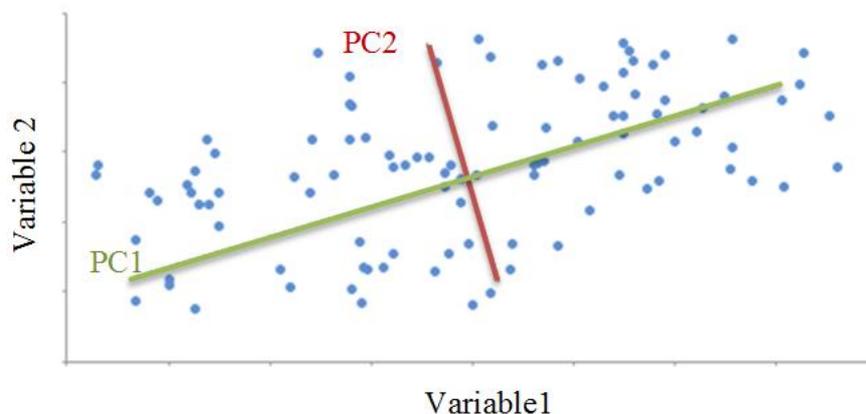


Figure 21. Exemple de deux composantes principales

Le nombre de composantes retenues est déterminé à partir d'une représentation graphique du diagramme d'éboulis (scree-plot) proposé par Cattell [80]. Ce diagramme correspondant à la représentation de décroissance des valeurs propres de chaque composante principale. Le principe consiste à rechercher, s'il existe, un « coude » sur la courbe et de ne conserver que les composantes jusqu'à ce « coude » (**Figure 22**).

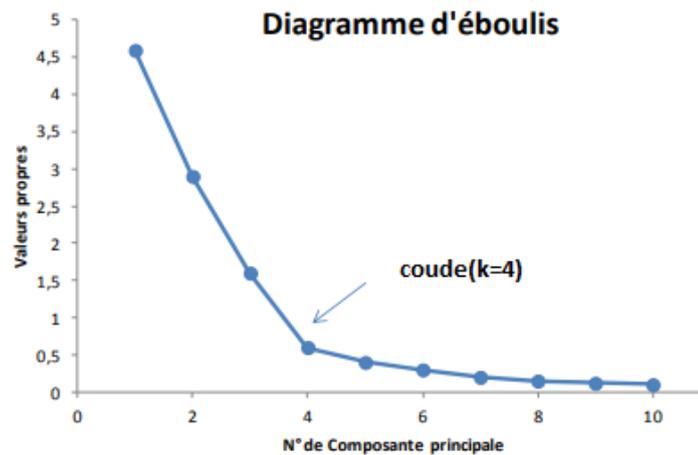


Figure 22. Exemple de diagramme d'éboulis

Une fois le nombre de composantes fixé, nous nous intéressons essentiellement aux variables les plus corrélées avec ces composantes.

2.2.1.3 Discussion

Il nous a semblé plus pertinent de réaliser une sélection de variables de façon disjointe de la classification. Nous avons donc choisi la méthode filtre en calculant le score Laplacien qui est le plus couramment utilisé [81]. Ce score est basé sur le respect de la structure d'un graphe de voisinage entre les individus dont l'hypothèse est que la structure des données dans l'espace d'entrée est localement préservée dans l'espace de sortie. Entre autres, nous avons opté pour l'application de l'ACP afin de sélectionner les mesures pertinentes du corps parce qu'elle a montré son intérêt dans le domaine anthropologique.

2.2.2 Normalisation des variables

La normalisation des variables est souvent requise avant l'application d'une méthode d'analyse qui présente une certaine sensibilité aux trop grandes dispersions des données. La normalisation permet ainsi d'assurer que la mesure de distance accorde le même poids à chaque variable [82]. Sans normalisation, la variable ayant la plus grande valeur dominera sur les autres mesures [83]. Ainsi, la normalisation permet d'ajuster l'ensemble des valeurs de chaque variable par une transformation mathématique pour rendre ces dernières comparables. Il existe différentes techniques de normalisation présentées ci-dessous.

2.2.2.1 La méthode Min-Max

Il s'agit de normaliser toutes les valeurs d'une variable en une série de valeurs comprises entre 0 et 1. La transformation se fait par la formule suivante [82] :

$$x_{normalisé} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Avec :

- x : la valeur à normaliser,
- x_{max} : la plus grande valeur observée pour la valeur x ,
- x_{min} : la plus petite valeur observée pour la valeur x .

2.2.2.2 La méthode Z-score

La transformation en Z-score (aussi appelé score standard) correspond à l'expression d'un écart à la moyenne de la population exprimée en fraction d'écart-type. La formule est la suivante [82]:

$$x_{normalisé} = \frac{x - \mu}{\sigma}$$

Avec :

- μ : la moyenne arithmétique,
- σ : l'écart-type de la population.

Les valeurs obtenues par cette transformation ont la particularité de toujours avoir une moyenne égale à 0 et un écart type égal à 1.

2.2.2.3 La méthode mise à l'échelle décimale

Cette méthode est basée sur le déplacement du point décimal de la valeur x . Le nombre de décimales déplacées dépend de la valeur absolue maximale de x . La formule est la suivante [82]:

$$x_{normalisé} = \frac{x}{10^j}$$

Avec :

- j : est le plus petit entier tel que $\max(|x_{normalisé}|) < 1$.

2.2.2.4 La méthode de la sphère minimale englobante

Dans le cadre de normalisation des objets 3D, la sphère minimale englobante est généralement utilisée [3, 4]. Cette technique s'avère particulièrement pertinente pour la normalisation des corps humains en amont d'une classification morphologique[85]. Elle permet de positionner l'objet 3D dans un repère intrinsèque et d'assurer l'invariance aux transformations géométriques suivantes : translation (normalisation de la position), rotation (normalisation de l'orientation) et changement d'échelle (**Figure 23**).

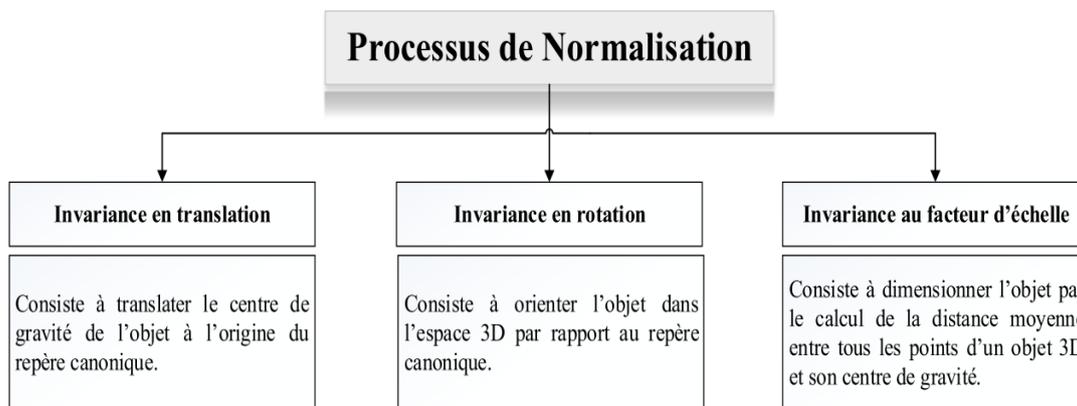


Figure 23. Processus de normalisation par la sphère minimale englobante [63].

Le principe de la méthode de la sphère minimale englobante consiste à trouver la plus petite sphère qui enveloppe l'objet 3D (**Figure 24**).

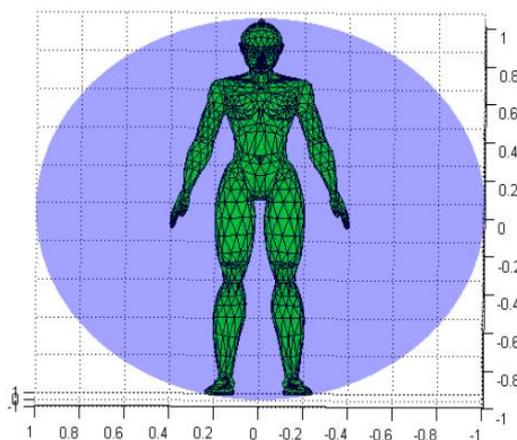


Figure 24. Sphère minimale englobante d'un modèle 3D humain [84].

Elle est définie par un rayon r_{max} donné par la distance maximale du centre de la sphère aux points représentatifs de la surface de l'objet 3D [86].

$$r_{max} = \max_{1 \leq i \leq N} \|g - p_i\|$$

Avec :

- g : le centre de gravité de l'objet 3D,
- p_i : les sommets des triangles du maillage.

Cette méthode a été initialement proposée par le mathématicien James Joseph Sylvester en 1857 par le calcul du cercle minimal englobant (dimension $d = 2$) [87]. Depuis, de nombreux algorithmes ont vu le jour afin d'étendre l'étude à des espaces de dimensions d supérieures. Megiddo [88] a proposé un algorithme basé sur la condition que le cercle minimum passe par deux ou trois points pertinents. Cet algorithme est utilisable seulement en dimension $d = 2$ et $d = 3$. L'algorithme de Gärtner [89] permet de résoudre le problème du calcul de la sphère minimale englobante en dimensions d quelconque. Il est basé sur une approche itérative qui cherche à réduire une sphère englobante initiale afin de la minimiser. L'algorithme de calcul de la sphère se résume alors en trois étapes :

- 1) Initialisation de la sphère en prenant comme centre C un point au hasard et comme rayon r . Le point le plus distant de C sera mémorisé dans un support T .
- 2) Déplacement du centre de la sphère C dans la direction du centre de gravité de T , jusqu'à ce que la sphère réduite touche un nouveau point p , nous insérons alors p dans T . Le support comporte désormais deux points. Répétition de cette étape jusqu'à obtenir 3 points dans T .
- 3) Vérification de la condition d'arrêt de l'algorithme qui est atteinte avec le centre de la sphère qui appartient à l'enveloppe convexe du T . Si cette condition n'est pas satisfaite, les points de T ayant un coefficient négatif seront supprimés de T et l'algorithme se répètera en reprenant le calcul à l'étape 2.

Afin de normaliser chaque objet 3D, les coordonnées de chaque sommet (x,y,z) deviennent alors :

$$x_{normalisé} = \frac{x-C_x}{r}, y_{normalisé} = \frac{y-C_y}{r}, z_{normalisé} = \frac{z-C_z}{r}$$

- r : rayon de la sphère minimale englobante,
- (C_x, C_y, C_z) : coordonnées du centre de la sphère minimale englobante.

2.2.2.5 Discussion

Ces méthodes de normalisation permettent de réduire l'espace de variation des valeurs. La méthode Min-max peut être appliquée quand les données varient dans des échelles différentes. Elle a l'avantage de préserver toutes les relations entre les données [3, 4]. La méthode Z-score peut être appliquée quand les données répondent à des distributions normales. Elle réduit l'effet des points aberrants dans les données mais elle peut modifier les données d'origine [3, 4]. La normalisation par la sphère minimale englobante avec l'algorithme de Gärtner offre des propriétés intéressantes pour décrire deux objets 3D similaires sur la forme [63]. Cette technique est particulièrement intéressante pour la normalisation des morphologies.

Dans notre contexte, il s'agit de normaliser les mesures clés du corps (les variables sélectionnées) afin de les rendre comparables et d'assurer la préservation des proportions du corps humain. Ainsi, nous proposons de normaliser ces mesures par le cercle minimal englobant en utilisant le même principe de la sphère minimale englobante.

2.2.3 Classification

La classification est utilisée dans de nombreux domaines tels que le domaine médical pour regrouper des patients [92], le domaine du marketing pour segmenter les clients [93] ou encore le domaine anthropologique pour distinguer les morphologies dans une population donnée [17].

Les méthodes de classification ont été initialement conçues pour analyser les données quantitatives. Ces données sont présentées sous la forme d'un tableau où les lignes sont les individus, appelés observations, et les colonnes sont les variables, appelés également attributs.

Nous distinguons deux types de classification (**Figure 25**):

- *supervisée* : les classes sont connues à priori. Elle consiste à affecter un nouvel individu à une classe préexistante.
- *non-supervisée* (en anglais clustering) : les classes sont fondées sur la structure, on parle alors de partitionnement des données.

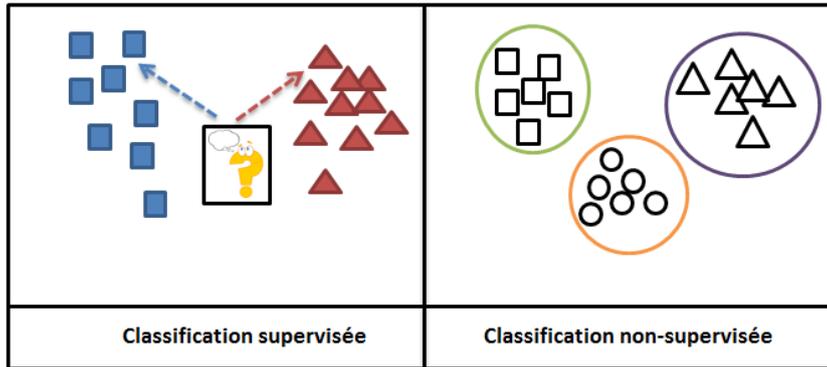


Figure 25. Illustration de deux types de classification

Le **Tableau 5** montre la différence entre les classifications supervisée et non-supervisée.

<i>Supervisé</i>	<i>Non-supervisée</i>
<ul style="list-style-type: none"> - Nombre de classes connues - Construction d'un modèle à partir d'une base d'apprentissage (training data set) - Utilisation de ce modèle pour prédire des données futures. 	<ul style="list-style-type: none"> - Nombre de classes inconnues - Pas de connaissance préalable nécessaire - Utilisé pour comprendre les données

Tableau 5. Classification Supervisée Vs Non-supervisée

2.2.3.1 Classification supervisée

Dans de nombreuses situations, il s'avère intéressant de coupler une classification non supervisée avec une classification supervisée. En effet, cela permet notamment d'assigner des nouveaux individus aux classes créées en définissant des relations entre les attributs utilisés pour la classification non supervisée et de nouveaux attributs ou nombre réduits d'attributs. Par exemple, il semble pertinent de pouvoir associer une personne à une morphologie, définie par un ensemble de mesures anthropométriques, avec seulement quelques paramètres facilement identifiables tels que le poids, la stature ou l'âge. Ainsi, nous proposons dans cette section, une présentation de la méthode de classification supervisée par arbre de décision, pour sa simplicité de mise en œuvre, pour sa large applicabilité et sa facilité d'interprétation [94]. Parmi les nombreuses méthodes existantes comme les machines à vecteurs de support (SVM), les k plus proches voisins et les forêts aléatoires.

L'arbre de décision est une méthode de référence en statistique pour prédire une variable Y à partir d'un ensemble de variables X . Le principe est de partitionner l'ensemble de données d'apprentissage en sous-groupes d'individus qui soient les plus homogènes possibles du point de vue de la variable à prévoir Y . La méthode CART (Classification And Regression Trees), proposé par Breiman et al. [95] est la plus connue car elle permet de construire un arbre binaire (découper une racine en deux nœuds fils) par le biais de partitionnement successives jusqu'à ce que chaque nœud soit homogène, c'est-à-dire que toutes les observations contenues dans le nœud partagent la même valeur pour Y [96].

L'algorithme CART se fait en deux étapes :

- Elaboration de l'arbre maximale : consiste à partitionner récursivement l'espace des données. La variable de partition choisie sera celle qui maximise la décroissance de l'impureté du nœud par rapport à la partition parent. Dans CART, cette impureté est souvent calculée par le biais de l'indice de Gini (plus l'indice est faible plus le nœud est considéré comme pur) :

$$I_{Gini}(t) = 1 - \sum_{k=1}^K f_k^2$$

Avec : f_k est la fréquence de la classe k dans le nœud t .

- Élagage : consiste à choisir le meilleur sous-arbre élagué de l'arbre maximal. Ce choix se fera grâce à la quantification de l'erreur de prédiction sur la base de test.

Il existe plusieurs méthodes de calcul de cette erreur, les plus utilisés sont: Mean Squared Error(MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) et Symmetric Mean Absolute Percent Error (SMAPE) comme l'illustre la **Figure 26** [97].

$$MSE = \frac{\sum_{k=1}^N (F_k - A_k)^2}{N}$$

$$RMSE = \sqrt{\frac{\sum_{k=1}^N (F_k - A_k)^2}{N}} \times 100$$

$$MAE = \frac{1}{N} \sum_{k=1}^N |F_k - A_k|$$

$$SMAPE = \frac{2}{N} \sum_{k=1}^N \frac{|F_k - A_k|}{F_k + A_k}$$

Where A_k is the actual value and F_k is the forecast value

Figure 26. Méthodes de calcul de l'erreur de prédiction [97]

La quantification de l'erreur de prédiction par une validation croisée permet de mieux tester la précision prédictive du modèle. La procédure de cette validation croisée illustrée dans la **Figure 27** consiste à partitionner la base de données en k blocs de même taille. Parmi les k blocs, un bloc est conservé comme base de test pour valider le modèle, et les autres $k-1$ blocs sont utilisés comme base d'apprentissage. Ce processus est ensuite répété k fois (itérations). Elle possède l'avantage de permettre une utilisation équilibrée des données à la fois pour l'apprentissage et la validation. Pour estimer la performance de modèle, il s'agit de calculer la moyenne des performances.

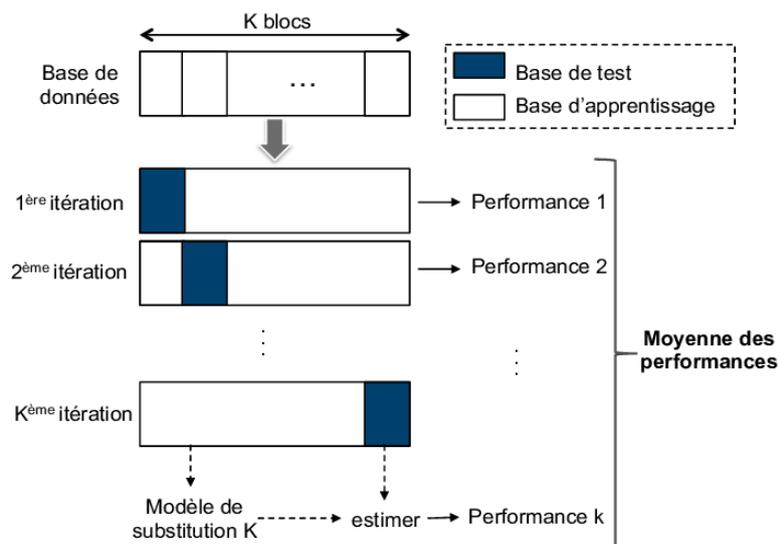


Figure 27. Procédure de la validation croisée

2.2.3.2 Classification non-supervisée

Nous nous sommes orientés vers la classification non-supervisée dans la mesure où nous souhaitons regrouper un ensemble d'individus présentant des caractéristiques morphologiques communes. Le principe de la classification non-supervisée est de trouver les structures des données en les organisant en groupes homogènes et distincts [98]. Les individus dans un même groupe doivent être similaires entre eux et différents des autres groupes. Afin de définir ce regroupement, il est nécessaire de mesurer l'éloignement entre deux individus en introduisant des distances [99].

2.2.3.2.1 Mesures d'éloignement

Une mesure d'éloignement entre individus peut être une similarité ou une dissimilarité.

Une mesure de similarité S est une fonction telle que: $\varphi \times \varphi \rightarrow \mathbb{R}_+$ avec $\varphi = \{i = 1, \dots, n\}$ l'ensemble des individus qui satisfait les trois propriétés suivantes [99] :

- Positivité : $\forall (x,y) \in \varphi, S(x,y) \geq 0$
- Symétrie : $\forall (x,y) \in \varphi, S(x,y) = S(y,x)$
- Maximalité : $\forall (x,y) \in \varphi, S(x,x) \geq S(x,y)$

Plus la mesure de similarité entre deux individus est élevée, plus ils se ressemblent, donc plus ils sont proches.

A l'inverse, une autre possibilité consiste à mesurer la dissimilarité entre individus.

Une mesure de dissimilarité D est une fonction telle que : $\varphi \times \varphi \rightarrow \mathbb{R}_+$ vérifiant les propriétés suivantes [98] :

- Identité: $\forall (x,y) \in \varphi, D(x,y) = 0 \Leftrightarrow x = y$
- Symétrie : $\forall (x,y) \in \varphi, D(x,y) = D(y,x)$

Contrairement à la similarité, moins les individus x et y se ressemblent, plus le score est élevé.

Une distance est une dissimilarité vérifiant en plus la propriété :

- Inégalité triangulaire : $\forall (x,y,z) \in \varphi, D(x,z) \leq D(x,y)+D(y,z)$

La mesure d'éloignement utilisée varie en fonction du type des données. La mesure de similarité est appliquée quand les données sont décrites à l'aide d'attributs binaires aussi

appelés présence/absence. En ce qui concerne les données numériques, la mesure de dissimilarité est appliquée [100].

Parmi les nombreuses mesures de distance de dissimilarité existantes, les plus fréquentes sont la distance Euclidienne et la distance de Manhattan qui sont des cas particuliers de la distance de Minkowski [100] définie pour un ordre $p \geq 1$ (**Tableau 6**).

La distance Euclidienne reste la distance la plus couramment utilisée. Cependant, la distance de Manhattan est considérée plus robuste que la distance euclidienne, vu qu'elle augmente moins rapidement et prend des valeurs plus faibles quand la base de données contient des données aberrantes [101].

Nom	Ordre	Fonction
distance de Minkowski	p	$d_{minkowski(x,y)} = \sqrt[p]{\sum_{i=1}^n x_i - y_i ^p}$
distance euclidienne	2	$d_{euclidienne(x,y)} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$
distance de Manhattan	1	$d_{manhattan(x,y)} = \sum_{i=1}^n x_i - y_i $

Tableau 6. Mesures de distances classiques

2.2.3.2.2 Méthodes de classification non-supervisée

Il existe une très large famille de techniques dédiées à la classification non-supervisée (Clustering). Saxena et *al.* [102] propose une taxonomie de ces techniques qui se déclinent en deux catégories principales : les méthodes hiérarchiques et les méthodes de partitionnement (**Figure 28**). Nous ajoutons à ces deux catégories la carte auto-organisatrice basée sur une méthode d'apprentissage non-supervisée.

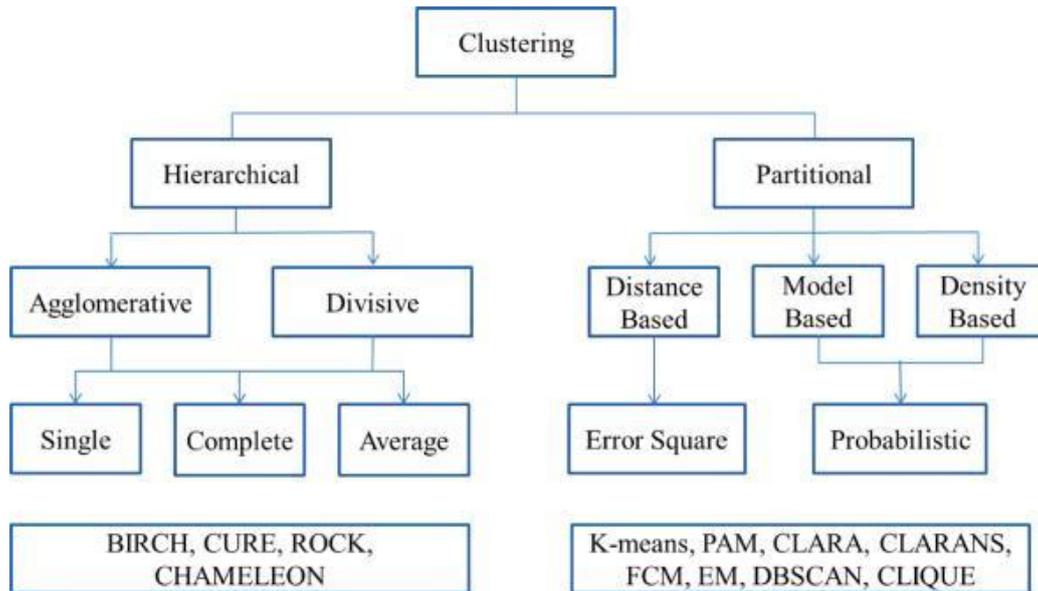


Figure 28. Taxonomie des approches de clustering [102].

A. Méthodes hiérarchiques

Les méthodes hiérarchiques fournissent un ensemble de stratégies de partitionnement des individus en un certain nombre de classes.

Il existe deux types d'algorithmes :

- **Ascendant**, aussi appelé agglomératif : les individus sont placés dans des classes séparées, à chaque étape d'évolution deux classes fusionnent, les itérations sont répétées jusqu'à former une seule classe.
- **Descendant**, aussi appelé divisif : les individus sont placés initialement dans la même et unique classe. Cette classe est ensuite divisée en sous-classes, les itérations sont répétées jusqu'à l'obtention de classes contenant chacune une seule variable.

Les partitions successivement obtenues sont hiérarchisées : elles sont emboîtées les unes dans les autres. De ce fait, le résultat d'une classification hiérarchique est représenté à l'aide d'une arborescence, aussi appelée un dendrogramme (**Figure 29**).

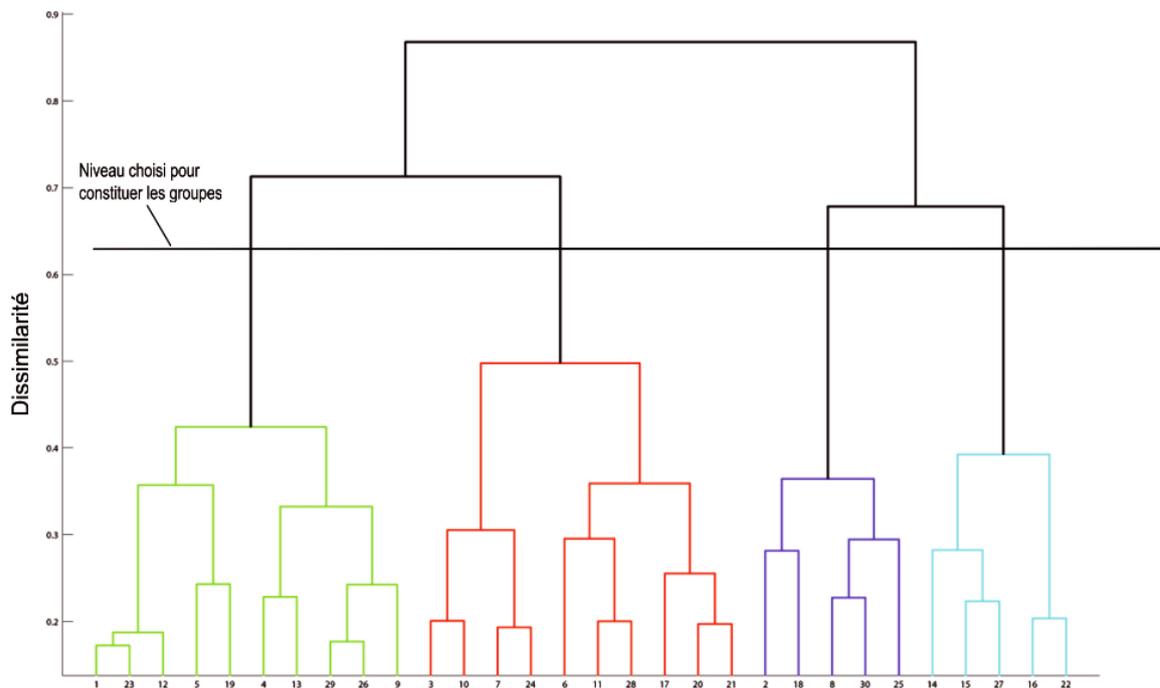


Figure 29. Dendrogramme obtenu après une classification ascendante hiérarchique avec quatre classes [63].

Le nombre de classes doit être déterminé à posteriori à partir de ce dendrogramme en jouant sur son niveau de dissimilarité. Nous pouvons alors explorer différentes possibilités et choisir le nombre de classes qui convient le mieux au problème considéré.

La mise en œuvre de la classification hiérarchique nécessite de définir une distance entre classes au lieu de la distance usuelle entre deux points. Il existe différentes façons de considérer une distance entre deux classes appelées les critères d'agrégation (méthodes de lien). Ces méthodes permettent de mesurer la dissimilarité entre deux classes.

Les critères d'agrégation les plus courants sont [103](**Figure 30**):

- **Single linkage**: la dissimilarité entre deux classes est mesurée par la dissimilarité minimum entre les variables des classes :

$$d(C_i, C_j) = \min_{x \in C_i, y \in C_j} d(x, y)$$

- **Complete linkage**: à l'inverse du single linkage, la dissimilarité entre deux classes est mesurée par la dissimilarité maximum entre les variables des classes :

$$d(C_i, C_j) = \max_{x \in C_i, y \in C_j} d(x, y)$$

- **Average linkage:** la dissimilarité est représentée par la dissimilarité moyenne entre les paires de points des deux classes :

$$d(C_i, C_j) = \frac{1}{\text{Card}(C_i) \times \text{Card}(C_j)} \sum_{x \in C_i, y \in C_j} d(x, y)$$

Avec : $\text{Card}(C_i)$, $\text{Card}(C_j)$, les effectifs respectifs des classes C_i et C_j .

- **Méthode des centroïdes:** la dissimilarité entre les classes est représentée par la dissimilarité entre leur centre:

$$d(C_i, C_j) = d(g_i, g_j)$$

g_i, g_j : les centres de gravité respectifs des classes C_i et C_j .

- **Méthode de Ward:** elle permet de réunir les deux classes dont le regroupement fera baisser l'inertie inter-classe le moins possible:

$$d(C_i, C_j) = \frac{\text{Card}(C_i) \times \text{Card}(C_j)}{\text{Card}(C_i) + \text{Card}(C_j)} d^2(g_i, g_j)$$

Le choix de la distance entre classes est bien souvent basé sur l'expertise de l'opérateur ou l'utilisateur et des données considérées. Hands et Everitt [104] ont comparé ces cinq critères d'agrégation et ils ont constaté que la méthode de Ward se révélait la meilleure dans l'ensemble comparé aux autres méthodes. Ferreira et Hitchcock [105] ont comparé la performance de quatre critères (single, complete, average et méthode de Ward). Ils ont utilisé l'indice de Rand pour comparer la performance de ces méthodes. Selon leur étude, la méthode de Ward était la meilleure, tandis que la méthode single linkage était la moins efficace. Single linkage est bien approprié quand nous nous attendons à des grappes de tailles relativement similaires et que l'ensemble des données ne contient pas de valeurs aberrantes [106].

La méthode de Ward est ainsi utilisée par défaut dans la plupart des programmes informatiques.

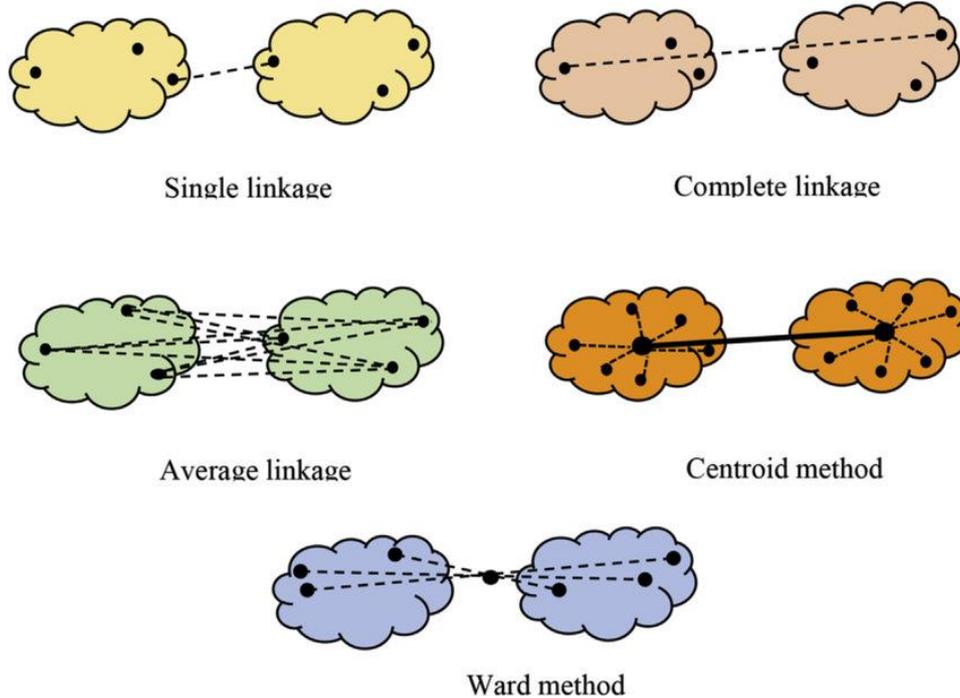


Figure 30. Illustration de différents critères d'agrégation [107].

L'algorithme ascendant qui est le plus utilisé est détaillé ci-dessous.

Algorithme : Classification Ascendante Hiérarchique

Chaque variable est placée dans une classe.

Calculer la dissimilarité entre chaque paire de variables.

Répéter

Calculer un critère d'agrégation entre chaque paire de classes.

Agréger les 2 classes les plus proches au sens du critère choisi.

Jusqu'à Agrégation en une seule classe

B. Méthodes de partitionnement

Les méthodes de partitionnement, appelées aussi méthodes de regroupement, permettent le partitionnement d'un ensemble de données en plusieurs sous-groupes pertinents défini a priori. Les données regroupées dans un même groupe doivent être similaires entre elles, contrairement aux données appartenant à des groupes différents.

Il existe trois principales familles de méthodes de partitionnement [23, 29] :

- 1) **Méthodes basées sur la distance** : elles reposent sur une mesure de distances entre les données. Ainsi, deux données qui se ressemblent auront une distance réduite, alors que deux données différentes auront une distance plus grande. Ces méthodes visent à minimiser en général la variance intra-classe:

$$\arg \min \sum_{i=1}^k \sum_{x \in C_i} \|x - g_i\|^2$$

Avec :

- k: l'ensemble des classes,
 - G_i : le centre de la classe C_i .
- 2) **Méthodes basées sur les graphes**: elles se basent sur la construction d'un graphe à partir de la mesure de similarité entre données. Dans ce graphe, chaque nœud représente une donnée et chaque poids associé à une arête représente la distance ou la similarité entre les deux données connectées. Les classes sont alors regroupées en minimisant la connectivité entre différentes classes (faibles valeurs associées aux arêtes) et en maximisant la connectivité au sein de chaque classe (grandes valeurs associées aux arêtes).
 - 3) **Méthodes basées sur la densité**: elles se basent sur une estimation de la densité des données dans l'espace de représentation en fonction de leur similarité. L'idée principale est que les classes à découvrir sont constituées d'un ensemble de points de forte densité qui forme le centre de la classe, entourée par des points en densité plus faible qui forment la frontière. Les zones de faibles densités définissent les limites entre les classes. Un des algorithmes les plus utilisés est DBSCAN (Density Based Spatial Clustering of Applications with Noise).

Nous nous limitons à expliquer l'une des méthodes basées sur la distance : K-moyennes (en anglais K-means) et la carte auto-organisatrice(en en anglais SOM : Self Organizing Map).

B.1 K-moyennes

L'algorithme k-moyennes est probablement l'un des algorithmes de partitionnement les plus répandus. Il a été largement utilisé pour classer les populations cibles en différentes morphologies sur la base de mesures anthropométriques (**section 1.2.3.4**). Il est relativement

simple et permet d'obtenir de bonnes performances grâce à la minimisation des distances qui favorise la découverte de classes compactes et bien séparées: les données d'une même classe sont proches les unes des autres et sont éloignées des données des autres classes. Néanmoins, Il n'est pas adapté aux données non convexes. De plus, le résultat dépend de l'initialisation de ces centres.

La **Figure 31** illustre le principe de la méthode K-moyenne qui se base sur la ré-allocation dynamique des individus à des centres de classes (nommé centroïde), eux-mêmes recalculés à chaque itération.

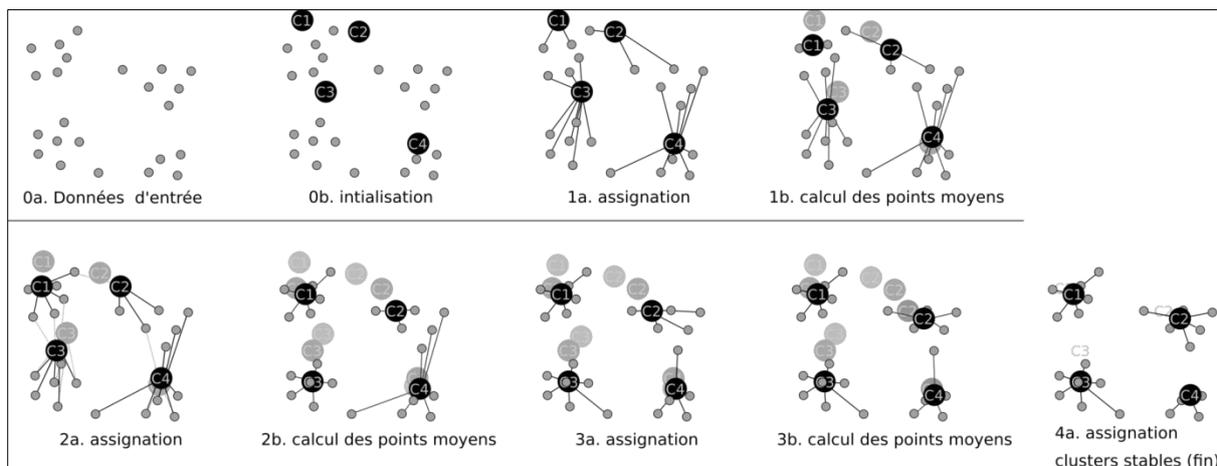


Figure 31. Illustration de déroulement de l'algorithme K-moyennes[37].

L'algorithme K-moyennes est détaillé ci-dessous.

Algorithme : K-moyennes

Initialisation: choisir aléatoirement K centres de classes parmi l'ensemble des variables

Répéter

Calculer la distance de chaque variable aux différents centres

Assigner chaque variable au groupe dont le centre est le plus proche

Recalculer les centres des classes

Jusqu'à convergence de la partition ou un nombre maximal d'itérations

De multiples variantes existent pour améliorer ses performances: kernel K-means [109] (K-means basée sur des méthodes à noyaux qui permet de travailler sur des problèmes non

linéairement séparables), k-means++ [110] (K-means qui propose une initialisation des centroïdes améliorant la probabilité d'obtenir la solution optimale).

B.1.1 Nombre optimal de classes

Le choix du nombre de classes est une étape cruciale dans les algorithmes de partitionnement. Un nombre de classes grand peut conduire à un partitionnement trop fragmenté des données. Par contre, un nombre de classes trop petit, conduira à avoir potentiellement des classes trop généralistes contenant beaucoup de données. La difficulté résidera donc à choisir un nombre de classes qui permettra de mettre en lumière des structures intéressantes entre les données. Cependant il n'existe pas de procédé automatisé pour trouver le nombre optimal de classes k , mais plutôt des indices qui permettent d'évaluer la qualité des partitions obtenues pour un ensemble de valeurs de k . Ces indices cherchent un compromis entre les mesures inter-classes et intra-classes. La mesure intra-classe quantifie la compacité des classes tandis que la mesure inter-classes quantifie la séparation entre les classes.

Dans ce qui suit, nous résumons les principaux indices d'évaluation du nombre de classes.

L'indice Silhouette (S): Il est introduit par Rousseeuw en 1987 [111]. Le but de cet indice est de vérifier si chaque élément i a été correctement classé. Il est décrit par la formule suivante [112] :

$$-1 \leq S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \leq 1$$

$a(i)$ représente la moyenne des distances entre l'élément i et tous les autres éléments de la même classe et $b(i)$ représente la moyenne de la distance entre l'élément i et tous les autres éléments de la classe la plus proche. Il est compris entre -1 et 1. Quand $S(i)$ est proche de 1, l'assignation de i à sa classe est satisfaisante. Par contre, si $S(i)$ est proche de -1, l'assignation de i à sa classe est insatisfaisante. Mais si $S(i)$ est proche de 0 alors il pourrait également être assigné dans la classe la plus proche.

Pour une classe C_i donnée, son indice de silhouette est défini par la moyenne des indices de silhouette des éléments qui lui appartiennent :

$$S(C_i) = \frac{1}{|C_i|} \sum_{i \in C_i} S(i)$$

L'indice de silhouette global S est donné par la moyenne de silhouettes de différentes classes $S(C_i)$:

$$S = \frac{1}{k} \sum_{i=1}^k S(C_i)$$

Le nombre optimal de classes retenues est alors celui qui permet d'obtenir une silhouette globale S maximale.

L'indice de Davies-Bouldin (DB) [113] : Il est basé sur la minimisation du rapport des distances intra-classe et inter-classe. L'indice DB est formulé de la façon suivante [112] :

$$DB = \frac{1}{K} \sum_{k=1}^K \max_{k \neq k'} \left\{ \frac{\delta_k + \delta_{k'}}{\Delta_{kk'}} \right\}$$

Avec:

- K : le nombre de classes,
- $\Delta_{kk'}$: la distance entre les centroïdes G^k et $G^{k'}$ respectivement des classes C_k et $C_{k'}$,
- δ_k : la moyenne des distances entre les points appartenant à la même classe C_k et leurs barycentres G^k .

La distance entre les deux barycentres est calculée par:

$$\Delta_{kk'} = \|G^{k'} - G^k\|$$

La distance moyenne δ_k est donnée par:

$$\delta_k = \frac{1}{n_k} \sum_{x \in C_k} \|x - G^k\|$$

Avec:

- n_k : le nombre d'éléments de la $k^{\text{ème}}$ classe et x est un élément de la classe G^k

La partition de meilleure qualité sera celle qui minimisera l'indice de Davies-Bouldin.

L'indice de Calinski-Harabasz (CH) [114]: Il est décrit par la formule suivante [112]:

$$CH = \frac{BGSS / (K - 1)}{WGSS / (N - K)} = \frac{N - K}{K - 1} \frac{BGSS}{WGSS}$$

Avec:

- BGSS (Between-Group Sum of Squares) et WGSS (Within-Group Sum of Squares) représentant les sommes carrés des inerties inter-classes et intra-classes telles que:

$$BGSS = \sum_{k=1}^K n_k \|G^k - G\|^2, \quad WGSS = \sum_{k=1}^K \sum_{x \in C_k} \|x - G^k\|^2$$

Avec:

- K : le nombre de classes,
- n_k : le nombre d'éléments de la classe,

- G^k : le centroïde de la classe,
- G : la moyenne globale de l'ensemble des données,
- x : un élément de l'échantillon de données.

La partition de meilleure qualité sera celle qui maximisera l'indice de Calinski-Harabasz.

C. La carte auto-organisatrice

La carte auto-organisatrice (en anglais SOM : Self Organizing Map), proposée par Kohonen [115], est une technique d'auto-organisation qui cherche à projeter des données multidimensionnelles sur un espace de dimension réduite. Elle permet d'obtenir un partitionnement des individus en groupes similaires, tout en conservant la structure topologique des données [116]. Cette carte est un type de réseau de neurones artificiels où chaque neurone représente une classe et chaque donnée est affectée au neurone dont le vecteur référent appelé aussi vecteur prototype W_{ij} qui est le plus proche. La dimension de ce vecteur est égale à la dimension de l'espace d'entrée. De plus, deux neurones voisins sur la carte représenteront des données qui sont proches dans l'espace d'entrée.

La **Figure 32** représente les deux couches qui constituent ce réseau : une couche d'entrée de dimension n sert uniquement à la présentation des données X et une couche de sortie (appelée aussi la couche de Kohonen) de dimension $m = p * q$ formée de neurones disposés suivant une topologie de distribution.

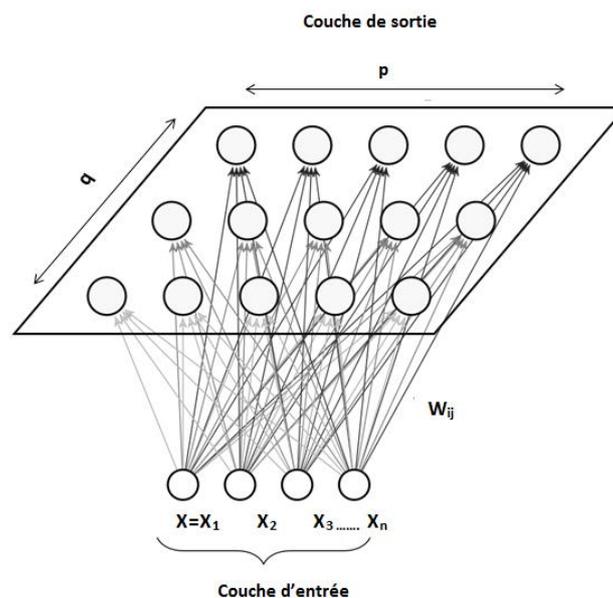


Figure 32. Architecture de la carte auto-organisatrice 2D.

C.1 Topologies de la carte auto-organisatrice

La topologie de distribution des neurones de la carte auto-organisatrice fait référence à la façon dont les neurones sont répartis sur la couche de Kohonen. Dans la littérature, nous distinguons trois grandes topologies: topologie 1D, topologie 2D rectangulaire ou hexagonal et topologie 3D (**Figure 33**) [117].

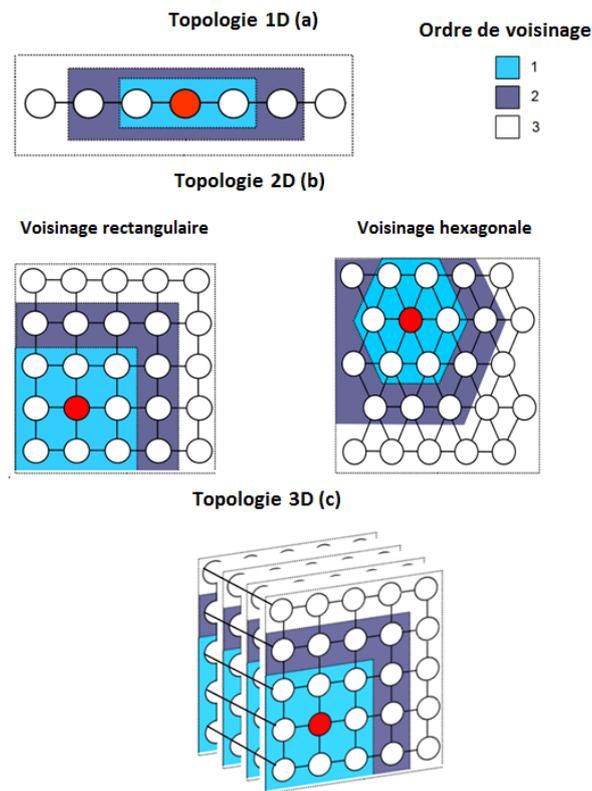


Figure 33. Illustration de différentes topologies de la carte auto-organisatrice.

- 1) **La topologie 1D**, comme le montre la **Figure 33 (a)**, est composée d'une ligne de neurones avec un lien direct entre les voisins. Ce type de topologie est assez simple à implémenter, mais il restreint le parallélisme des connexions des réseaux de neurones.
- 2) **La topologie 2D** forme la distribution la plus utilisée pour la mise en place de la carte auto-organisatrice. Cette distribution présentée à la **Figure 33 (b)**, est composée d'un ensemble de neurones répartis sur un plan bidimensionnel. Chaque neurone est identifié par sa position en termes de ligne et colonne sur la carte. Pour cette topologie bidimensionnelle, il existe deux principales méthodes pour l'arrangement des neurones (la manière d'interconnexion des neurones sur la couche de Kohonen) :
 - *Voisinage rectangulaire* où chaque neurone possède 4 voisins de premier ordre,
 - *Voisinage hexagonale* où chaque neurone possède 6 voisins de premier ordre.

- 3) **La topologie 3D** est illustrée par la **Figure 33 (c)**. Les neurones de la couche de Kohonen sont répartis sur un plan tridimensionnel. Chaque neurone est identifié par trois indices : couche, ligne et colonne. Elle est généralement utilisée pour traiter des données présentées sur des espaces d'entrée en 3D.

C.2 Algorithme de la carte auto-organisatrice

L'algorithme SOM se base sur une approche itérative et compétitive : les neurones entrent en compétition pour sélectionner le neurone gagnant (en anglais BMU : Best Matching Unit) ayant la distance la plus petite avec le vecteur d'entrée [117]. Ainsi, au cours de l'apprentissage compétitif, l'adaptation du neurone gagnant, ainsi que ceux de ses voisins, se fait en modifiant leurs vecteurs prototypes de façon à ce qu'ils se rapprochent de la valeur du vecteur d'entrée. Cette stratégie est appelée « Winner Takes ALL », c'est-à-dire tout au gagnant [108].

L'algorithme suivant décrit les trois phases de SOM [117] :

- **La phase d'initialisation** : cette phase consiste à initialiser les valeurs des éléments du vecteur prototype de chaque neurone sur la carte.
- **La phase de compétition** : cette phase consiste à calculer la distance euclidienne entre le vecteur prototype et le vecteur d'entrée X pour chaque neurone. Le neurone gagnant correspond au neurone qui a obtenu la distance minimale. L'indice de ce neurone sera ensuite utilisé au cours de la phase d'adaptation.
- **La phase d'adaptation** : chaque neurone calcule son taux de voisinage par rapport au neurone gagnant en utilisant la fonction gaussienne. Ensuite, les vecteurs prototypes des neurones concernés seront mis à jour.

L'objectif de cet algorithme est alors d'adapter les vecteurs prototypes des neurones, à la topologie de l'espace d'entrée, tout en gardant la disposition de voisinage définie au préalable de la carte auto-organisatrice. Après plusieurs itérations, la carte converge vers une représentation réduite de l'espace d'entrée.

Algorithme : SOM

Entrées :

- $X = \{\text{La matrice des individus } x\}$
- $n = \text{nombre d'observations}$
- $m = \text{nombre de neurones (dimensions de la carte = } lig * col)$
- $lig, col: \text{nombre de lignes, nombre de colonnes}$
- $T = \text{nombre total d'itérations}$
- $\sigma_{init}, \sigma_{final} = \text{valeurs initiale et finale du rayon de voisinage}$
- $\alpha_{init} = \text{valeur initiale du pas d'apprentissage}$

Initialisation: initialiser les vecteurs prototypes de chaque neurone W_k avec des valeurs aléatoires, $k \in \{1..m\}$
 $t \leftarrow 0$;

Répéter

$$\sigma(t) \leftarrow \sigma_{init} \left(\frac{t}{T} \right) (\sigma_{final} - \sigma_{init})$$

$$\alpha(t) \leftarrow \alpha_{init} \left(1 - \frac{t}{T} \right)$$

Compétition : calculer le neurone gagnant d'indice c

Pour i de 1 à n Faire
 $d_2(x_i, w_c) \leftarrow \min_{k \in \{1..m\}} d_2(x_i, w_k(t))$
Fin pour

Adaptation : calculer le taux de voisinage

Pour k de 1 à m Faire
 $h_{kc}(t) \leftarrow \exp\left(-\frac{\|r_c - r_k\|^2}{2\sigma^2(t)}\right)$ OÙ r_c et r_k sont l'emplacement des neurones c et k
 $\sigma(t)$ rayon de voisinage à l'instant t
Fin pour

Mettre à jour les vecteurs prototypes :

Pour k de 1 à m Faire
 $w_k(t+1) \leftarrow \begin{cases} w_k(t) + \alpha(t) h_{kc}(t) [x_i - w_k(t)] & \text{si } \|r_c - r_k\|^2 \leq \sigma(t) \\ w_k(t) & \text{sinon} \end{cases}$
Fin pour
 $t \leftarrow t + 1$

Jusqu'à $t > T$

Sorties : $W_k(T)$

C.3 Paramètres de la carte auto-organisatrice

Les cartes auto-organisatrices sont souvent critiquées pour le nombre de paramètres à régler, notamment du fait qu'un mauvais choix d'un de ces paramètres peut générer des résultats incohérents. Par exemple, si le pas α d'apprentissage diminue trop rapidement, le réseau n'a pas le temps de s'organiser de manière convenable.

Les principaux paramètres sont :

- **La dimension de la carte :** une méthode considère que puisque la carte auto-organisatrice est destinée à la classification des données, sa dimension sera choisie de sorte que le nombre total de neurones correspond au nombre de classes désirées [116]. Garcia et González [118] propose une autre méthode où le nombre de neurones m de la carte dépend du nombre d'observations de vecteur d'entrée n :

$$m = 5\sqrt{n}$$

- **Les vecteurs prototypes initiaux :** l'initialisation aléatoire des vecteurs prototypes peut influencer le résultat produit par l'algorithme et le temps de convergence. Une solution à ce problème consiste à exécuter plusieurs fois l'algorithme avec des vecteurs initiaux différents, et à choisir les vecteurs qui permettent de minimiser au plus l'erreur moyenne de quantification (en anglais mqe =mean quantization error) qui est définie par la moyenne des distances euclidiennes entre le neurone gagnant c et le vecteur d'entrée $X=\{x_1 x_2 \dots x_n\}$ [116]:

$$mqe = \frac{\sum_{i=1}^n d_2(x_i, w_c)}{n}$$

- **Le rayon de voisinage initial σ_{init} :** ce rayon correspond à la zone autour du neurone gagnant dont les neurones seront affectés par l'adaptation et leurs vecteurs prototypes seront mis à jour. L'idée est d'initialiser le rayon avec une grande valeur qui décroît à chaque itération. Cette valeur est souvent initialisée par la valeur du diamètre D de la carte [117] :

$$D = \frac{lig + col}{2}$$

- **Le pas d'apprentissage initial α_{init} :** il représente le coefficient d'adaptation du vecteur prototype. Ainsi, l'adaptation de ce vecteur au cours de l'apprentissage est réduite par le pas $\alpha(t)$ (où $0 \leq \alpha \leq 1$) qui tend vers 0 [117].

2.2.3.3 Discussion

Les méthodes de classification non-supervisée sont divisées en trois grandes familles: les approches hiérarchiques, les approches par partitionnement et la carte auto-organisatrice.

L'algorithme de l'approche hiérarchique a l'avantage d'être stable étant donné que des exécutions multiples donnent toujours le même résultat. Une fois le dendrogramme obtenu, il est nécessaire de choisir un niveau d'acceptation pour obtenir les classes. Le choix de ce niveau reste un problème. De plus, l'inconvénient principal est sa complexité car de grands jeux de données, le coût de stockage et de calcul de cet algorithme peut devenir prohibitif. Autre inconvénient, le critère de partitionnement n'est pas global et dépend des classes obtenues, c'est-à-dire, deux variables placées dans des classes différentes ne seront jamais plus comparées à l'étape suivante. Contrairement aux approches hiérarchiques, les approches par partitionnement cherchent la meilleure partition en k classes disjointes des données, le nombre de classes étant fixé à priori. Ce nombre n'est pas forcément intuitif. Pour pallier ce problème, les indices d'évaluation d'une partition résultante d'un algorithme de classification sont utilisés pour déterminer le nombre optimal de classe, en exécutant plusieurs fois l'algorithme avec à chaque fois un nombre k de classes différent. Il nous semble donc approprié d'utiliser l'algorithme de partitionnement K-moyennes avec ces indices pour identifier les morphologies dans une population donnée.

La carte auto-organisatrice présente un grand intérêt dans la classification non-supervisée des données notamment grâce à la préservation de la topologie des données. Cette technique s'avère très intéressante pour un pré-traitement des données en vue d'une classification ou d'un processus d'initialisation.

2.3 Optimisation

Dans cette partie, nous présenterons le concept d'optimisation, ainsi que les notions de dominance Pareto. Enfin, nous expliquerons les principales méthodes de résolution existantes. Ces outils seront principalement utilisés pour la construction du système de tailles.

2.3.1 Formulation d'un problème d'optimisation

Dans le cadre d'une optimisation mono-objectif, il s'agit de chercher l'optimum d'une fonction f appelée fonction objectif ou fonction coût qui peut être un minimum ou un maximum présentant toujours une solution unique. Cependant, dans la plupart des problèmes

réels, il s'agit de satisfaire des besoins souvent contradictoires à travers plusieurs fonctions coûts, ce qui étend le problème d'optimisation au cadre multi-objectif.

Un problème d'optimisation multi-objectif peut être exprimé comme [119] :

$$y = f(x) = \left\{ \begin{array}{l} \min (f_1(x), f_2(x), f_3(x) \dots, f_m(x)) \\ \text{tel que } x \in X \text{ et } m \geq 2 \end{array} \right\}$$

Avec:

- X : l'ensemble des solutions
- m : le nombre de fonctions objectif
- $x = (x_1, x_2, x_3, \dots, x_n)$: le vecteur des paramètres représentant les variables de décision.

A chaque solution $x \in X$, est associé un vecteur objectif $y \in Y$, où Y est l'ensemble des points réalisables de l'espace objectif et $y = (y_1, y_2, y_3, \dots, y_n)$ est le vecteur des objectifs.

Ainsi, la résolution d'un problème d'optimisation multi-objectif conduit à l'obtention d'un assortiment de solutions. Il n'existe généralement pas de solution qui serait à la fois optimale pour chaque objectif, étant donné qu'aucun critère de préférence n'est incorporé dans l'étape d'optimisation [120]. Ces solutions fournissent alors un compromis entre toutes les fonctions à optimiser. Un nombre réduit des meilleures solutions dites « solutions efficaces » va être retenu par un décideur. Ce choix des solutions efficaces est basé sur la relation de dominance au sens de Pareto qui permet de comparer les solutions potentielles deux à deux dont nous rappelons sa notion dans la section suivante.

2.3.2 Notion de dominance Pareto

L'objectif de l'optimisation multi-objectif est de fournir l'ensemble des solutions efficaces tel qu'il n'existe aucune solution qui soit meilleure qu'une autre sur tous les objectifs, aussi appelé ensemble de Pareto qui doit son nom à l'économiste italien Vilfredo Pareto [121]. Pour lui, cet ensemble constitue « un équilibre tel que l'on ne peut pas améliorer un critère sans détériorer au moins un des autres critères ». De manière plus concrète, l'ensemble de Pareto est constitué des solutions non dominées et sa représentation graphique est le front de Pareto qui forme une frontière d'optimalité, comme l'illustre la **Figure 34**.

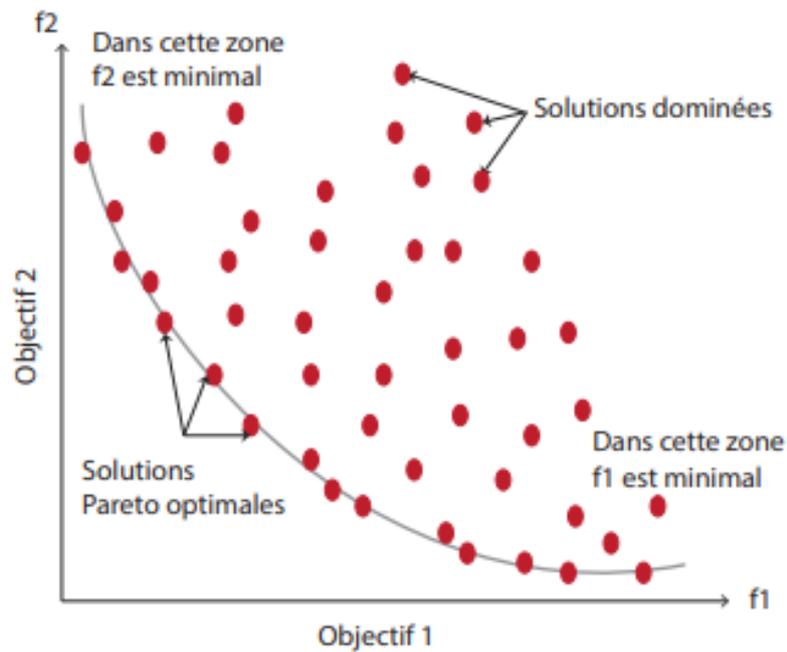


Figure 34. Illustration de front de Pareto pour un problème de minimisation bi-objectif [122]

La notion de dominance peut se traduire par la définition suivante [123]:

Définition : une solution x_1 **domine au sens de Pareto** une solution x_2 si et seulement si :

- x_1 est au moins aussi bon que x_2 dans tous les objectifs,
- et x_1 est strictement meilleur que x_2 dans au moins un objectif.

$$\left\{ \begin{array}{l} \forall i \in \{1, \dots, n\}, f_i(x_1) \leq f_i(x_2) \\ \exists i \in \{1, \dots, n\}, f_i(x_1) < f_i(x_2) \\ x_1 \neq x_2 \end{array} \right\}$$

$n = \text{nombre de fonctions objectifs}$

2.3.3 Les méthodes de résolution

Dans cette section, nous présentons les méthodes les plus utilisées dans les problèmes d'optimisation mono et multi-objectif. La **Figure 35** présente une possible classification de ces méthodes donnée par Liefoghe [124].

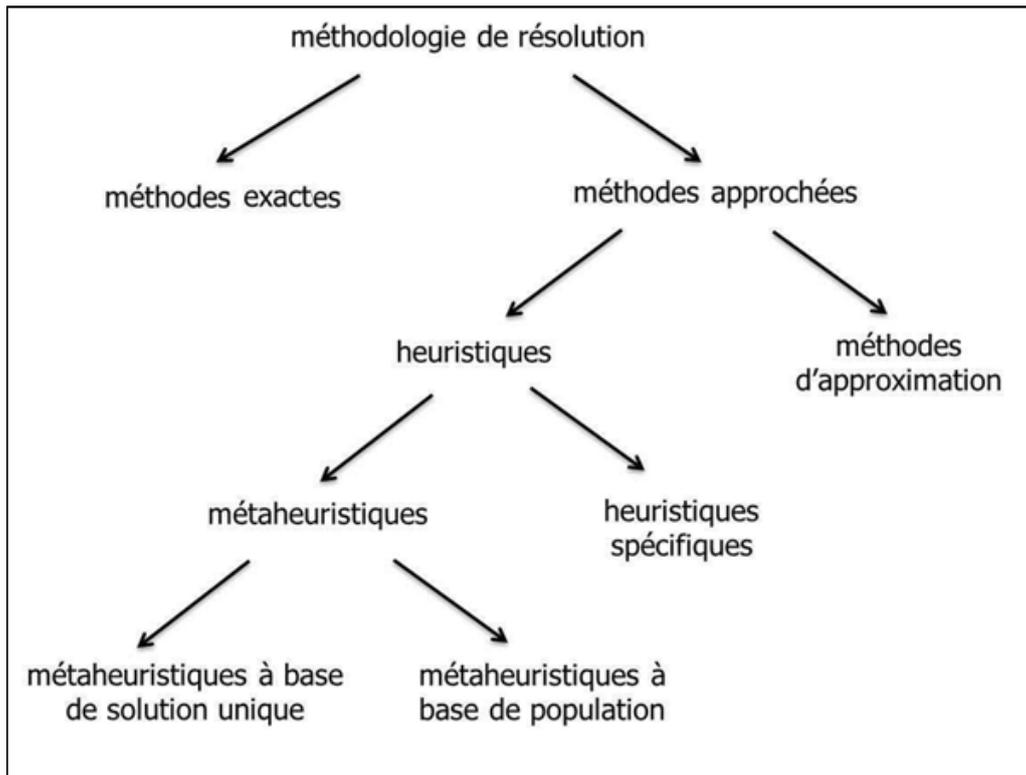


Figure 35. Classification des méthodes de résolution [124]

Selon la complexité du problème d'optimisation à résoudre, deux catégories de méthodes peuvent être distinguées : les méthodes exactes et les méthodes approchées. Les méthodes exactes permettent d'obtenir des solutions dont l'optimalité est garantie, mais sont souvent coûteuses en temps d'exécution. En revanche, les méthodes approchées permettent de générer des solutions satisfaisantes pour des problèmes complexes de grandes tailles en un temps de calcul raisonnable, mais sans garantie d'optimalité [42, 43]. Ces méthodes sont elles-mêmes divisées en deux sous-catégories : les heuristiques et les méthodes d'approximation. Les heuristiques se déclinent en deux méthodes : les heuristiques spécifiques à un problème donné et les métaheuristiques qui seront considérées dans le cadre de cette thèse. Elles ont montré leur efficacité dans divers domaines d'application en délivrant des solutions acceptables. Cela découle du fait qu'elles peuvent être appliquées à tout problème pouvant être exprimé sous la forme d'un problème d'optimisation [126].

Les métaheuristiques sont basées sur la recherche du meilleur compromis entre l'exploration de l'espace de recherche (aussi appelée diversification : recherche des nouvelles solutions pour trouver les potentiels optimums locaux) et l'exploitation des résultats (aussi appelée intensification : sélectionne le sous-espace de recherche le plus promoteur). Elles appartiennent à la famille des méthodes stochastiques [125], c'est-à-dire que l'algorithme ne

renvoie pas la même solution entre deux exécutions distinctes à cause de l'aspect aléatoire présent dans ces méthodes. Deux types de métaheuristiques existent (**Figure 36**):

- **Les métaheuristiques à base de solution unique** : ces méthodes manipulent une seule solution durant le processus de recherche telle que la méthode du recuit simulé, la méthode de recherche tabou, ... Elles sont généralement axées sur l'exploitation. Le principe de fonctionnement commence par une solution unique, puis, à chaque itération, la solution courante est déplacée vers une meilleure solution dans son voisinage, sans considérer les autres solutions appartenant à toute l'espace, jusqu'à ce qu'aucune amélioration supplémentaire ne soit possible. L'inconvénient principal de ce principe de fonctionnement est qu'il se trouve piégé dans le premier optimum local rencontré [126].
- **Les métaheuristiques à base de population** : Contrairement aux algorithmes partant d'une seule solution, ces méthodes opèrent simultanément avec un ensemble de solutions courantes appelé population améliorée, au fur et à mesure des itérations. Il existe souvent une interaction, directe ou indirecte, entre les solutions, pour faire évoluer cette population, ce qui la rend moins déficiente en matière de convergence vers les optimums locaux [126]. Ces méthodes englobent les algorithmes génétiques, l'optimisation par essais particuliers, Elles sont généralement exploratoires.

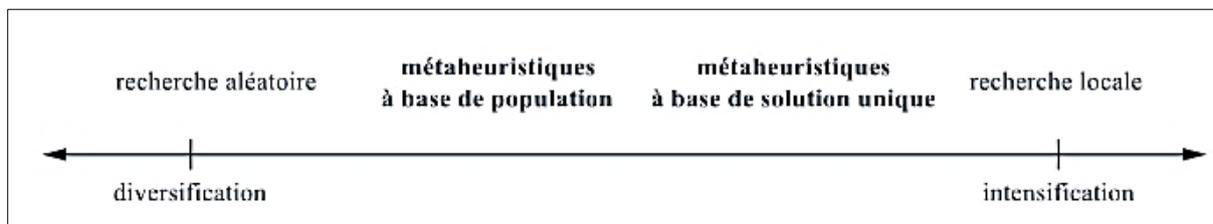


Figure 36. Critères contradictoires des métaheuristiques [124]

Nous nous sommes intéressés plus particulièrement aux algorithmes génétiques qui appartiennent aux métaheuristiques à base de population car ce sont des méthodes qui simulent le processus d'évolution d'une population et sont capables de trouver le minimum global d'une fonction [127]. Ils présentent par ailleurs des facilités de mise en œuvre pour les spécificités de notre problème, comme nous le verrons dans le chapitre 3. Dans ce qui suit, nous décrivons brièvement leur principe.

2.3.3.1 Les algorithmes génétiques

Les algorithmes génétiques (noté AG) s'inspirent de la théorie de l'évolution de Darwin (1869) dont le but est de résoudre un problème d'optimisation. Ils cherchent à permettre à l'ordinateur d'imiter les mécanismes d'évolution des êtres vivants. C'est la combinaison de deux domaines la biologie et l'informatique. Dans ce mécanisme, nous commençons par définir la terminologie de la génétique suivante [128]:

- Genèse ou initialisation de la population : c'est la phase d'initialisation de l'algorithme, il s'agit de définir une population initiale de solutions potentielles de n individus.
- Chromosome : c'est une chaîne représentant les caractéristiques de l'individu.
- Gène : c'est une caractéristique d'un individu.
- Evaluation : c'est la phase d'analyse des individus basée sur le calcul de la fonction de performance (fitness) permettant de les classer.
- Sélection : c'est le choix des individus les mieux adaptés qui vont se reproduire pour former la population suivante.
- Croisement: c'est la phase de reproduction des individus choisis. Ils échangent des leurs gènes.
- Mutation: c'est la modification d'un gène dans le but de modifier les caractéristiques de l'individu et ainsi contrer les phénomènes de convergence dans les optimums locaux.

La **Figure 37** présente l'organigramme du fonctionnement d'un algorithme génétique.

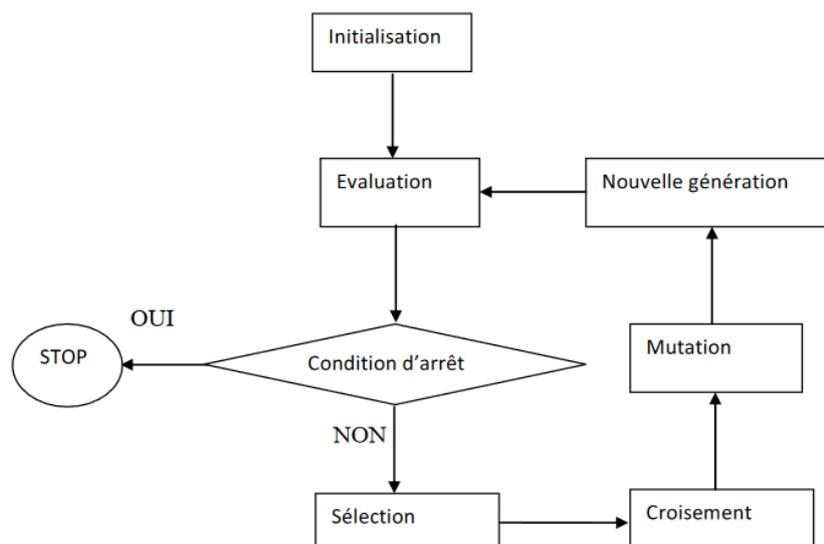


Figure 37. Fonctionnement général d'un AG [128]

L'algorithme débute par le choix, généralement aléatoire, d'une population initiale de n chromosomes. Par la suite, la performance relative de chaque chromosome est estimée à l'aide d'une fonction fitness ou objectif qui permet de déterminer son degré de qualité : il s'agit d'établir une mesure commune d'évaluation afin de le comparer aux autres chromosomes. Sur la base de ces performances, la génération d'une nouvelle population est effectuée par les trois opérateurs génétiques : la sélection, le croisement et la mutation. Ce cycle d'évolution se répète jusqu'à ce que le critère d'arrêt soit atteint, qui peut être le nombre de générations fixé depuis le début.

2.3.3.1.1 Opérateurs génétiques

Les trois opérateurs génétiques jouent un rôle important dans la réussite de l'évolution de l'AG [129]. Nous présentons brièvement chaque opérateur.

A. Opérateur de Sélection

Cet opérateur permet de choisir les chromosomes pour la reproduction afin de garantir la convergence vers l'optimum global. C'est le *principe d'adaptation*. Il existe plusieurs méthodes de sélection dans la littérature, à savoir [128]:

- **Sélection par rang :** Cette méthode consiste à attribuer à chaque chromosome c_i , avec $i \in [1, \dots, n]$, une probabilité de sélection proportionnelle à son classement (rang) en fonction de la valeur fitness. Cette sélection assure que tous les chromosomes ont une chance d'être sélectionnés. Cependant, elle conduit à une convergence plus lente vers la bonne solution.

$$prob(c_i) = \frac{rang(c_i)}{\sum_{i=1}^n rang(c_i)}$$

- **Sélection par roulette :** Cette méthode consiste à associer à chaque chromosome c_i une probabilité de sélection proportionnelle à sa valeur fitness $f(c_i)$. Elle s'inspire des roues de loterie. Les chromosomes ayant une grande valeur fitness ont donc plus de chance d'être sélectionnés. Il s'agit de la méthode la plus courante, mais elle présente un inconvénient quand la valeur de fitness des individus est extrêmement variée. En effet, il y a un risque d'aboutir à une situation de stagnation dans l'évolution où la roulette est attribuée au chromosome ayant la grande probabilité par exemple 90%, alors les autres chromosomes auront une très faible probabilité d'être sélectionnés.

$$prob(c_i) = \frac{f(c_i)}{\sum_{i=1}^n f(c_i)}$$

- **Sélection par Tournoi** : Cette méthode utilise la sélection proportionnelle sur des paires de chromosomes, puis choisit parmi ces paires le vainqueur du tournoi qui a la meilleure valeur fitness.
- **Sélection uniforme** : la sélection se fait aléatoirement, uniformément et sans intervention de la valeur fitness.

B. Opérateur de croisement

Cet opérateur permet d'effectuer une combinaison des gènes des parents sélectionnés. C'est le *principe d'hérédité*. Nous présentons deux types de croisement classiques (**Figure 38**) [130]:

- **Croisement 1 point** : consiste à fusionner les gènes de deux parents envisagés à partir d'un point de coupure choisi au hasard, afin de générer deux enfants. Ce type de croisement est le plus simple et le plus utilisé.
- **Croisement 2 points** : est une généralisation du croisement 1 point qui repose sur le même principe avec deux points de coupure.

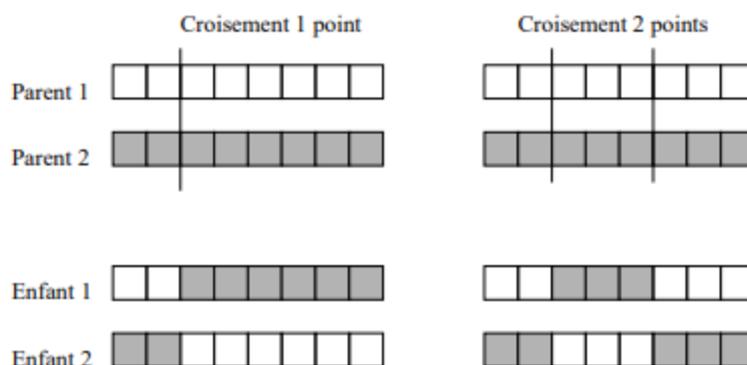


Figure 38. Exemple de types de croisement [130]

C. Opérateur de mutation

Cet opérateur permet de faire apparaître de nouveaux chromosomes. La mutation consiste généralement à tirer aléatoirement un gène dans le chromosome et à le remplacer par une valeur aléatoire (**Figure 39**) [131]. Elle intervient en tant que bruit et prévient l'évolution de se figer. C'est le *principe de variation*.



Figure 39. Exemple d'une mutation binaire[131]

2.3.4 Discussion

Pour les problèmes d'optimisation, les algorithmes génétiques semblent être plus utiles à notre problème car ils sont capables de déterminer plusieurs solutions optimales en préservant une certaine diversité entre eux par les opérateurs de variation. Ils permettent également de résoudre des problèmes d'optimisation complexes lorsqu'il est difficile ou impossible d'obtenir une modélisation mathématique du système. Le principe des AG est de faire évoluer une population dans le but d'améliorer les individus à l'aide des trois opérateurs principaux : la sélection, le croisement et la mutation, qui permettent d'éviter de converger dans un optimum local. Cependant, l'efficacité de ces algorithmes est très sensible au choix des paramètres comme la taille de la population, le nombre maximal des générations, l'initialisation de la population. En effet, ces paramètres conditionnent fortement la rapidité et/ou la qualité de la solution obtenue par l'algorithme.

L'optimisation d'un système de tailles présente les caractéristiques d'un problème multi-objectif : réduction du nombre de tailles, amélioration du bien aller, maximisation du taux de couverture. Le recours au principe de dominance de Pareto permettra d'orienter le choix final pour l'obtention du système le plus adapté.

2.4 Conclusion

Dans ce chapitre, nous avons passé en revue les différentes méthodes existantes pour la sélection, la normalisation des mesures du corps humain ainsi que les méthodes de classification qui se distinguent en deux catégories : supervisées et non-supervisées. Nous avons présenté l'arbre de décision pour la classification supervisée. Pour la classification non-supervisée qui est basée sur les mesures d'éloignement entre les individus. Nous avons parlé des méthodes hiérarchiques et de partitionnements. Nous avons décrit la méthode k-moyennes avec les différents indices existants dans la littérature pour évaluer le nombre

optimal de classes. Nous avons également expliqué la carte auto-organisatrice qui forme une classe de réseau de neurones artificiels fondée sur une méthode d'apprentissage non-supervisée en détaillant l'algorithme et le choix des paramètres d'apprentissage. Enfin, nous avons évoqué la problématique d'optimisation, nous avons expliqué les métaheuristiques à base de solution unique, puis les métaheuristiques à population de solutions dont le recours à des algorithmes génétiques. Tous ces fondements théoriques relatifs aux méthodes que nous avons sélectionnées parmi d'autres méthodes choisies ont permis de mettre en place une méthodologie spécifique à notre problématique décrite dans le chapitre suivant.

Chapitre 3 Contributions méthodologiques

3.1 Introduction

Ce chapitre présente la méthodologie générale proposée pour l'optimisation des systèmes de tailles par classe morphologique. Le travail au niveau des classes morphologiques permet de développer des systèmes de tailles performants et d'assurer le bien-être des vêtements pour chaque individu de la population considérée. Ainsi, notre méthodologie est composée en deux étapes. La première étape porte sur la détection des morphologies représentatives d'une population donnée à travers une classification non-supervisée. Cette classification est précédée par une phase de sélection de mensurations clés suivie d'une normalisation. La deuxième étape porte sur le développement de méthodes d'optimisation des systèmes de tailles par morphologie. Trois modèles de construction de systèmes de tailles sont développés, chacun proposant des degrés d'intégration du client et de facilité de mise en œuvre industrielle différents. Ce chapitre est organisé comme suit : la première section présente la méthodologie proposée pour la classification des morphologies à partir de mensurations. La deuxième section décrit les trois modèles de construction des systèmes de tailles et plus précisément les fonctions objectifs et les algorithmes d'optimisation employés. Une conclusion, incluant un schéma global de la méthodologie, est proposée dans la dernière section de ce chapitre.

3.2 Classification des morphologies

La première étape de notre méthodologie consiste à partitionner l'ensemble d'une population décrite par les mensurations de ses individus en classes de telle sorte que les individus d'une même classe aient des morphologies les plus proches possibles et que les individus de deux classes différentes aient des morphologies très distinctes. Ce partitionnement requiert une sélection des mensurations pertinentes, puis une normalisation spécifique avant de pouvoir les exploiter dans un algorithme de classification non-supervisée (**Figure 40**).

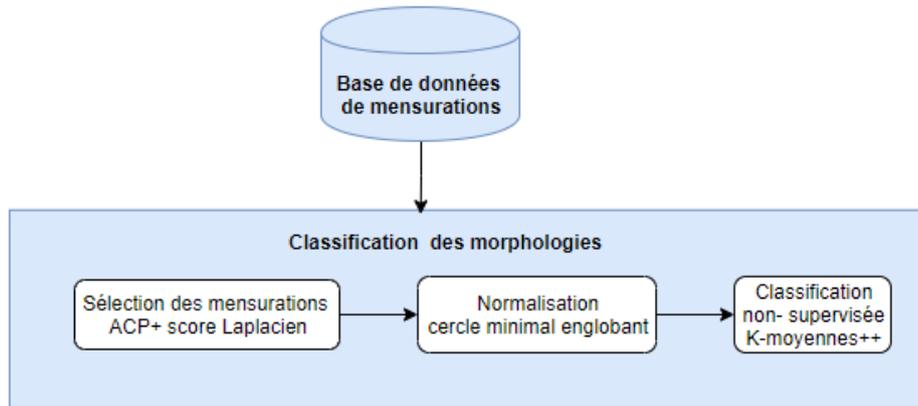


Figure 40. Processus de classification non-supervisée de morphologies.

3.2.1 Sélection des mensurations

La sélection des mensurations est importante car elle a pour objectif de définir les attributs de chaque individu de manière pertinente pour identifier leur morphologie. Dans la littérature, deux méthodes sont souvent utilisées pour évaluer la pertinence d'un sous-ensemble de variables sélectionnées : l'ACP et le calcul des scores Laplacien (**section 2.2.1**).

Nous avons choisi de tester ces deux méthodes sous condition qu'elles soient utilisées en adéquation avec les connaissances du domaine d'application. En effet, les experts du prêt-à-porter ont l'habitude de classer les différents types de morphologies en classes distinctes et symbolisées par des chiffres ou des lettres décrivant le torse (**Figure 41**). Ainsi, les mensurations considérées dans notre sélection concernent uniquement les mesures relatives au torse. Les mensurations sélectionnées seront ensuite normalisées puis classées par un algorithme de classification non-supervisée.

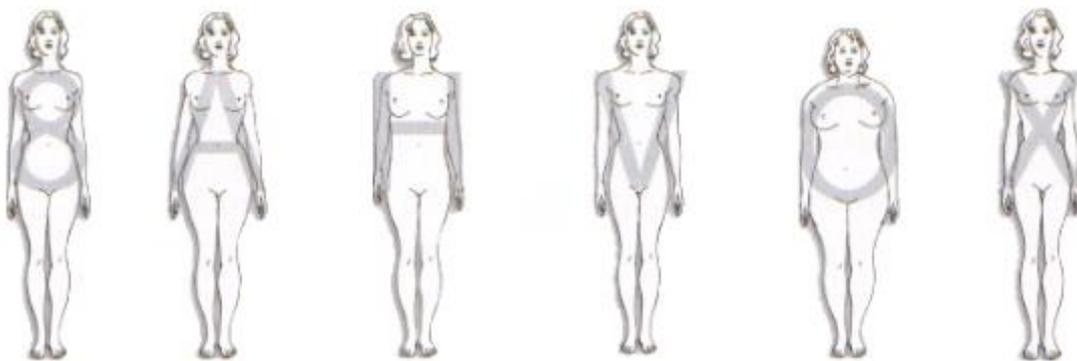


Figure 41. Exemple de différents types de morphologies [132].

3.2.2 Normalisation des mensurations

Afin d'obtenir une classification pertinente des morphologies à partir des mensurations, il est nécessaire d'effectuer une étape préalable de normalisation des données. Cette normalisation est destinée à rendre les individus comparables entre eux et d'assurer la préservation des proportions du corps humain de chaque individu. En effet, une classification directe sur les mensurations pourrait privilégier une dimension par rapport à une autre, c'est à dire conduire à définir des groupes d'individus selon leur stature et non selon leurs proportions (ratio stature/volume), comme illustré en **Figure 42**.

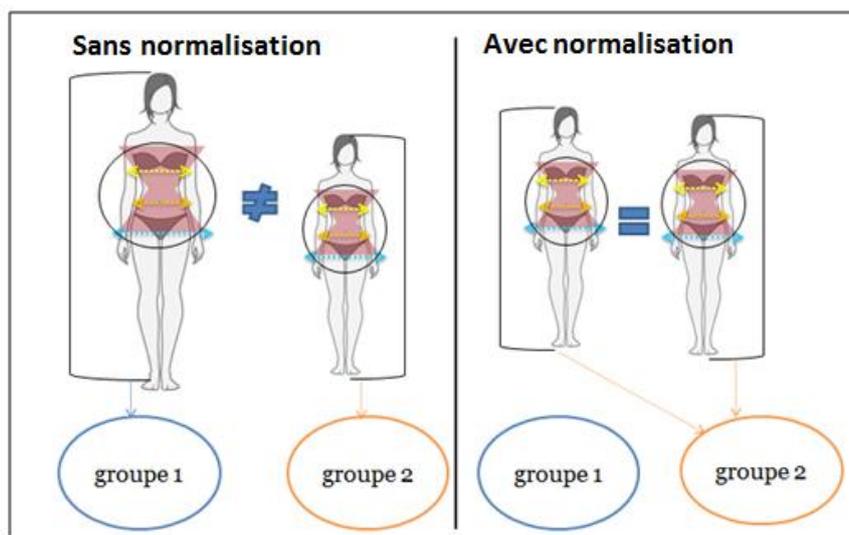


Figure 42. Exemple de deux morphologies similaires normalisées par le cercle minimum englobant.

Ainsi, pour éviter ce problème, une normalisation des mensurations est réalisée avec la méthode du cercle englobant (**section 2.2.2.4**).

La mise en œuvre de cette méthode nécessite de placer l'ensemble des points de mesures caractérisant la morphologie du torse dans un espace à 2 dimensions. Les points choisis pour représenter le torse doivent être représentatifs des deux dimensions verticales et horizontales afin de conserver les rapports anthropométriques de chaque individu entre ces deux dimensions. Ainsi, la dimension verticale est représentée par deux points anthropométriques : l'un situé sur la 7^{ème} vertèbre cervicale, le second situé au point le plus bas de l'entrejambe. La deuxième dimension est représentée par d'autres points obtenus grâce aux mensurations suivantes : largeur d'épaules, tour de poitrine, tour de tailles et tour de hanches. Pour la largeur d'épaule, nous prenons les points extrêmes situés sur les acromions. Pour les contours morphologiques, nous prenons la demi-mesure de chaque contour pour définir les extrémités

des points de ces mensurations. Bien entendu, cet exemple oriente déjà un choix possible des variables qu'il sera nécessaire de valider dans le chapitre 4.

Le problème de la détection du cercle minimum englobant pour un ensemble de points consiste à trouver le cercle le plus petit contenant ces points. L'algorithme appliqué à un individu suit la procédure suivante (**Figure 43**) :

- Positionner les deux points extrêmes du torse relatif à la dimension verticale tels que le point de l'entrejambe se trouve à l'origine (0,0) du repère et le point de la nuque soit placé verticalement à une distance D ($D =$ mensuration entre ces deux points). Cette étape de transformation prépare l'alignement des données dans la dimension verticale sur l'axe Y du repère (invariance des données en translation et rotation).
- Placer les deux points extrêmes de chaque mensuration relative à la dimension horizontale dans le repère prédéfini. Ces points sont alignés deux à deux pour chaque mensuration suivant l'axe X. La valeur des coordonnées en $\pm X$ d'une mensuration est fixée par la moitié de la demi-mesure (pour les contours morphologiques) ou la mesure (distance entre deux points). La valeur des coordonnées en Y d'une mensuration est fixée par l'écart entre cette mensuration et le point de l'entrejambe (D_1, D_2, D_3, D_4).
- Chercher le centre et le rayon r du cercle minimum englobant en utilisant l'algorithme de Gärtner (**section 2.2.2.4**).

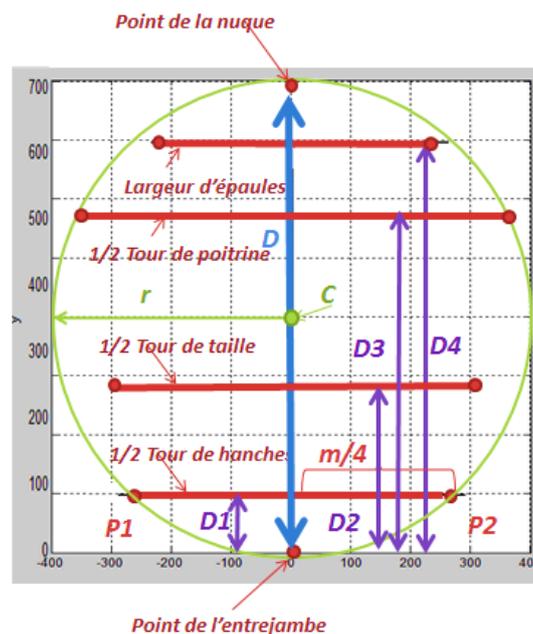


Figure 43. Illustration d'algorithme de cercle minimal englobant.

Chaque morphologie est alors représentée par son propre cercle englobant donnant une valeur spécifique au rayon. Ce rayon constitue le ratio de la normalisation des mensurations de chaque individu (invariance des données au facteur d'échelle).

3.2.3 Algorithme de classification non-supervisée

Les critères de choix de l'algorithme de classification non-supervisée ont été discuté dans (**section 2.2.3.3**). Pour la classification des morphologies, nous avons sélectionné l'algorithme des k-moyennes++. Une des difficultés de mise en œuvre de cet algorithme réside dans le choix du nombre de partitions k. Par conséquent, nous exécutons l'algorithme avec différents k afin de choisir la meilleure des partitions obtenues au sens de l'inertie intra-classe et l'inertie inter-classes. Cette évaluation est faite par deux indices d'évaluation des partitions : l'indice de Davies-Bouldin et l'indice silhouette décrits dans (**section B.1.1**).

A l'issue de la classification non-supervisée appliquée sur les mesures normalisées du corps, les centres de chacune des classes sont calculés par la moyenne arithmétique de toutes les mesures appartenant à la classe. Ainsi, pour déterminer la morphologie associée à un centre de classe, nous sélectionnons la personne ayant les mesures du corps la plus proche de ce centre de classe. La section suivante présente la deuxième étape de la méthodologie proposée : l'optimisation des systèmes de tailles par morphologie.

3.3 Optimisation des systèmes de tailles par morphologies

Dans la **section 1.3.3** du chapitre 1, nous avons présenté les différents concepts de création d'un système de tailles de vêtements à partir de la littérature. La première étape consiste à développer le système sous la forme d'une grille de tailles dont chaque case représente une taille potentielle de vêtement comme illustrée par **Figure 44** . Le système de tailles est ensuite construit à partir de l'évaluation de la perte totale et du taux de couverture (**section 1.3.3**).

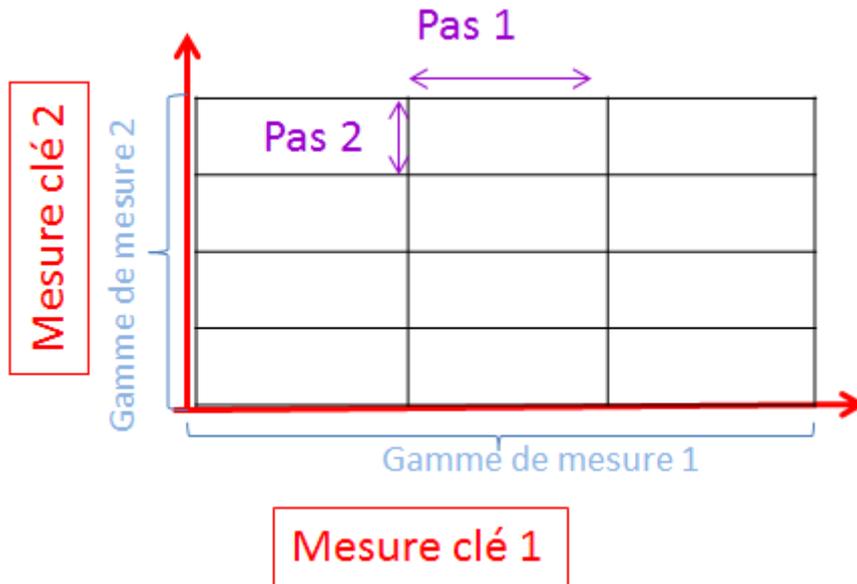


Figure 44. Illustration de grille de tailles.

Le système obtenu dépend fortement de la valeur des deux pas fixés pour les deux mesures clés considérées. Si la valeur du pas est trop petite, le concept génère trop de tailles qui seront économiquement désavantageuses. Inversement, une valeur grande génère peu de tailles avec pour inconvénient une insatisfaction en matière d'ajustement du vêtement. Toutefois, cette technique permet une mise en œuvre industrielle du vêtement plus aisée notamment en assurant une gradation linéaire des tailles.

Les trois modèles que nous proposons pour l'élaboration d'un système de tailles s'inspirent de cette technique très courante dans la littérature. Afin, de palier aux inconvénients décrits précédemment, nous avons apporté diverses évolutions, implémentées différemment dans trois modèles. Le premier modèle est basé sur un critère de performance spécifique pour optimiser la construction du système de tailles par un choix approprié des pas de la grille. Les deux autres modèles intègrent deux évolutions orientées plus fortement sur l'analyse des données, notamment des données clients, mais plus disruptives d'un point de vue métier.

3.3.1 Modèle GBSS: Grid Based Sizing System

Le modèle GBSS, Grid Based Sizing System, s'inspire fortement de la méthodologie de construction des systèmes de tailles rencontrés dans la littérature. Cette technique, décrite dans la **section 3.3**, permet de conserver un processus métier classique avec une gradation linéaire. Cependant, le problème d'optimisation du système de tailles nécessite de définir une fonction objectif pertinente. Généralement, dans la littérature, un système de tailles est évalué par le pourcentage de couverture de la population et la perte globale quantifiant le bien-être

(section 1.3.3). Les méthodes d'optimisation proposées consistent alors à améliorer un des deux critères en fixant l'autre critère à une valeur cible. Afin d'obtenir une optimisation plus globale du système de tailles, nous avons choisi de définir un nouveau critère de performance, appelé FL (Fit Level), pour évaluer un système de tailles par rapport à une population donnée. L'objectif est d'obtenir un seul et unique critère pour quantifier à la fois le bien-être et le taux de couverture. Le principe consiste à classer les individus de la population dans trois catégories (**Figure 45**) :

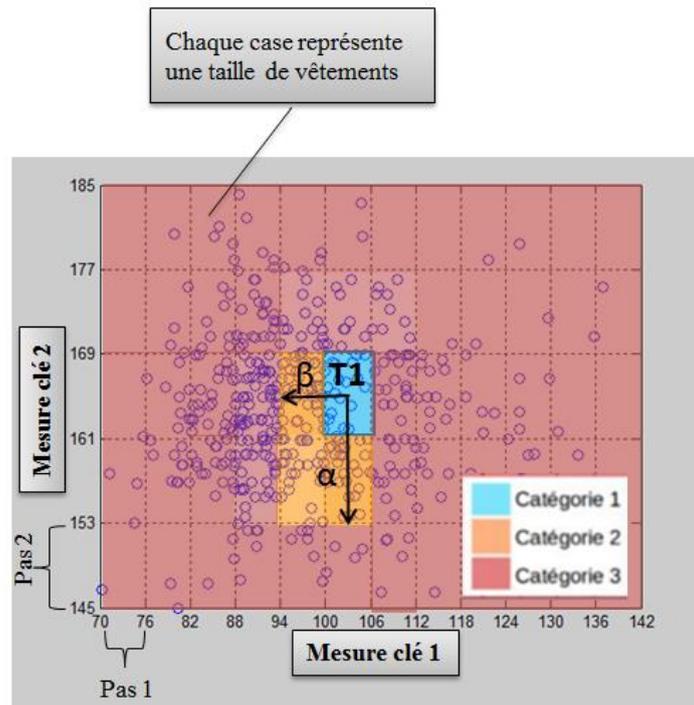


Figure 45. Catégories pour l'évaluation du système de tailles.

- **Catégorie 1** les individus de cette catégorie correspondent aux clients parfaitement couverts par le système de tailles. Pour cette catégorie, le critère de performance $FL_1(c)$ pour un client c affecté à la taille s est :

$$FL_1(c) = dist_m(s, c) \quad (1)$$

Avec :

- $dist_m$ la distance de Manhattan. La distance de Manhattan est privilégiée pour notre problème pour les raisons indiquées en chapitre 2, **section 2.2.3.2.1**.
- **Catégorie 2** les clients de cette catégorie pourront s'habiller dans une taille proposée par le système, mais avec des vêtements choisis plus grands. Par conséquent, le bien-être est dégradé dans les limites de tolérances fixées, nommées α et β sur la **Figure 45**, pour chacune

des mensurations. Ces tolérances doivent être fixées par une expertise métier. Pour cette catégorie, le critère de performance $FL_2(c)$ pour un client c affecté à la taille s est :

$$FL_2(c) = dist_m(s, c) \quad (2)$$

Avec :

- s la taille la plus proche des mensurations du client selon la distance de Manhattan et dans les limites des tolérances.
- **Catégorie 3** les clients non couverts par le système de taille, c'est-à-dire avec des mensurations trop grandes pour les tailles proposées ou au-delà des tolérances α et β , sont affectés à la catégorie 3. Le critère $FL_3(c)$ pour un client c doit donc être plus pénalisant pour le système. Cette pénalité est calculée en fonction des tolérances fixées :

$$FL_3(c) = \alpha + \beta \quad (3)$$

Finalement, le critère de performance global du système de tailles par rapport à une population donnée est :

$$FL = \sum_{c \in cat1} FL_1(c) + \sum_{c \in cat2} FL_2(c) + \sum_{c \in cat3} FL_3(c) \quad (4)$$

Le modèle GBSS est développé à partir d'un algorithme d'optimisation incrémental qui s'appuie sur cet indicateur (**Figure 46**). Le principe est de partir d'une grille de tailles définie avec deux pas (p, q) fixés couvrant l'ensemble de la population donnée et de chercher la taille générant le minimum de FL. Elle constitue le système de tailles initial. Ensuite, il suffit d'ajouter incrémentalement les tailles parmi les voisins des tailles existantes dans le système en fonction de la variation de la valeur du critère. L'incrémental sur les tailles voisines permet de conserver un système de tailles compact et d'éviter les « trous » entre deux tailles. Ainsi, les contraintes métiers, notamment en termes de gradation, sont garanties.

Cet algorithme permet de générer, pour chaque combinaison de pas (p, q) , un système de tailles pour différent nombre de tailles (n) . Enfin, la simulation des différentes combinaisons de pas (p, q) permet de générer et d'évaluer un ensemble de solutions optimums représentant chacune des combinaisons (p, q, n) .

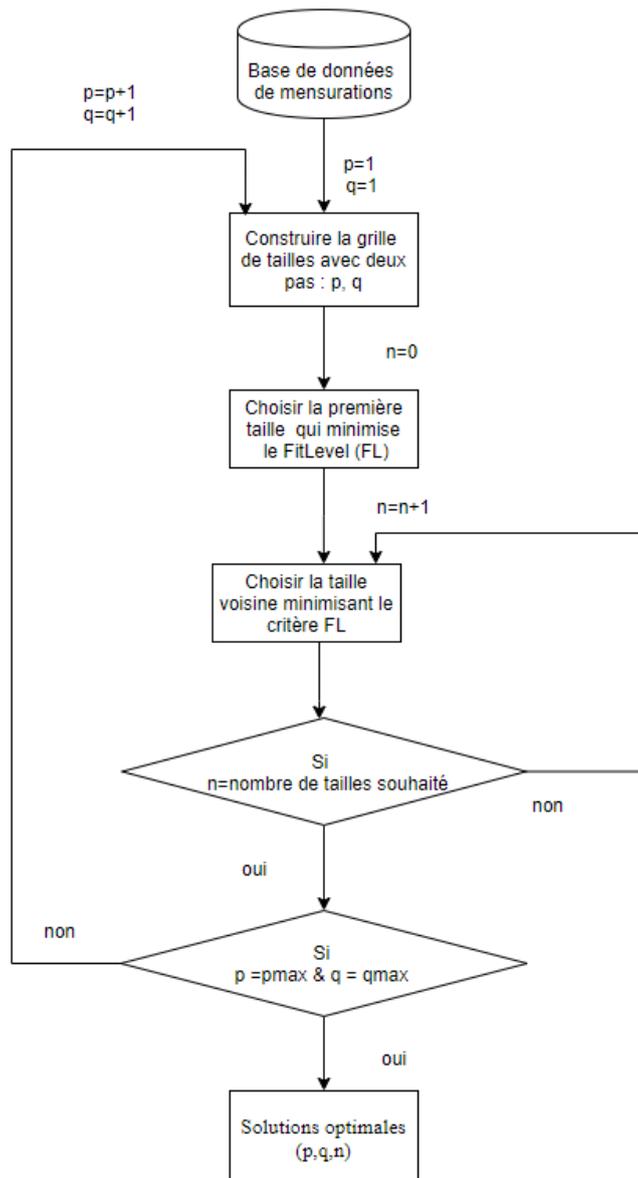


Figure 46. Processus de construction d'un système de tailles optimisé du modèle GBSS

Il est alors possible de représenter cet ensemble de systèmes de tailles optimums en fonction du critère nombre de tailles n et du critère FL. Le choix du système de tailles optimal parmi cet ensemble devient un problème d'optimisation multi-critère avec ces deux critères conflictuels. Ainsi, la représentation de l'ensemble des solutions dans le plan (n, FL) permet de définir un front de Pareto. Pour chaque nombre de tailles n , les pas (p, q) optimums de la grille sont donnés par le front de Pareto, qui constituent les meilleurs compromis entre les deux critères (**Figure 47**).

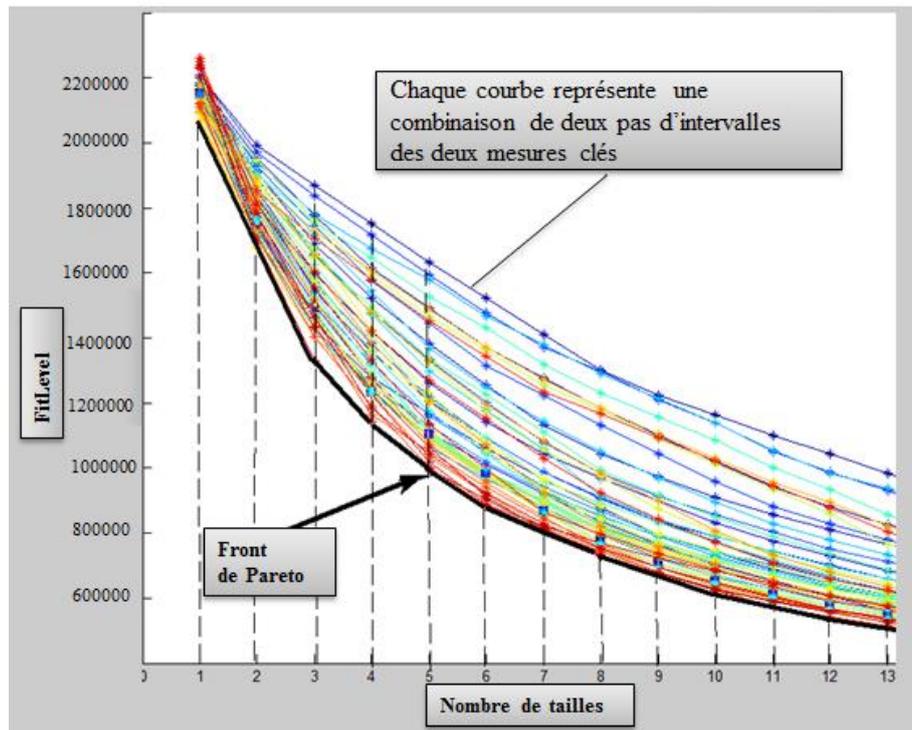


Figure 47. Illustration d'un front de Pareto obtenu avec un modèle GBSS

Cette méthode permet ainsi d'obtenir une vision globale des systèmes de tailles optimums pour une population donnée. Cela constitue un outil d'aide à la décision intéressant afin de définir le nombre de tailles et les pas de la grille. Cette méthode permet également de conserver une gradation linéaire et une mise en œuvre directe dans un processus classique de conception de vêtements. Cependant, plusieurs points d'amélioration peuvent être soulevés.

Le premier problème du modèle GBSS est dû aux valeurs fixes des tolérances α et β qui représentent l'aisance acceptable du vêtement. Les valeurs de ces tolérances sont assez difficiles à définir car très subjectives et fluctuantes en fonction du consommateur. Nous pouvons également facilement imaginer que les tolérances seront différentes sur les petites tailles et les grandes tailles (vêtements plus amples ou bien trop serrés).

Le deuxième problème est la façon de calculer le critère FL qui est basé sur la somme des distances Manhattan entre la taille de vêtement et les mesures du client. Ce calcul suppose que les mesures ont une influence similaire sur le bien aller, alors que certaines mensurations peuvent être plus ou moins privilégiées par le client. Par exemple, les clients peuvent accorder plus d'importance au tour de poitrine que la longueur de bras lors de l'achat d'un chemisier ou chemise.

Un autre problème de la méthode proposée réside dans l'algorithme incrémental. En effet, cette approche peut converger vers des optimums locaux, notamment à cause de

l'initialisation (système à une taille) et de la contrainte de voisinage. Enfin, la gradation linéaire peut également constituer une limite dans l'optimisation du critère FL.

Afin de remédier à ces différents problèmes, nous proposons différentes évolutions de notre méthode :

- Pour les problématiques liées aux tolérances et sensibilités des consommateurs, il est important d'intégrer les données issues d'études clients dans la construction du modèle. Ainsi, nous nous sommes appuyés sur un sondage de satisfaction client sur les tailles des vêtements d'une enseigne de prêt à porter féminin pour compléter notre modèle.
- Concernant les limites de notre algorithme, nous nous sommes orientés vers des techniques de type méta-heuristiques pour assurer une meilleure optimisation du système de tailles sur les données de la population ciblée.

Ces évolutions sont décrites dans les deux modèles présentés dans les sections suivantes :

- Le modèle BBSS (Box Based Sizing System), basé sur le modèle GBSS, intègre les tolérances exprimées par les clients pour porter la taille de vêtement concernée, ce qui permet de converser des contraintes "métier" telles que la définition précise de la couverture des tailles.
- Le modèle FFSS (Free Form Sizing System), basé sur la distribution statistique des tolérances, permet une meilleure prise en compte de la sensibilité des clients par rapport aux tailles mais il est plus disruptif par rapport à la méthodologie usuelle dans l'habillement.

Ces deux modèles utilisent également un algorithme génétique spécifique pour la construction du système de tailles.

3.3.2 Modèle BBSS: Box Based Sizing System

L'objectif de ce modèle est de remplacer le concept de pas de grille du modèle précédent par des paramètres traduisant la sensibilité des clients. Pour ce faire, il est nécessaire d'estimer les tolérances des clients en fonction des mensurations, tailles et du type de vêtement. Souvent, pour refléter ce type de données, il est nécessaire de s'appuyer sur des informations recueillies au moyen d'enquête dont l'objectif est d'investiguer le niveau de satisfaction des clients sur les tailles de vêtements existantes d'une enseigne. La section suivante décrit brièvement l'exploitation des données issues d'une telle enquête menée spécifiquement pour notre problématique.

3.3.2.1 Génération des tolérances clients

3.3.2.1.1 Enquête client

Afin d'obtenir les informations relatives aux tolérances des clients, nous nous appuyons sur une enquête menée en ligne par une enseigne de prêt à porter féminin en 2019.

Les questions relatives aux tailles de vêtements, intégrées dans l'enquête, permettront d'identifier les attentes des clients en termes de bien aller d'articles (exemples de questions sont illustrés en **Annexe 2**). A partir des réponses des clients interrogés, nous avons cherché à établir un lien entre les mensurations réelles des clients et les mensurations décrivant les tailles. Certaines mensurations n'ont pas pu être obtenues en raison soit de la tendance des clients à se sentir inconfortables à les reporter, soit des contraintes liées à la difficulté de les mesurer, soit du changement de leur valeur d'une enseigne à une autre. Le tour de poitrine en est un exemple. Ainsi, il est nécessaire de prédire ces mensurations à partir des informations déclarées dans l'enquête telles que l'âge, le poids et la stature.

3.3.2.1.2 Prédiction des mensurations clients

L'objectif est de développer un modèle prédictif pour estimer les valeurs des mensurations manquantes des clients qui ont répondu à l'enquête à partir des attributs disponibles : poids, âge, stature (**Figure 48**). La base de données de mensurations utilisée précédemment dans la classification morphologique (**Figure 40**) servira pour l'apprentissage des modèles prédictifs. Cette prédiction sera d'autant plus performante si elle est effectuée par type de morphologie. Ainsi, une nouvelle classification non supervisée des données de cette base est effectuée en amont. L'objectif de cette classification diffère de la classification morphologique décrite en **section 3.2**. En effet, son objectif est d'améliorer la prédiction des mensurations à partir de la classification. Ainsi, cette classification intègre les trois paramètres "âge, poids, stature" et le nombre optimal de classes est défini en fonction des performances du modèle de prédiction. Le modèle prédictif choisi est une régression polynomiale car elle permet de trouver une fonction adéquate pour ces paramètres non linéaires. Le nombre optimal de classes est celui qui minimise l'erreur moyenne de la prédiction M_i . Cette erreur est estimée par validation croisée (**section 2.2.3.1**). Par la suite, la prédiction des mensurations des clients participant à l'enquête sera faite à partir du meilleur modèle avec un nombre de classes optimales. Pour assigner les clients issus de l'enquête à la classe la plus proche, et par conséquent au bon modèle prédictif, à partir des attributs disponibles dans l'enquête (stature, poids, âge), nous avons introduit un arbre de décision (**section 2.2.3.1**). Cet arbre de décision permet par une

classification supervisée d'affecter ces nouveaux clients à l'une des classes obtenues à partir d'un sous ensemble d'attributs utilisés par l'algorithme de partitionnement. Cet arbre de décision est créé à partir des données d'apprentissage suite à la génération des classes et s'appuie sur le critère de Gini et l'algorithme CART proposé par Breiman et al. [95].

A l'issue de cette prédiction des mensurations, des statistiques sur le nombre de personnes satisfaites pour porter une taille de vêtement permettent de calculer les écarts-types représentant les clients satisfaits par les mensurations des tailles de l'enseigne. Ces écarts-types sont ensuite utilisés pour estimer les valeurs de tolérances.

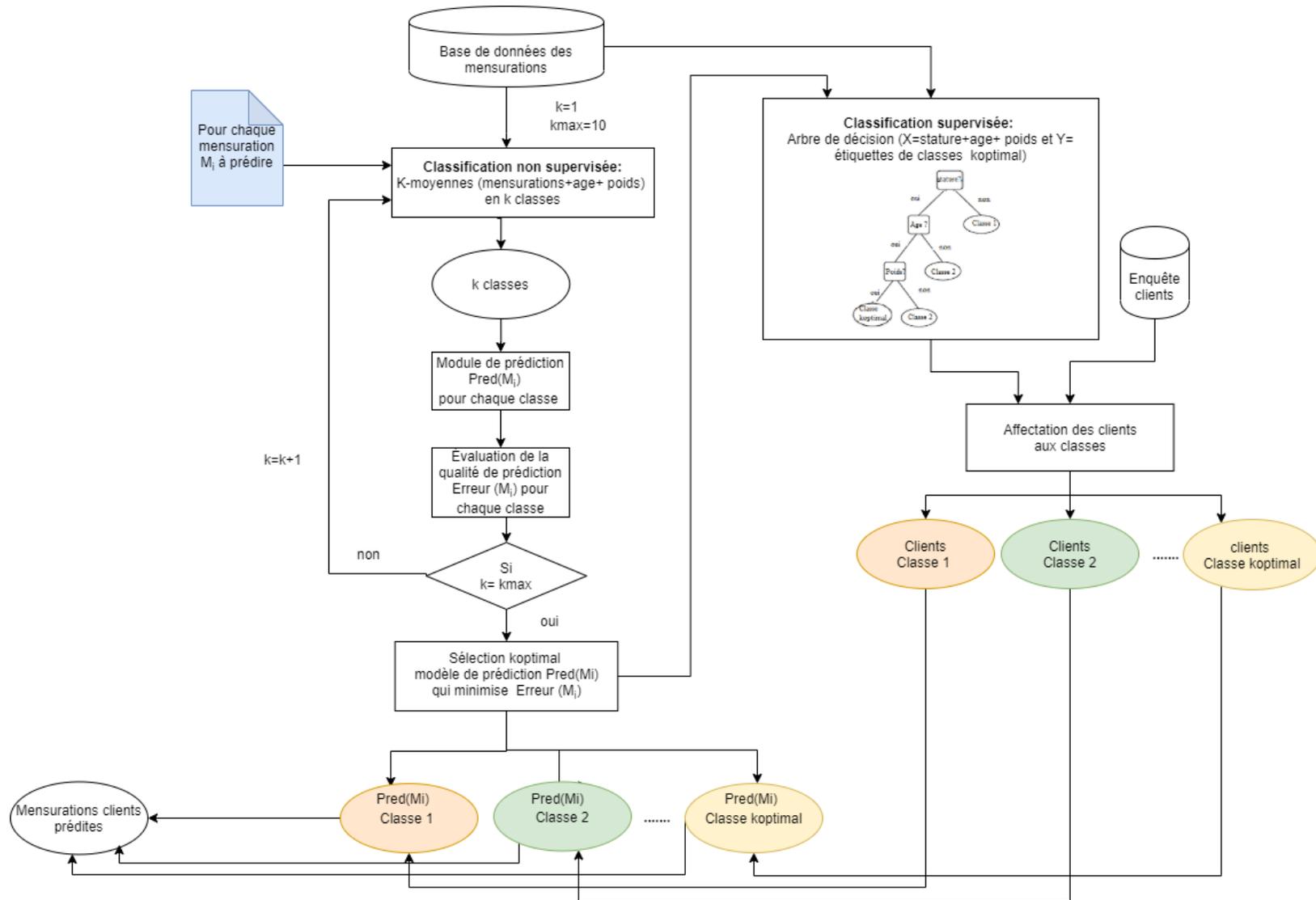


Figure 48. Processus de prédiction des mensurations

3.3.2.1.3 Calcul des tolérances des clients

La modélisation des tolérances des clients est faite par le calcul de deux écarts-types pour chaque mensuration clé de taille de vêtement de l'enseigne (**Figure 49**). La notion d'écart type est couramment utilisée dans le problème de classification des données de la population (**section 1.2.3.1**). Ainsi, nous considérons ici que les clients aux mensurations inférieures ou supérieures à une taille donnée et qui sont au-delà d'une tolérance basée sur les écarts types issus de l'enquête. Ils ne sont donc plus couverts par la taille. Pour chaque mensuration de chaque taille, un premier écart-type, noté σ_l , présente l'étalement à gauche des individus ayant des mensurations plus petites que la taille qu'ils choisissent et qui leur donne satisfaction. Selon le même principe, un deuxième écart-type, noté σ_r , présente l'étalement à droite des individus ayant des mensurations plus grandes que la taille qu'ils choisissent et qui leur donne satisfaction. Ce choix de deux écarts types, dû à la distribution asymétrique, permet de traduire des tolérances clients différentes pour des vêtements un peu trop grands ou un peu trop petits. Ces écarts types représentent plus fidèlement les sensibilités des clients que les tolérances définies précédemment dans le modèle GBSS par les experts du métier.

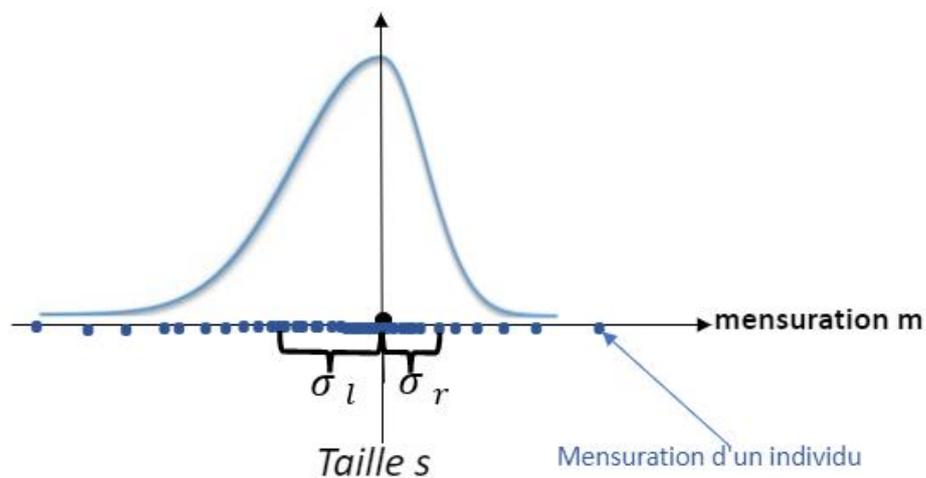


Figure 49. Illustration du calcul de deux écarts-types pour taille de vêtement

Les écarts types σ_l et σ_r illustrent uniquement les tolérances pour les tailles des vêtements de l'enseigne qui a effectué l'enquête. En effet, les questions portaient sur les satisfactions clients pour rapport aux vêtements de l'enseigne. Lors de la construction de notre système, les nouvelles tailles générées seront différentes des tailles proposées par l'enseigne. Par conséquent, il est nécessaire d'estimer les tolérances pour ces nouvelles tailles. Ces estimations sont effectuées à l'aide d'un modèle de régression construit à partir des écarts types relatifs aux données de l'enquête. L'analyse des résultats, effectuée en chapitre 4,

montreront qu'une régression linéaire sera plus appropriée notamment à cause du nombre de points limités pour la prédiction. Ce principe, illustré en **Figure 50**, nous permet de gérer les différentes tolérances, ou exigences, des clients en fonction des tailles. En effet, un centimètre d'écart sera perçu différemment sur une taille XS ou XL.

Finalement, cette prédiction des tolérances à partir de l'enquête client permet de différencier les sensibilités des clients :

- En fonction des différents types de mensuration,
- Pour des vêtements plus grands ou plus petits Gauche droite
- Proportionnellement aux tailles

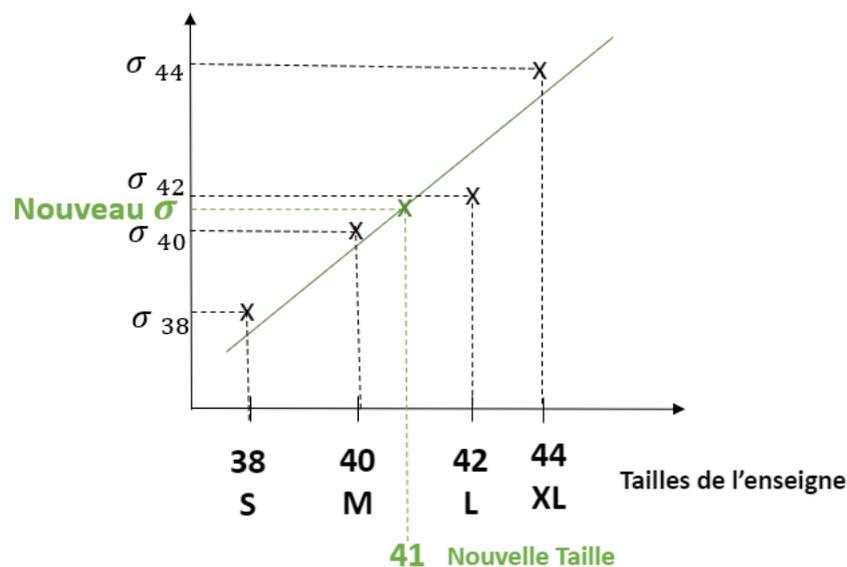


Figure 50. Illustration de la régression linéaire utilisée pour estimer les écarts types pour différentes tailles

3.3.2.2 Mise en place de Modèle BBSS

Le modèle BBSS est une évolution du modèle GBSS décrit dans la **section 3.3.1** intégrant les tolérances estimées par le modèle de régression de tolérances décrit ci-dessus.

Le principe consiste à définir chaque taille, non plus à partir d'une grille pré-établie (modèle GBSS), mais en fonction des écarts types issues des tolérances pour chacune des mensurations. Ainsi, une taille définie pour une mensuration m couvrira les individus possédant de mensuration $m-2\sigma_l$ et $m+2\sigma_r$, avec σ_l et σ_r étant les écarts types calculé précédemment. Le facteur 2 garantit de couvrir 95% de la population dans le cas d'une loi normale standard (**section 1.2.3.1**). Sur un système de tailles en deux dimensions, chaque

taille S_i est alors représentée par une boîte de dimensions variables comme illustré en (**Figure 51**). Ce principe génère un système de tailles non linéaire. Un individu est affecté à la taille la plus proche parmi les tailles du système qui le couvre. Ce modèle peut produire des chevauchements entre deux tailles, c'est-à-dire que certains individus sont couverts par plusieurs tailles. Dans cette situation, les individus sont affectés à la taille la plus proche selon la distance décrite dans la section suivante.

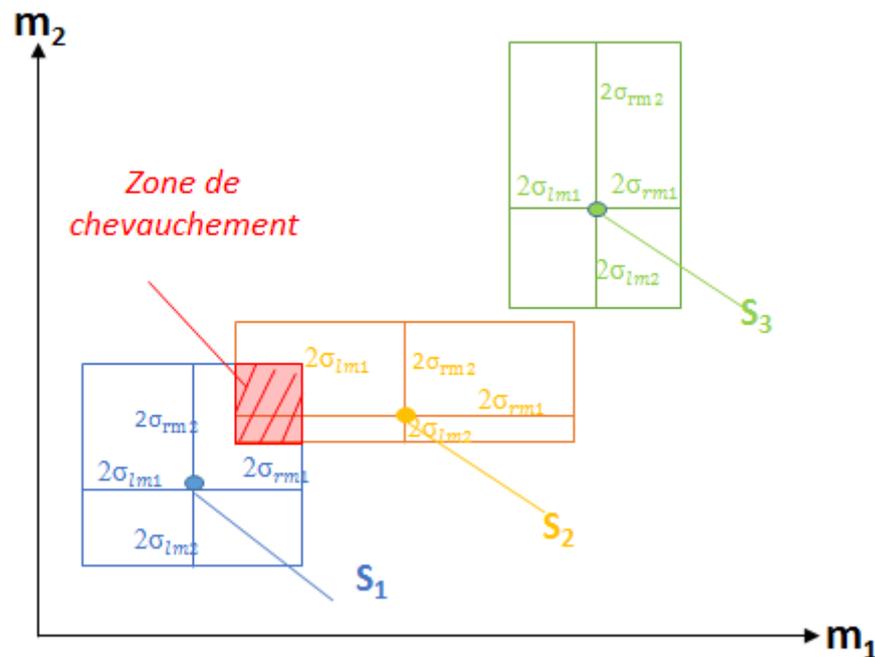


Figure 51. Illustration d'un système de tailles générées par le modèle BBSS.

3.3.2.3 Fonction objectif

A l'instar du modèle précédent, l'optimisation du système de tailles du modèle BBSS doit permettre de proposer, d'une part, des tailles compactes pour assurer le meilleur bien aller, et d'autre part, couvrir au maximum la population. Ainsi, la fonction objectif utilisée pour le modèle BBSS est constituée par deux fonctions appelées L et Cov et décrites ci-dessous.

- 1) La première fonction appelée L vise à évaluer et minimiser la distance entre les clients potentiels et leur taille la plus proche s .

$$L = \sum_{s \in S} \sum_{c \in C_s} dist(s, c) \quad (5)$$

Avec :

- S = l'ensemble des tailles,
- s = une taille de l'ensemble S ,
- C = l'ensemble des clients c ,
- C_s = l'ensemble des clients c affecté à la taille s ,
- $\text{dist}(s,c)$ = distance entre s et c

2) La deuxième fonction appelée Cov vise à évaluer et minimiser le nombre de clients non couverts par les tailles proposées. Ainsi, la formule suivante permet de calculer le nombre de clients qui ne sont pas dans les boites représentant les tailles du système considéré :

$$Cov = Card(C) - \sum_{s \in S} Card(C_s) \quad (6)$$

La fonction objectif associe les deux fonctions L et Cov , tel que :

$$P = L + \gamma Cov \quad (7)$$

Une des problématiques de cette fonction objectif est de pénaliser le système par le nombre d'individus non couverts en évitant de prendre en compte les points aberrants. Ainsi, une des pénalités envisagées consiste à affecter à chaque individu non couvert, la distance maximale affecté à un individu couvert (c'est à dire la distance maximale entre le centre de taille la plus étendue du système et l'extrémité de la boite de cette taille)

Ainsi, γ est donné par la formule suivante :

$$\gamma = \max_{s \in S} (\max_{c \in C_s} (\text{dist}(s, c))) \quad (8)$$

Comme évoqué en fin de section 3.3.1, l'étude client confirme que :

- Toutes les mensurations n'ont pas la même importance pour le client.
- Pour une même mensuration, les tolérances supérieure et inférieure sont différentes,
- Les tolérances diffèrent en fonction des tailles (un même écart n'a pas le même impact pour une taille XS et une taille XL)

Ainsi, nous avons proposé une évolution de la distance de Manhattan utilisée pour la construction du système de tailles pour permettre de mieux prendre en compte les sensibilités des clients. Cette nouvelle distance est pondérée par les écarts types obtenus par le modèle de régression présenté dans la **section 3.3.2.1.3**.

$$dist(s_m, c_m) = \left\{ \begin{array}{l} \frac{1}{\sigma_l(s_m)} \cdot |s_m - c_m|, \text{ si } c_m \leq s_m \\ \frac{1}{\sigma_r(s_m)} \cdot |s_m - c_m|, \text{ sinon} \end{array} \right\} \quad (9)$$

Avec :

- s_m = la mesure de la mensuration m pour la taille s ,
- c_m = la mesure de la mensuration m pour le client c ,
- $\sigma_l(s_m)$ et $\sigma_r(s_m)$ = les écarts-types prédits en utilisant le modèle de régression linéaire pour la mesure de la mensuration m pour la taille s .

Ainsi, la distance entre la taille s et le client c est la somme des distances pour chacune des mensurations m :

$$dist(s, c) = \sum_{m \in S} dist(s_m, c_m) \quad (10)$$

Le modèle BBSS est une évolution intéressante des systèmes de tailles existants qui intègre les tolérances clients. Ce système constitue un compromis intéressant entre :

- les contraintes métiers qui nécessitent notamment des tailles avec des couvertures bien définies (boîtes),
- les sensibilités des consommateurs en intégrant une étude client dans la construction du système,
- l'adaptabilité du système de tailles par rapport à une population cible avec un algorithme basé sur une distance et une fonction objectif spécifique. Cet algorithme de construction du système de tailles par méta-heuristique sera décrit dans la **section 3.3.4**.

A ce stade, il nous semble intéressant d'explorer plus profondément les capacités des algorithmes d'optimisation à proposer un système de tailles encore plus adapté aux mensurations de la population et aux tolérances des clients. Ainsi, nous avons développé le modèle FFSS qui est une approche exclusivement basée sur les données (de mensurations et d'études clients). Cette méthode peut paraître assez disruptive pour le métier de la confection,

notamment par la définition de tailles basées sur des distributions statistiques, mais permet d'obtenir un système totalement orienté vers le client.

3.3.3 Modèle FFSS: Free Form Sizing System

L'analyse des distributions des préférences des clients pour les différentes tailles de vêtements se rapproche d'une loi normale (section 4.3.2.1.2). Nous proposons ainsi un modèle qui utilise la densité de probabilité de la loi normale (ou loi gaussienne) qui est donnée par la fonction φ :

$$\varphi(c) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{c-s}{\sigma}\right)^2} \quad (11)$$

FFSS est un modèle stochastique basé sur un calcul des densités de probabilité de la loi normale avec des écarts types estimés. Pour chaque système de tailles développé avec ce modèle, $2 \times \text{card}(S) \times \text{card}(M)$ lois normales sont paramétrées, avec S l'ensemble des tailles et M l'ensemble des attributs morphologiques (mensurations). En effet, deux lois normales sont définies par taille et par attribut, une pour les clients plus grands qu'une taille s pour un attribut m et une pour les clients plus petits. La moyenne d'une loi normale pour une taille s et un attribut m est la mesure de la taille s pour cet attribut. L'écart type de la loi est prédit selon la méthode décrite en section 3.3.2.1.3.

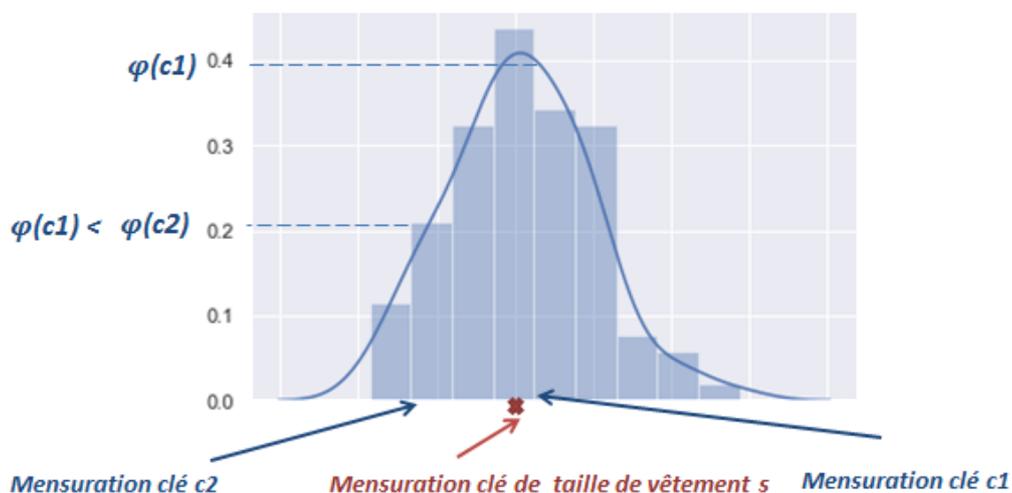


Figure 52. Illustration de la loi normale.

Ce modèle optimise les probabilités que les clients soient satisfaits par leur taille la plus proche. Ce calcul remplace le calcul de distances effectuées pour les modèles précédents.

Le modèle FFSS se démarque des méthodes basées sur les distances par les trois caractéristiques suivantes :

1. Le calcul de la probabilité de lois normales sur plusieurs mensurations se base sur des composantes multiplicatives. En d'autres termes, la probabilité qu'un client c soit satisfait par une taille s est obtenue en multipliant une probabilité par attribut morphologique m (mensuration). Ainsi, cette probabilité est maximisée si toutes les mensurations d'un client se rapprochent des mesures de s , et peut rapidement se dégrader si l'une des mensurations s'éloigne de cette taille. Les méthodes basées sur les distances permettent la compensation entre différents attributs morphologiques, où un client proche d'une taille sur plusieurs attributs morphologiques peut être affecté à cette dernière même s'il est très distant sur l'un des attributs. Ainsi, l'utilisation d'un modèle à composantes multiplicatives permet de mieux capturer le comportement des clients étant donné qu'il est peu probable qu'ils soient satisfaits d'un produit si une de ses mensurations ne lui correspond pas (trop petit ou trop grand).
2. La densité de probabilité d'une loi normale est robuste face aux données aberrantes. Les méthodes qui minimisent des distances nécessitent l'identification de ces données aberrantes qui seront ensuite traitées différemment selon la méthode utilisée (individus considérés comme non couverts dans les modèles GBSS et BBSS). La probabilité de densité d'une donnée aberrante se rapproche de 0 et a ainsi peu d'effet sur la génération du système de tailles.
3. La distribution des clients satisfaits par une taille ressemble à une distribution normale (**Figure 52**). Les distances utilisées dans les modèles GBSS et BBSS sont linéaires, augmentant la pénalité d'un système de taille proportionnellement à la proximité de ses tailles des clients. La distribution normale affecte des probabilités qui augmentent exponentiellement quand les clients sont de 2 à 1 écart type de la taille, ce qui semble correspondre au comportement des clients de l'enquête. Cette distribution donne également des variations de probabilités de satisfaction similaires pour les clients très proches ou très distants de la taille.

3.3.3.1 Fonction objectif

Le modèle FFSS s'appuie sur une fonction objectif Q à maximiser, basée sur la densité de probabilité de la loi normale φ :

$$Q = \sum_{c \in C} \max_{s \in S} (\varphi(c)) = \sum_{c \in C} \max_{s \in S} \left(\prod_{m \in S} \varphi(c_m) \right) \quad (12)$$

Ainsi, en maximisant la somme des probabilités de satisfaction des clients par système de tailles (Q), cette fonction optimise à la fois le nombre de clients pouvant être habillé par le tailleur et le bien aller des clients.

Le modèle FFSS permet ainsi d'associer l'ensemble des individus de la population à chacune des tailles du système selon les densités de probabilités. Par conséquent, le système de tailles généré à une prise en compte plus exhaustive de la population ciblée. En effet, il permet un meilleur compromis entre l'augmentation du nombre de clients couverts et la minimisation des distances entre un tailleur et les clients en se basant sur la probabilité de satisfaction de ces derniers. En contrepartie, la couverture des tailles n'est pas clairement définie. Ce dernier point pourrait compliquer la mise en œuvre de cette approche pour l'industrie de l'habillement traditionnelle.

3.3.4 Algorithme de construction du BBSS et FFSS

Les modèles BBSS et FFSS, avec leurs fonctions objectif respectives, s'appuient sur un algorithme d'optimisation similaire. Cet algorithme a pour objectif d'éviter la convergence vers des optimums locaux en respectant les spécificités et contraintes de la construction d'un système de tailles. Compte tenu également de la complexité du problème, qui s'apparente à un problème de partitionnement couplé à un problème de couverture maximale, notre choix s'est orienté vers les méthodes méta-heuristiques. En effet, il semble peu raisonnable d'essayer de formaliser un tel problème par un modèle mathématique et d'envisager une résolution par des techniques d'optimisation exactes.

Nous détaillons dans cette partie le principe de l'algorithme utilisé. Notre choix s'est porté sur un algorithme génétique, dont les avantages sont présentés dans le chapitre 2, **section 2.3.4**, auquel plusieurs modifications ont été apportées pour répondre à notre problème. Ces modifications portent sur les étapes de génération de la population initiale, d'évaluation de la population, de sélection des individus et de croisement, décrites dans cette section.

Chaque chromosome de l'algorithme génétique décrit un système de tailles pour un nombre de tailles fixé (**Figure 53**). Ainsi, pour un système à s tailles sur m mensurations, les chromosomes possèdent $s \times m$ gènes.

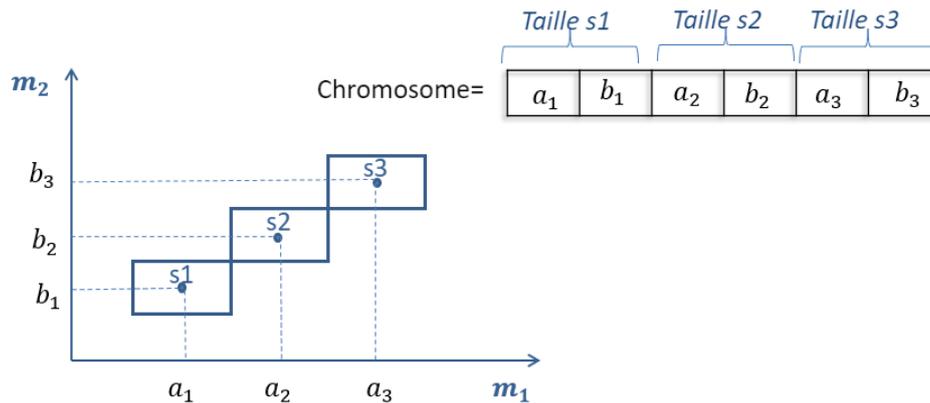


Figure 53. Exemple de codage d'un chromosome.

Nous exposons dans cette partie les principales caractéristiques de l'algorithme génétique (AG) proposé :

- **Génération de la population initiales** : La population initiale est généralement générée aléatoirement. Cette initialisation peut influencer les performances de l'algorithme génétique, notamment le temps de convergence vers l'optimum global. Afin d'améliorer les performances, nous proposons de réaliser une classification par carte auto-organisatrice SOM (**section C**) en amont. En effet, les SOM sont reconnus pour résumer un ensemble d'observations en un nombre réduit de nœuds. Ainsi, nous utilisons un SOM pour réaliser un pré-traitement de l'espace des solutions. Chaque nœud du SOM représente une taille possible pour constituer les chromosomes initiaux. La population initiale est finalement construite à partir d'une sélection aléatoire de chromosomes formés à partir des nœuds du SOM.
- **Evaluation** : Cette phase consiste à calculer la fonction objectif, appelée également fitness, de chaque chromosome. Deux fonctions sont proposées pour les deux modèles BBSS (7) et FFSS (12).
- **Sélection** : Cet opérateur permet de sélectionner les individus (chromosomes) qui constitueront la population suivante. Nous utilisons une sélection par tournoi (**section 2.3.3.1.1A**). Le principe consiste à choisir itérativement un groupe de taille m de la population parents et de sélectionner le meilleur individu (au sens de la fonction fitness) pour la population suivante.
- **Croisement** : Nous utilisons un croisement classique à 1 point de coupure à tirage aléatoire (**section 2.3.3.1.1B**). Cependant, sur notre problème, cette technique peut

conduire à la génération de chromosomes de très faible qualité comme illustré en **Figure 54**.

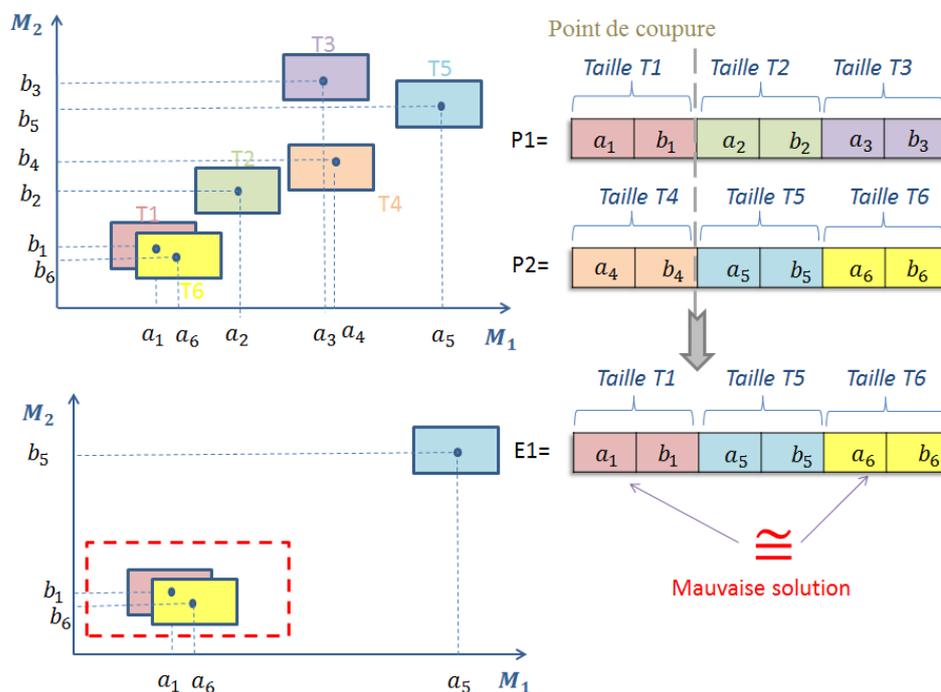


Figure 54. Problème de l'opérateur de croisement classique

En effet, il faut éviter que le croisement produise un système avec des tailles similaires. Ainsi, nous avons introduit une méthode de vérification lors du croisement. L'idée est d'effectuer un test de similarité de chaque gène des parents candidats au croisement avec les gènes qui constituent l'enfant. Si un des gènes candidats est similaire à un des gènes de l'enfant, le gène candidat est rejeté. La méthode commence par définir pour chaque gène d'un parent son plus proche voisin parmi les gènes de l'autre parent. Ensuite, les gènes du premier parent à gauche de la coupure constituent les premiers gènes de l'enfant. Les gènes du deuxième parent sont ajoutés à droite de la coupure en vérifiant qu'ils ne sont pas similaires avec les premiers gènes de l'enfant. Deux gènes de deux parents différents sont considérés s'ils sont les plus proches voisins de l'un et l'autre (**Figure 55**).

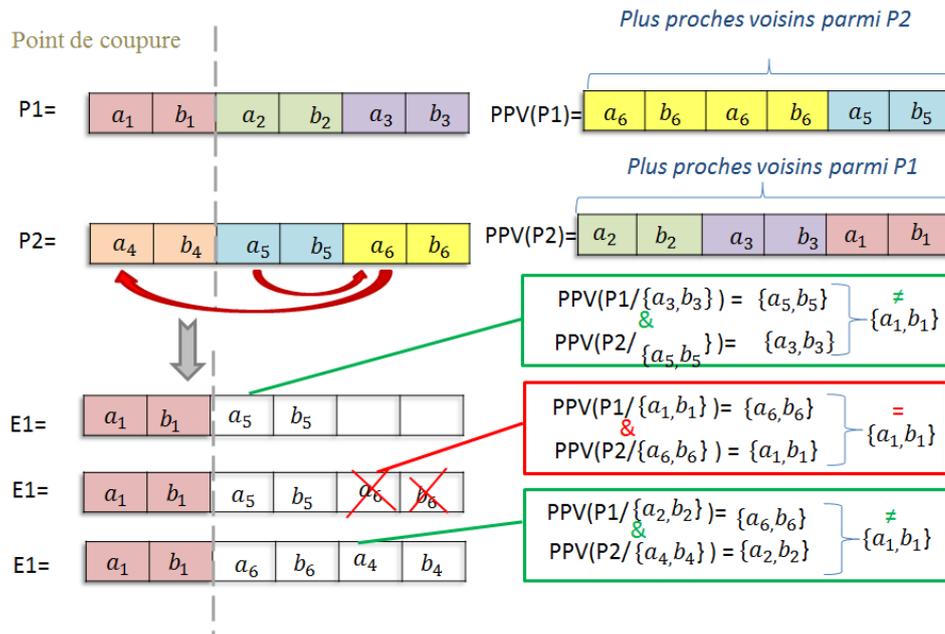


Figure 55. Principe de l'opérateur de croisement proposé

L'algorithme génétique est exécuté pour chaque nombre de tailles. Il est alors possible de représenter l'ensemble des performances des systèmes de tailles obtenus par l'AG en fonction du critère nombre de tailles n (**Figure 56**)

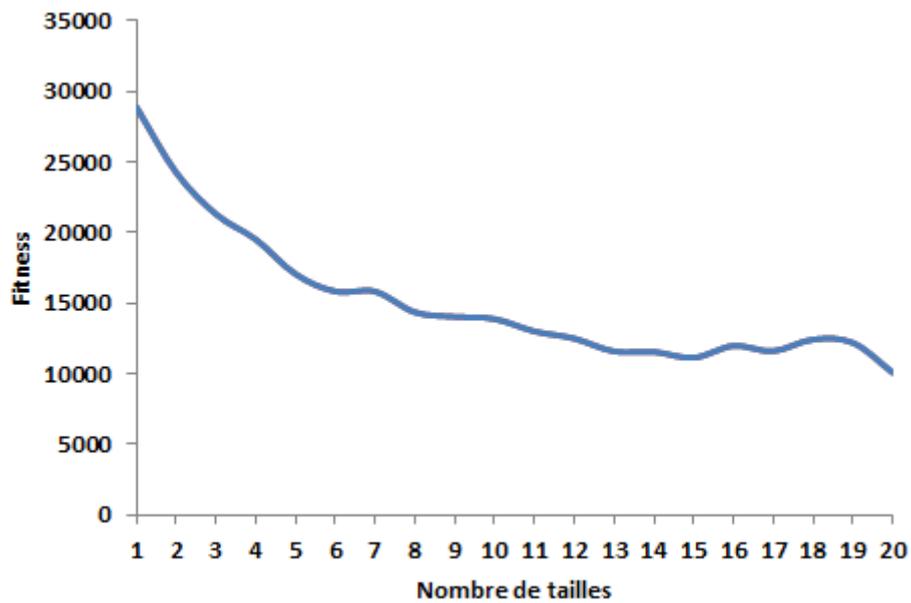


Figure 56. Illustration de la variation de Fitness en fonction nombre de tailles.

3.4 Conclusion

Dans ce chapitre, nous avons présenté la méthodologie proposée pour optimiser les systèmes de tailles à partir d'une classification morphologique. Cette méthodologie décrite par la **Figure 58** et composée de trois parties. Dans un premier temps, nous avons décrit la démarche de la classification non-supervisée des morphologies humaines à partir d'une base de données de mensurations. Cette classification s'appuie sur une sélection des mensurations puis sur la normalisation des mesures à partir de la technique du cercle englobant. Cette normalisation est importante pour garantir la comparaison des proportions du corps et ainsi, mieux considérer les caractéristiques morphologiques des individus lors du processus de classification.

Dans un deuxième temps, nous avons proposé trois modèles pour l'optimisation des systèmes de tailles. Le premier modèle GBSS inspiré de la littérature repose sur la création d'une grille de tailles qui permet d'obtenir un système avec une gradation linéaire. Cette démarche classique a été améliorée en introduisant un nouveau critère de performance. Ce critère permet de définir une fonction objectif. L'optimisation consiste alors à construire le système de tailles par une incrémentation du nombre de tailles en ajoutant un voisin qui minimise cette fonction. En examinant les différentes combinaisons possibles du pas de la grille, nous définissons un front de Pareto qui constitue l'ensemble des systèmes de tailles optimaux. Cependant, la subjectivité des tolérances fixées, la gradation linéaire et la convergence possible du système vers un optimum local restent problématiques. Pour résoudre ces problèmes, deux évolutions ont été proposées.

La première évolution concerne l'intégration de la sensibilité des clients dans la construction du système de tailles, qui constitue la troisième partie. Un sondage de satisfaction client réalisé avec une enseigne de mode a permis de mieux estimer les tolérances des clients par rapport aux tailles des vêtements et également d'améliorer le calcul de la fonction objectif pour l'optimisation du système de tailles. La seconde évolution consiste à utiliser un algorithme génétique pour la construction du système de tailles. Certaines opérations de cet algorithme génétique ont été adaptées pour répondre spécifiquement à notre problématique.

Ces deux évolutions sont implémentées dans deux modèles, appelés BBSS et FFSS, qui diffèrent par leur fonction objectif et le type de système de tailles obtenu. Le modèle BBSS génère un système de tailles non linéaire mais avec une définition précise de la couverture de chaque taille. Le modèle FFSS se concentre sur une optimisation maximale du système de

tailles par rapport aux données clients (mensurations et sensibilités par rapport au bien aller). Ce dernier modèle constitue une évolution majeure de la méthodologie de construction de taillant et des pratiques dans les métiers de la confection. Cela s'intègre parfaitement dans une stratégie d'amélioration industrielle centrée sur le client.

Les apports des trois modèles proposés sont schématisés en **Figure 57**. Comparaison entre les modèles. Le chapitre suivant présentera la mise en œuvre de cette méthodologie sur des données réelles issues d'un ensemble de campagnes de mensurations sur plusieurs pays et d'une enquête client.

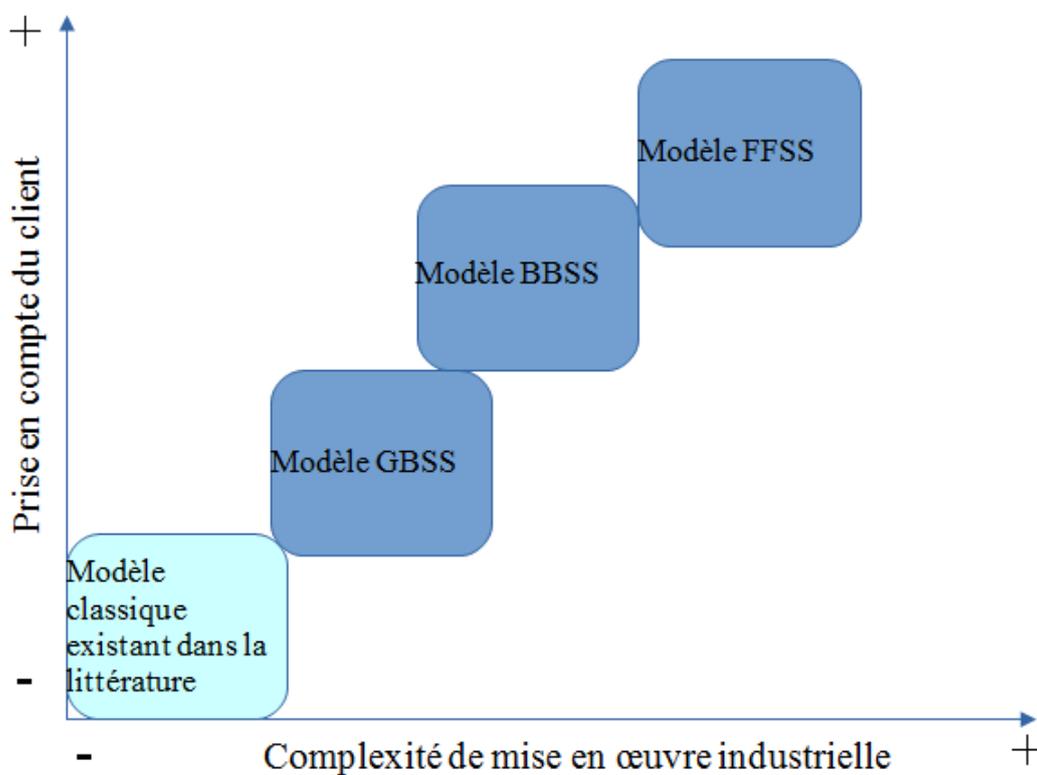


Figure 57. Comparaison entre les modèles.

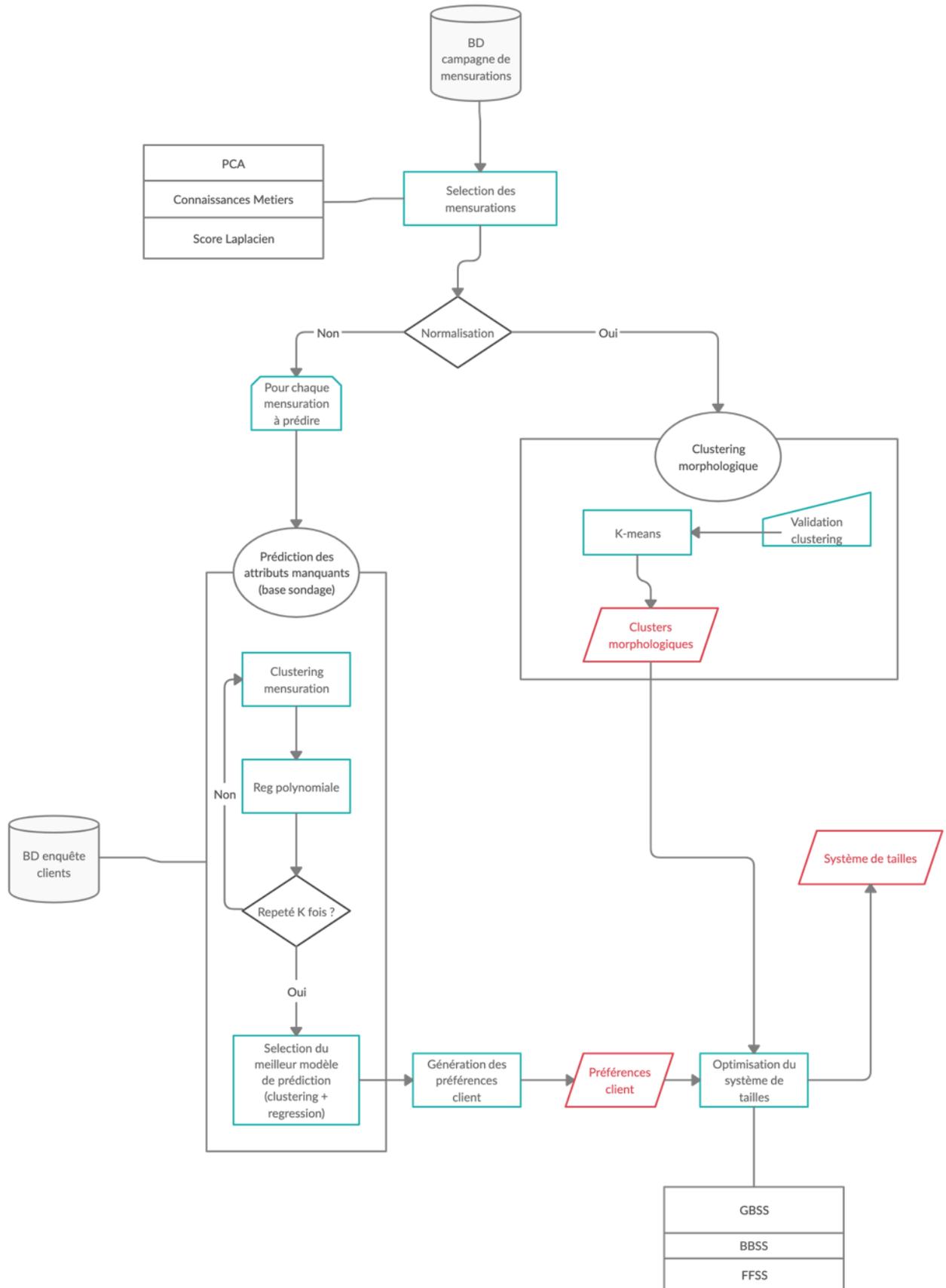


Figure 58. Principe général de la méthodologie proposée

Chapitre 4 Mise en œuvre et résultats

4.1 Introduction

Dans ce chapitre, nous mettons en œuvre la méthodologie proposée dans le chapitre 3 sur des données réelles issues de la campagne de mensurations CAESAR et d'une enquête de satisfaction client menée par une enseigne de prêt à porter féminin. Nous présentons d'abord les résultats de la classification non-supervisée des morphologies appliquées sur les mensurations normalisées du corps humain. Ensuite, nous appliquons les trois modèles de construction de système de tailles sur l'ensemble des données de la base de mensurations puis sur chacune des classes morphologiques. Les résultats obtenus sont finalement analysés et discutés.

4.2 Classification non supervisée de morphologies

La classification non supervisée cherche à définir des groupements d'individus similaires dans une population donnée. Nous présentons dans ce qui suit les résultats des différentes étapes du processus proposé (**section 3.2**), c'est à dire : la première étape concerne le prétraitement des données, la deuxième étape consiste à effectuer une classification non supervisée sur les variables sélectionnées et les données normalisées.

4.2.1 Description de la base de données

La base de données de mensurations utilisée dans ce travail de recherche est un échantillon représentatif de 2267 de femmes issues du projet CAESAR. Ce projet regroupe les mensurations de trois pays à savoir : Italie (388), Pays-Bas (619), l'Amérique du Nord (1260). Chaque individu de la base de données est décrit par 148 attributs relatifs aux mensurations et aux données sociales (âge, nombres d'enfant, ...).

4.2.2 Prétraitement de données

Une première analyse de la base de données CAESAR montre que certains individus ne possèdent pas toutes les mensurations requises. Certaines mensurations pour un individu donné peuvent aussi présenter des incohérences anthropométriques. Ces données doivent être filtrées afin d'éviter la génération de résultats erronés. Un prétraitement spécifique a donc été mis en œuvre sur l'ensemble des données brutes de la base de données afin d'obtenir une

base de travail cohérente et non bruitée par ces erreurs ou ces manques de mesure. La base de données utile est donc passée de 2267 femmes à 2196.

4.2.3 Sélection des variables

L'identification des variables les plus représentatives constitue un élément décisif pour la réussite de la classification. Elle permet également de réduire la complexité de l'algorithme de classification. La réduction des dimensions est la solution la plus intuitive pour réduire le nombre de variables requises et de choisir celles les plus représentatives. Nous procédons alors par appliquer l'ACP sur 30 mensurations sélectionnées (**Annexe 1**). Par la suite, nous nous intéressons au calcul de score Laplacien sur ces mêmes mensurations.

Le **Tableau 7** ci-dessous représente les coefficients de corrélation linéaire entre ces mensurations et les cinq premières composantes.

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
Age (years)	0.05344492	0.39906383	-0.582255094	0.248335186	-0.24957019
Acromial Ht Stand Max (mm)	0.89893740	-0.36645432	0.030068390	0.009912791	0.01928944
Axilla Ht Max(mm)	0.85045349	-0.43494226	0.035361884	0.067884853	0.05265721
Bustpoint Brth (mm)	0.45913223	0.61304393	0.022145122	0.137187885	0.20526963
Cervicale Ht (mm)	0.89877368	-0.36466343	0.002726468	0.029020591	-0.06820349
Chest Ht Stand (mm)	0.80182812	-0.50895586	0.057952623	0.019473484	0.03320854
Neck Ht (mm)	0.12855182	-0.44315804	0.425490393	0.251829212	0.51231102
Sleeve Outseam Len Max (mm)	0.83252054	-0.34926448	-0.136309060	-0.195098309	-0.03948778
Trochanterion Ht Max (mm)	0.76415561	-0.46477849	-0.112054390	-0.157168596	0.02590337
Waist Back (mm)	0.43645003	-0.29648670	0.602038538	0.036569557	-0.44211006
Acromial Height, Sitting (mm)	0.70456967	-0.11318844	0.333561364	0.199664758	0.07523677
Spine-to-Shoulder (mm)	0.51740242	0.22450071	-0.060811500	0.628862584	-0.29690652
Spine-to-Elbow (mm)	0.86724072	-0.07118791	-0.108198598	0.204271082	-0.12550217
Arm Length (Spine to Wrist) (mm)	0.88064950	-0.15470703	-0.153277339	-0.034604106	-0.17724043
Arm Length (Shoulder to Wrist) (mm)	0.81464341	-0.27487289	-0.154005523	-0.303221774	-0.08342161
Arm Length (Shoulder to Elbow) (mm)	0.79608877	-0.23512969	-0.101700408	-0.134913149	0.02599639
Armscye Circumference (Scye Circ Over Acro...	0.57566224	0.66766349	0.038062521	-0.039141956	0.06818514
Chest Circumference (mm)	0.54636291	0.79233312	-0.016490430	-0.018789948	0.02896506
Bust/Chest Circumference Under Bust (mm)	0.57859410	0.75660341	-0.002183331	0.031215081	0.02783370
Crotch Height (mm)	0.76301196	-0.53019787	-0.069945716	-0.003161853	0.12187377
Hip Circumference, Maximum (mm)	0.59316652	0.67766711	0.003184985	-0.161604553	0.07341345
Hip Circ Max Height (mm)	0.80736766	-0.05264237	-0.211752843	-0.125421164	-0.01750759
Neck Base Circumference (mm)	0.54303523	0.41399983	0.145422884	0.212152076	0.22831306
Shoulder Breadth (mm)	0.61297269	0.58329379	-0.010722562	0.065632363	0.02215995
Stature (mm)	0.87574878	-0.40641401	0.054975533	0.104708766	0.03428522
Thigh Circumference (mm)	0.55002217	0.64182231	0.014980236	-0.189141551	0.11676970
Waist Circumference, Pref (mm)	0.55612473	0.75030685	0.135844749	-0.116902003	-0.04747516
Waist Front Length (mm)	0.45351984	0.26084600	0.652064564	-0.184088892	-0.26562952
Waist Height, Preferred (mm)	0.81175342	-0.31906539	-0.339625184	0.082982919	0.16987936
Weight (kg)	0.70127650	0.67219147	0.020976068	-0.080447326	0.04020296

Tableau 7. Coefficients de corrélation.

C'est à partir de la représentation "Scree plot" des valeurs propres des composantes principales que nous avons choisi le nombre de composantes à retenir dans notre ACP (**Figure 59**). Cette courbe présente un coude à partir de la troisième composante et une décroissance régulière au-delà de celle-ci. Ce phénomène met en évidence deux composantes principales. La première composante représente 47,4% de l'inertie totale. La deuxième composante retient 22,5% de l'inertie totale. Ces deux composantes conduisent à un taux d'inertie expliqué de 69,9%.

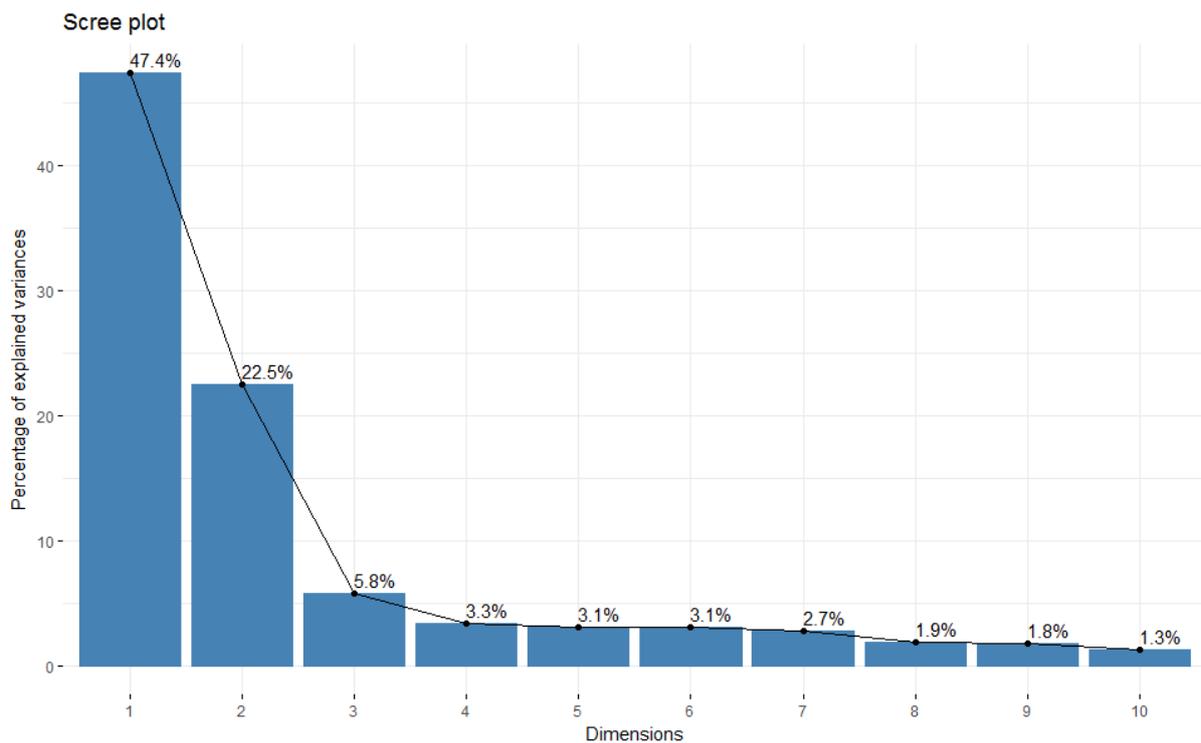


Figure 59. Scree plot des composantes principales de l'ACP.

Il est possible visualiser les individus et les variables de deux premières composantes principales à travers un biplot (**Figure 60**). Cette représentation permet de visualiser que la première composante caractérise principalement les longueurs du corps, alors que la deuxième composante représente majoritairement les circonférences du corps.

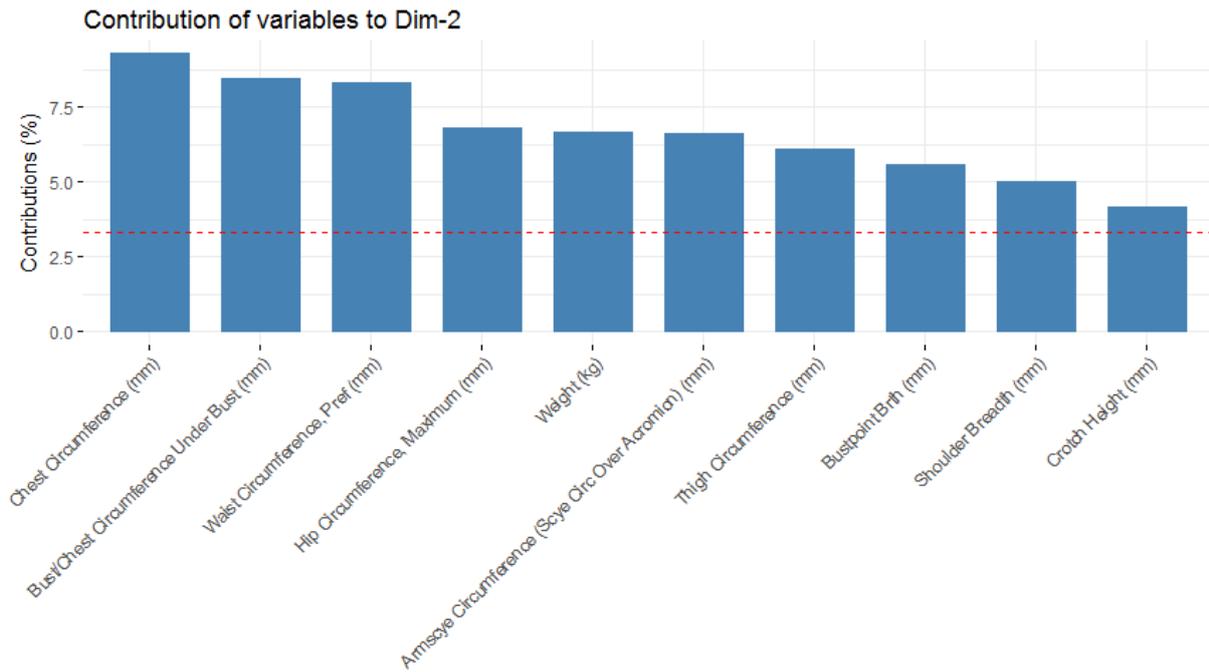


Figure 61. Contribution des variables sur la deuxième composante.

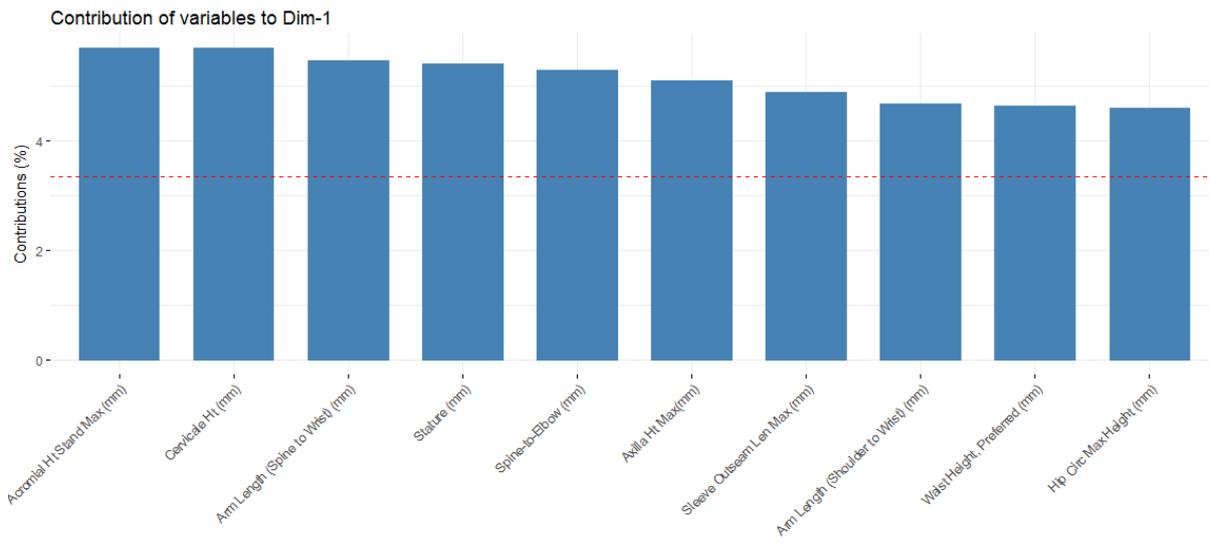


Figure 62. Contribution des variables sur la première composante.

Nous avons ensuite utilisé le score Laplacien pour la sélection des variables dans un contexte non-supervisé. Ce score sélectionne les variables pertinentes qui préservent au mieux la structure de données et qui maximisent la variance (**Figure 63**).

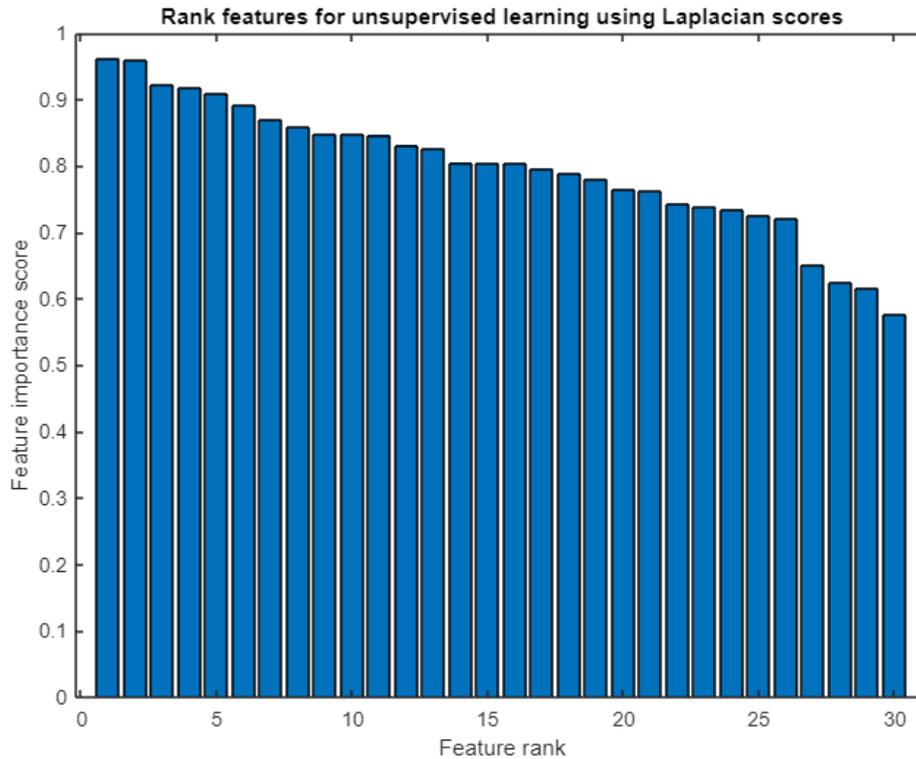


Figure 63. Score Laplacien

Le **Tableau 8** montre les résultats de classement des dix variables par le score Laplacien. Le score le plus élevé est alors le plus pertinent pour la sélection.

Id	Score	Variables
1	0,9623	'Age (years)'
30	0,9600	'Weight (kg)'
18	0,9219	'Chest Circumference (mm)'
27	0,9181	'Waist Circumference, Pref (mm)'
19	0,9085	'Bust/Chest Circumference Under Bust (mm)'
21	0,8917	'Hip Circumference, Maximum (mm)'
26	0,8697	'Thigh Circumference (mm) '
7	0,8593	'Neck Ht (mm) '
25	0,8476	'Stature (mm)'
2	0,8474	' Acromial Ht Stand Max (mm) '

Tableau 8. Les variables ordonnées selon le score Laplacien

Nous remarquerons que les quatre premières circonférences sont les mêmes variables qui ont été perçues par l'ACP pour la deuxième composante principale. De même, nous retrouvons aussi des variables communes relatives à la première composante principale.

Afin de respecter une cohérence entre les deux dimensions, un choix de substitution de variables doit être réalisé. Ce choix doit s'appuyer sur la notion de rapport anthropométrique pour créer un lien utile entre les deux dimensions et obtenir un descripteur de forme 2D respectant la morphologie de chaque individu. En d'autres termes, nous ne pouvons pas choisir une variable de la dimension 2, représentant par exemple un contour morphologique, sans l'associer à sa variable analogue de la dimension 1, qui positionne ce contour verticalement. Ainsi, le dessous poitrine sera remplacé par la largeur d'épaule que l'on justifie aisément car la position verticale de l'acromion représente la variable la plus pertinente dans la dimension 1. Les deux premières variables sont alors conservées alors que les deux suivantes doivent être remplacées par les variables analogues des trois autres contours, c'est à dire : hauteur poitrine/entrejambe, hauteur taille/entrejambe, hauteur hanche/entrejambe. Ces données représentent des couples car chacune d'elles est référencée par rapport au sol.

4.2.4 Normalisation des variables

Nous proposons de ramener à la même échelle les neuf mensurations sélectionnées grâce au rayon du cercle minimum englobant la description ponctuelle du torse. La **Figure 64** montre la première étape qui est de positionner les paires de points représentant ces mensurations en fonction de la méthode décrite en **section 3.2.2**.

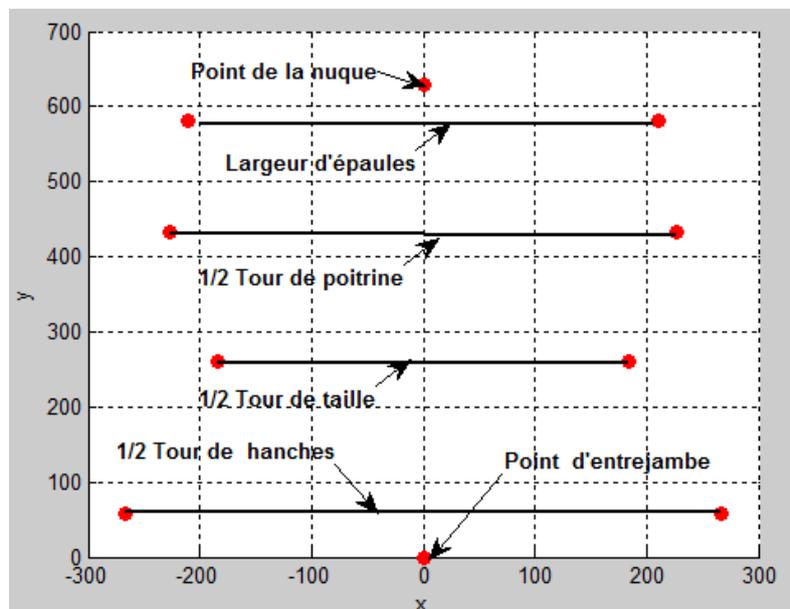


Figure 64. Descripteur de forme 2D du torse par ces points de mensurations.

Une fois les points positionnés, nous cherchons le cercle et le rayon du cercle minimum englobant (**Figure 65**).

Chaque individu possède alors son propre cercle et son propre rayon relativement à sa morphologie. Ce rayon permet de normaliser les mesures de cet individu en ramenant chaque individu dans un cercle de rayon $r = 1$ (**Figure 66**).

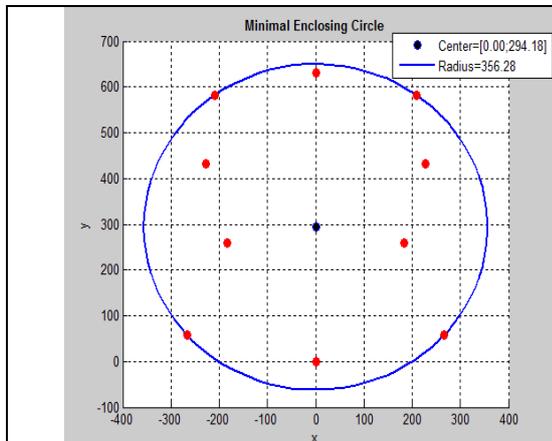


Figure 65. Cercle minimum englobant.

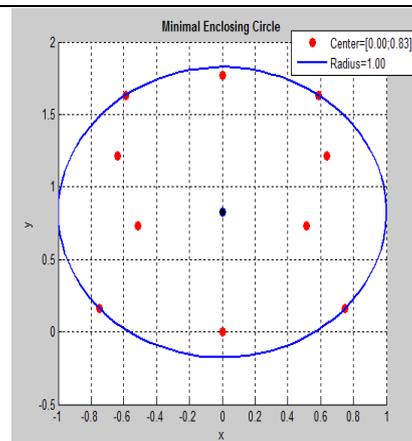


Figure 66. Cercle minimum englobant de rayon $r=1$.

Nous avons appliqué à nouveau l'ACP sur les données normalisées pour identifier l'effet de la normalisation (**Figure 67**). Cet effet est caractérisé par un recentrage des variables sur les deux premières composantes. La proximité relative des mesures sélectionnées par rapport aux deux dimensions justifie nos choix.

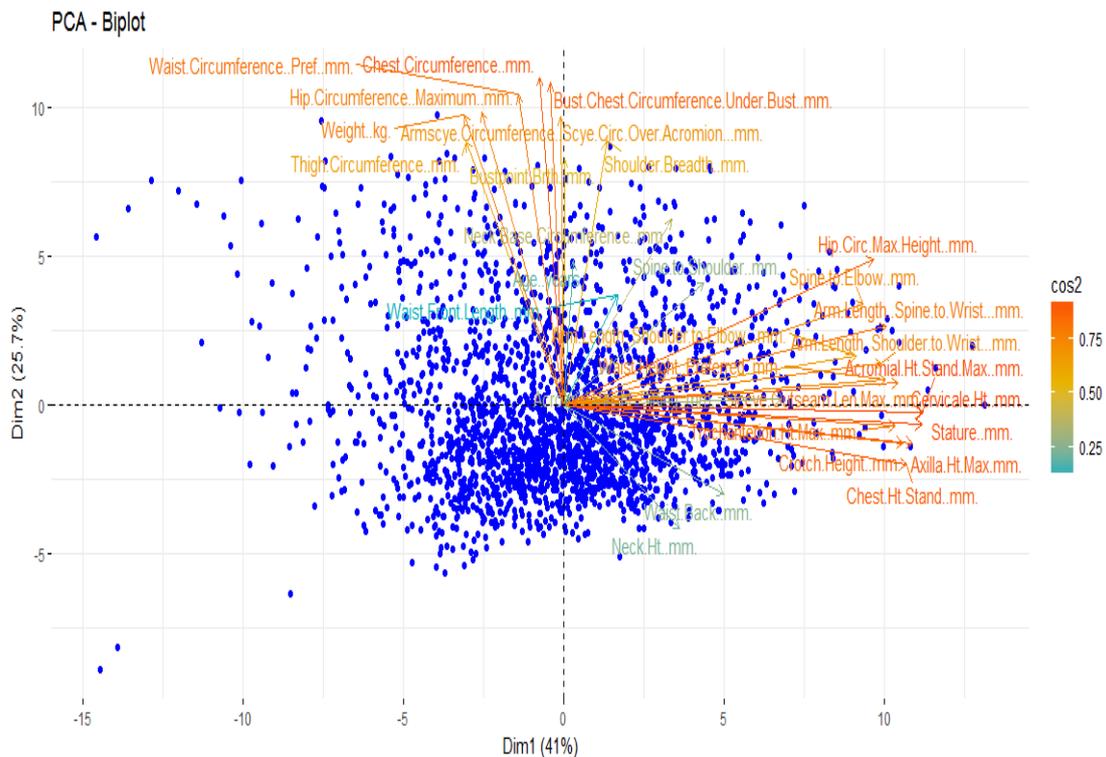


Figure 67. Biplot de l'ACP après la normalisation

4.2.5 Classification non supervisée

4.2.5.1 Choix du nombre de classes

Le partitionnement de la population est réalisé par application de l’algorithme K-moyennes++ avec une initialisation aléatoire des centroïdes introduit en **section B.1**. L’algorithme a été exécuté avec un nombre de classes k variant de 1 à 10. Afin d’éviter la convergence vers des optimums locaux, nous avons répété la procédure 300 fois pour chaque valeur de k . La qualité de la classification est ensuite évaluée par deux indices : Davies-Bouldin et Silhouette. La **Figure 68** montre que l’augmentation du nombre de classes a pour effet de dégrader les deux indices. Ces courbes tendent à montrer que les classes morphologiques sont difficiles à identifier distinctement par les données de mensurations. En d’autres termes, il n’y a pas de séparation claire entre les classes. Cela signifie qu’il faudrait des informations complémentaires pour déterminer avec sûreté le nombre de classes, telles que des scans 3D des individus [85] ou des connaissances métier.

Faute des données 3D disponibles pour notre base de données, nous fixons pour la suite la valeur à $k=2$. Ce choix est motivé par les valeurs des indices qui sont optimums pour $k=2$ mais aussi par les contraintes “métier” qui demande à limiter le nombre de classes et de systèmes de tailles associés.

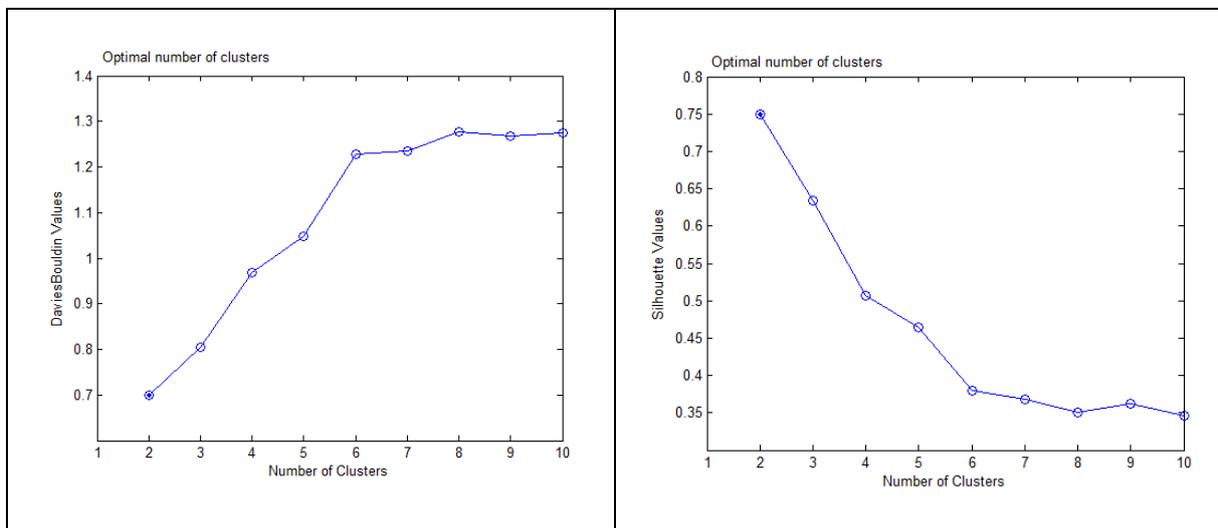


Figure 68. Indices Davies-Bouldin et Silhouette en fonction du nombre de classes

4.2.5.2 Résultats de la classification non-supervisée

Le **Tableau 9** résume le nombre d'individus par classe.

	Classe 1	Classe 2
Nombre d'individus	566	1630

Tableau 9. Nombre d'individus pour les deux classes

Le **Tableau 10** présente la description des principales mensurations des deux centroïdes.

	Tour de poitrine (mm)	Tour de sous-poitrine (mm)	Tour de taille (mm)	Tour de hanches (mm)	Stature (mm)
Centroïde 1	1111	879	942	1145	1629
Centroïde 2	913	787	754	1015	1633

Tableau 10. Mensurations des deux centroïdes

Afin de définir les deux morphotypes de deux classes, nous nous sommes procuré les scans 3D des individus les plus proches des centroïdes (**Figure 69**). Le morphotype 1 représente une morphologie ronde, dite aussi « Morphologie en O ». Ce sont des femmes à forte corpulence dont la masse graisseuse est assez prononcée au niveau des hanches ainsi qu'au niveau de l'abdomen. Nous sommes en présence de personnes en surpoids de forte poitrine avec une carrure plus importante. Le morphotype 2 représente une morphologie équilibrée, appelée « Morphologie en X ou sablier ». Ce sont des femmes plutôt minces avec des épaules de largeur proche de la largeur de hanches.

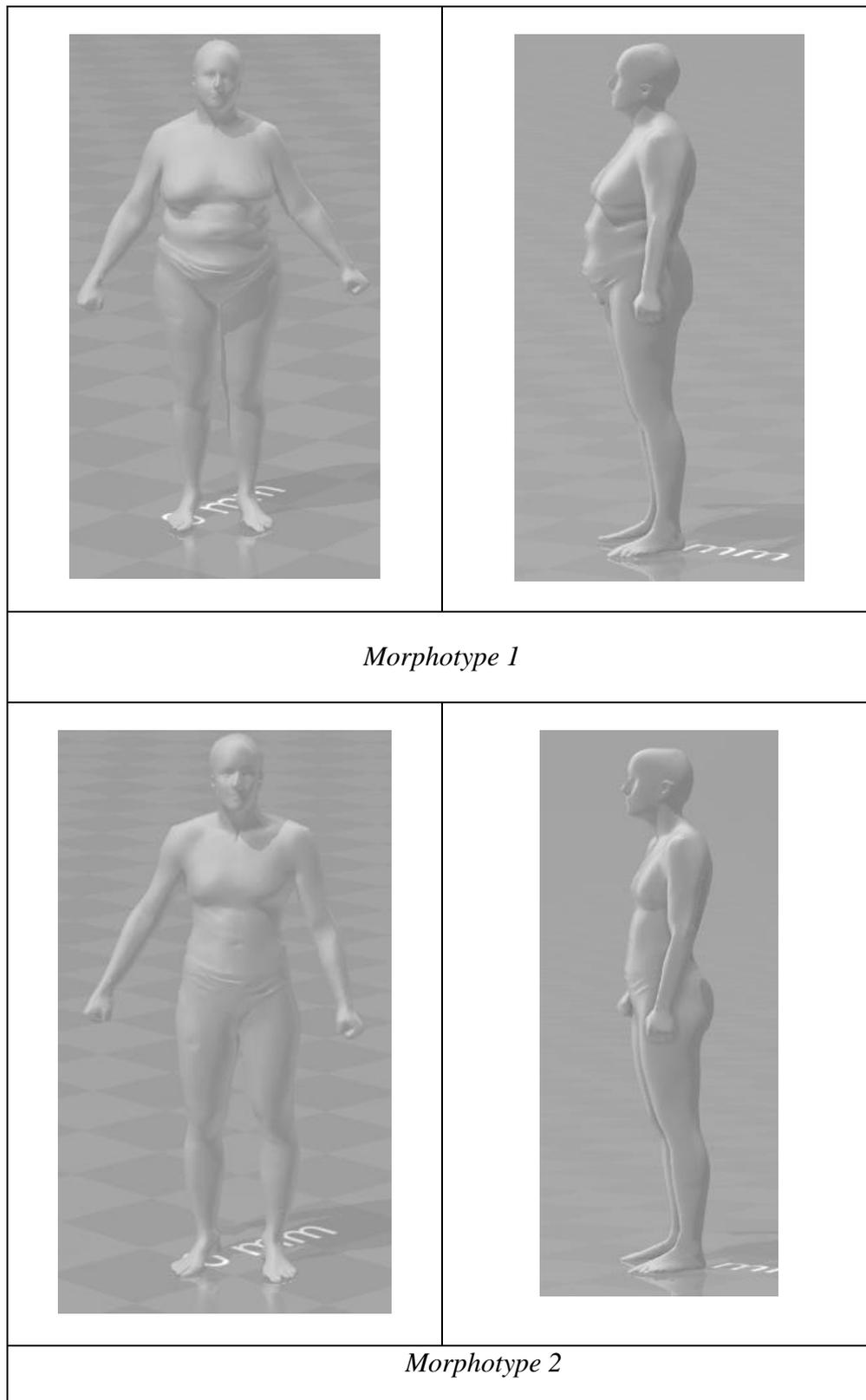


Figure 69. Les deux morphotypes représentant les deux classes morphologiques.

4.3 Optimisation système de tailles

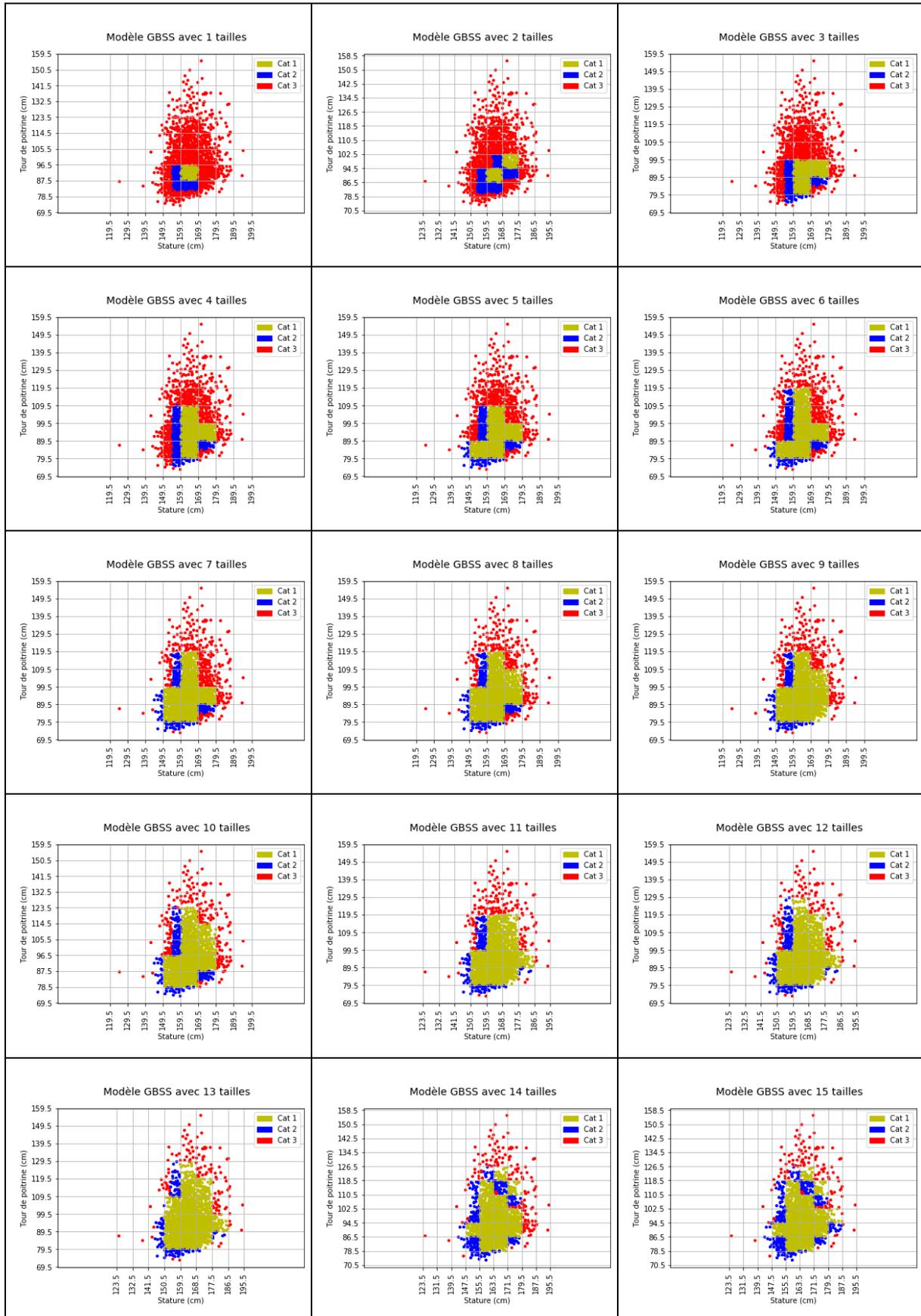
4.3.1 Modèle GBSS

Le processus de construction du système de tailles du modèle GBSS débute avec la création d'une grille de tailles en fixant deux pas d'intervalles pour les deux mesures clés choisies (P pour la première mensuration et Q pour la seconde). Ce modèle s'appuie sur le critère de performance FL (Fit Level) pour évaluer les systèmes de tailles générés (voir **section 3.5.1**). Nous avons implémenté ce modèle sur l'ensemble de la population de la base de données CAESAR, puis sur chacune des deux classes morphologiques, pour un nombre de tailles variant de 1 à 20. Le modèle GBSS requiert la définition de paramètres de tolérances, α pour la première mensuration et β pour la seconde, pour identifier les individus hors du système de tailles en cours d'évolution mais pouvant toutefois s'habiller avec les tailles proposées (catégorie 2). Ces deux paramètres α et β ont été fixés à 10cm en accord avec les professionnels du métier. L'ensemble des combinaisons des pas P et Q de la grille variant entre 1cm et 10cm, représentant au total 100 combinaisons, a été testé pour toutes les évolutions du nombre de tailles. Les résultats obtenus permettent de définir un front de Pareto qui montre l'ensemble des systèmes optimums en termes de critère FL pour chaque nombre de tailles.

Nous présentons dans cette section les résultats obtenues pour les systèmes de tailles sur les mensurations stature/poitrine et poitrine/hanche. Les résultats relatifs aux autres combinaisons de mensurations sont donnés en **Annexes 4, 5 et 6**.

4.3.1.1 Système Stature/Poitrine

La **Figure 70** représente les différents systèmes de tailles générés par le modèle GBSS pour les mensurations stature et tour de poitrine sur la population totale. Nous pouvons remarquer que les systèmes proposés par ce modèle restent bien compacts jusqu'à une certaine limite définie par le nombre de tailles (13 tailles dans cet exemple). Au-delà de cette limite, certains systèmes peuvent présenter quelques "trous" dans les tailles proposés (pour les nombres de tailles > 13 dans cette exemple).



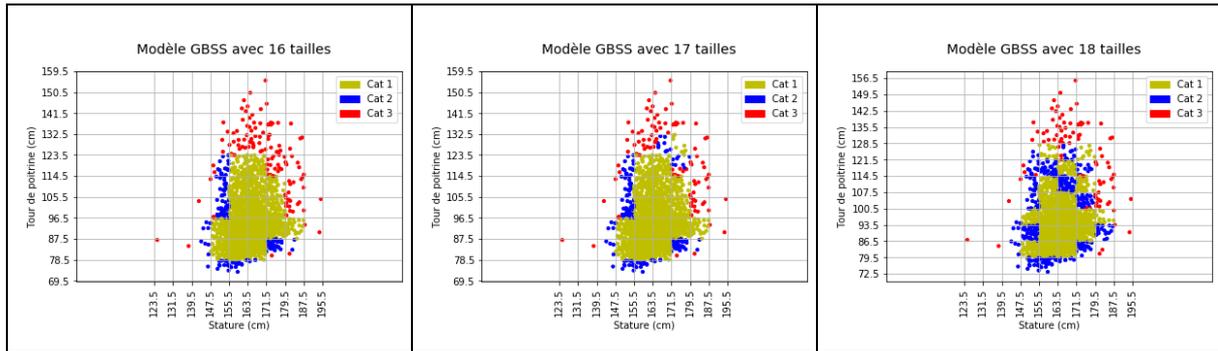


Figure 70. Systèmes de tailles générés par le modèle GBSS sur la population totale pour les mensurations stature/tour de poitrine

L'ensemble des résultats obtenus pour chaque combinaison de pas et le front de Pareto qui en résulte sont illustrés en **Figure 71(a)**. La courbe obtenue est classique : le critère de bien aller FL calculé avec la distance de Manhattan diminue de façon continue lorsque le nombre de tailles augmente. Cependant, il est possible de détecter des “coudes” sur ce front qui permettent d’orienter le choix du nombre de tailles à privilégier (4 et 6 tailles pour l'exemple de la **Figure 71 (b)**). Le front de Pareto donne les systèmes de tailles optimums par chaque nombre de tailles avec les pas de grille correspondants donnés en **Figure 72**.

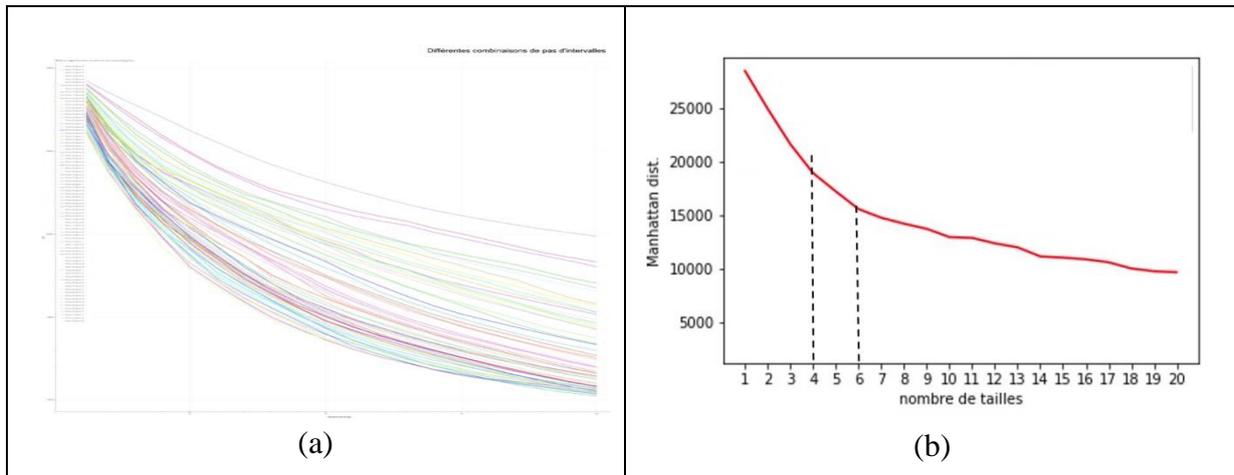


Figure 71. Résultats obtenus pour l'ensemble des combinaisons de pas (a) et le front de Pareto correspondant (b) pour le système de tailles poitrine/hanche

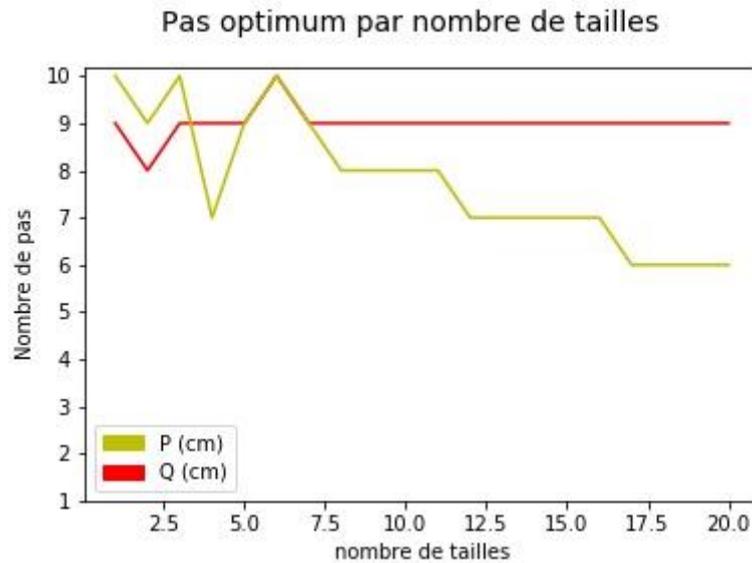


Figure 72. Pas de la grille pour les systèmes de tailles du front de Pareto

Le processus de conception des systèmes de tailles a été appliqué sur les populations de chacune des classes morphologiques obtenues en section 4.2. Les trois fronts de Pareto relatifs aux systèmes optimisés sur la population totale et les deux classes morphologiques de ont été générés sont présentés en **Figure 73**. Il apparaît sur cette figure que, quel que soit le nombre de tailles, les systèmes de tailles proposés par le modèle GBSS sont toujours de meilleure qualité en termes de bien aller (*FL*) pour la morphologie 2. Les systèmes obtenus pour la population totale et la morphologie 1 sont équivalents en termes de critère *FL*. La morphologie 1 représente les individus à forte corpulence. Il apparaît donc qu'il est plus compliqué de créer un système de tailles adapté à cette morphologie, notamment à cause des proportions entre la stature et le tour de poitrine plus disparates.

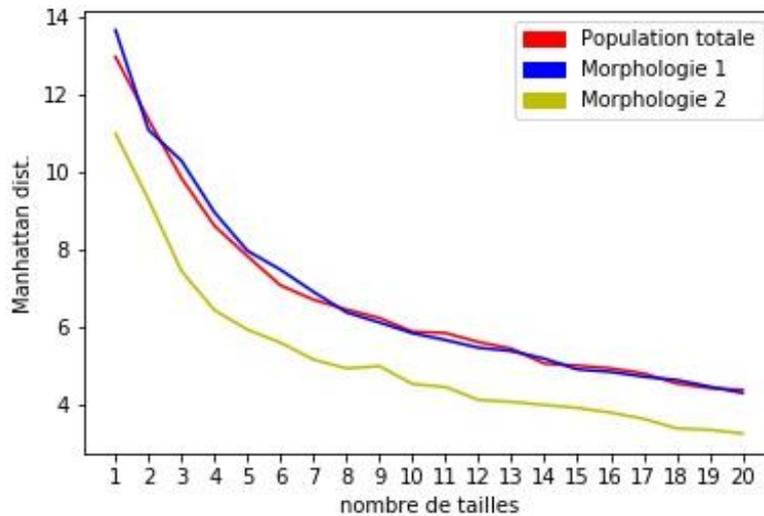


Figure 73. Front de Pareto obtenu sur la population totale et les deux classes morphologiques pour les systèmes stature/poitrine

Néanmoins, il faut prendre en compte ici le nombre de personnes constituant chacune des populations, pour rappel : 556 individus pour les morphologies 1, 1630 individus pour la morphologie 2 et 2196 pour la population totale. La morphologie 1 est minoritaire dans la population totale. Il est donc intéressant d'étudier plus précisément quelles sont les performances des systèmes de tailles sur les différentes populations. Ainsi, nous avons évalué le bien aller (critère *FL*) obtenu avec les systèmes de tailles générés sur la population totale sur les individus de chacune des morphologies. La **Figure 74** illustre les comparaisons des performances obtenues sur une morphologie donnée par des systèmes de tailles optimisés d'une part sur la population totale et d'autre part sur la population de la morphologie considérée. Comme attendu, il apparaît clairement que les systèmes générés spécifiquement pour une morphologie ont de meilleurs résultats que les systèmes relatifs à la population totale. Cela confirme la capacité de notre modèle à s'adapter aux mensurations d'une population donnée. Plus précisément, cette comparaison montre que les bénéfices sont plus significatifs sur la population de la morphologie 1, c'est à dire la population la plus minoritaire et aux morphologies plus corpulentes. Enfin, il est intéressant d'étudier si une enseigne doit s'appuyer sur un seul système de tailles global défini pour l'ensemble de sa clientèle ou développer des systèmes de tailles plus modestes pour chacune des morphologies rencontrées parmi sa clientèle. En d'autres termes, sur notre population et pour un nombre de tailles n donné, est-il plus intéressant de produire deux systèmes de tailles pour chacune des deux morphologies, avec des nombres de tailles n_1 et n_2 tels que $n_1+n_2=n$, ou chercher à adapter les n tailles sur l'ensemble de la population. La **Figure 75** montre sur les

mesurations stature/poitrine, nous obtenons les mêmes performances avec un système à 12 tailles optimisé sur la population totale (**Figure 76a**) que deux systèmes à 6 tailles optimisés sur chacune des morphologies (**Figure 76b** et **Figure 76c**) : le bien aller des vêtements obtenus par ces deux stratégies est globalement similaire pour l'ensemble des individus. En dessous d'un total de 12 tailles, le système global semble plus performant, notamment à cause de la morphologie 2, au-delà de 12 tailles, il semble plus bénéfique d'utiliser des systèmes de tailles spécifiques aux deux morphologies.

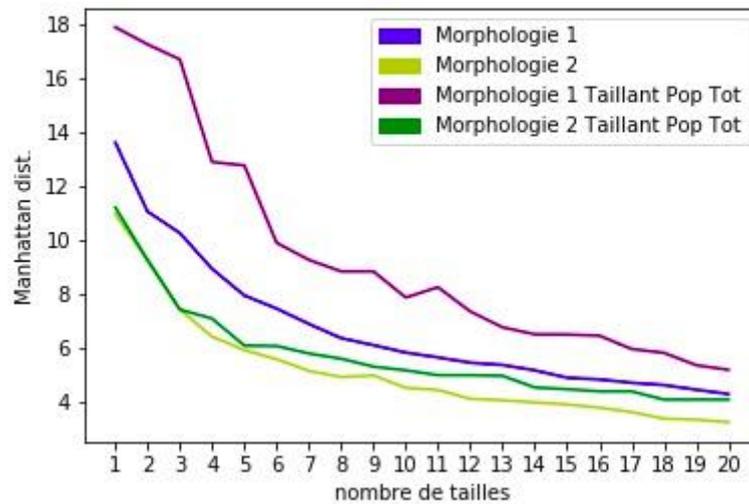


Figure 74. Comparaison des performances des systèmes de tailles générés pour la population totale et pour les deux morphologies

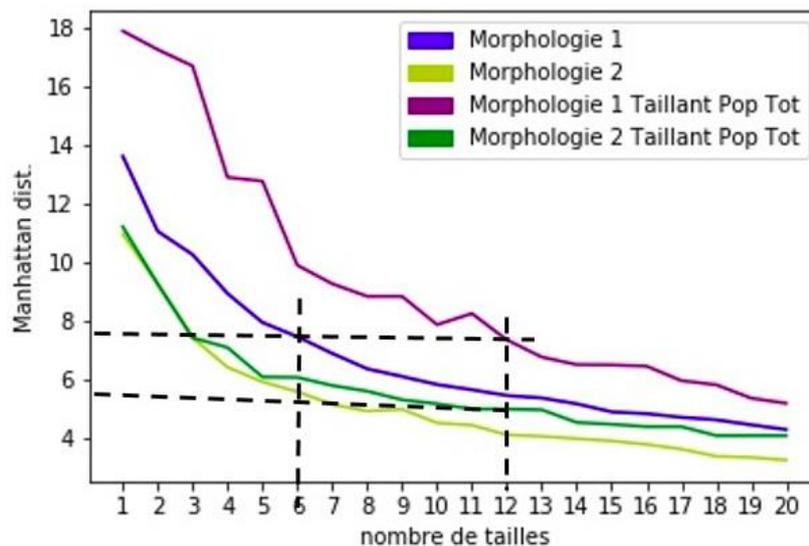


Figure 75. Comparaison des performances d'un système à 12 tailles pour la population totale et des deux systèmes à 6 tailles pour chacune des morphologies

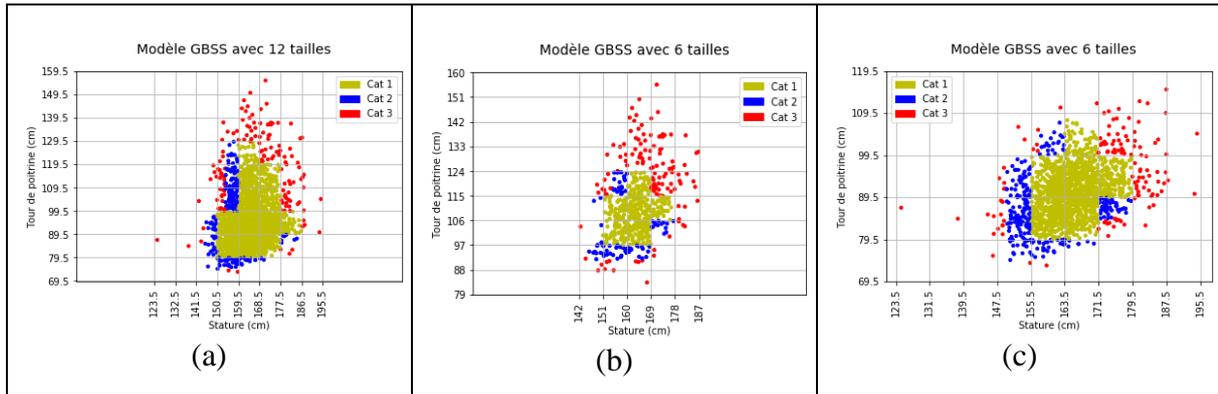
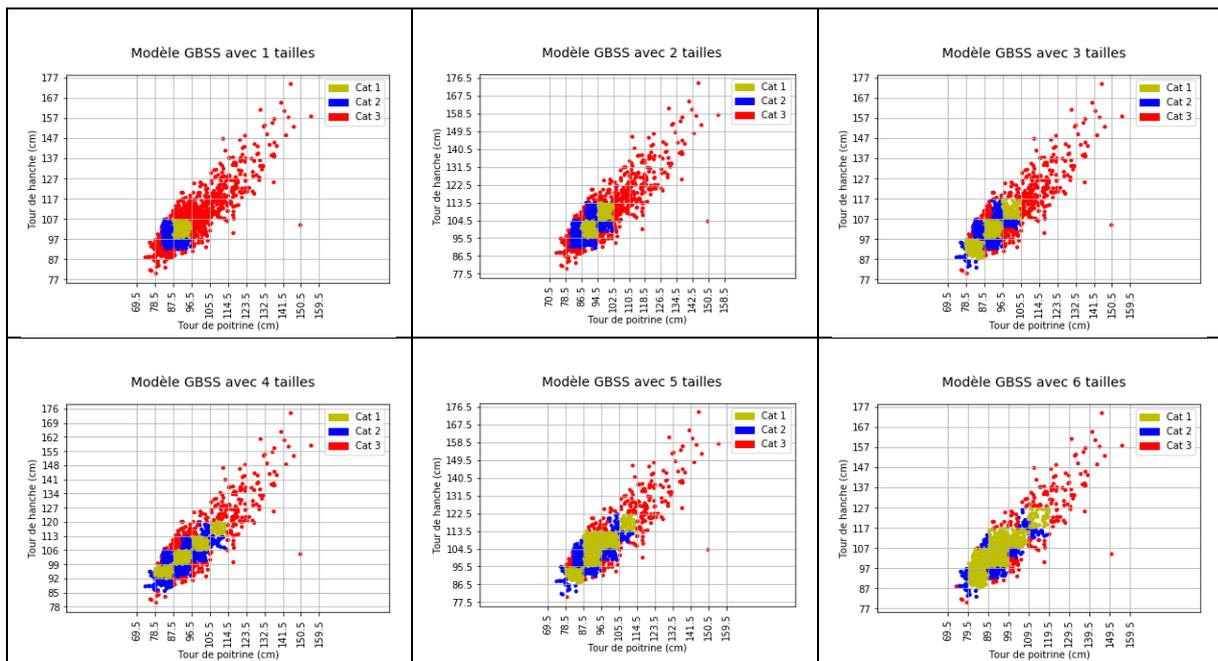


Figure 76. Systèmes obtenus avec 12 tailles pour la population totale (a), 6 tailles pour la morphologie 1 (b) et 6 tailles pour la morphologie 2 (c)

4.3.1.2 Système Poitrine/Hanche

La **Figure 77** représente les différents systèmes de tailles générés par le modèle GBSS pour les mensurations tour de poitrine et tour de hanche sur la population totale. Comme pour les systèmes stature/poitrine, nous pouvons remarquer que les systèmes générés par ce modèle restent également bien compacts jusqu'à 11 tailles. La **Figure 77** montre également que l'algorithme de construction des systèmes de tailles permet de bien prendre en compte la forme du nuage de points formé par les mensurations de chaque individu de la population. Notamment à partir de 5 tailles, les tailles ajoutées ne complètent pas uniquement le système de tailles dans la diagonale du plan, c'est à dire la direction principale du nuage de points, mais également de part et d'autre de cette direction selon les densités du nuage des points non couverts par le système de taille en cours d'évolution.



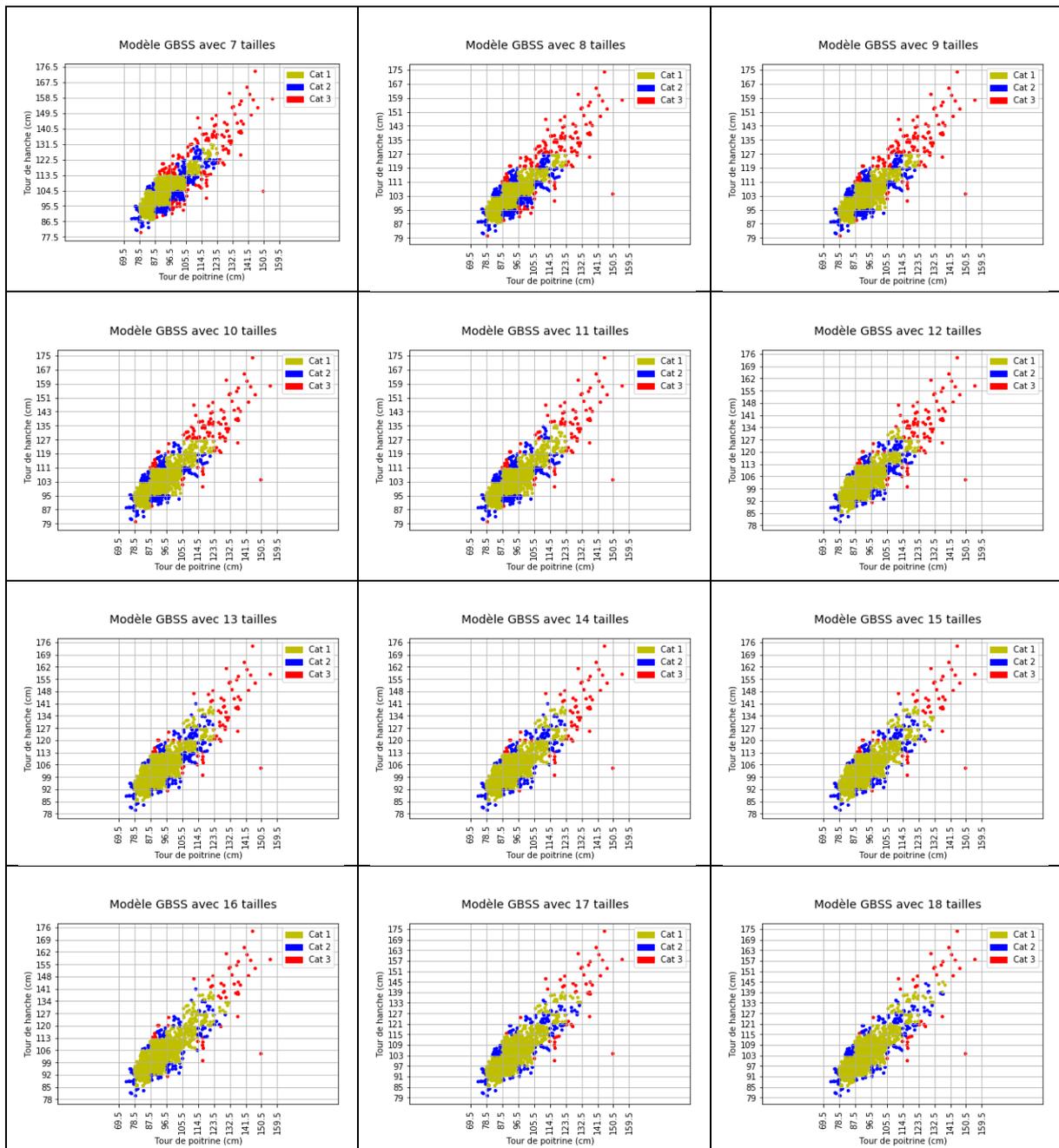


Figure 77. Systèmes de tailles générés par le modèle GBSS sur la population totale pour les tour de hanche/tour de poitrine

Les trois fronts de Pareto obtenus sur la population totale et les morphologies 1 et 2 sont présentés en **Figure 78**. L'optimisation du système de tailles pour la morphologie 1 est moins performante : les évolutions des proportions entre les deux mensurations considérées, poitrine/hanche, sont très irrégulières et plus difficile à couvrir par le système de tailles.

A noter également sur ces courbes, des cassures apparaissent plus clairement que pour les systèmes basés sur les mensurations stature/poitrine. La prise de décision sera donc plus aisée.

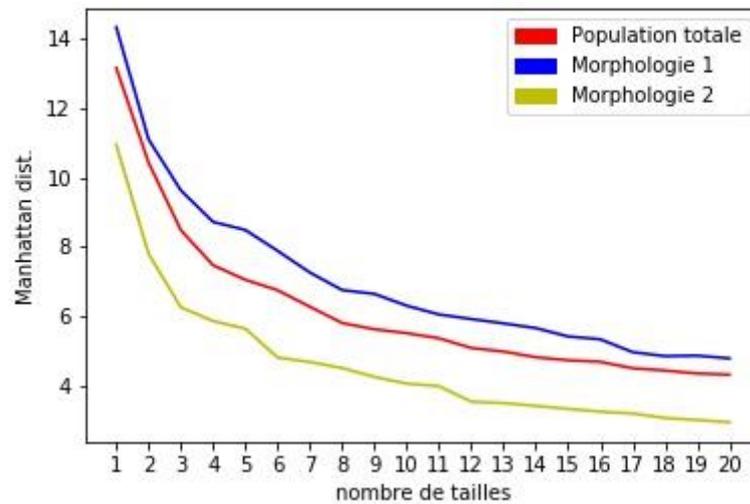


Figure 78. Front de Pareto obtenu sur la population totale et les deux classes morphologiques pour les systèmes hanche/poitrine

L'analyse par morphologie (**Figure 79**) confirme les résultats obtenus précédemment sur les mensurations stature/hanche : l'optimisation par des systèmes par morphologie garantit de meilleures performances en termes de bien aller. A partir d'un nombre total de tailles égal à 12, il s'avère également pertinent de choisir deux systèmes de tailles, un par morphologie, plutôt qu'un système global (**Figure 80**).

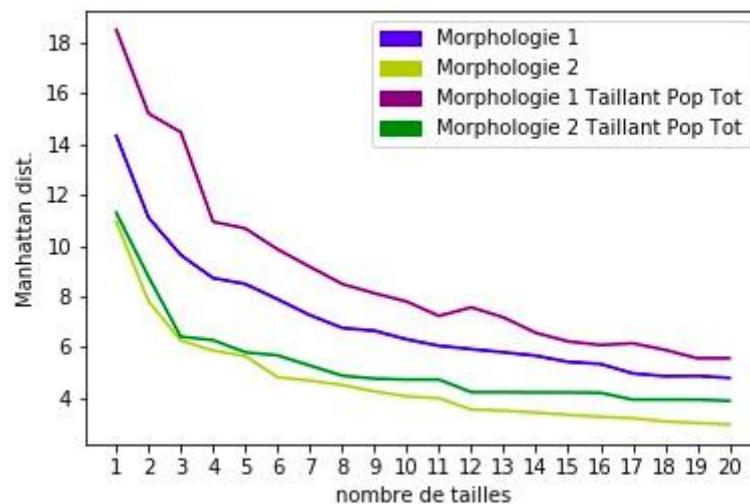


Figure 79. Comparaison des performances des systèmes de tailles générés pour la population totale et pour les deux morphologies

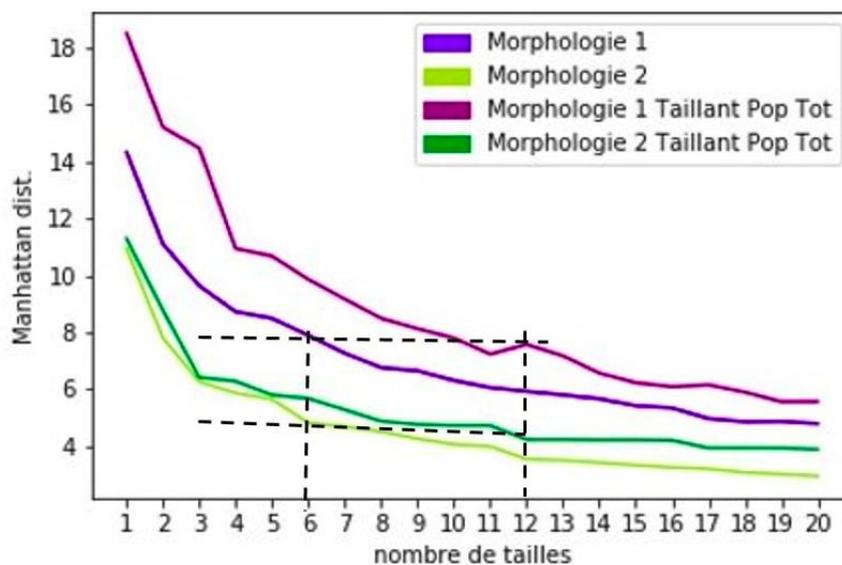


Figure 80. Comparaison des performances d'un système à 12 tailles pour la population totale et des deux systèmes à 6 tailles pour chacune des morphologies

4.3.1.3 Discussion sur le modèle GBSS

L'analyse des résultats obtenus montre que les systèmes de tailles générés par le modèle GBSS sur les mensurations stature/poitrine et hanche/poitrine sont particulièrement performants. Notamment, l'algorithme de construction du système de tailles permet une bonne prise en compte de la forme du nuage de points des mensurations de la population considérée. Il s'avère également que la classification morphologique en amont de la construction du système de taille est intéressante dans la mesure où elle permet d'améliorer le bien aller des vêtements. Cela est particulièrement vrai sur les morphologies de forte corpulentes. La question du choix entre un système de tailles pour une population globale ou plusieurs systèmes de tailles pour chacune des morphologies de la population reste posée. Cependant, il semble qu'à partir d'un certain nombre de tailles, 12 sur les données étudiées ici, un ensemble de systèmes spécifiques à chaque morphologie avec un nombre réduit de tailles devient plus performant qu'un système unique pour toute la population.

Néanmoins, certains aspects du modèles GBSS méritent d'être amélioré. Le modèle GBSS nécessite de définir des tolérances α et β a priori. Cela peut être effectué à partir des connaissances métier mais il semble plus judicieux d'intégrer le client dans la définition de ces tolérances. L'algorithme de construction montre également certaines limites. Lorsque le nombre de tailles augmente, des "trous" apparaissent dans le système à cause notamment des

contraintes imposées à la grille. Par ailleurs, l'incrémentation des tailles avec les contraintes de voisinage peut être une source de convergence vers des optimums locaux.

Ainsi, les modèles BBSS et FFSS mis en œuvre dans la section suivante visent à pallier ces points faibles.

4.3.2 Modèles orientés clients (BBSS et FFSS)

Les modèles BBSS et FFSS intègrent deux évolutions majeures par rapport au modèle GBSS : la prise en compte des tolérances clients en termes de bien aller et l'algorithme de construction du système basé sur un algorithme génétique.

4.3.2.1 Tolérances clients

Les informations sur les tolérances clients par rapport aux différentes mensurations et tailles de vêtements sont obtenues à partir d'une enquête de satisfaction menée par une enseigne de prêt à porter féminin. Les résultats de cette enquête permettent de calculer les tolérances des clients qui seront intégrées dans la construction des systèmes de tailles des modèles BBSS et FFSS.

4.3.2.1.1 Enquête de satisfaction client

L'enquête de satisfaction a été réalisée en ligne sur la base de données clients d'une enseigne Française de prêt à porter féminin. Les clients ont été invités à répondre à des questions sur leur morphologie et leur satisfaction vis-à-vis des produits, et plus particulièrement la taille des vêtements. Différents types de vêtements étaient considérés dans le questionnaire : T-Shirt, pull, pantalon, Au total, 3685 clients ont répondu à l'enquête. La majorité des répondants, 1741 femmes, a apporté des réponses complètes et exploitables (c'est à dire à la fois sur leurs mensurations et leur satisfaction) pour les T-Shirts uniquement. Les tolérances calculées par la suite et les systèmes de tailles associés sont donc relatifs aux T-Shirts.

L'enseigne qui a mené le sondage propose quatre tailles de vêtement (T1, T2, T3, T4) caractérisés par les trois mensurations présentées dans le **Tableau 11**. Les clientes interrogées mentionnent leur taille préférée parmi ces quatre tailles.

	T1	T2	T3	T4
Tour de poitrine (cm)	90	96	102	110
Tour de taille (cm)	70	76	82	91
Tour de hanches (cm)	96	102	108	115

Tableau 11. Mensurations clés des quatre tailles de l’enseigne.

Les informations déclarées par les répondants sur leur morphologie restent cependant assez limitées. En effet, seules les indications sur l’âge, le poids et la stature des personnes sont disponibles de façon fiables. Ainsi, le modèle de prédiction composé successivement d’une classification non-supervisée par les k-moyennes++, une classification supervisée par arbre de décision CART (**section 3.3.2.1.2**) et une régression polynomiale est appliqué pour estimer les mensurations manquantes et nécessaires pour l’élaboration du système de tailles : tour de taille, tour de hanche et tour de poitrine.

Le nombre optimal de classes utilisé dans ce modèle prédictif est estimé pour chacune des mensurations en fonction des erreurs de prévision sur les données d’apprentissage (base de données CAESAR) par validation croisée (k-folds avec k= 10). Les erreurs SMAPE (Symetric Mean Absolute Percentage Error) obtenues pour chaque mensuration en fonction du nombre de classes sont présentées dans le **Tableau 12** . Pour le tour de poitrine, le nombre optimal de classe est k=3. Le tour de sous-poitrine et le tour de taille, le nombre optimal de classe est k=1 tandis que pour le tour de hanches k=4. Un exemple d’un arbre de décision permettant d’affecter les clients de l’enquête aux classes en fonction des informations du sondage (âge, poids, stature) est présenté en **Annexe 3**.

	SMAPE pour k=1	SMAPE pour k=2	SMAPE pour k=3	SMAPE pour k=4	SMAPE pour k=5	SMAPE pour k=6	SMAPE pour k=7	SMAPE pour k=8	SMAPE pour k=9	SMAPE pour k=10
Prédiction tour de poitrine	0.03420647	0.03403049	0.03378834	0.03442228	0.03494491	0.03488492	0.03525261	0.03593829	0.03581115	0.03613023
Prédiction tour de sous-poitrine	0.03464282	0.03485708	0.0347976	0.03509964	0.03576162	0.03606737	0.03597085	0.03652031	0.03659107	0.03678317
Prédiction tour de taille	0.05647465	0.0567701	0.05679155	0.05744387	0.05851893	0.05885341	0.05922304	0.06068689	0.05992769	0.06050699
Prédiction tour de hanches	0.03211756	0.03158797	0.0315058	0.03147751	0.03150381	0.0315405	0.03179104	0.03153081	0.03169876	0.03169903

Tableau 12. SMAPE sur les mensurations prédites.

Ces prédictions complètent les informations du sondage et permettent ainsi d'obtenir, d'une part, le choix des clients en termes de tailles de vêtement et leur niveau de satisfaction pour la taille choisie, et d'autre part, les mensurations nécessaires pour la construction du système de taille de ces mêmes clients.

4.3.2.1.2 Calcul des tolérances des clients

Les informations retenues pour le calcul des tolérances concernent uniquement les clients qui ont déclaré un bon niveau de satisfaction pour une taille, c'est à dire correspondant à "satisfaisant" ou "très satisfaisant" dans le sondage. Les **Figure 81** et **Figure 82** illustrent les distributions de la mensuration tour de poitrine des clients satisfaits ou très satisfaits par les tailles de l'enseigne.

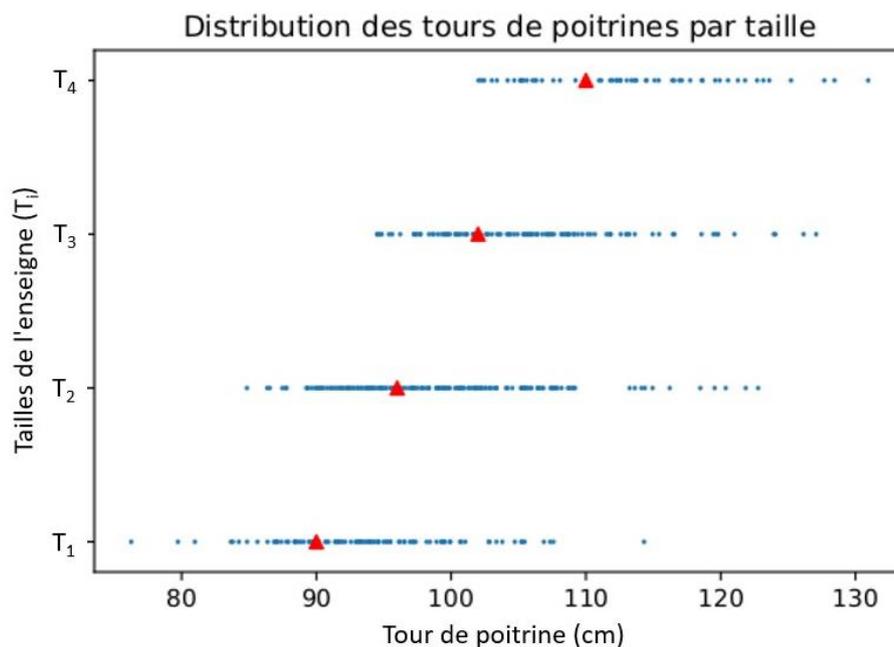


Figure 81. Distributions des mesures du tour de poitrine des clients parmi les tailles de l'enseigne T_i ($i = 1, \dots, 4$)

La **Figure 81** représente les distributions des mesures de tour de poitrine des différents clients pour leur taille préférée parmi les tailles de l'enseigne T_i ($i = 1, \dots, 4$). Nous pouvons observer que ces mesures suivent des distributions semi-normales de moyenne T_i et de différents écarts types en fonction de la valeur inférieure ou supérieure de la mensuration du client par rapport à T_i . Nous pouvons également constater que les écarts types des distributions semi-normales augmentent avec la valeur des tailles T_i . Il est en effet naturel que les mesures des clients soient plus étendues lorsque leur taille préférée est grande et que les tolérances soient plus faibles pour les petites tailles.

Ainsi, nous définissons des lois semi-normales à gauche et à droite de chaque taille T_i . Les écarts types des lois semi-normales à gauche, respectivement à droite, sont calculés à partir des clients qui ont une mesure inférieure, respectivement supérieure, à T_i .

La **Figure 82** présente les lois semi-normales obtenues pour le tour de poitrine. La partie droite a toujours un écart type plus grand que la partie gauche. La raison principale est que les clients tolèrent plus facilement de porter un vêtement plus ample alors que les tolérances pour les vêtements serré sont plus faibles. Comme expliqué en chapitre 3, nous calculons ici des écarts types pondérés par le nombre de répondants. En effet, la distribution de la population n'est pas homogène sur l'ensemble des tailles T_i . Par exemple, il peut y avoir plus de clients dont la mesure est à 5cm de la taille S et qui indiquent qu'ils préfèrent cette taille, que des clients situés à 1cm de cette taille S pour la simple raison qu'il y a plus d'individu situés à 5cm dans la population totale.

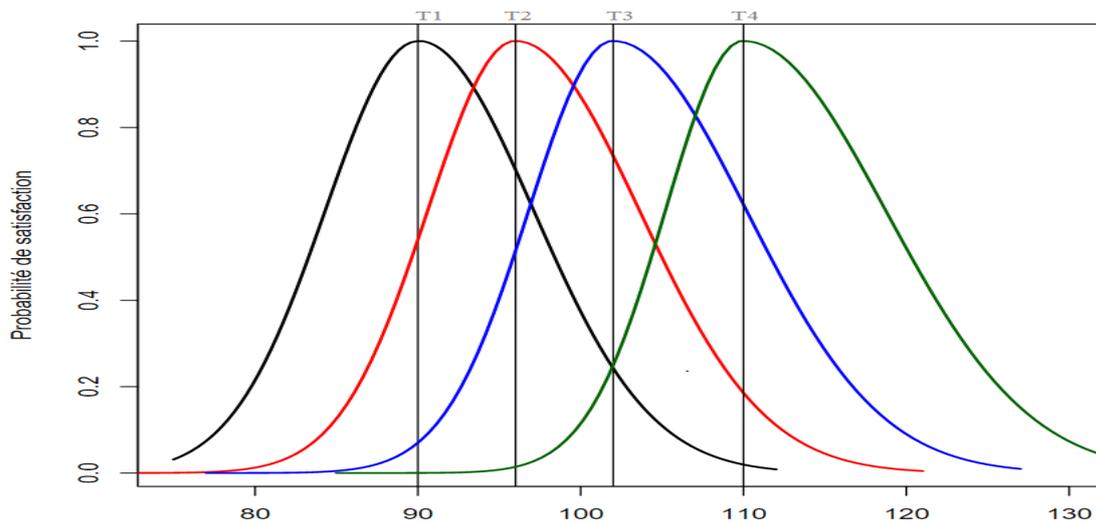


Figure 82. Lois semi-normales obtenues à partir des tours de poitrine des clients et de moyenne T_i ($i = 1, \dots, 4$)

A partir des écarts types calculés avec les distributions semi-normales obtenues pour chacune des tailles T_i , nous construisons un modèle pour prédire les écarts types pour toute autre taille potentiellement générée par le système. Comme observer en **Figure 83**, une simple régression linéaire permet de modéliser correctement l'évolution des écarts types en fonction des tailles. Par ailleurs, le faible nombre de tailles (4) ne permet pas d'envisager la mise en œuvre de modèle plus complexe afin d'éviter les problèmes de sur-apprentissage.

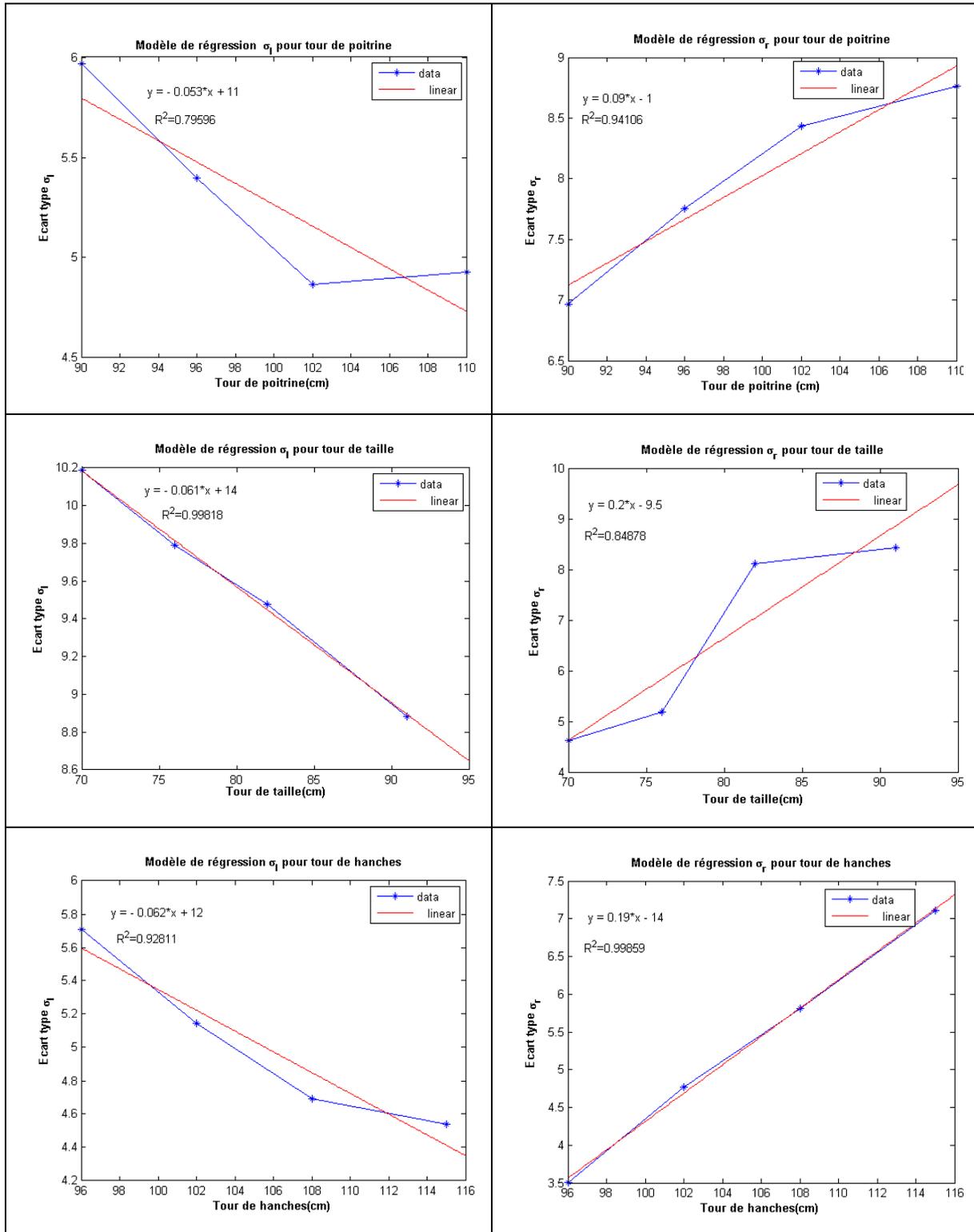


Figure 83. Prédiction des écarts types pour chacune des tailles par régression linéaire

4.3.2.2 Algorithme génétique

L'algorithme d'optimisation du système de tailles est un algorithme génétique qui est commun aux deux modèles BBSS et FFSS. Les spécificités de ces algorithmes sont décrites

en **section 3.3.4** et les paramètres utilisés lors de l'expérimentation sont donnés dans le **Tableau 13**.

Taille de la population	30
Nombre maximal de générations	200
Taux de mutation	5%
Taux de croisement	50%
Taille du tournoi	3

Tableau 13. Paramètres de l'algorithme génétique

L'initialisation de la population initiale est effectuée par un SOM de 200 neurones. Le résultat obtenu avec le SOM sur le système stature/poitrine est illustré en **Figure 84**.

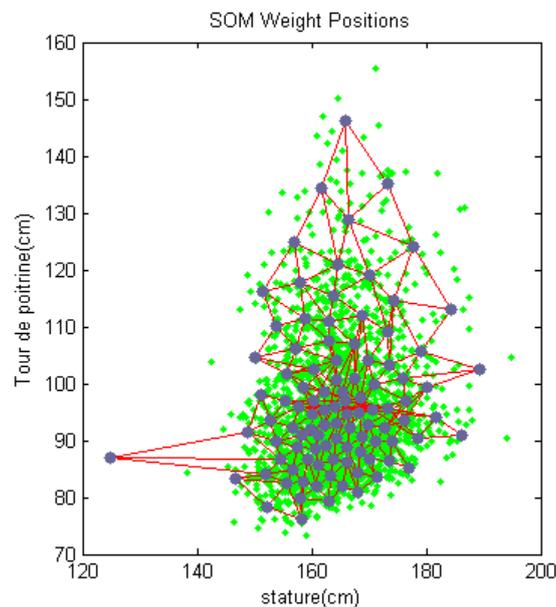


Figure 84. Exemple de SOM pour l'initialisation de la population de l'algorithme génétique sur un système stature/poitrine

Les **Figure 85(a)** et **Figure 85(b)** illustrent la convergence de l'algorithme génétique pour les modèles BBSS et FFSS. On remarque la stagnation de la fonction fitness à partir 120-140 générations. L'algorithme FFSS converge plus lentement que l'algorithme BBSS. Cette lenteur est due à la mesure optimisée Q (a composantes multiplicatives) contrairement à la mesure P optimisée par BBSS.

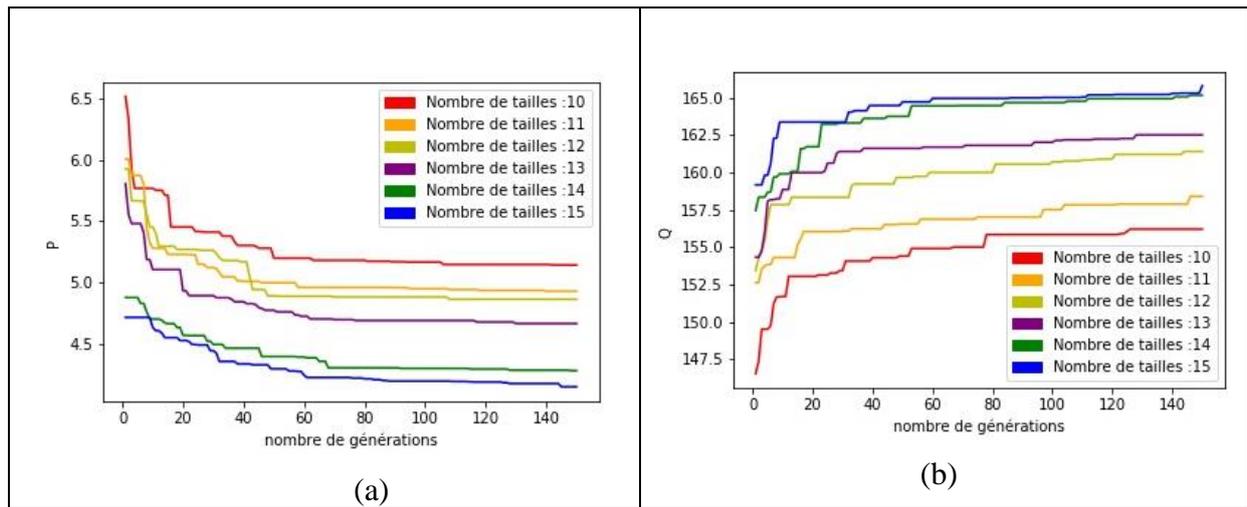


Figure 85. Convergence de la fonction fitness du modèle BBSS (a) et le modèle FFSS (b)

4.3.2.3 Modèle BBSS

Le modèle s'appuie sur la construction d'un système de tailles qui couvrent les individus contenus dans des boîtes de dimensions proportionnelles aux tolérances clients (**section 3.3.2.2**). L'optimisation de la fonction objective du modèle (**section 3.3.2.3**) est réalisée par l'algorithme génétique décrit en **section 3.3.4**.

Nous montrons ici les mêmes expérimentations que nous avons présentées pour le modèle BBSS, c'est à dire les systèmes de tailles générés pour les mensurations stature/poitrine et hanche/poitrine. L'ensemble des résultats obtenus pour les autres couples de mensurations et par morphotype est donné en **Annexe 4, 5 et 6**.

4.3.2.3.1 Système Stature/Poitrine

Les systèmes de tailles générés par le modèle BBSS sur les mensurations stature/poitrine sont présentés sur la **Figure 86**. La principale différence avec le modèle précédent est qu'il n'y a pas de grille prédéfinie pour la localisation des tailles. Ainsi, les systèmes obtenus s'adaptent plus précisément à la forme du nuage de points formés par les mensurations. De ce fait, lorsque le nombre de tailles augmente, le système obtenu devient assez complexe avec notamment des gradations non linéaires. La mise en œuvre industrielle pourrait donc s'avérer plus délicate. On notera cependant que le phénomène de "trous" dans le système de tailles, identifié sur le modèle précédent, n'est plus présent avec le modèle BBSS.

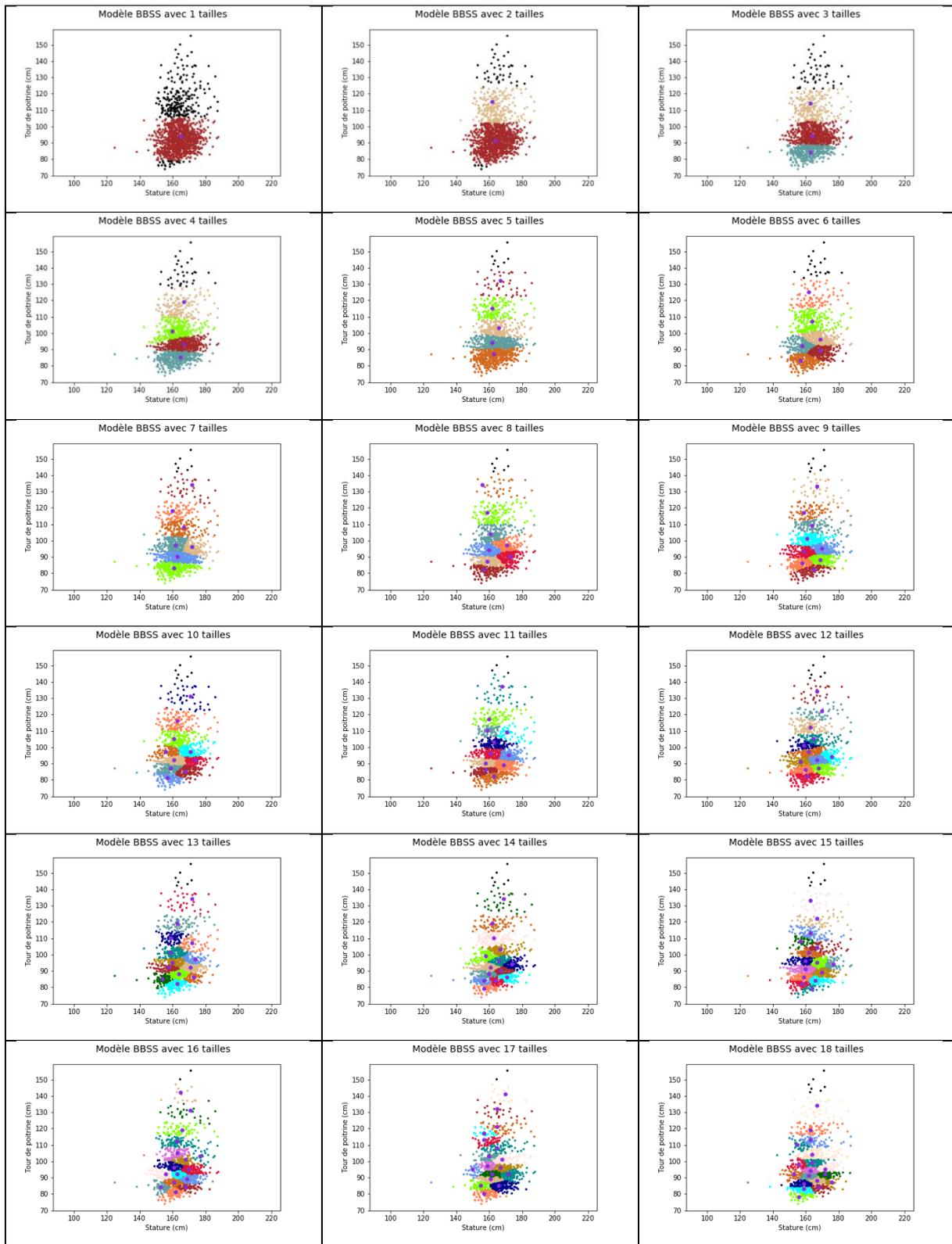


Figure 86. Systèmes de tailles générés par le modèle BBSS sur la population totale pour les mensurations stature/tour de poitrine

L'analyse des performances des systèmes de tailles du modèles BBSS sur la population et les deux morphologies confirme les résultats du modèle GBSS : les systèmes de tailles pour la morphologie 1 obtiennent des résultats moins satisfaisants (**Figure 87**). Pour ce modèle, deux

critères de performance sont présentés : le critère basé sur la distance de Manhattan, similaire au critère FL utilisé pour le modèle GBSS, et le critère P utilisé comme fonction objectif du modèle BBSS. Les courbes obtenues avec le critère de Manhattan présentent des points d'inflexion très prononcés : le choix du nombre de tailles est donc facilité.

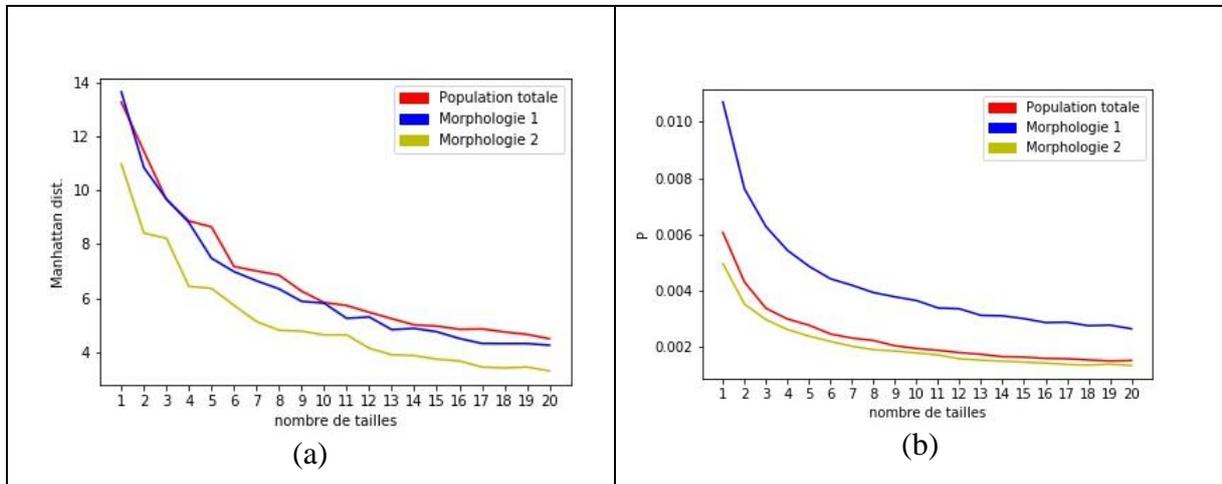


Figure 87. Evaluation par la distance de Manhattan (a) et la fonction objectif P (b) des systèmes des tailles obtenus avec le modèle BBSS pour les mensurations stature/poitrine. La **Figure 88** permet de comparer les performances des systèmes sur une même population. Les conclusions sont similaires au modèle GBSS : les systèmes développés spécifiquement pour les morphologies, et plus particulièrement pour la morphologie 1, obtiennent de meilleures performances. On notera également l'apparition d'optimum sur les courbes relatives au système global, et plus précisément une opposition des optimums entre les morphologies 1 et 2 (**Figure 89**). Cela pourrait traduire que le système parvient difficilement à optimiser le bien aller des vêtements sur les deux morphologies simultanément.

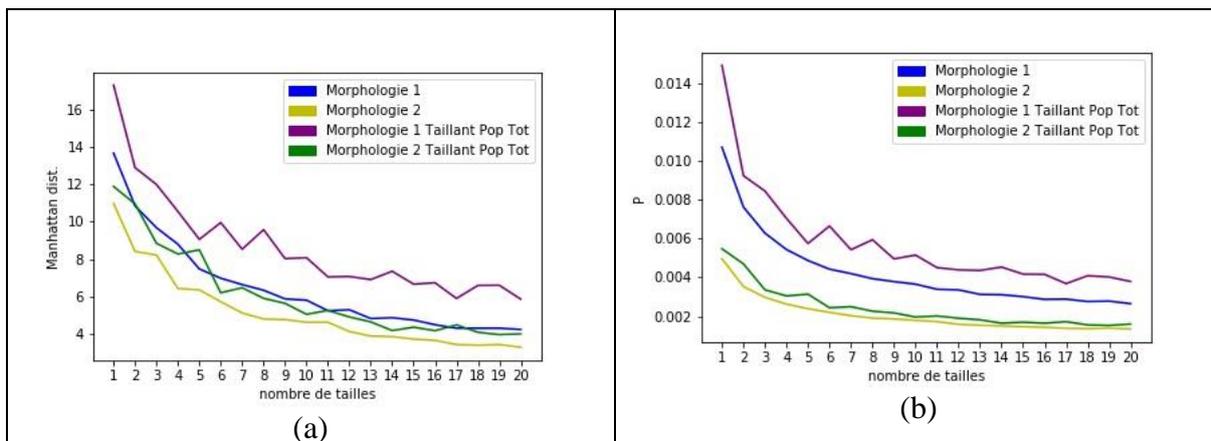


Figure 88. Comparaison des performances des systèmes de tailles générés pour la population totale et pour les deux morphologies avec les critères Manhattan (a) et P (b)

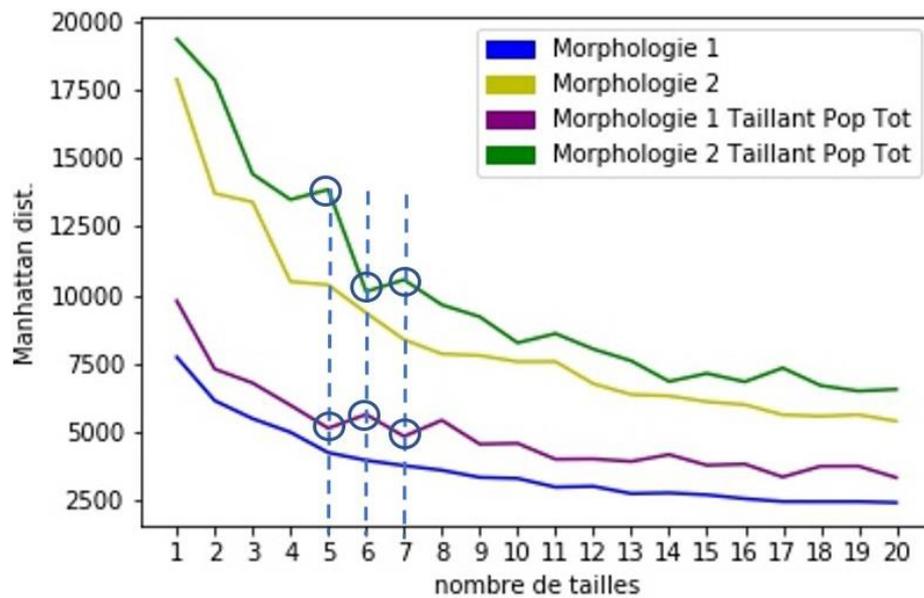
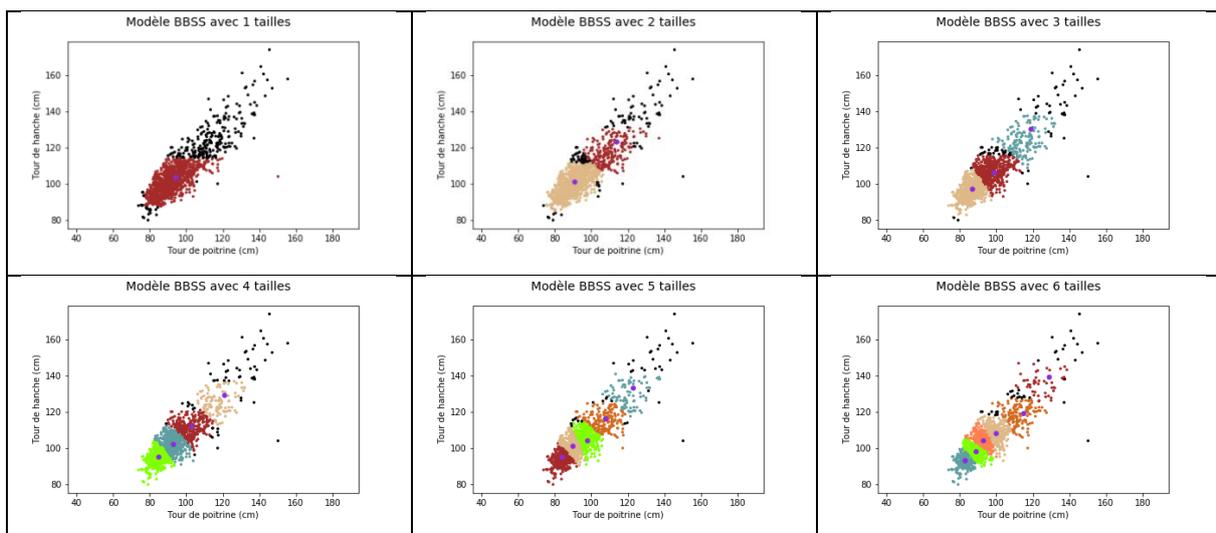


Figure 89. Exemples d’optimums locaux sur les morphologies 1 et 2 avec les systèmes de tailles générés sur la population totale

4.3.2.3.2 Système Poitrine/hanche

Cette section présente les systèmes de tailles obtenus avec le modèle BBSS sur les mensurations tour de hanche et tour de poitrine (**Figure 90**). L’analyse des résultats est sensiblement identique aux analyses précédentes : la capacité de l’algorithme à bien s’adapter aux caractéristiques du nuage de points des mensurations, l’absence de “trou” dans les systèmes de tailles, des systèmes de tailles plus performants lorsqu’ils sont générés par morphologie.



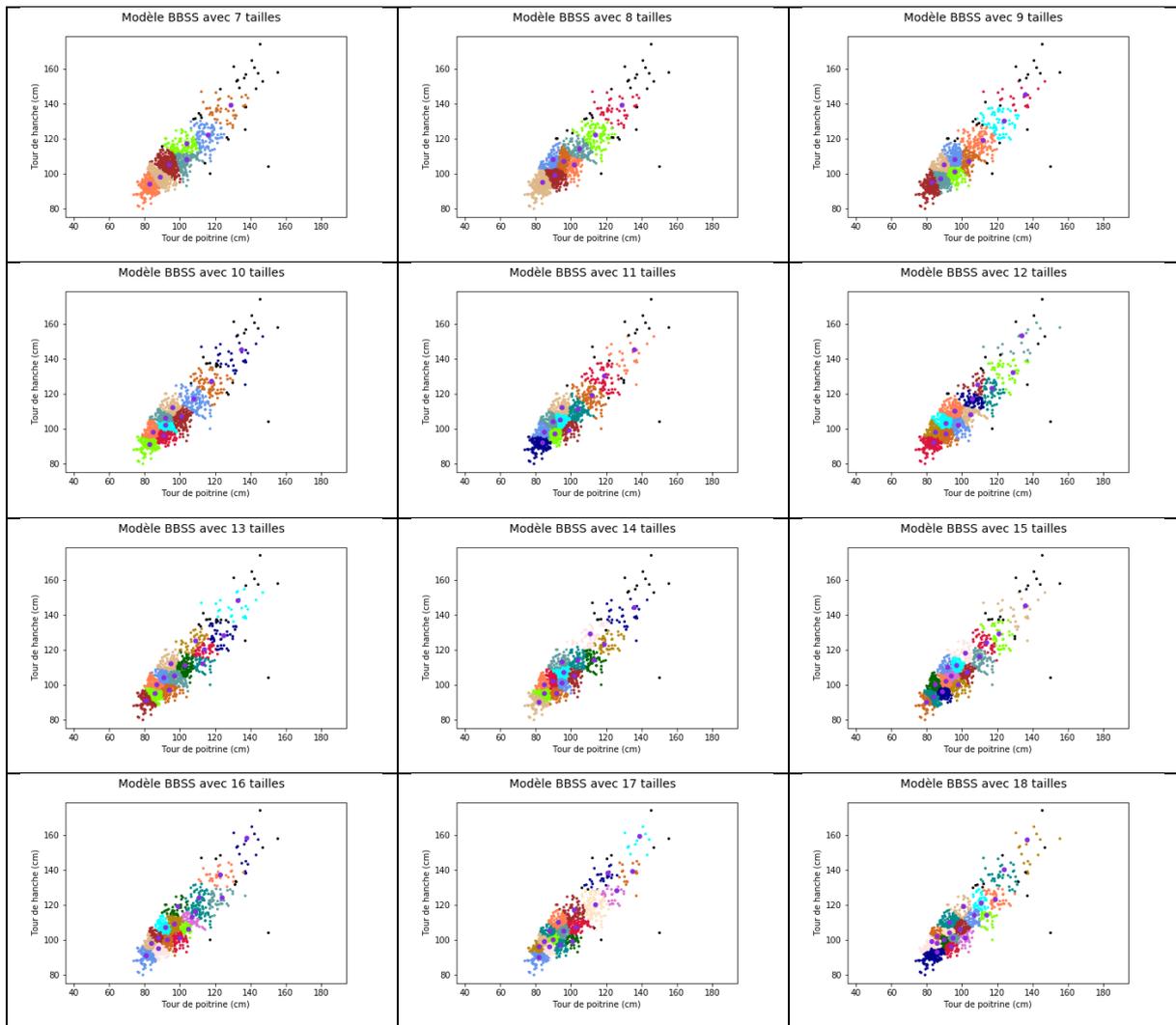


Figure 90. Systèmes de tailles générés par le modèle BBSS sur la population totale pour les mensurations tour de hanche/tour de poitrine

Les performances du système pour les mensurations tour de hanche/tour de poitrine (**Figure 91** et **Figure 92**) sont similaires à celle des mensurations stature/tour de poitrine.

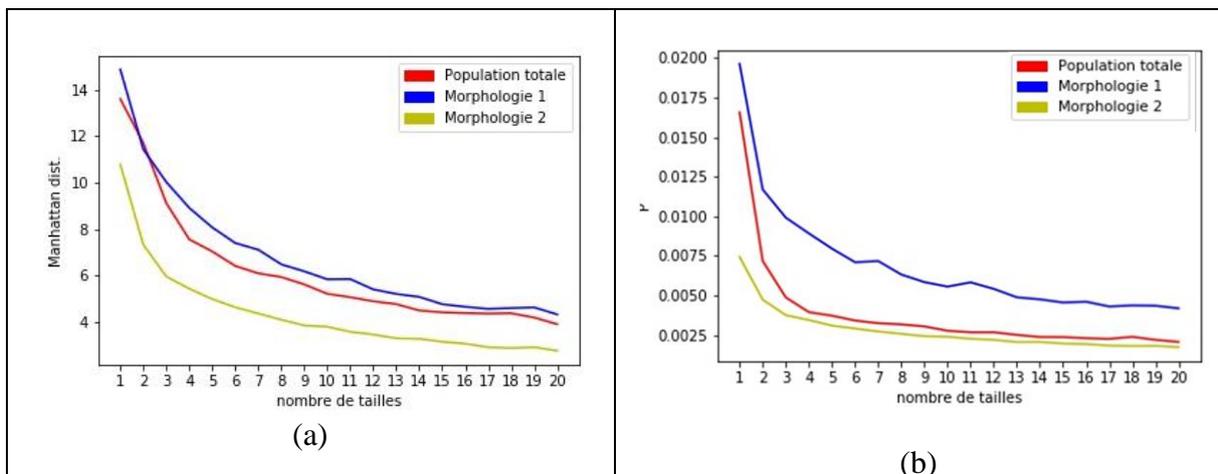


Figure 91. Évaluation par la distance de Manhattan (a) et la fonction objectif P (b) des systèmes des tailles obtenus avec le modèle BBSS pour les mensurations hanche/poitrine

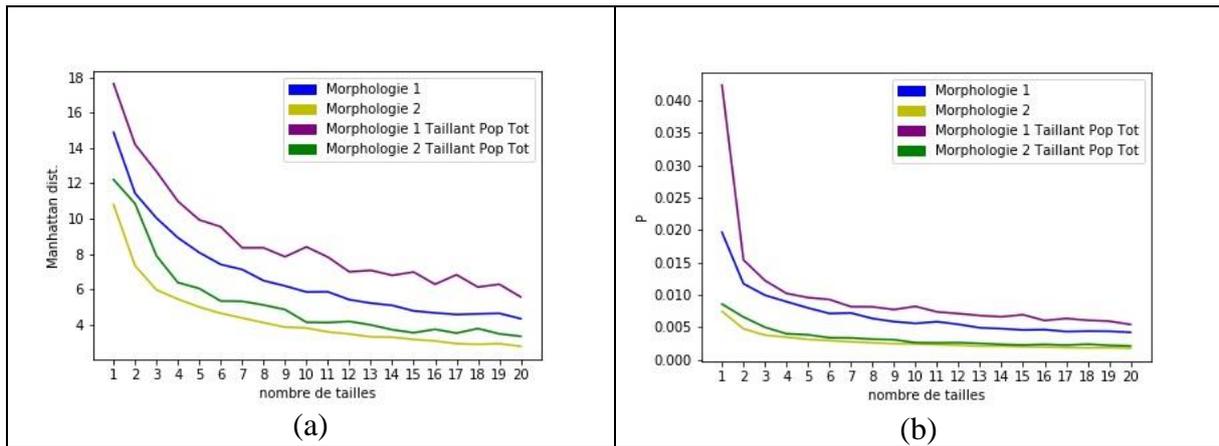


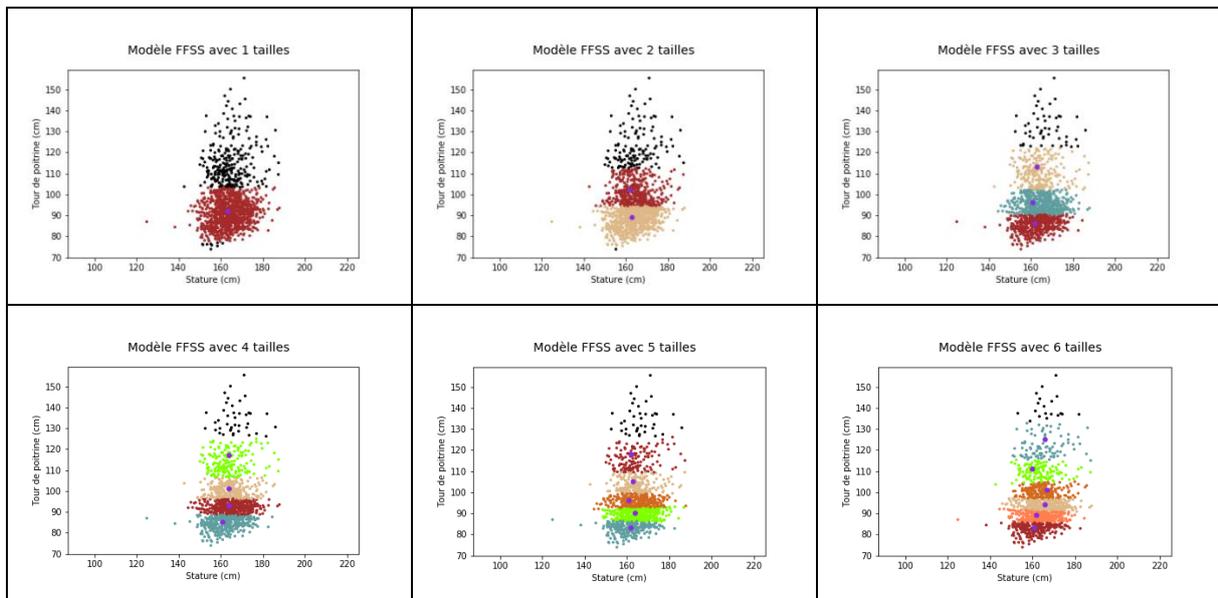
Figure 92. Comparaison des performances des systèmes de tailles générés pour la population totale et pour les deux morphologies avec les critères Manahattan (a) et P (b)

4.3.2.4 Modèle FFSS

Le modèle FFSS est également appliqué sur les mêmes données que le modèle BBSS. A nouveau, les résultats (**Figure 93** et **Figure 94**) confirment les analyses réalisées sur le modèle BBSS. Les sections 4.3.2.4.1 et 4.3.2.4.2 sont donc principalement illustratives.

A noter que pour ce modèle, la fonction fitness utilisée, Q , doit être maximisée (**Figure 95(a)** et **Figure 95 (b)**).

4.3.2.4.1 Système Stature/Poitrine



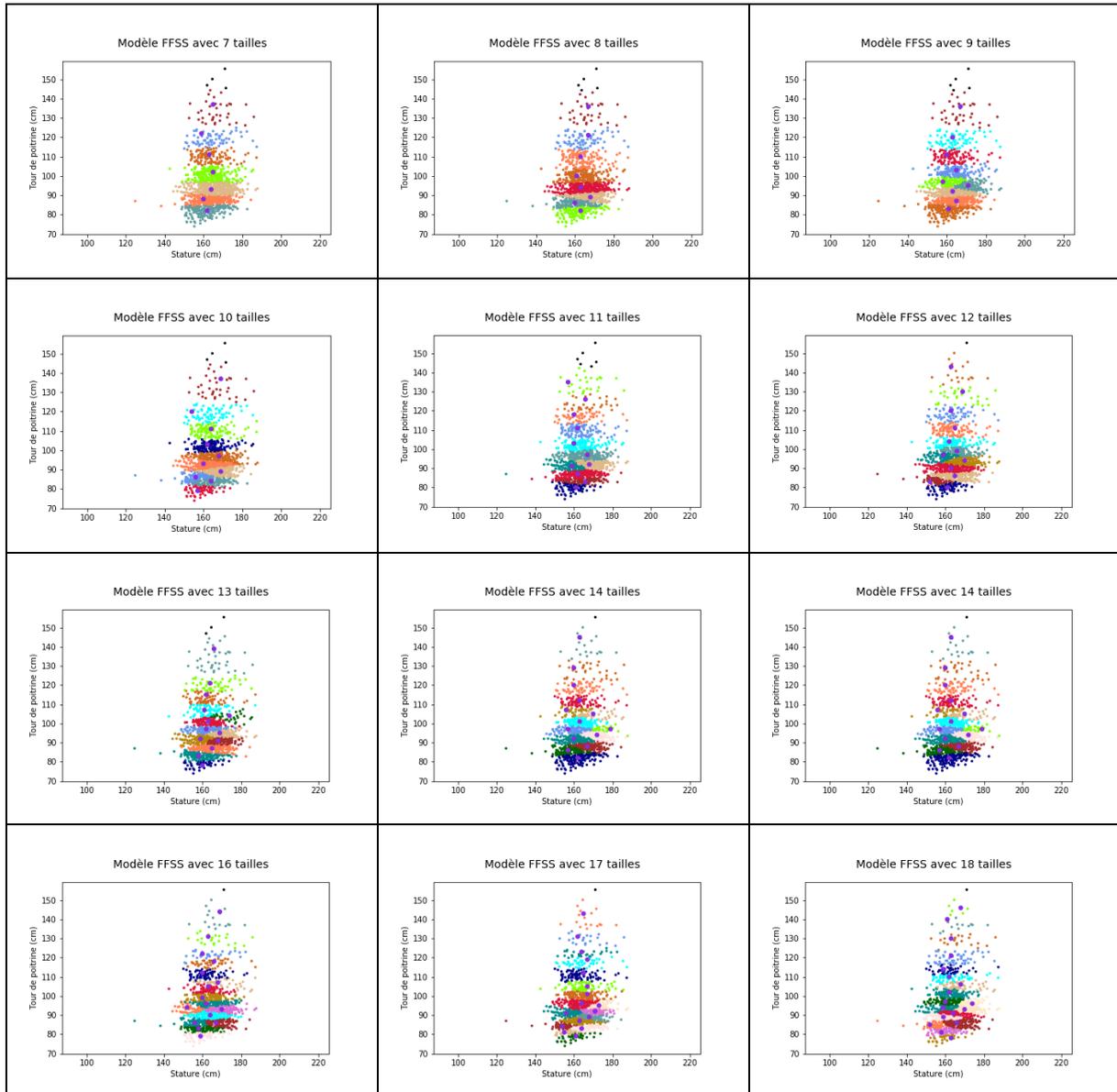


Figure 93. Systèmes de tailles générés par le modèle FFSS sur la population totale pour les mensurations stature/tour de poitrine

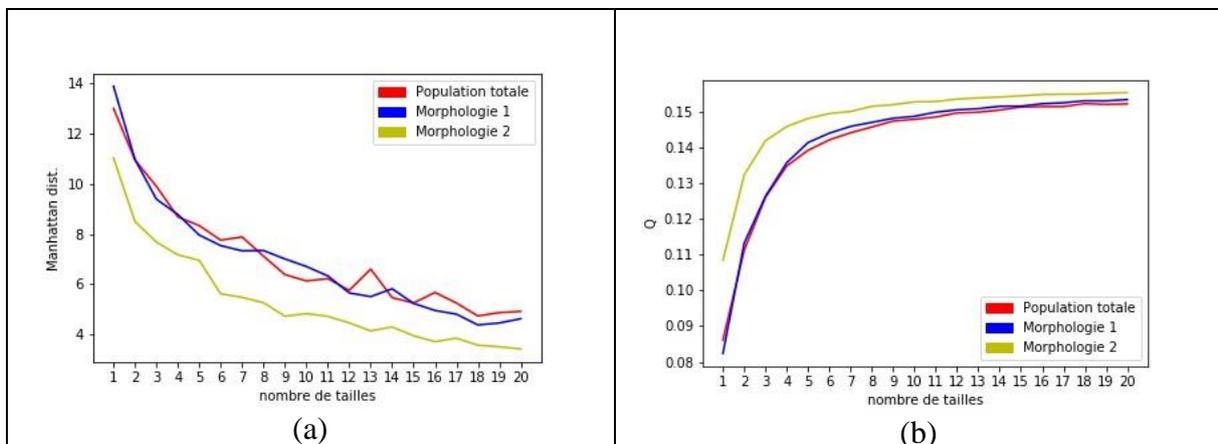


Figure 94. Évaluation par la distance de Manhattan (a) et la fonction objectif Q (b) des systèmes des tailles obtenus avec le modèle FFSS pour les mensurations stature/poitrine

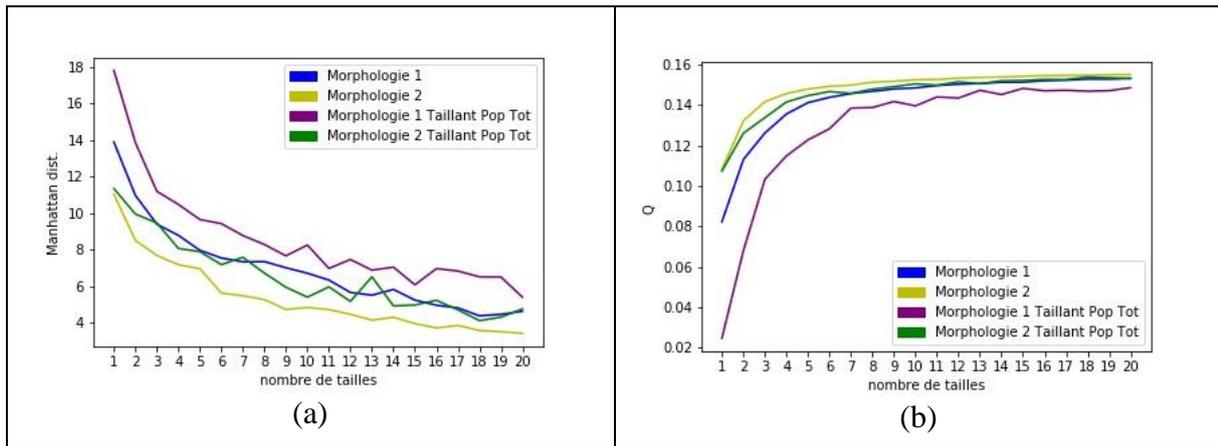
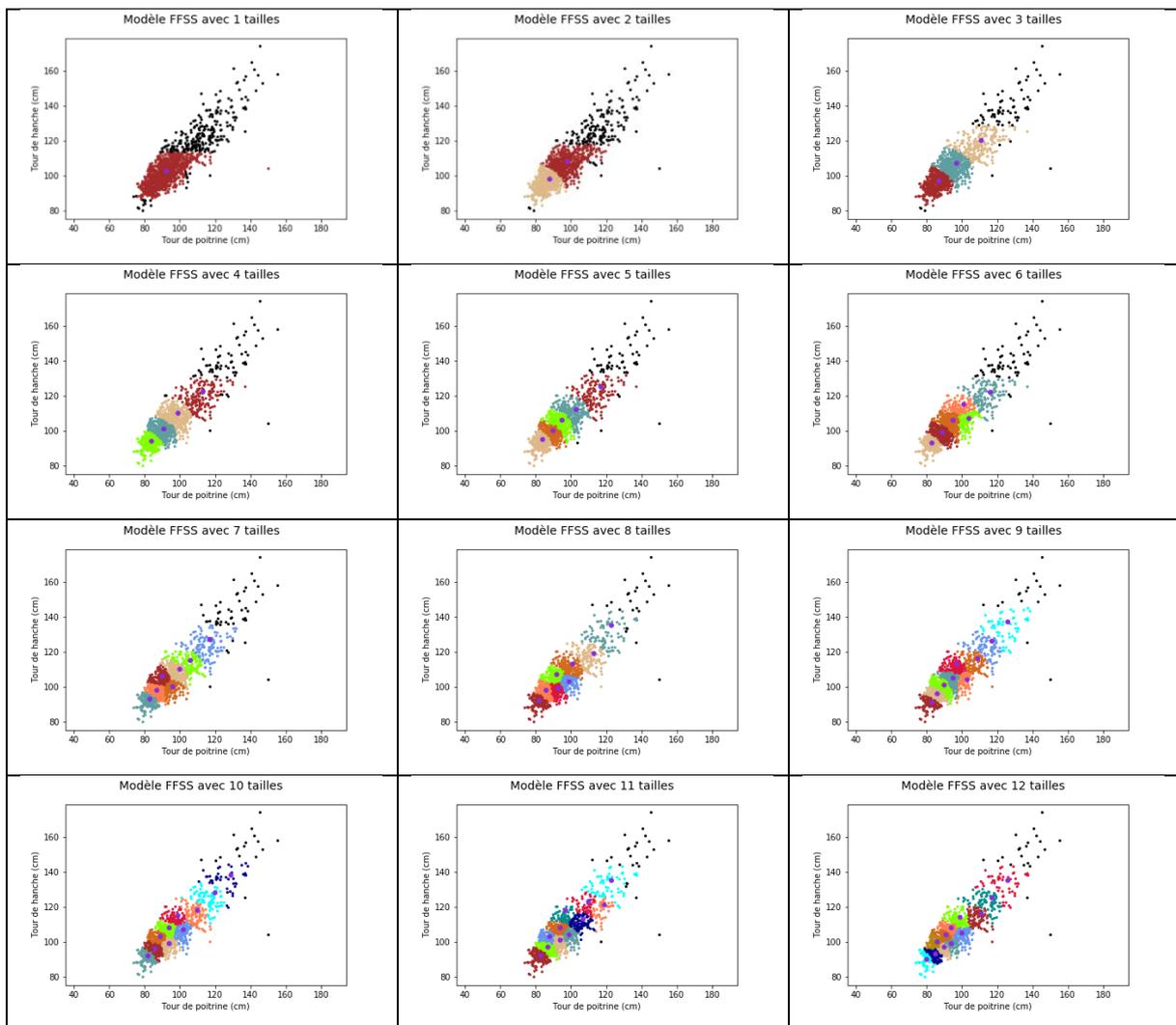


Figure 95. Comparaison des performances des systèmes de tailles générés pour la population totale et pour les deux morphologies avec les critères Manahattan (a) et Q (b)

4.3.2.4.2 Système Poitrine/Hanche



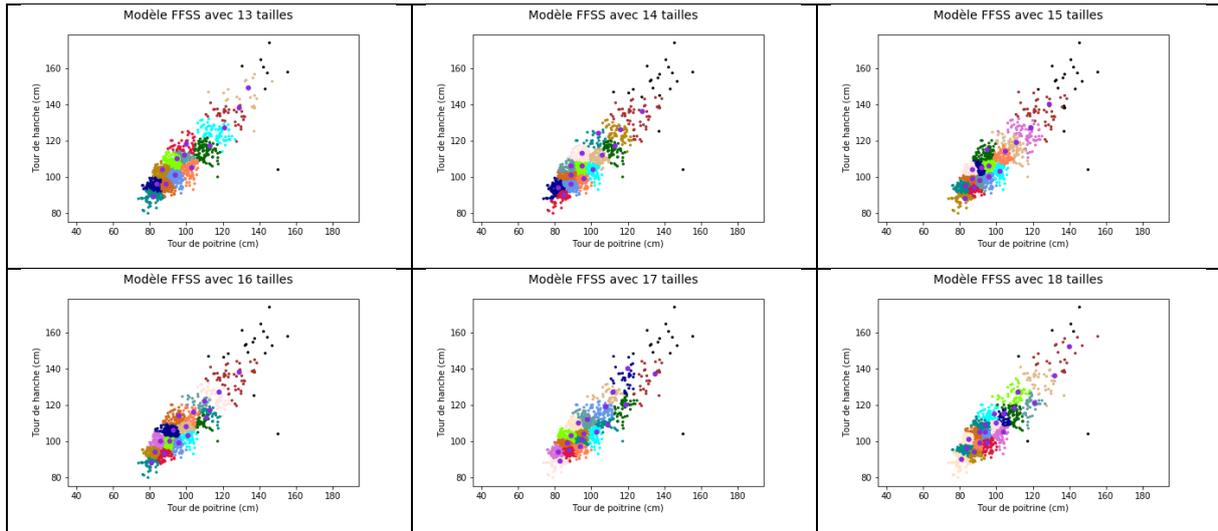


Figure 96. Systèmes de tailles générés par le modèle FFSS sur la population totale pour les mensurations tour de hanche/tour de poitrine

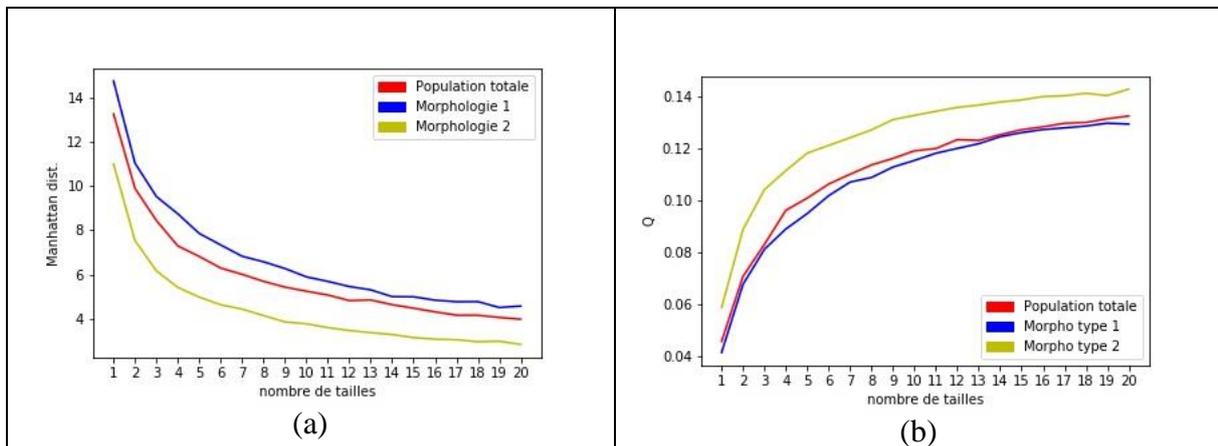


Figure 97. Evaluation par la distance de Manhattan (a) et la fonction objectif Q (b) des systèmes des tailles obtenus avec le modèle FFSS pour les mensurations hanche/poitrine

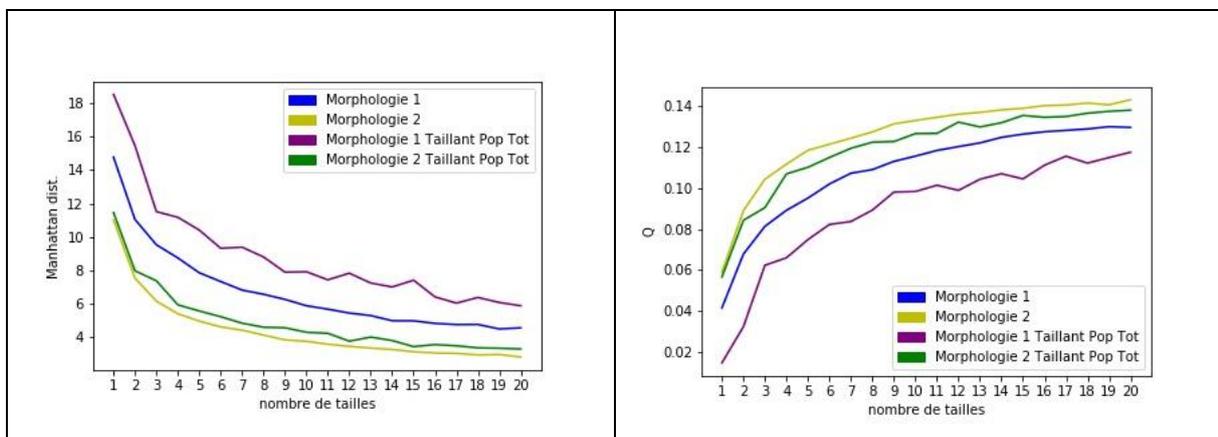


Figure 98. Comparaison des performances des systèmes de tailles générés pour la population totale et pour les deux morphologies avec les critères Manhattan (a) et Q (b)

4.3.2.5 Discussion sur les modèles BBSS et FFSS

Les modèles BBSS et FFSS apportent deux évolutions majeures : l'intégration des tolérances clients et l'optimisation de la construction du système de tailles.

Les bénéfices de la prise en compte des tolérances clients sont difficilement quantifiables à notre niveau puisque nos évaluations portent sur la capacité du système à répondre aux mensurations d'une population globale. La mesure des effets de cette approche nécessiterait de mettre en place les systèmes de tailles proposés et d'effectuer une nouvelle enquête de satisfaction.

Les avantages de l'algorithme d'optimisation utilisé sur les modèles BBSS et FFSS peuvent être constatés en étudiant les systèmes proposés pour différentes tailles : cet algorithme permet de s'adapter au nuage de points des mensurations considérées en évitant le phénomène de trou dans lors de la construction du système. Par ailleurs, il est possible de quantifier les gains en comparant les différents critères d'évaluation des trois modèles proposés. Cette comparaison est effectuée dans la section suivante.

Enfin, il faut noter que les conclusions sur l'impact de la classification morphologique en amont du système de tailles ne dépendent pas du modèle. Seules les capacités des modèles BBSS et FFSS à mieux s'adapter aux caractéristiques d'une population peuvent assurer un léger avantage à la création de systèmes spécifiques aux morphologies plutôt qu'un système global.

4.3.3 Comparaison des modèles GBSS, BBSS et FFSS

Les trois modèles proposés peuvent être comparés qualitativement, comme réalisé dans le chapitre 3 en **Figure 57**: le modèle GBSS présente des avantages de mise en œuvre industrielle alors que le modèle FFSS permet une meilleure prise en compte des données clients.

Les diverses simulations réalisées sur nos données permettent également une comparaison quantitative des modèles. Chaque modèle possède sa propre fonction objectif. Il n'est pas possible d'affirmer qu'une fonction permet une meilleure évaluation d'un système de tailles. En effet, chacune présente ses avantages, discutés dans le chapitre 3. Nous avons donc comparé l'ensemble des modèles avec l'ensemble des fonctions objectif sur chacune des populations : morphologie 1, morphologie 2 et population totale.

Les résultats des modèles BBSS et FFSS sont très similaires en termes de performance quels que soient le critère choisi et la population considéré. De façon

générale, le modèle GBSS obtient également des résultats assez proches mais toutefois légèrement inférieurs notamment sur les tailles médianes, de 3 à 12.

Les **Figure 99**,**Figure 100**,**Figure 101** justifient ces remarques pour différentes combinaisons de mensurations sur différentes populations.

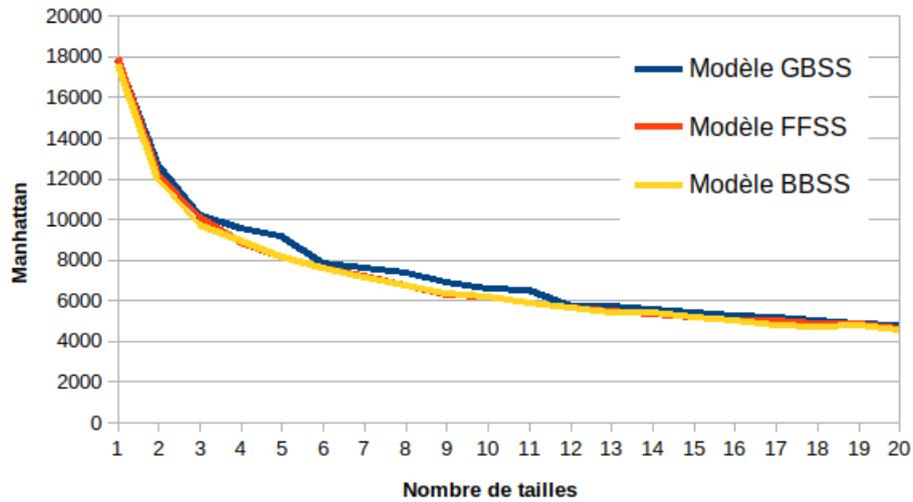


Figure 99. Comparaison des trois modèles avec le critère Manhattan sur la morphologie 2 et les mensurations hanche/poitrine

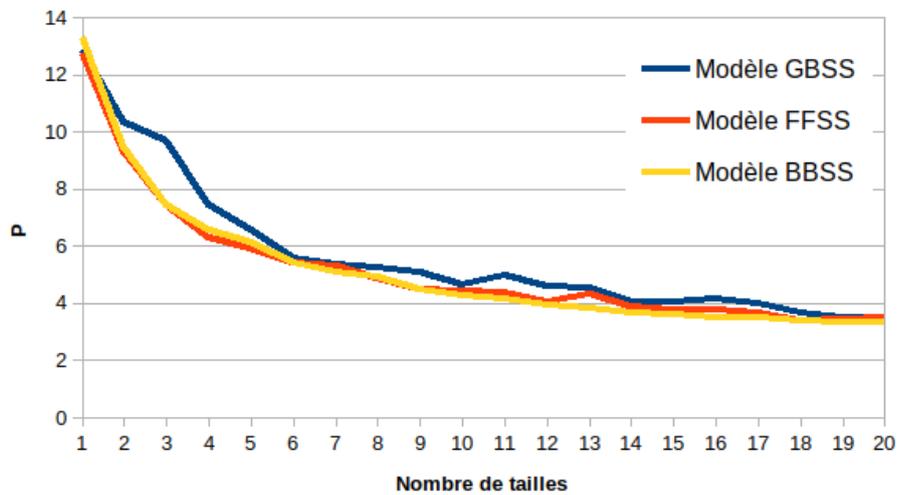


Figure 100. Comparaison des trois modèles avec le critère P sur la population totale et les mensurations stature/poitrine

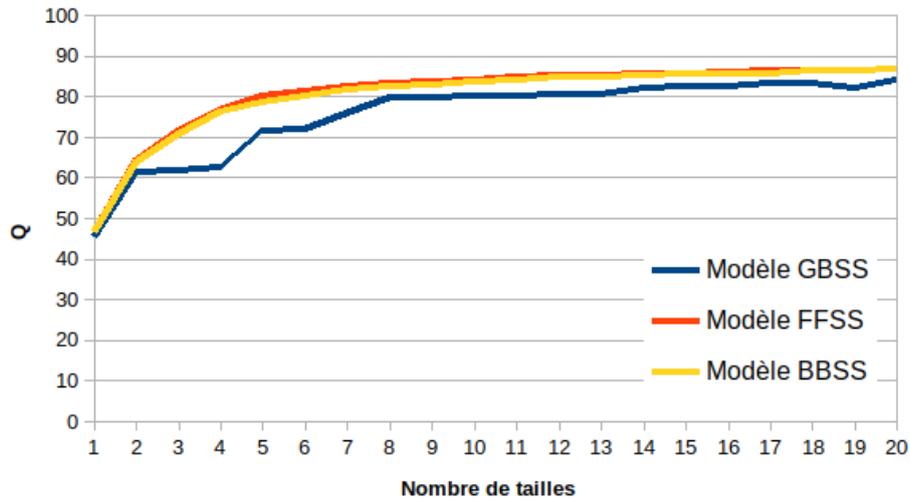


Figure 101. Comparaison des trois modèles avec le critère Q sur la morphologie 1 et les mensurations stature/poitrine

4.4 Conclusion

Dans ce chapitre, nous avons mis en œuvre l'ensemble de la méthodologie proposée sur des mensurations issues d'une base de données du projet CAESAR. Notre méthodologie s'appuie sur deux étapes principales : la classification morphologique de la population et l'optimisation du système de tailles.

Pour la classification morphologique, nous avons procédé, dans un premier temps, à une sélection des attributs afin d'identifier les mensurations les plus pertinentes pour caractériser une morphologie. Les résultats obtenus par ACP montrent que deux catégories d'attributs se distinguent nettement : les mensurations basées sur les circonférences et les mensurations caractérisant les hauteurs. Ainsi, en accord avec les pratiques du métier, nous avons sélectionné les mensurations dans ces deux catégories : tour de poitrine, tour de hanche, tour de taille et stature. Ceci constitue une première contribution par rapport aux pratiques courantes de développement de système de tailles qui n'utilise pas la stature comme paramètres du système. Dans un deuxième temps, nous avons appliqué une classification non supervisée avec la méthode des k-moyennes++ sur les attributs sélectionnés et normalisés. Le nombre optimal de classes sur notre jeu de données selon les indices de David-Bouldin et Silhouette est de 2. Cependant, il apparaît qu'il n'y a pas de séparation claire entre les classes. Les mensurations ne sont pas suffisantes pour distinguer clairement les morphologies et d'autres sources d'informations, telles que des données 3D pourraient permettre une meilleure classification. Néanmoins, les deux morphologies obtenues peuvent être

considérées comme satisfaisantes, notamment pour les experts du métier, avec d'une part des individus à fortes corpulences et d'autres part des individus à corpulences de moyennes à minces.

Pour le développement de systèmes de tailles, nous avons testé les trois modèles proposés au chapitre 3 : GBSS, BBSS et FFSS. Ces modèles ont été appliqués sur l'ensemble de la population puis sur chacune des morphologies identifiées précédemment. Différentes combinaisons de mensurations sont testées et présentées dans ce chapitre ou en annexe : poitrine/stature, poitrine/hanche, poitrine/taille.

Le modèle GBFF, qui s'appuie sur une grille définie à partir de l'étendue des valeurs des mensurations considérées, nécessite de simuler différentes combinaisons de pas. L'algorithme développé pour ce modèle est basé sur l'optimisation d'une fonction objectif, appelée FL, qui quantifie à la fois le bien aller des individus et le taux de couverture de la population. L'application de cet algorithme sur les combinaisons de pas permet d'obtenir un front de Pareto qui définit l'ensemble des systèmes de tailles proposant le meilleur compromis entre le critère FL et le nombre de tailles du système. L'analyse des systèmes obtenus montre que notre modèle permet de bien prendre en compte la forme du nuage de points formé par les mensurations des individus de la population considérée. Cependant, à partir d'un certain nombre de tailles, le système présente un phénomène de trou dans l'évolution des tailles, ce qui pourrait nuire à la mise en œuvre industrielle notamment en termes de gradation. Enfin, nous avons montré que la génération de système de tailles spécifiques à chaque morphologie améliore le critère FL quel que soit le nombre de tailles. Ce gain, par rapport à un système de tailles global, est très significatif sur la population des individus de morphologie très corpulente. Cela peut s'expliquer par le fait que cette morphologie est assez minoritaire dans la population globale (environ 25%) et surtout par les proportions entre les mensurations de cette morphologie qui sont très irrégulières.

Les modèles BBSS et FFSS apportent deux évolutions principales au modèle précédent : ils intègrent des tolérances issues de satisfactions clients et s'appuient sur un algorithme d'optimisation plus performant, totalement basé sur les données, sans définition de grille et plus robuste aux phénomènes d'optimums locaux. Afin de définir les tolérances clients, nous avons utilisé les données issues d'une enquête de satisfaction client menée par une enseigne de prêt à porter féminin. A partir de la distribution des mensurations des client satisfaits par rapport à leurs tailles choisies, nous avons défini des probabilités de bien aller à partir de lois semi-normales. En effet, les données de l'enquête confirment que les tolérances des clients

dépendent de la taille choisie et de la mensuration du client par rapport à cette taille. En d'autres termes, les clients sont plus tolérants sur les grandes tailles et lorsque le vêtement est légèrement trop grand. Enfin, les écarts types des densités de probabilités des lois semi-normales, qui caractérisent les tolérances, sont utilisés pour le développement des systèmes de tailles. Les résultats obtenus par les deux modèles sont relativement similaires. Il apparaît que l'algorithme génétique utilisé pour l'optimisation du système de tailles permet une meilleure prise en compte des mensurations de la population considérée en évitant le phénomène de trou lors de la construction du système.

Finalement, l'analyse globale des résultats montre que le bénéfice de créer des systèmes de tailles par morphologie est réel. Cependant, le choix entre l'utilisation d'un système de tailles pour une population globale et la création d'un système de tailles par morphologie reste à discuter en fonction du nombre total (ou maximum) de tailles de vêtements souhaité et de la population considérée (nombre et type de morphologies présentes dans la population).

Conclusion générale

Les enseignes de prêt-à-porter sont de plus en plus confrontées à un dilemme très fort qui est d'adapter les produits aux besoins des consommateurs et d'éviter une surproduction ou des retours sur invendus. De nos jours, l'acte d'achat du consommateur peut être assuré si le produit lui va bien car le style du produit est en général validé en amont de la vente lorsqu'il a choisi le produit. Le bien aller d'un produit dépend de la matière, du confort qu'il apporte, mais aussi de son taillant, c'est-à-dire sa capacité à suivre la morphologie du client selon les critères définis par le styliste et de ses propres critères. Le système de tailles permet d'adapter le vêtement à la morphologie du consommateur selon des règles définies par le styliste et son équipe du bureau d'étude, donc propre à la marque.

En pratique, la plupart des enseignes créent leurs modèles sur une morphologie standard et utilise ce système de tailles pour altérer ce modèle vers les autres tailles à partir des patronages 2D. Le problème qui se pose est qu'aucun lien n'existe entre cette morphologie standard et ce système de taille. Entre autres, deux cas peuvent se poser : soit que l'enseigne n'utilise qu'un système de tailles évoluant de taille en taille pour une stature unique et moyenne, soit qu'elle prend en compte en plus la variation de la stature ce qui amène autant de système de tailles en plus que de statures envisagées. Le premier cas limite la vente du produit à une population ciblée par le bien aller du produit. Le deuxième cas permet d'élargir la population en assurant ce bien aller, mais conduit à créer des produits qui ne seront peut-être jamais achetés si l'on reproduit toutes les tailles du premier système de tailles pour toutes les statures envisagées. Ceci amène forcément à une surproduction ou à des retours de produits.

L'objectif de cette thèse a donc été de mettre en place une méthodologie qui combine l'analyse morphologique des consommateurs et l'optimisation de systèmes de tailles. Cela se traduit par l'extraction de morphotypes à partir d'une population et la mise en place de systèmes de tailles propres à chaque morphotype.

Le premier chapitre dresse, dans un premier temps, un état de l'art sur les méthodes d'analyse anthropométriques et plus précisément les techniques de caractérisation des morphologies humaines. Dans un second temps, les méthodes de construction de systèmes de tailles existantes dans la littérature sont présentées. A la lumière de cet état de l'art, nous constatons que ces méthodes ne permettent pas d'intégrer de façon efficace les spécificités des morphologies dans les systèmes de tailles actuels. Ainsi, nous proposons à la fin de ce

chapitre, les deux axes de notre méthodologie : la classification morphologique et l'optimisation des systèmes de tailles.

Le second chapitre présente les outils nécessaires à la mise en œuvre de notre méthodologie. L'extraction de morphologie à partir d'une population et plus particulièrement à partir d'une base de données de mensurations requiert l'utilisation de différentes techniques : la sélection des attributs pertinents, la normalisation des données et la classification non supervisée. Pour le développement des systèmes de tailles, nous nous sommes orientés vers les techniques d'optimisation multicritère, et plus particulièrement les concepts d'optimum de Pareto et de méta-heuristiques. A la connaissance de ces fondements théoriques, nous avons sélectionnées les méthodes les plus pertinentes pour mettre en place une méthodologie spécifique à notre problématique.

Le troisième chapitre regroupe l'ensemble des outils mathématiques qui a permis de résoudre différents problématiques. Le premier problème a été de regrouper les clients de morphologies similaires par une méthode de classification non-supervisée de type K-moyennes ++ avec au préalable une sélection des mensurations pertinentes par une ACP dont les résultats sont validés et justifiés par le score Laplacien. Le second problème a été créé des systèmes de tailles pour chaque classe morphologique. L'optimisation de ces systèmes de tailles a été effectuée selon trois modèles.

Le premier modèle GBSS conduit à un système de taille linéaire défini par des tolérances de tailles subjectives. Ce système est optimisé par la mesure du bien aller des vêtements. Le deuxième modèle BBSS intègre en complément une enquête clients pour prendre en compte leurs exigences sur le bien aller du vêtement. Ce facteur intervient dans la définition des tolérances des tailles. Un algorithme génétique est utilisé pour améliorer la convergence des résultats vers un optimal global. Ce modèle conduit à un système de taille non linéaire. Le dernier modèle FFSS utilise également l'algorithme génétique du modèle BBSS. Cependant, il s'appuie sur une fonction objectif basée sur une distribution semi-normale centrée sur la taille du vêtement pour définir les tolérances de tailles. Ce modèle permet ainsi une prise en compte plus exhaustive des mensurations de la population.

Le quatrième chapitre présente l'ensemble des résultats obtenus sur les données issues de la base de mensurations CAESAR et d'une enquête de satisfaction client effectuée par une enseigne de prêt à porter féminin. Sur ces données, notre processus de classification morphologique a permis d'identifier deux classes morphologiques : une première classe regroupe les individus de forte corpulence, la seconde classe regroupe les individus de

corpulence moyenne à mince. Les trois modèles proposés sont ensuite testés et évalués sur la population totale de la base de données, puis sur chacune des morphologies. Le premier modèle GBSS, montre de bonnes performances en termes de prise en compte des mensurations de la population cible. Il permet également de montrer que la mise en œuvre des systèmes de tailles par morphologies s'avère pertinente notamment dans le cas des morphologies à forte corpulence. Les deux autres modèles, BBSS et FFSS, sont similaires en termes de performances et permettent de pallier les inconvénients du modèle GBSS, à savoir : l'apparition de "trous" dans le système de tailles et la convergence vers un optimum local. Cela s'explique principalement par les capacités de l'algorithme génétique développé pour l'optimisation des systèmes de tailles. L'intégration des tolérances clients issues de l'enquête de satisfaction permet également d'obtenir des systèmes de tailles moins subjectifs et plus orientés vers les exigences des consommateurs. Enfin, l'ensemble des résultats obtenus montrent que la classification morphologique en amont de la construction du système de tailles apporte des bénéfices systématiques en termes de bien aller. Cependant, en fonction du nombre total de tailles à utiliser, le choix entre un système global pour l'ensemble de la population et plusieurs systèmes de tailles pour chacune des morphologies est à discuter. En effet, nous montrons sur nos données qu'en dessous d'un certain nombre de tailles, la multiplication des systèmes n'est pas un choix pertinent.

Ainsi, ces travaux de recherche ont apporté les contributions suivantes :

- Une classification non-supervisée des morphologies humaines à partir d'une base de données de mensurations. Cette classification s'appuie sur une sélection des mensurations puis une normalisation des mesures relatives au torse basée sur la technique du cercle englobant avant d'être exploitables dans l'algorithme k-moyennes avec les indices d'évaluations de partitions pour choisir le nombre optimal de classes k.
- Trois modèles de construction des systèmes de tailles qui diffèrent par leur fonction objectif. Les modèles GBSS et BBSS s'appuient sur la minimisation des distances entre les tailles générées et les clients. Ils permettent de converser des contraintes métier telles que la définition précise de la couverture des tailles. Le modèle FFSS se concentre sur la maximisation de probabilités que les clients soient satisfaits par leur taille générée la plus proche. Il permet une meilleure prise en compte de la sensibilité des clients. C'est une évolution centrée clients.

- L'utilisation du front de Pareto comme un outil d'aide à la décision afin de choisir le nombre de tailles en fonction du bien aller et inversement.
- Une investigation sur le niveau de satisfaction des clients sur les tailles de vêtements existantes d'une enseigne par le biais d'une enquête. Pour cela un modèle prédictif pour estimer les valeurs des mensurations manquantes et un modèle d'estimations des tolérances ont été développés.
- Des améliorations apportées dans les étapes de l'algorithme génétique : L'utilisation de la carte auto-organisatrice dans l'étape d'initialisation des chromosomes initiaux pour générer des tailles admissibles de bonne qualité et l'ajout d'une étape de vérification dans l'opérateur de croisement de l'algorithme génétique afin de maintenir la diversification des tailles générées.

Plusieurs perspectives de recherche peuvent être envisagées suite à ces travaux.

Pour la classification morphologique :

- Les mensurations ne permettent pas d'identifier clairement les classes. Il serait pertinent d'utiliser d'autres sources d'information pour caractériser les morphologies telles que par exemple des données 3D du corps humain.
- Le choix du nombre optimal de classes pourrait être réalisé en association avec la construction du système de tailles. Ainsi, cette méthode permettrait une optimisation globale de l'ensemble des étapes de notre méthodologie.
- L'augmentation du nombre de tailles dans l'enquête clients permettrait de créer un modèle de régression plus précis pour quantifier les tolérances.

Pour l'optimisation des systèmes de tailles :

- La construction des systèmes de tailles pourrait s'appuyer sur trois mensurations simultanément afin de mieux caractériser les évolutions morphologiques.
- Le système de tailles doit être déployé sur la partie basse du corps afin d'élargir la gamme de produits.
- L'intégration de critères économiques pourrait également orienter la construction du modèle. Par exemple, le gain potentiel de client actuellement non couvert et/ou la réduction du nombre de retours pourrait être intégré dans les fonctions objectif du système.

Enfin, l'ensemble de ces travaux pourraient mener au développement des modèles adaptatifs de mannequin basé sur les morphotypes de chaque classe et paramétrés par les systèmes de

tailles [133]. Ces mannequins seraient très utiles au bureau de création pour valider les prototypes des modèles de vêtements.

Liste des publications

Revue internationale à comité de lecture (publié) :

- 1) B. Hamad, M. Hamad, S. Thomassey, and P. Bruniaux, “3D Adaptive Morphotype Mannequin for Target Population,” *J. Ergon.*, vol. 8, no. 2, 2018.

Revue internationale à comité de lecture (en cours de soumission) :

- 1) B. Hamad, O. Jafoor, M. Hamad, S. Thomassey, and P. Bruniaux, “Consumer-Oriented Sizing System with Morphology Clustering,” *IEEE Technology and Engineering Management*, 2020.

Conférences internationales à comité de lecture (publié) :

- 1) M. K.Mishra, B. Hamad, D. Crepin, and P. Bruniaux, “Studying foot anthropometric measures using 3d scanning method,” *CLOTECH2017*, 11th International Conference on Innovative Materials & Technologies in Made-up Textiles Articles, Protective Clothing and Footwear, Lodz, Pologne, 11-14 Octobre,2017.
- 2) B. Hamad, S. Thomassey, and P. Bruniaux, “ Classification of women’s body based on extracting anthropometric measurements from 3D body scanning”, *ITMC2017*,International Conference on Intelligent Textiles and Mass Customisation, Ghent, Belgium, October 16-18, 2017.
- 3) B. Hamad, S. Thomassey, and P. Bruniaux, “A method to reduce the number of sizes in apparel industry”, *ITMC2019*,International Conference on Intelligent Textiles and Mass Customisation, Marrakech, Morocco, November 13-14, 2019.

Conférences nationale :

- 1) B. Hamad, S. Thomassey, and P. Bruniaux, “Une méthodologie de construction de système de tailles optimisé,” *JRDA2019* ,6 ème Journée Régionale des Doctorants en Automatique, Lille ,4 Juillet 2019.

BIBLIOGRAPHIE

- [1] H. A. M. Daanen and A. Psikuta, *3D body scanning*, vol. 1, no. 2017. Elsevier Ltd, 2017.
- [2] N. D'Apuzzo, "3D body scanning technology for fashion and apparel industry," 2007, vol. 6491, p. 649100.
- [3] J. Bougourd and P. Treleaven, "National size and shape surveys for apparel design," *Anthr. Appar. Sizing Des.*, pp. 141–166, Jan. 2014.
- [4] "CAESAR: Civilian American and European Surface Anthropometry Resource Project." [Online]. Available: <http://store.sae.org/caesar/>. [Accessed: 14-Jun-2019].
- [5] IFTH, "Résultats de la Campagne National de Mensuration," 2006.
- [6] "SizeGERMANY - Startseite." [Online]. Available: <https://portal.sizegermany.de/SizeGermany/pages/home.seam>. [Accessed: 14-Jun-2019].
- [7] "National Size Survey." [Online]. Available: <http://www.size.org/>. [Accessed: 14-Jun-2019].
- [8] "Size North America - Bienvenue." [Online]. Available: http://www.sizenorthamerica.com/cms/front_content.php?idart=114&lang=3. [Accessed: 11-Jun-2019].
- [9] Z. BEN AZOUZ, "ANALYSE DE LA VARIABILITÉ DE LA FORME DU CORPS HUMAIN À PARTIR DE DONNÉES ANTHROPOMÉTRIQUES TRIDIMENSIONNELLES," ÉCOLE DE TECHNOLOGIE SUPÉRIEURE UNIVERSITÉ DU QUÉBEC, 2005.
- [10] J.-B. DJIRE, "EXTRACTION AUTOMATIQUE DE MESURES SUR DIFFÉRENTES POSTURES DE MANNEQUINS 3D DE LA BASE CAESAR," 2016.
- [11] K. M. Robinette, S. Blackwell, H. Daanen, M. Boehmer, and S. Fleming, "Civilian American and European Surface Anthropometry Resource (CAESAR), Final Report. Volume 1. Summary," 2002.
- [12] W. Lee *et al.*, "A 3D anthropometric sizing analysis system based on North American CAESAR 3D scan data for design of head wearable products," *Comput. Ind. Eng.*, vol. 117, pp. 121–130, Mar. 2018.
- [13] K. M. Robinette and H. A. M. Daanen, "Precision of the CAESAR scan-extracted measurements," *Appl. Ergon.*, vol. 37, no. 3, pp. 259–265, 2006.
- [14] A. Godil and S. Ressler, "Retrieval and clustering from a 3D human database based on body and head shape," *arXiv Prepr. arXiv1105.2800*, 2011.
- [15] M. Hamad, "Classification non-supervisée de morphologies 3D de corps humain pour la mise en œuvre de mannequins morphotypes adaptatifs," <http://www.theses.fr>, Jun. 2015.
- [16] T. Xiaohui, P. Xiaoyu, L. Liwen, and X. Qing, "Automatic human body feature extraction and personal size measurement," *J. Vis. Lang. Comput.*, vol. 47, pp. 9–18,

Aug. 2018.

- [17] D. Gupta and N. Zakaria, *Anthropometry, apparel sizing and design*. .
- [18] M. TIWARI, “SIZE CHART STANDARDIZATION OF BOTTOM-WEAR FOR INDIAN MALE YOUTH (18-29 YEARS),” NATIONAL INSTITUTE OF FASHION TECHNOLOGY, 2017.
- [19] “ISO 8559:1989(fr), Confection des vêtements et relevés anthropométriques — Mesures du corps humain.” [Online]. Available: <https://www.iso.org/obp/ui#iso:std:iso:8559:ed-1:v1:fr>. [Accessed: 17-Jun-2019].
- [20] A. Baccini, “Statistique Descriptive Elémentaire,” 2010.
- [21] J. CLARKSON, “HUMAN CAPABILITY AND PRODUCT DESIGN,” *Prod. Exp.*, pp. 165–198, Jan. 2008.
- [22] “Loi normale.” [Online]. Available: <http://www.ilovestatistics.be/probabilite/loi-normale.html>. [Accessed: 17-Jun-2019].
- [23] D. Gupta and B. R. Gangadhar, “A statistical model for developing body size charts for garments,” *Int. J. Cloth. Sci. Technol.*, vol. 16, no. 5, pp. 458–469, 2004.
- [24] M. L. Mpampa, P. N. Azariadis, and N. S. Sapidis, “A new methodology for the development of sizing systems for the mass customization of garments,” *Int. J. Cloth. Sci. Technol.*, vol. 22, no. 1, pp. 49–68, 2010.
- [25] A. Petrova, “Creating sizing systems,” *Sizing Cloth. Dev. Eff. sizing Syst. ready-to-wear Cloth.*, pp. 57–87, 2007.
- [26] S. P. Ashdown and E. Textile Institute (Manchester, *Sizing in clothing: developing effective sizing systems for ready-to-wear clothing*. Woodhead Pub. in association with the Textile Institute, 2007.
- [27] R. Otieno, “Approaches in researching human measurement,” *EuroMed J. Bus.*, vol. 3, no. 1, pp. 63–82, May 2008.
- [28] D. Gupta, “Anthropometry and the design and production of apparel: an overview,” *Anthr. Appar. Sizing Des.*, pp. 34–66, Jan. 2014.
- [29] P. Vithanage, T. Jayawardene, and N. Niles, ““The Development of a Pant Sizing System for Sri Lankan Females,”” *Int. J. Res. Soc. Sci.*, vol. 1, pp. 30–38, 2013.
- [30] C.-H. Hsu, “Data mining to improve industrial standards and enhance production and marketing: An empirical study in apparel industry,” *Expert Syst. Appl.*, vol. 36, no. 3, pp. 4185–4191, Apr. 2009.
- [31] M. CHERIF Yamina, B. El Habib, D. Mansour, A. Khaldia, and A. Rachida, “APPLICATION DE L’ACP À UNE BASE DE DONNÉES DE LIQUIDES IONIQUES,” in *Third International Conference on Energy, Materials, Applied Energetics and Pollution*, 2016.
- [32] A. Baccini, “Statistique Descriptive Multidimensionnelle (pour les nuls).”
- [33] A. Saxena *et al.*, “A review of clustering techniques and developments,” *Neurocomputing*, vol. 267, pp. 664–681, Dec. 2017.
- [34] P. Robinet and S. Carrier, “Male and female consumers: segmenting consumers in the apparel market by body shape and other factors,” *Des. Appar. Consum.*, pp. 221–234, Jan. 2014.

- [35] H. L. Viktor, E. Paquet, and H. Guo, “Measuring to Fit: Virtual Tailoring Through Cluster Analysis and Classification,” 2006, pp. 395–406.
- [36] M. M. Jeyasingh and K. Appavoo, “Information Technology and Computer Science,” *Inf. Technol. Comput. Sci.*, vol. 6, pp. 12–17, 2012.
- [37] N. Zakaria, “Sizing system for functional clothing-Uniforms for school children,” *Indian J. Fibre Text. Res.*, vol. 36, pp. 348–357, 2011.
- [38] R. Ünlü and P. Xanthopoulos, “Estimating the number of clusters in a dataset via consensus clustering,” *Expert Syst. Appl.*, vol. 125, pp. 33–39, Jul. 2019.
- [39] R. Rakotomalala, “Tanagra Data Mining 1 Objectif,” 2016.
- [40] J. Shahrabi, E. Hadavandi, and M. S. Esfandarani, “Developing a hybrid intelligent model for constructing a size recommendation expert system in textile industries.”
- [41] M.-J. Chung, H.-F. Lin, and M.-J. J. Wang, “The development of sizing systems for Taiwanese elementary- and high-school students,” *Int. J. Ind. Ergon.*, vol. 37, no. 8, pp. 707–716, Aug. 2007.
- [42] N. Zakaria, *Clothing for children and teenagers : anthropometry, sizing and fit*. Woodhead Publishing, 2016.
- [43] M. T. Fralix, “Mass customization using the internet,” in *Proceedings of the 80th World Conference of the Textile Institute, Manchester*, 2000, pp. 16–19.
- [44] IFTH, “Tableaux mensurations ‘Prêt à porter femme’. Campagne nationale de mensuration.” 2009.
- [45] N. Z. Aydinoglu and A. Krishna, “Imagining thin: Why vanity sizing works,” *J. Consum. Psychol.*, vol. 22, no. 4, pp. 565–572, Oct. 2012.
- [46] Fitizzy, “Le guide des tailles est mort, vive le guide des tailles | Fitizzy.” [Online]. Available: <https://www.fitizzy.com/le-guide-des-tailles-est-mort-vive-le-guide-des-tailles/>. [Accessed: 05-Jul-2019].
- [47] ClicknDress, “Tailles de vêtements et tailles des françaises : une offre adaptée à la réalité ?,” 5 juillet 2016. [Online]. Available: <https://clickndress.com/news/2016/7/5/tailles-de-vetements-et-tailles-des-francaises-une-offre-adaptee-a-la-realite>. [Accessed: 05-Jul-2019].
- [48] “ISO 7250-1:2017 - Définitions des mesures de base du corps humain pour la conception technologique -- Partie 1: Définitions des mesures du corps et repères.” [Online]. Available: <https://www.iso.org/fr/standard/65246.html>. [Accessed: 02-Jul-2019].
- [49] “ISO/TR 7250-2:2010 - Définitions des mesures de base du corps humain pour la conception technologique -- Partie 2: Résumés statistiques des mesurages du corps de populations nationales.” [Online]. Available: <https://www.iso.org/fr/standard/41249.html>. [Accessed: 02-Jul-2019].
- [50] “ISO 7250-3:2015 - Définitions des mesures de base du corps humain pour la conception technologique -- Partie 3: Gammes de conception régionales et mondiales pour utilisation dans les normes de produits.” [Online]. Available: <https://www.iso.org/fr/standard/64237.html>. [Accessed: 02-Jul-2019].
- [51] “ISO 8559-1:2017 - Désignation des tailles de vêtements -- Partie 1: Définitions anthropométriques pour la mesure du corps.” [Online]. Available:

- <https://www.iso.org/fr/standard/61686.html>. [Accessed: 02-Jul-2019].
- [52] “ISO 8559-2:2017 - Désignation des tailles de vêtements -- Partie 2: Indicateurs des dimensions principales et secondaires.” [Online]. Available: <https://www.iso.org/fr/standard/64075.html>. [Accessed: 02-Jul-2019].
- [53] “ISO 8559-3:2018 - Désignation des tailles des vêtements -- Partie 3: Méthodologie de création de barèmes de mensuration du corps et des intervalles.” [Online]. Available: <https://www.iso.org/fr/standard/67334.html>. [Accessed: 02-Jul-2019].
- [54] “NF EN 13402-1 - Juin 2001.” [Online]. Available: <https://www.boutique.afnor.org/norme/nf-en-13402-1/designations-des-tailles-de-vetements-partie-1-termes-definitions-et-procedes-de-mesurage-du-corps/article/659359/fa114014>. [Accessed: 02-Jul-2019].
- [55] “NF EN 13402-2 - Février 2003.” [Online]. Available: <https://www.boutique.afnor.org/norme/nf-en-13402-2/designation-des-tailles-de-vetements-partie-2-mesures-primaires-et-secondaires/article/779513/fa106580>. [Accessed: 02-Jul-2019].
- [56] “NF EN 13402-3 - Septembre 2017.” [Online]. Available: <https://www.boutique.afnor.org/norme/nf-en-13402-3/designation-des-tailles-de-vetements-partie-3-etiquetage-de-la-taille-sur-la-base-des-mesures-corporelles-et-des-intervalles/article/856990/fa061109>. [Accessed: 02-Jul-2019].
- [57] C. E. Mcculloch, B. Paal, and S. P. Ashdown, “An optimisation approach to apparel sizing,” *J. Oper. Res. Soc.*, vol. 49, no. 5, pp. 492–499, May 1998.
- [58] C. J. Salusso, J. J. Borkowski, N. Reich, and E. Goldsberry, “An alternative approach to sizing apparel for women 55 and older,” *Cloth. Text. Res. J.*, vol. 24, no. 2, pp. 96–111, 2006.
- [59] R. Hrženjak, K. Doležal, and D. Ujević, “Sizing system for girls aged 13–20 years based on body types,” *Text. Res. J.*, vol. 85, no. 12, pp. 1293–1304, 2015.
- [60] A. H. Doustaneh, M. Gorji, and M. Varsei, “Using Self Organization Method to Establish Nonlinear Sizing System,” vol. 9, no. 12, pp. 1359–1364, 2010.
- [61] M. Vadood, M. S. Esfandarani, and M. S. Johari, “Developing a New Suit Sizing System Using Neural Network,” *J. Eng. Fiber. Fabr.*, vol. 10, no. 2, p. 155892501501000, 2018.
- [62] J. YANG, C. K. CHAN, and A. LUXIMON, “A SURVEY ON 3D HUMAN BODY MODELING FOR INTERACTIVE FASHION DESIGN,” *Int. J. Image Graph.*, vol. 13, no. 4, p. 1350021, Oct. 2013.
- [63] M. Hamad, “Classification non-supervisée de morphologies 3D de corps humain pour la mise en œuvre de mannequins morphotypes adaptatifs,” 2015.
- [64] D. A. Hirshberg *et al.*, “Evaluating the Automated Alignment of 3D Human Body Scans,” in *Proceedings of the 2nd International Conference on 3D Body Scanning Technologies, Lugano, Switzerland, 25-26 October 2011*, 2011, pp. 76–86.
- [65] F. Cordier, Hyewon Seo, and N. Magnenat-Thalmann, “Made-to-measure technologies for an online clothing store,” *IEEE Comput. Graph. Appl.*, vol. 23, no. 1, pp. 38–48, Jan. 2003.
- [66] H. Seo and N. Magnenat-Thalmann, “An automatic modeling of human bodies from sizing parameters,” in *Proceedings of the 2003 symposium on Interactive 3D graphics*

- *SI3D '03*, 2003, p. 19.

- [67] S. Agnieszka, “« Contribution à la modélisation et à la simulation de vêtements sur mannequin adaptatif »,” *Distribution*, pp. 1–179, 2008.
- [68] C. C. L. Wang, “Parameterization and parametric design of mannequins,” *Comput. Des.*, vol. 37, no. 1, pp. 83–98, Jan. 2005.
- [69] C. C. L. Wang, “Parameterization and parametric design of mannequins,” *Comput. Des.*, vol. 37, no. 1, pp. 83–98, Jan. 2005.
- [70] S.-W. Hsiao and R.-Q. Chen, “A study of surface reconstruction for 3D mannequins based on feature curves,” *Comput. Des.*, vol. 45, no. 11, pp. 1426–1441, Nov. 2013.
- [71] Y. Zeng, J. Fu, and H. Chao, “3D Human Body Reshaping with Anthropometric Modeling,” Springer, Singapore, 2018, pp. 96–107.
- [72] H. Ltifi *et al.*, “Démarche centrée utilisateur pour la conception de SIAD basés sur un processus d’ECD, application dans le domaine de la santé,” 2010.
- [73] N.-Q. Doan, H. Azzag, M. Lebbah, U. Paris, and P. Cité, “Sélection de variables non supervisée sous contraintes hiérarchiques,” 2013.
- [74] S. El Ferchichi, “Sélection et extraction d’attributs pour les problèmes de classification,” Lille 1, 2013.
- [75] U. M. Khaire and R. Dhanalakshmi, “Stability of feature selection algorithm: A review,” *J. King Saud Univ. - Comput. Inf. Sci.*, Jun. 2019.
- [76] V. Gardeux, “Conception d’heuristiques d’optimisation pour les problèmes de grande dimension: application à l’analyse de données de puces à ADN,” 2011.
- [77] M. Carricano, F. Poujol, and L. Bertrandias, *Analyse de données avec SPSS®*. Pearson Education France, 2010.
- [78] M.-F. Harkat, “Détection et localisation de défauts par analyse en composantes principales,” Institut National Polytechnique de Lorraine-INPL, 2003.
- [79] E. J. Pedhazur and L. P. Schmelkin, *Measurement, design, and analysis : an integrated approach*. Lawrence Erlbaum Associates, 1991.
- [80] R. B. Cattell, “The Scree Test For The Number Of Factors,” *Multivariate Behav. Res.*, vol. 1, no. 2, pp. 245–276, Apr. 1966.
- [81] X. He, D. Cai, and P. Niyogi, “Laplacian score for feature selection,” in *Advances in neural information processing systems*, 2006, pp. 507–514.
- [82] J. Han, M. Kamber, and J. Pei, “Data Preprocessing,” *Data Min.*, pp. 83–124, Jan. 2012.
- [83] D. Virmani, S. Taneja, and G. Malhotra, “Normalization based K means Clustering Algorithm.”
- [84] R. Jin and J. Kim, “A method for 3D model identification using histogram normalization based on conformal factor,” in *2016 IEEE International Conference on Electronic Information and Communication Technology (ICEICT)*, 2016, pp. 360–362.
- [85] M. Hamad, S. Thomassey, and P. Bruniaux, “A new sizing system based on 3D shape descriptor for morphology clustering,” *Comput. Ind. Eng.*, vol. 113, pp. 683–692, Nov. 2017.
- [86] M. Chaouch, “3D Shape-content based retrieval,” *Técom ParisTech*, 2009.

- [87] J. J. Sylvester, “A question in the geometry of situation,” *Q. J. Pure Appl. Math.*, vol. 1, no. 1, pp. 79–80, 1857.
- [88] N. Megiddo, “Linear-Time Algorithms for Linear Programming in \mathbb{R}^3 and Related Problems,” *SIAM J. Comput.*, vol. 12, no. 4, pp. 759–776, Nov. 1983.
- [89] K. FISCHER and B. GÄRTNER, “THE SMALLEST ENCLOSING BALL OF BALLS: COMBINATORIAL STRUCTURE AND ALGORITHMS,” *Int. J. Comput. Geom. Appl.*, vol. 14, no. 04n05, pp. 341–378, Oct. 2004.
- [90] A. Adeyemo and H. Wimmer, “Effects of Normalization Techniques on Logistic Regression in Data Science,” 2018.
- [91] P. Walid Hasen Atomi, “THE EFFECT OF DATA PREPROCESSING ON THE PERFORMANCE OF ARTIFICIAL NEURAL NETWORKS TECHNIQUES FOR CLASSIFICATION,” 2012.
- [92] C. Lopez, S. Tucker, T. Salameh, and C. Tucker, “An unsupervised machine learning method for discovering patient clusters based on genetic signatures,” *J. Biomed. Inform.*, vol. 85, pp. 30–39, Sep. 2018.
- [93] S. Tripathi, A. Bhardwaj, and P. E, “Approaches to Clustering in Customer Segmentation,” *Int. J. Eng. Technol.*, vol. 7, no. 3.12, p. 802, Jul. 2018.
- [94] R. Genuer and J. Poggi, “Arbres CART et Forêts aléatoires Importance et sélection de variables Table des matières,” 2017.
- [95] L. Breiman, A. D. Gordon, J. H. Friedman, R. A. Olshen, and C. J. Stone, “Classification and Regression Trees,” *Biometrics*, vol. 40, no. 3, p. 874, Sep. 1984.
- [96] A. Poterie, “Arbres de décision et forêts aléatoires pour variables groupées,” Rennes, INSA, 2018.
- [97] B. Doucouré, “Proposition, intégration dans un système de gestion de réseau intelligent et validation expérimentale d’une méthode de prédiction pour un système d’énergies renouvelables,” Université du Québec à Trois-Rivières, 2015.
- [98] E. Lebarbier and T. Mary-Huard, “Classification non supervisée.”
- [99] M. RIFQI, “Mesures de similarité, raisonnement et modélisation de l’utilisateur,” 2010.
- [100] M.-J. Lesot and M. Rifqi, “Degrés d’équivalence de mesures de comparaison pour données binaires et pour données numériques,” 2014.
- [101] J. Han, M. Kamber, J. Pei, J. Han, M. Kamber, and J. Pei, “Getting to Know Your Data,” *Data Min.*, pp. 39–82, Jan. 2012.
- [102] A. Saxena *et al.*, “A review of clustering techniques and developments,” *Neurocomputing*, vol. 267, pp. 664–681, Dec. 2017.
- [103] Q. Grimonprez, “Sélection de groupes de variables corrélées en grande dimension,” Lille 1, 2016.
- [104] S. Hands and B. Everitt, “A Monte Carlo Study of the Recovery of Cluster Structure in Binary Data by Hierarchical Clustering Techniques,” *Multivariate Behav. Res.*, vol. 22, no. 2, pp. 235–243, Apr. 1987.
- [105] L. Ferreira and D. B. Hitchcock, “A Comparison of Hierarchical Methods for Clustering Functional Data,” *Commun. Stat. - Simul. Comput.*, vol. 38, no. 9, pp. 1925–1949, Oct. 2009.

- [106] E. Mooi and M. Sarstedt, “Cluster analysis. A concise guide to market research,” *Berlin Springer-Verlag. PBOT plan b{á}sico ordenamiento Territ. Sylvania,(2000) acuerdo*, vol. 22, p. 33, 2011.
- [107] A. Kijewska and A. Bluszcz, “Analysis of greenhouse gas emissions in the European Union member states with the use of an agglomeration algorithm,” *J. Sustain. Min.*, vol. 15, no. 4, pp. 133–142, 2016.
- [108] G. Cabanes, “Classification non supervisée à deux niveaux guidée par le voisinage et la densité,” Paris 13, 2010.
- [109] I. Dhillon, Y. Guan, B. K.-P. of the tenth A. SIGKDD, and undefined 2004, “Kernel k-means: spectral clustering and normalized cuts,” *dl.acm.org*.
- [110] D. Arthur, S. V.-P. of the eighteenth annual ACM, and undefined 2007, “k-means++: The advantages of careful seeding,” *dl.acm.org*.
- [111] P. J. Rousseeuw, “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis,” *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, Nov. 1987.
- [112] B. Desgraupes, “Clustering indices,” *Univ. Paris Ouest-Lab Modal’X*, vol. 1, p. 34, 2013.
- [113] D. L. Davies and D. W. Bouldin, “A Cluster Separation Measure,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, no. 2, pp. 224–227, Apr. 1979.
- [114] T. Calinski and J. Harabasz, “A dendrite method for cluster analysis,” *Commun. Stat. - Theory Methods*, vol. 3, no. 1, pp. 1–27, 1974.
- [115] T. Kohonen, “Self-organized formation of topologically correct feature maps,” *Biol. Cybern.*, vol. 43, no. 1, pp. 59–69, 1982.
- [116] C. Hajjar, “Cartes auto-organisatrices pour la classification de données symboliques mixtes, de données de type intervalle et de données discrétisées,” ECOLE DOCTORALE STITS, 2014.
- [117] M. Abadi, “Réalisation d’un réseau de neurones ‘SOM’ sur une architecture matérielle adaptable et extensible à base de réseaux sur puce ‘NoC,’” l’Université de Lorraine, 2018.
- [118] H. López García and I. Machón González, “Self-organizing map and clustering for wastewater treatment monitoring,” *Eng. Appl. Artif. Intell.*, vol. 17, no. 3, pp. 215–225, Apr. 2004.
- [119] N. Jozefowicz, “Optimisation combinatoire multi-objectif: des méthodes aux problèmes, de la Terre à (presque) la Lune,” Institut National Polytechnique de Toulouse, 2013.
- [120] S. Jacquin, “Hybridation des métaheuristiques et de la programmation dynamique pour les problèmes d’optimisation mono et multi-objectif: application à la production d’énergie,” Lille 1, 2015.
- [121] V. Pareto, “Cours d’économie politique,” *Genève, Droz*, 1896.
- [122] I. Boussaid, “Perfectionnement de métaheuristiques pour l’optimisation continue,” Université Paris-Est, 2013.
- [123] Q. Mercier, “Optimisation multicritère sous incertitudes : un algorithme de descente stochastique,” Université Côte d’Azur, 2018.

- [124] A. Liefooghe, “Métaheuristiques pour l’optimisation multiobjectif: approches coopératives, prise en compte de l’incertitude et application en logistique,” Lille 1, 2009.
- [125] A. Ghoumari, “Métaheuristiques adaptatives d’optimisation continue basées sur des méthodes d’apprentissage,” Université Paris-Est, 2018.
- [126] A. N. Benaichouche, “Conception de métaheuristiques d’optimisation pour la segmentation d’images: application aux images IRM du cerveau et aux images de tomographie par émission de positons,” Université Paris-Est, 2014.
- [127] M.-E. Marmion, “Recherche locale et optimisation combinatoire: de l’analyse structurelle d’un problème à la conception d’algorithmes efficaces,” Université des Sciences et Technologie de Lille , 2011.
- [128] R. Zaghdoud, “Hybridation d’algorithme génétique pour les problèmes des véhicules intelligents autonomes: applications aux infrastructures portuaires de moyenne taille,” Ecole centrale de Lille, 2015.
- [129] T. Vallée and M. Yildizoglu, “Présentation des algorithmes génétiques et de leurs applications en économie,” *Rev. d’'économie Polit.*, vol. 114, no. 6, pp. 711–745, 2004.
- [130] L. Jourdan, “Métaheuristiques pour l’extraction de connaissances: application à la génomique,” Lille 1, 2003.
- [131] Y. Elmir, “Optimisation multi objectif par les algorithmes génétiques et approche Pareto de paramètres d’un contrôleur PID,” Université des Sciences et de la Technologie d’Oran Mohamed Boudiaf, 2016.
- [132] “Quelle est ma morphologie? - Cristina Cordula.” [Online]. Available: <https://www.cristinacordula.com/2015-10-morphologie-2/>. [Accessed: 22-Jan-2020].
- [133] B. Hamad, M. Hamad, S. Thomassey, and P. Bruniaux, “3D Adaptive Morphotype Mannequin for Target Population,” *J. Ergon.*, vol. 8, no. 2, 2018.

Annexe 1 : Description mesures anthropométriques

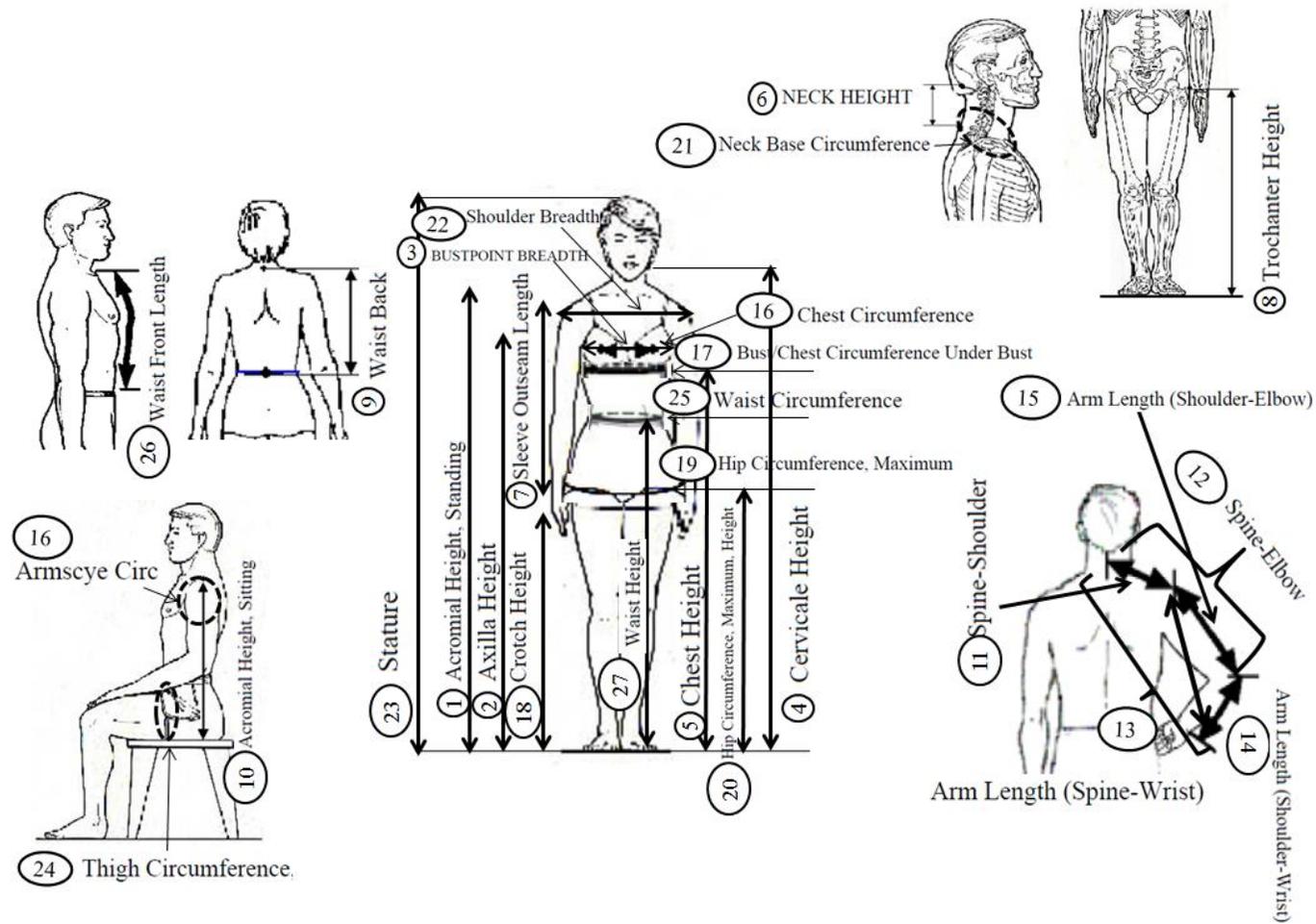


Figure 102. Illustration mensurations CAESAR.

Annexe 2 : Exemples de questions posés lors de l'enquête clients

QUESTIONNAIRE DE SATISFACTION

Afin de connaître votre avis sur nos articles et de l'améliorer au mieux, nous vous proposons de répondre à ce questionnaire .

Question 1

Quel est votre age ?

Question 2

Quel est votre poids?

Question 3

Quelle est votre taille?

Question 4

Pouvez-vous préciser votre taille de vêtements pour les t-shirts et les pulls?

- XS
- S
- M
- L
- XL
- XXL
- Je ne souhaite pas répondre

[décocher](#)

Question 5

Pouvez-vous préciser votre taille de vêtements pour les pantalons?

- XS
- S
- M
- L
- XL
- XXL

[décocher](#)

Question 6

Comment trouvez-vous les tailles de vêtements proposées?

- Très satisfaisant
- satisfaisant
- Moyennement satisfaisant
- Insatisfaisant

[décocher](#)

Figure 103. Questionnaire de satisfaction clients

Annexe 3 : Exemple de l'arbre de décision généré

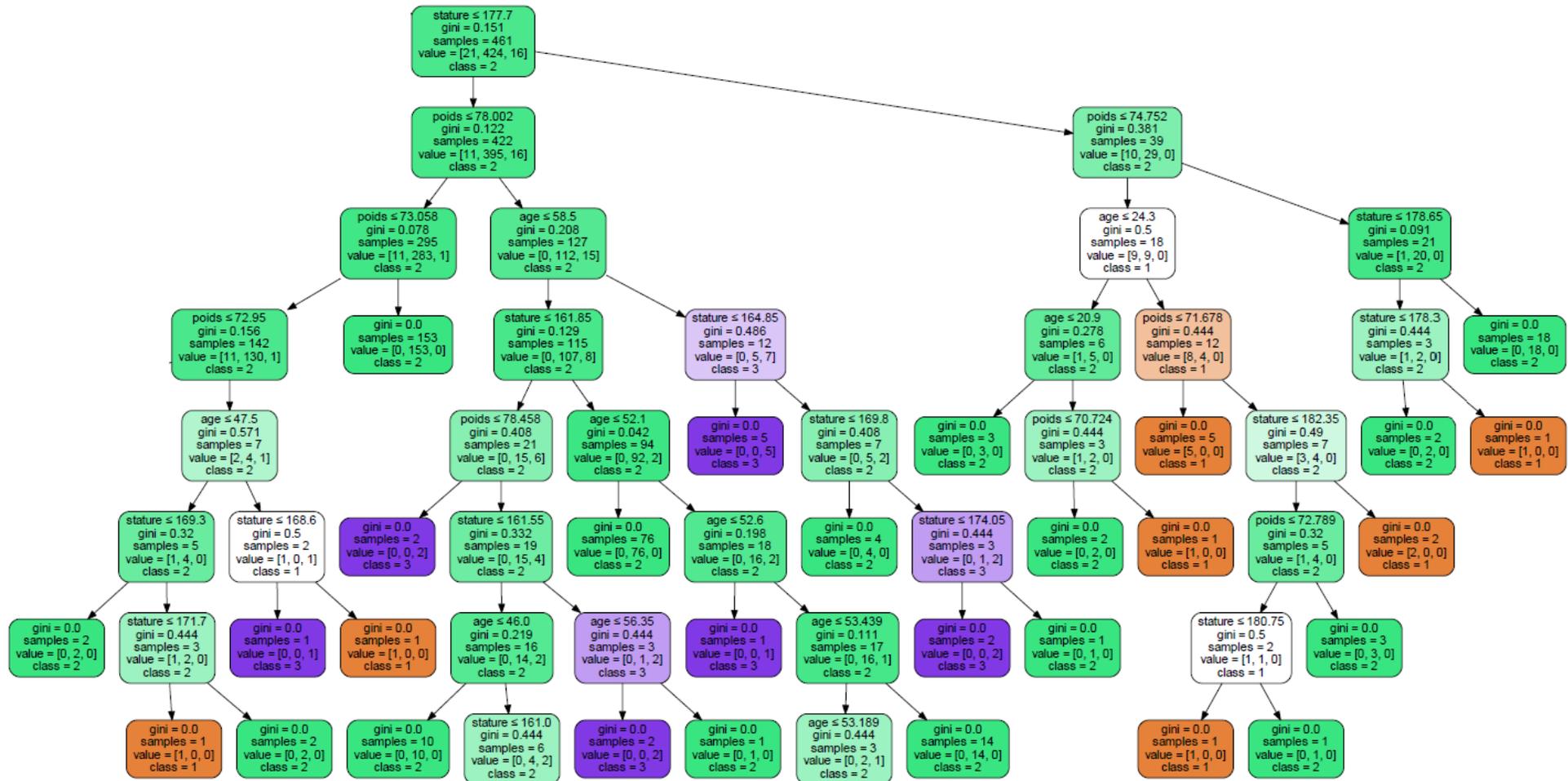
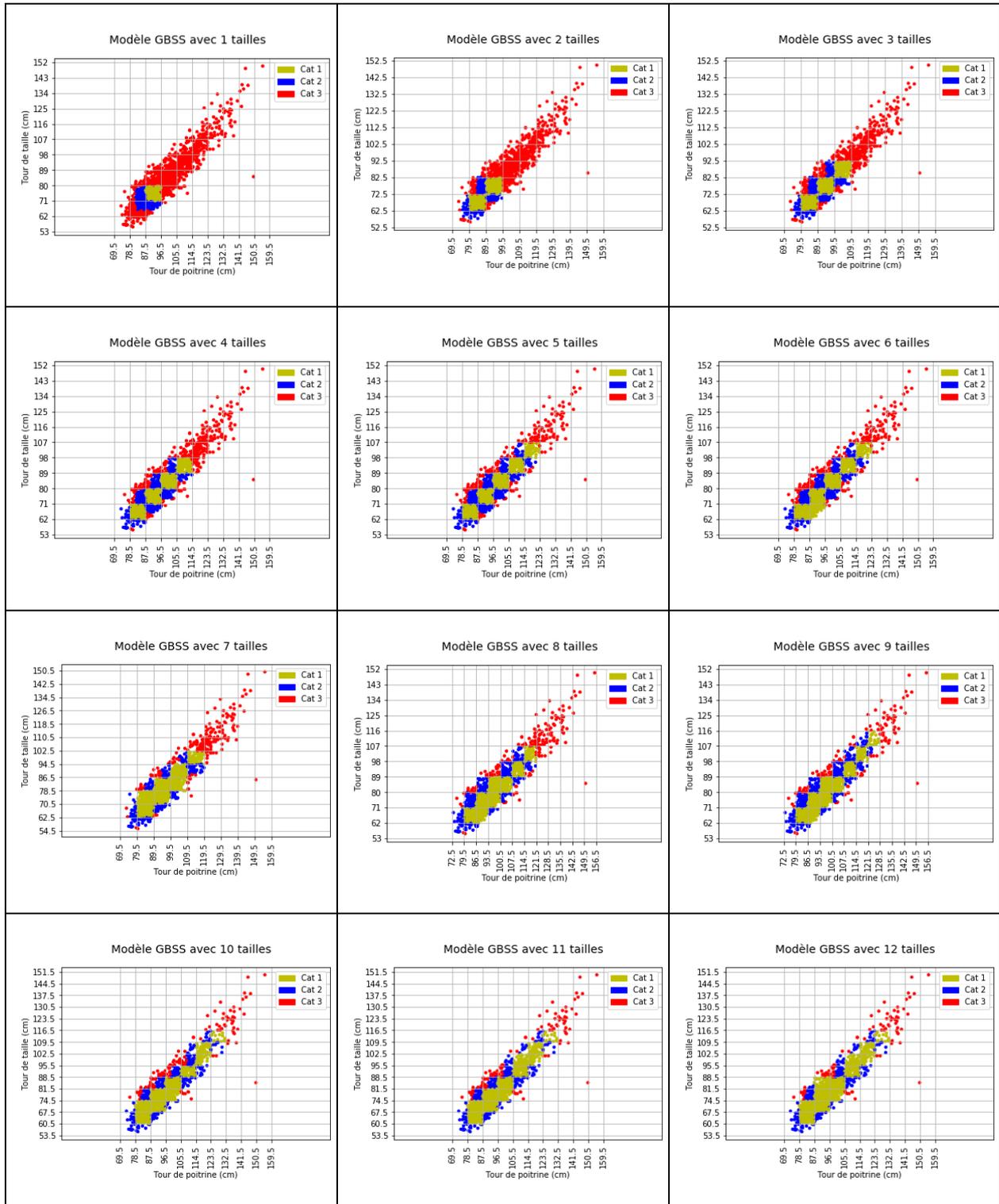


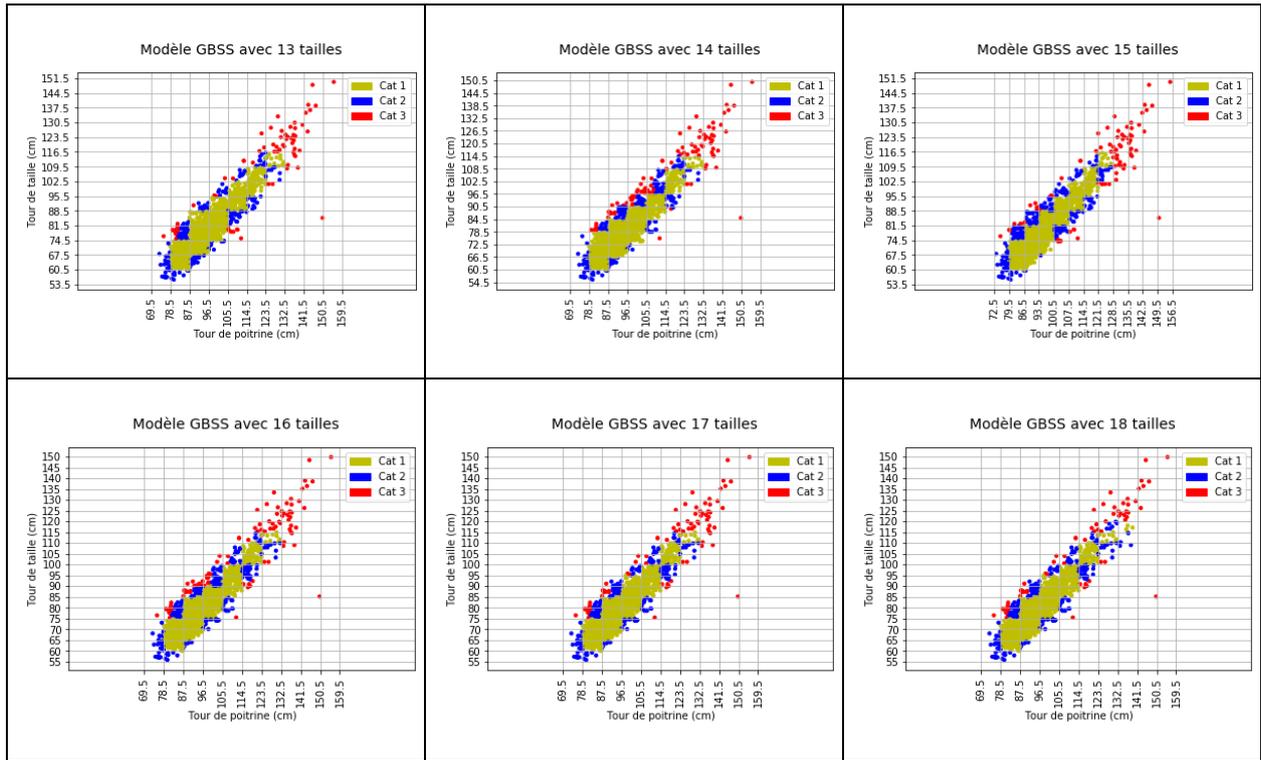
Figure 104. Exemple d'une partie d'arbre CART généré pour le tour de poitrine.

Annexe 4 : Systèmes de tailles pour les mensurations

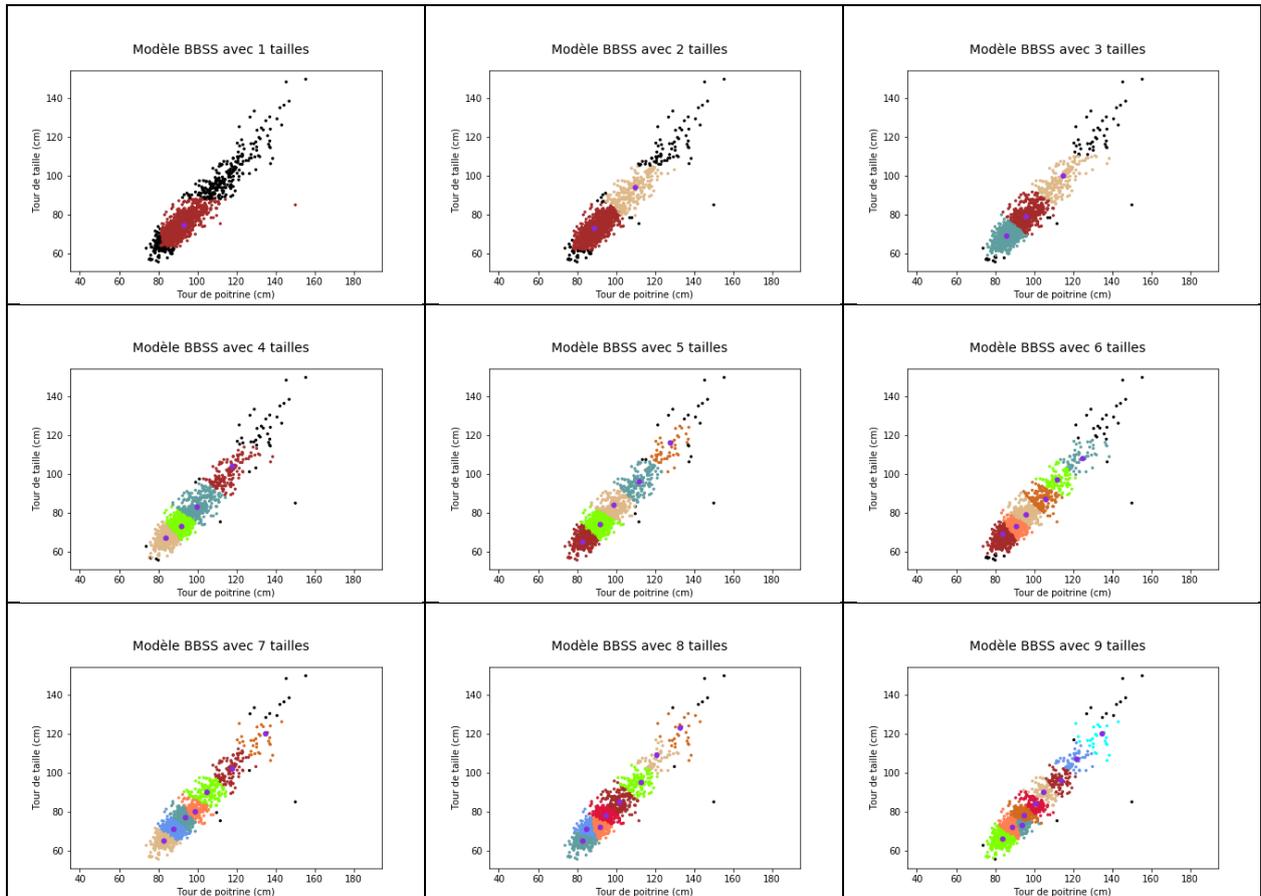
Poitrine/Taille sur la population totale

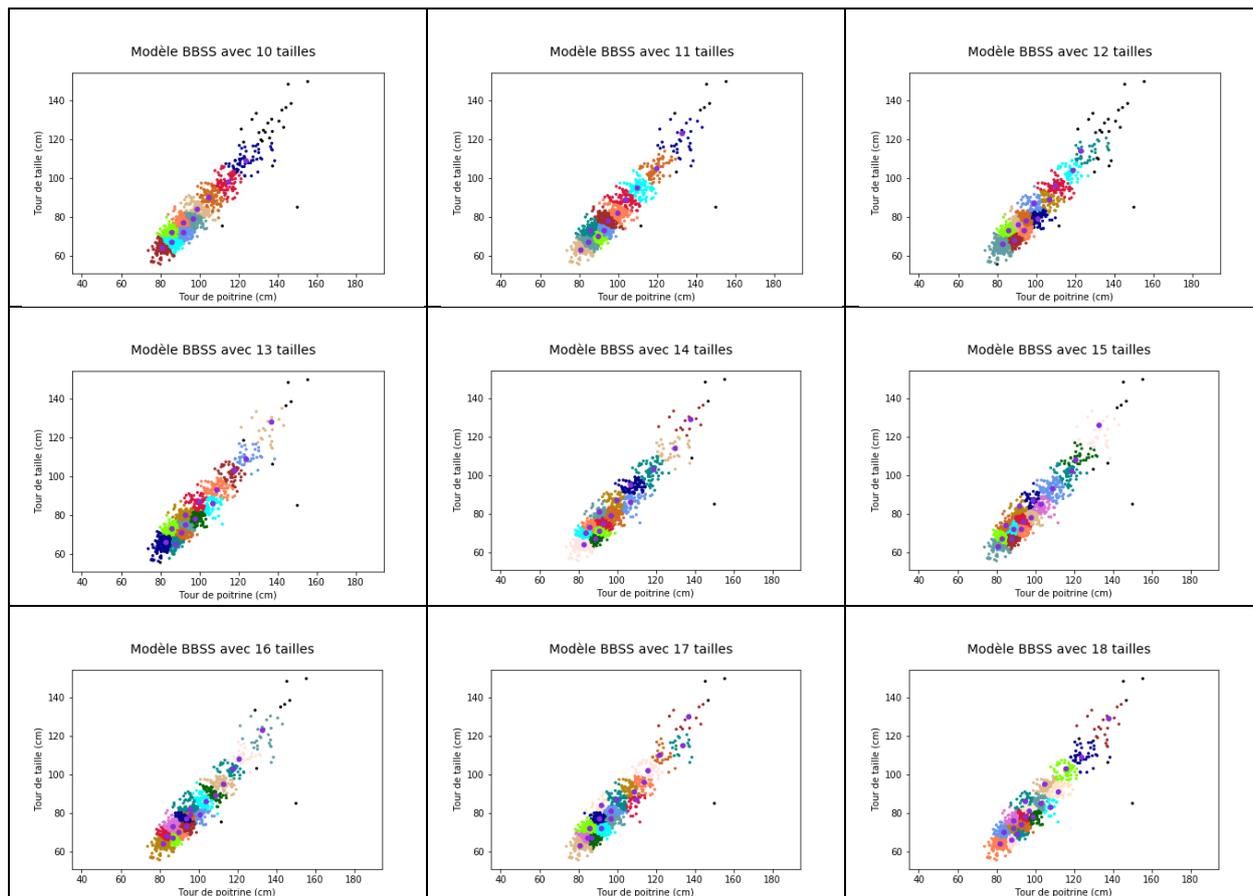
Modèle GBSS



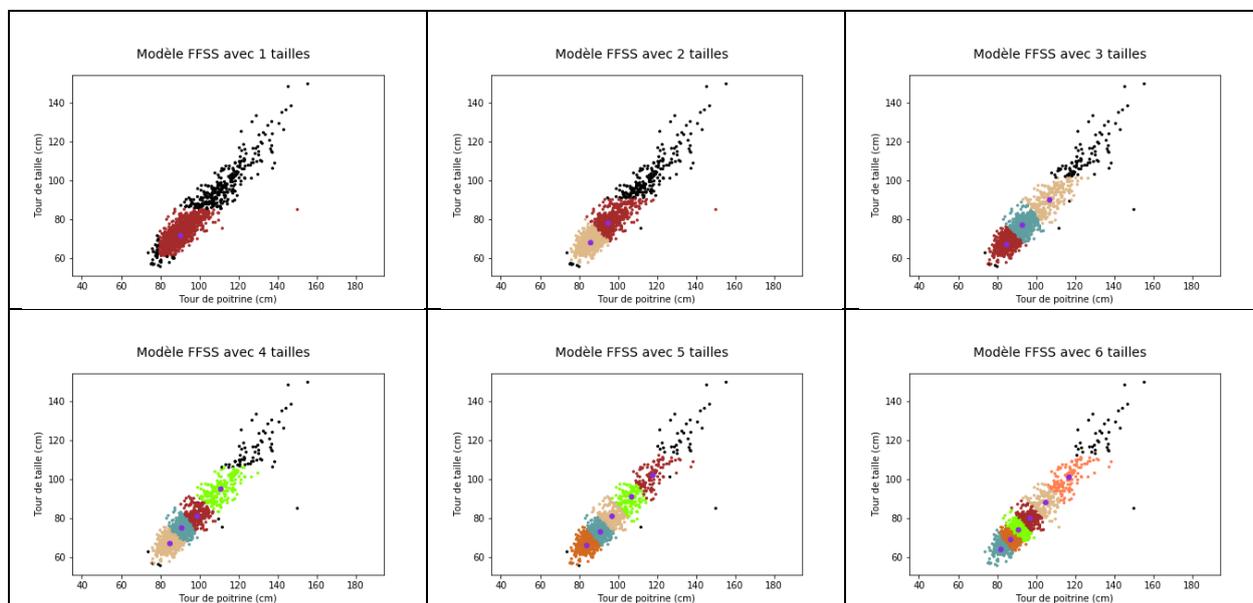


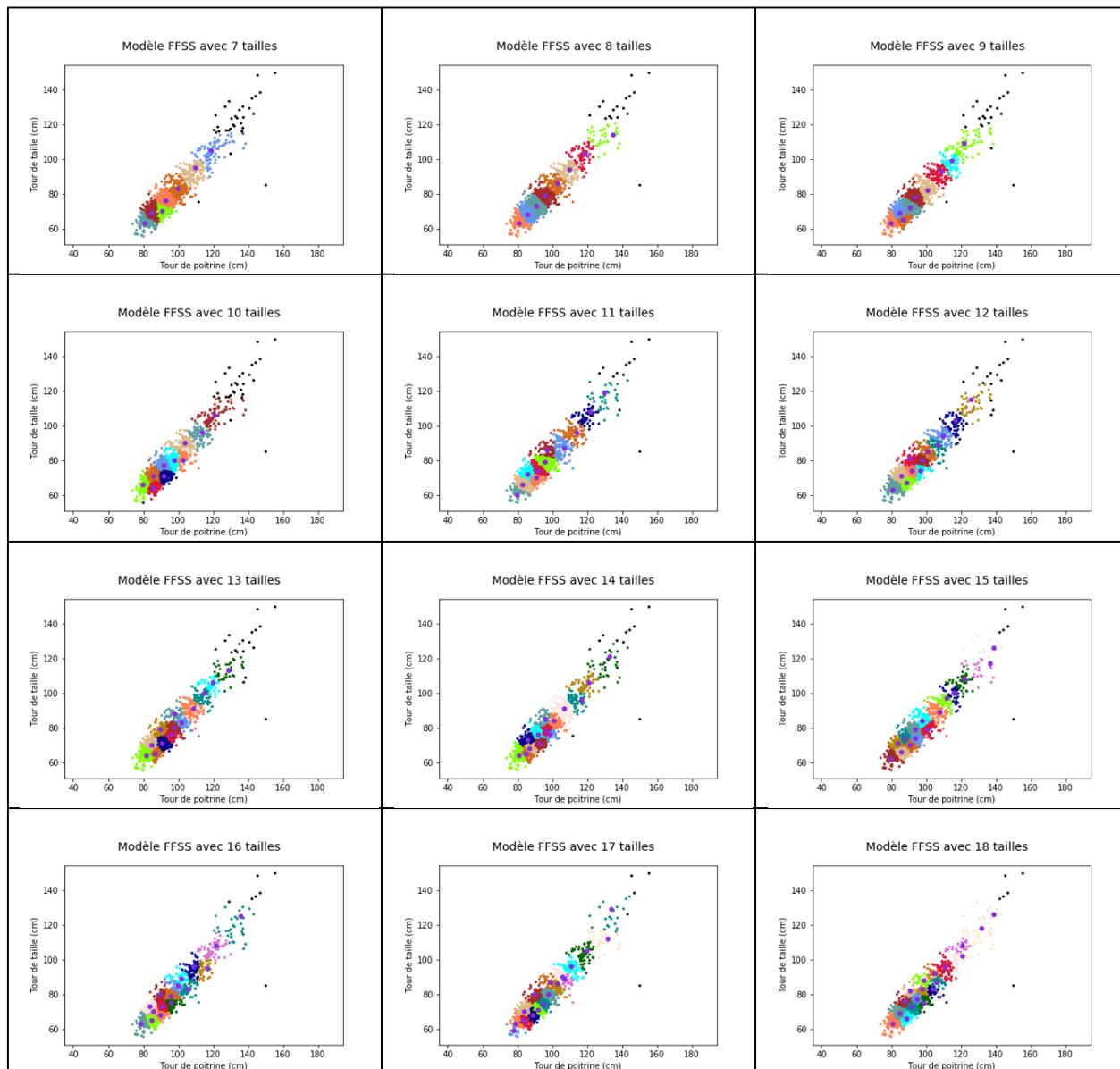
Modèle BBSS





Modèle FFSS

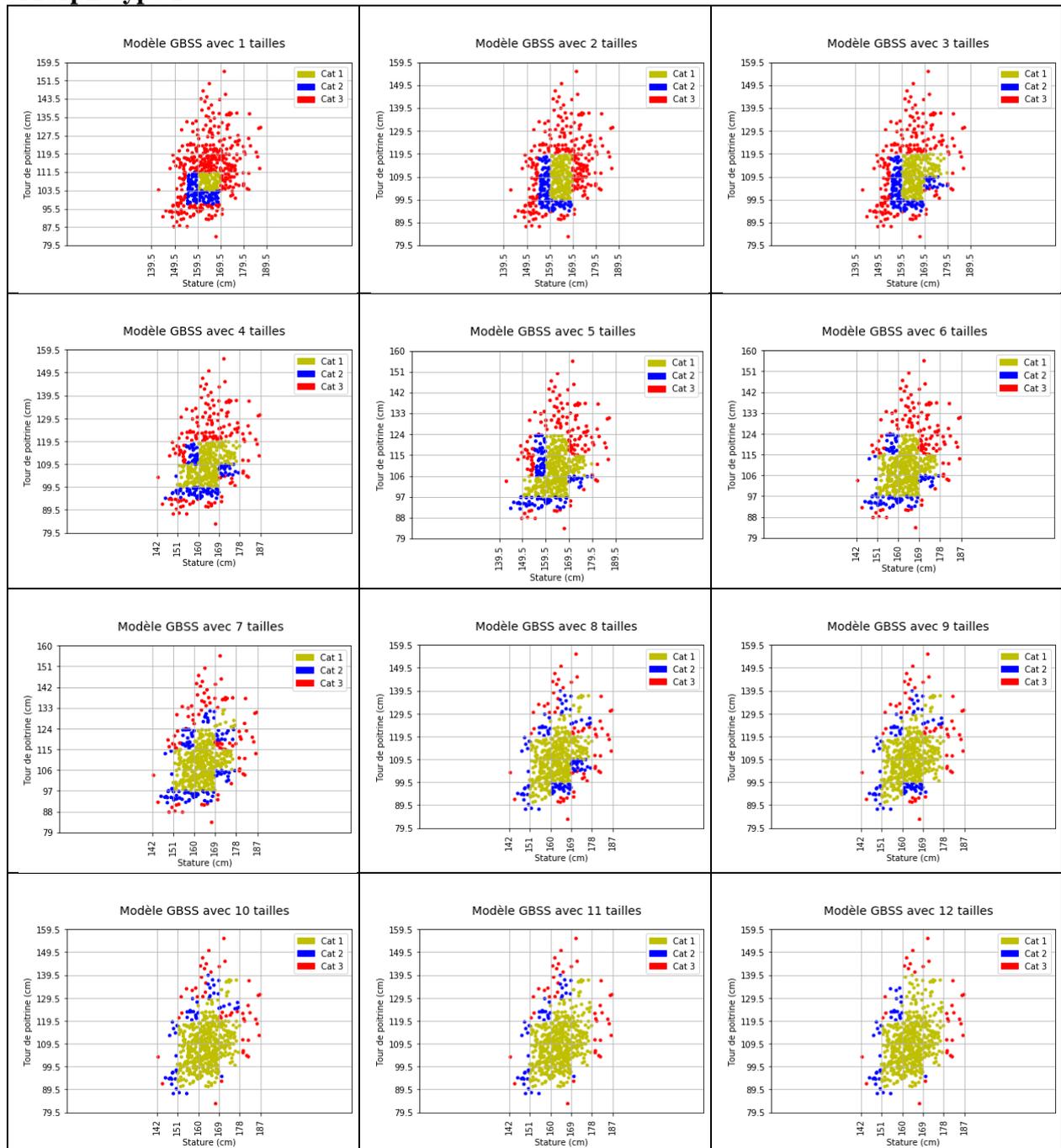


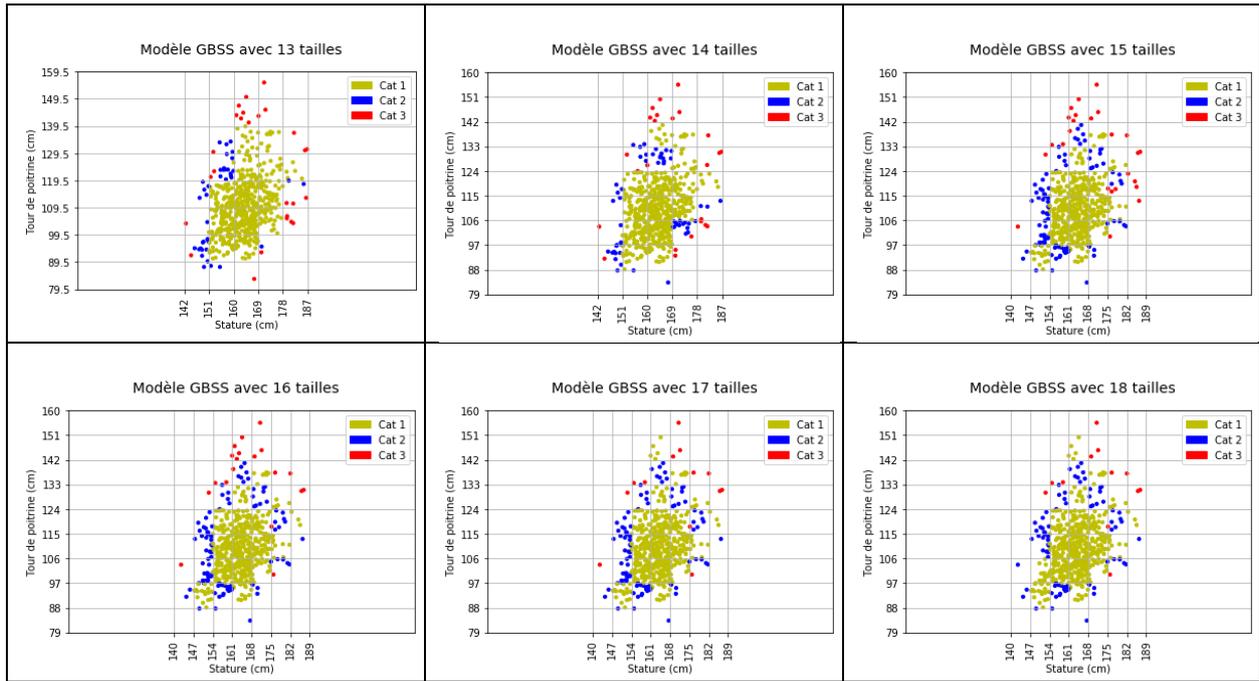


Annexe 5 : Systèmes de tailles pour les mensurations

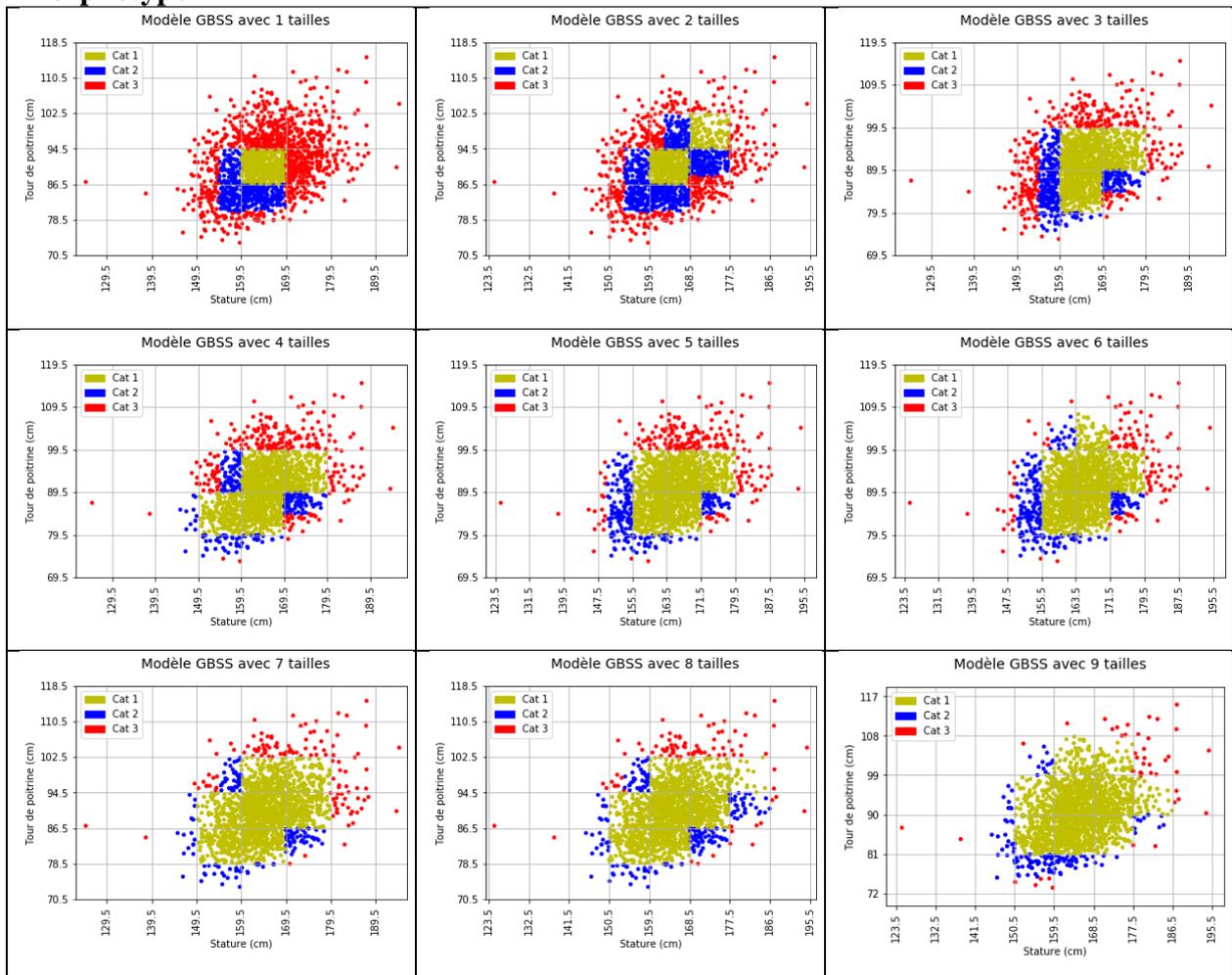
Stature/Poitrine par morphotype

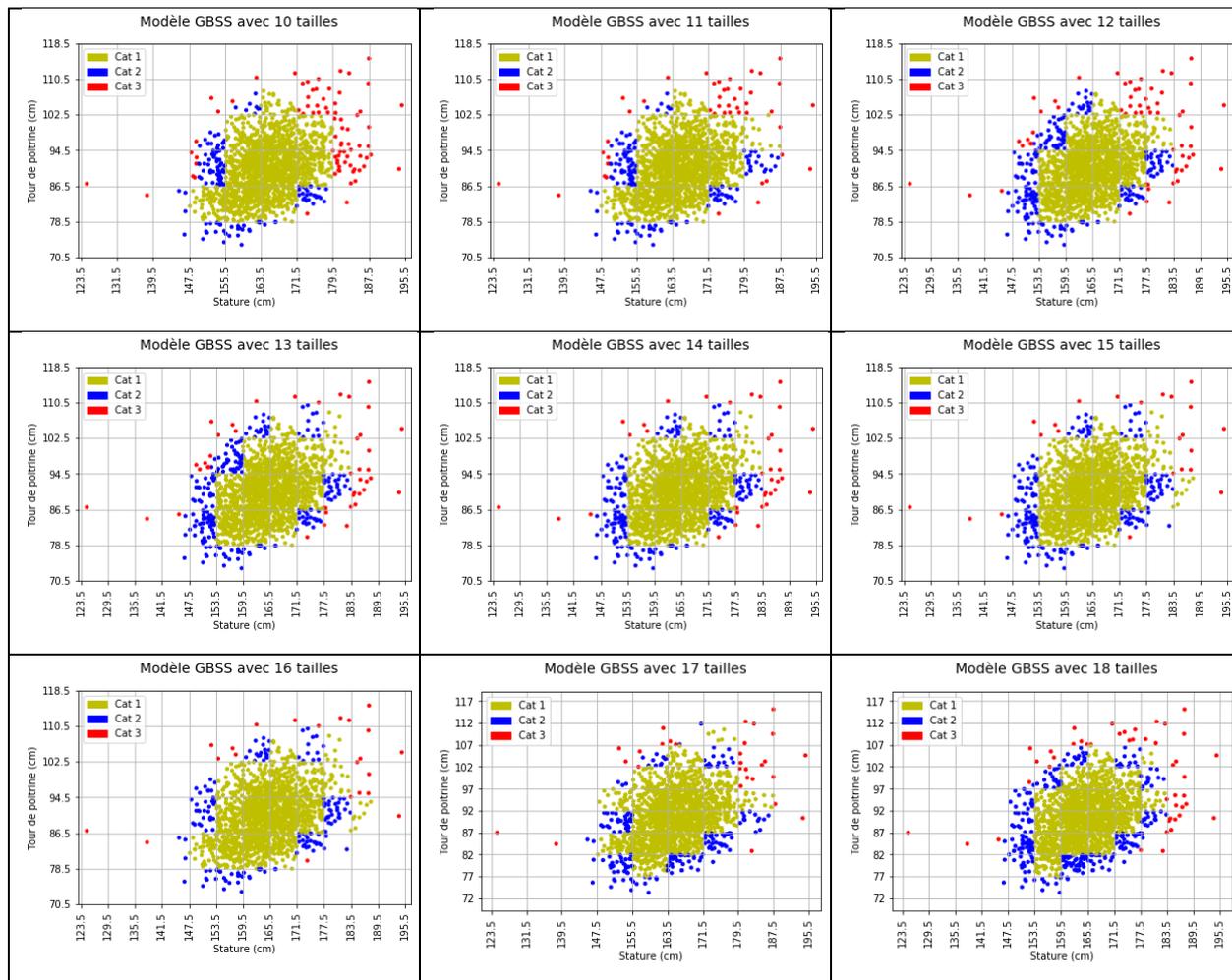
GBSS Morphotype 1



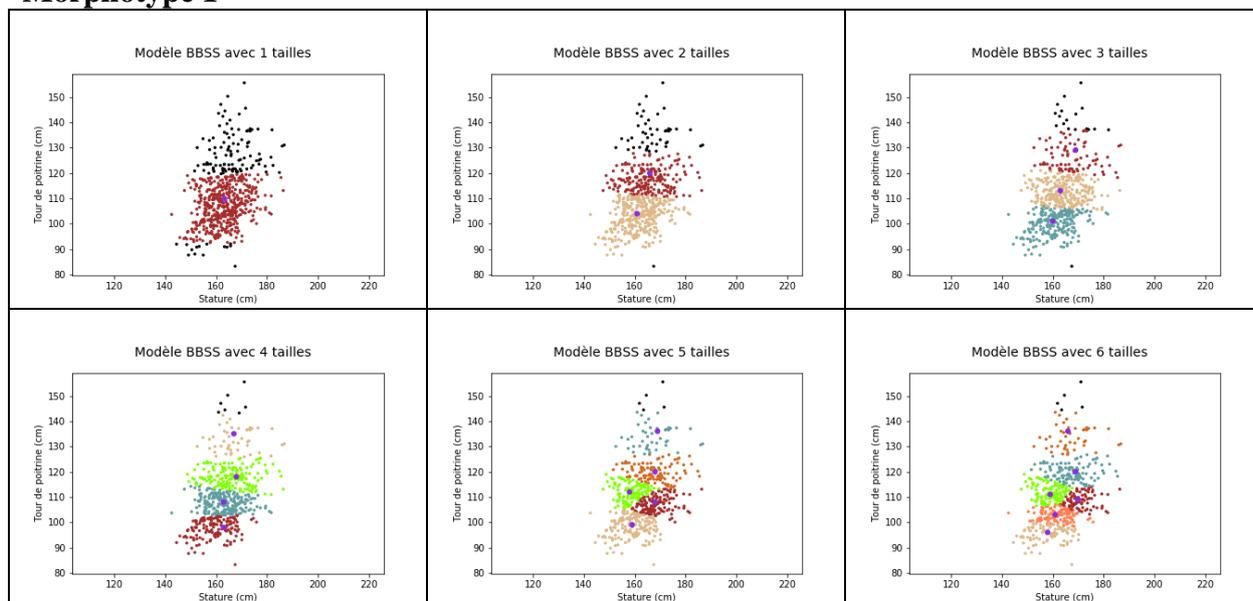


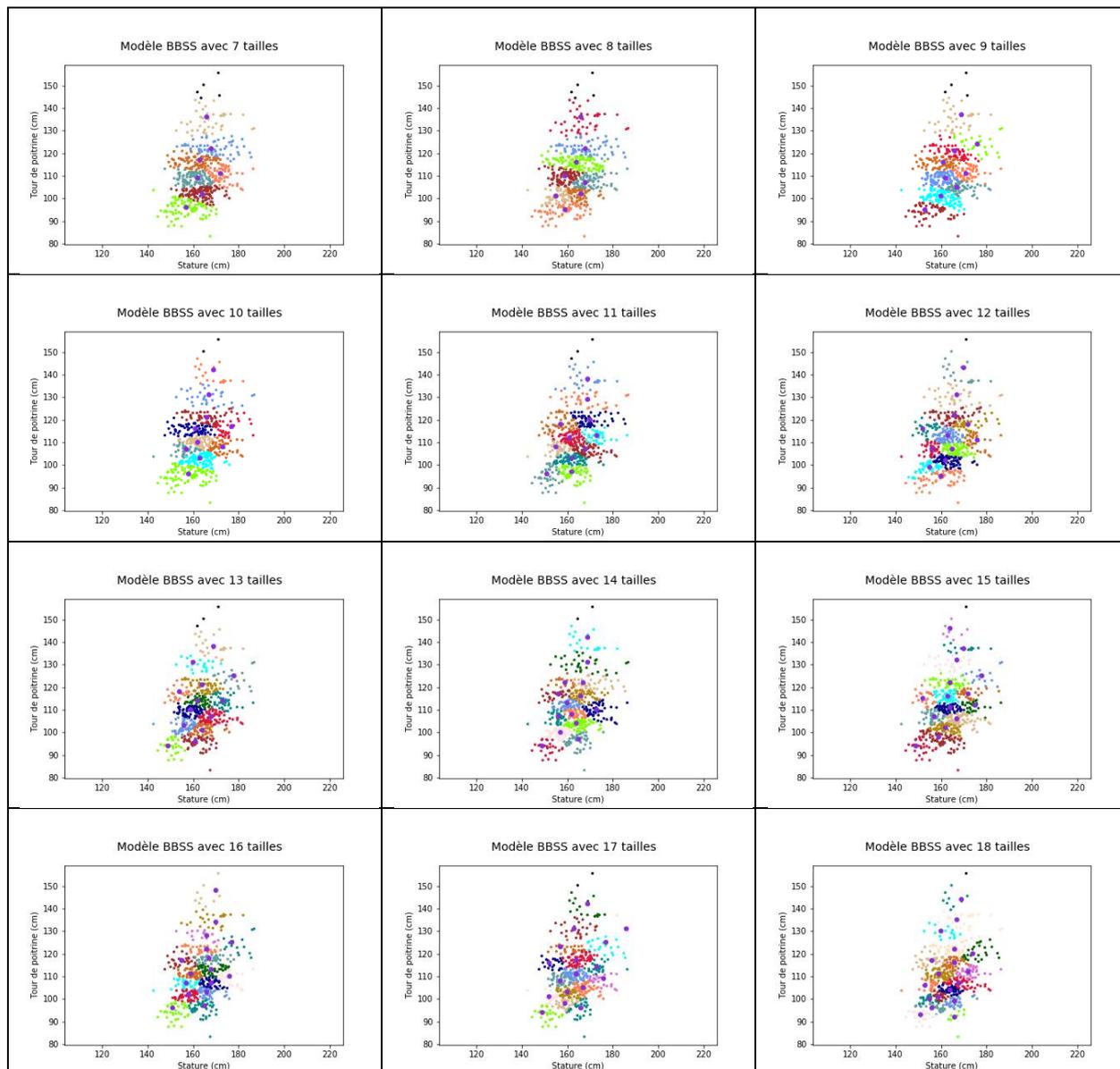
Morphotype 2



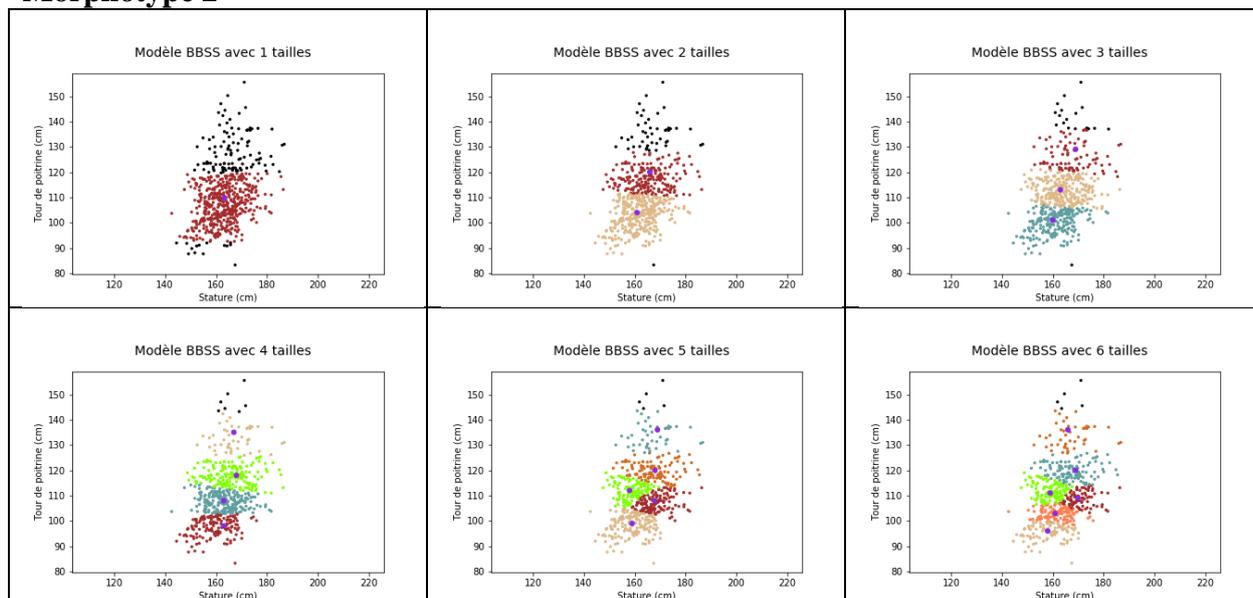


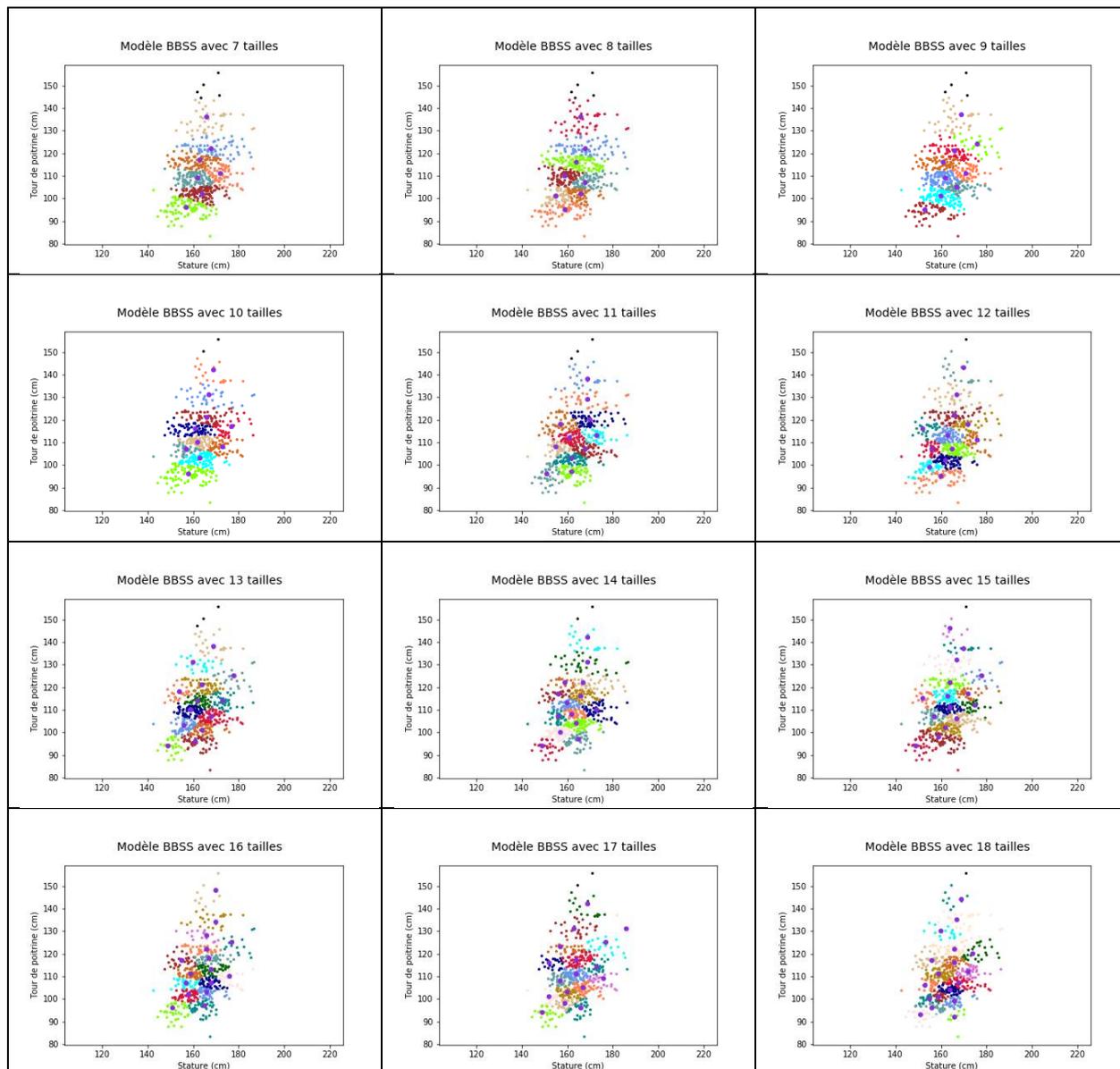
BBSS Morphotype 1



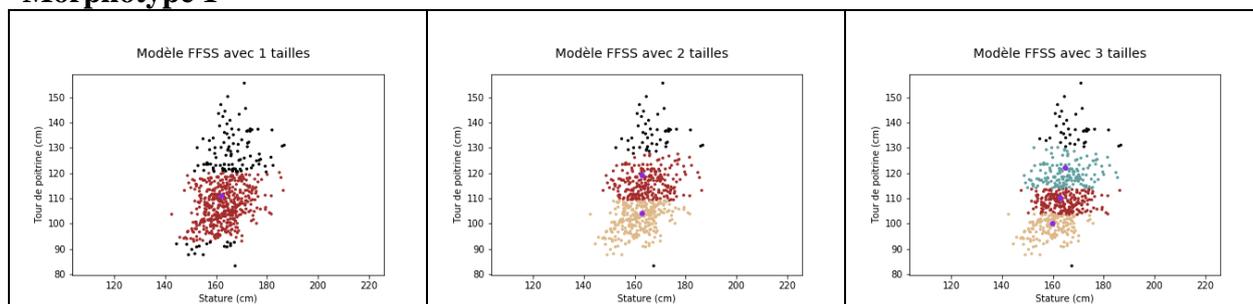


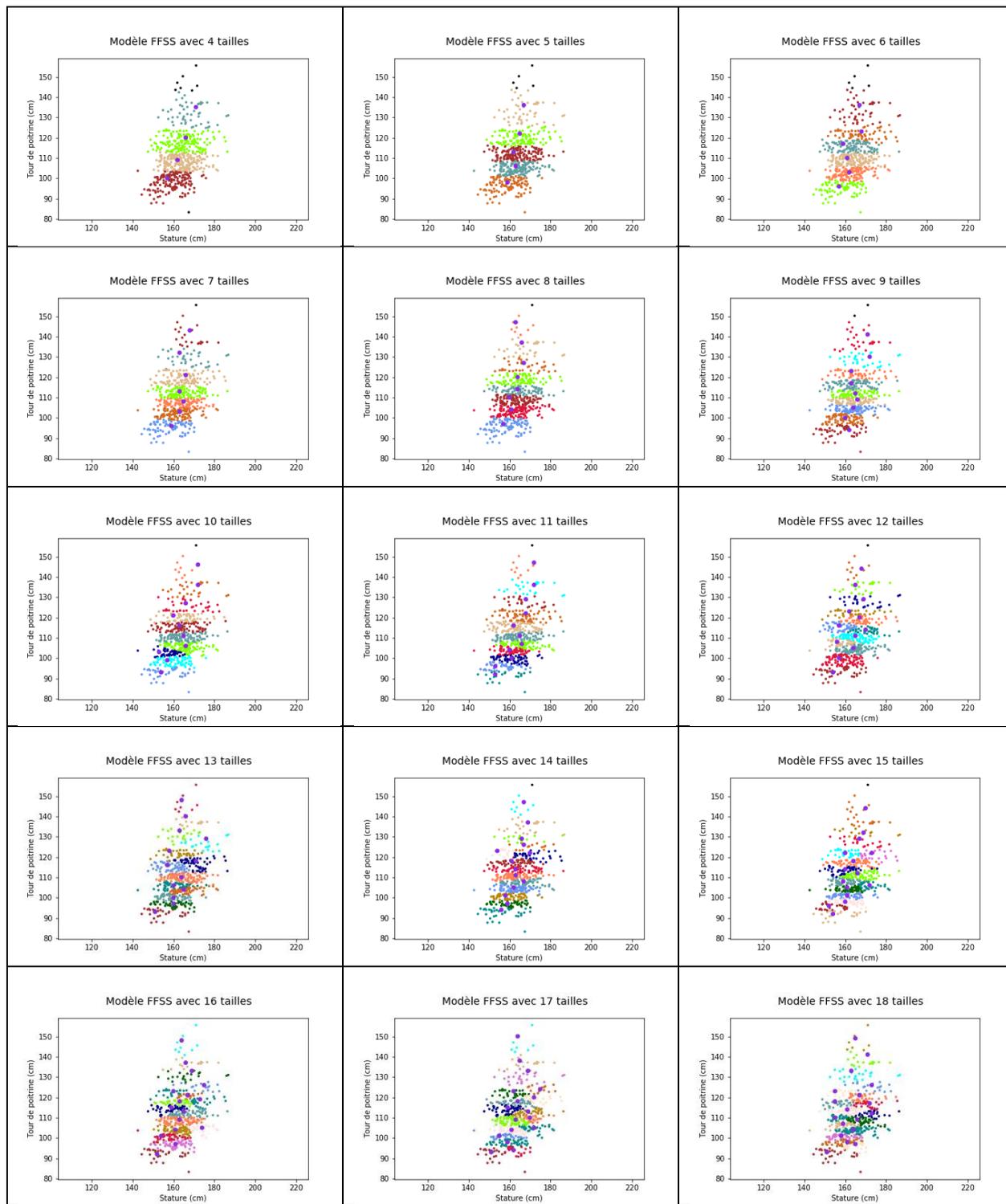
Morphotype 2



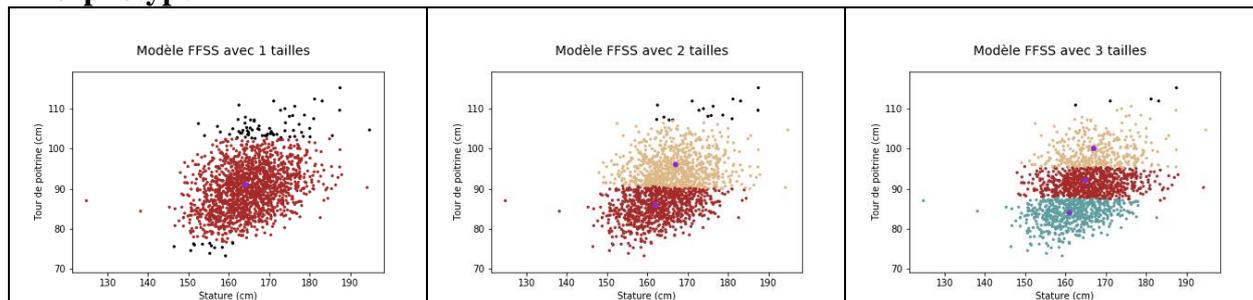


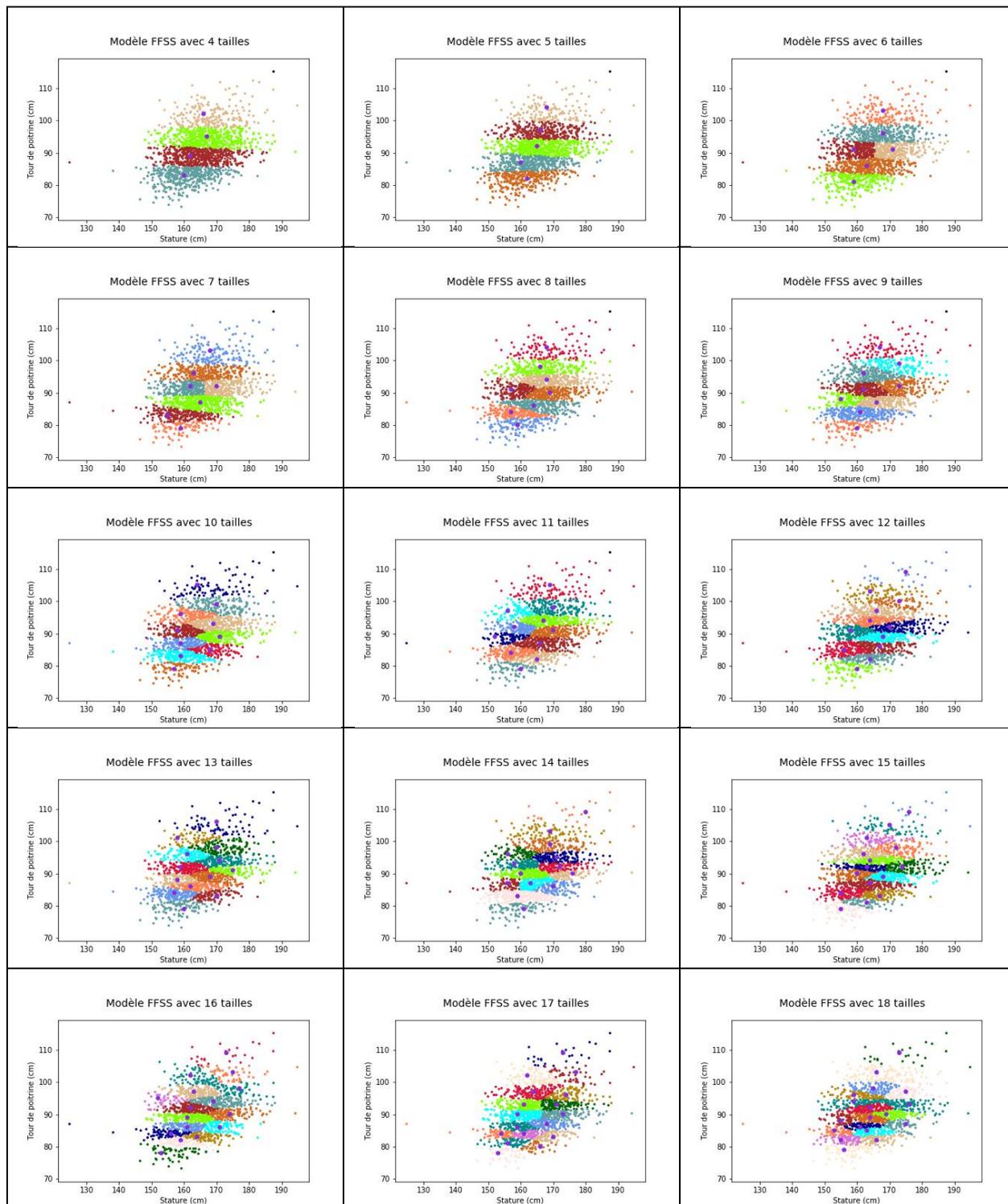
FFSS Morphotype 1





Morphotype 2

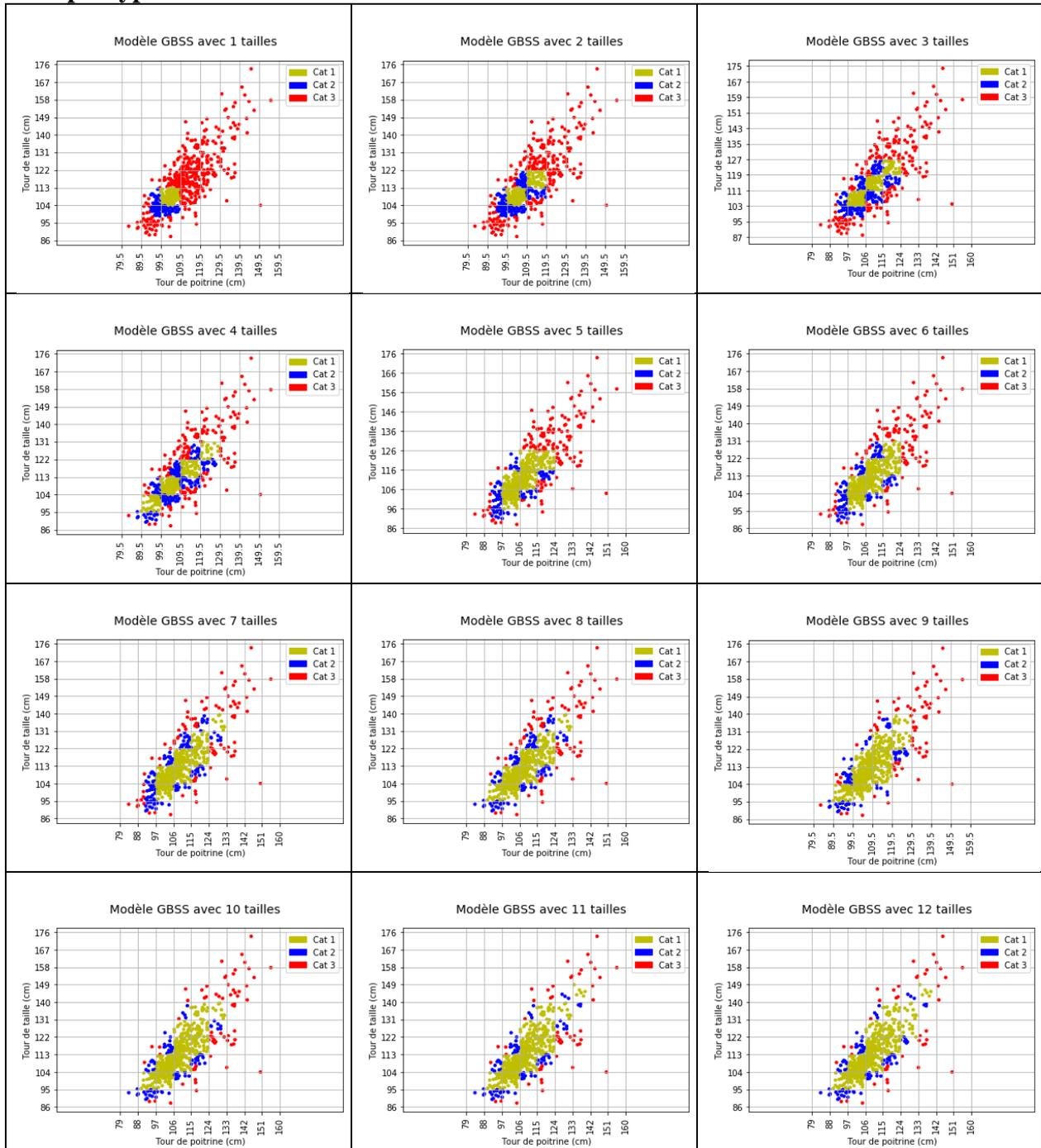


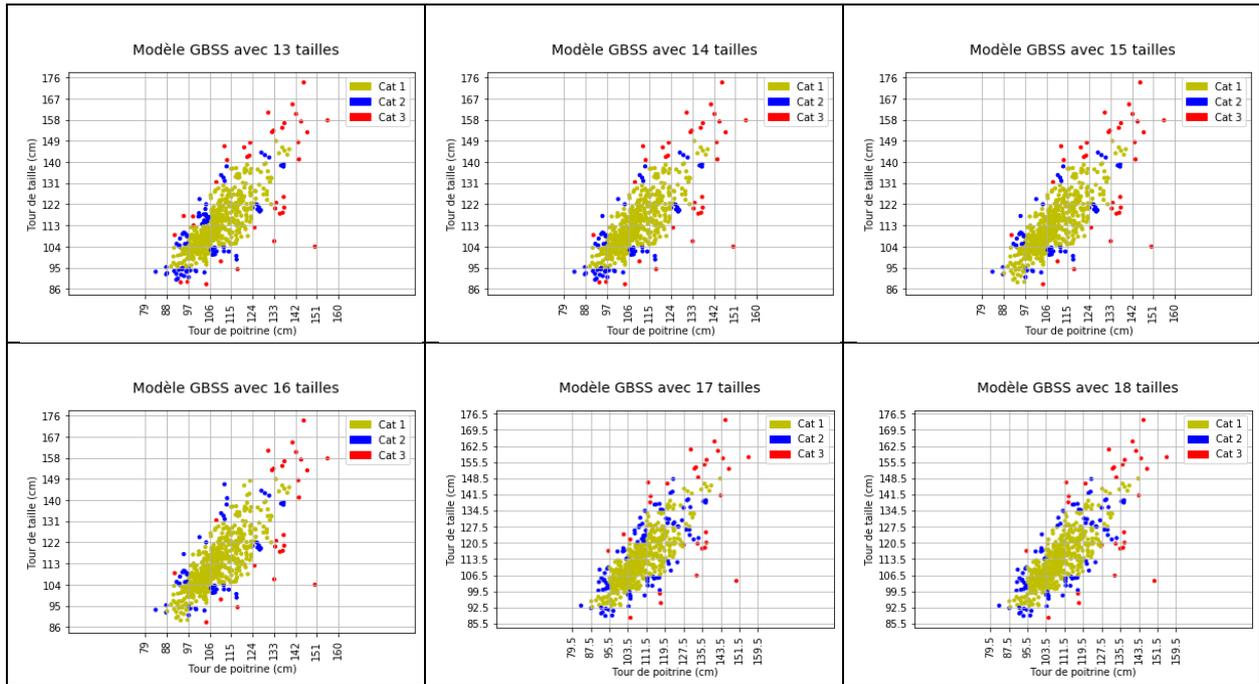


Annexe 6 : Systèmes de tailles pour les mensurations

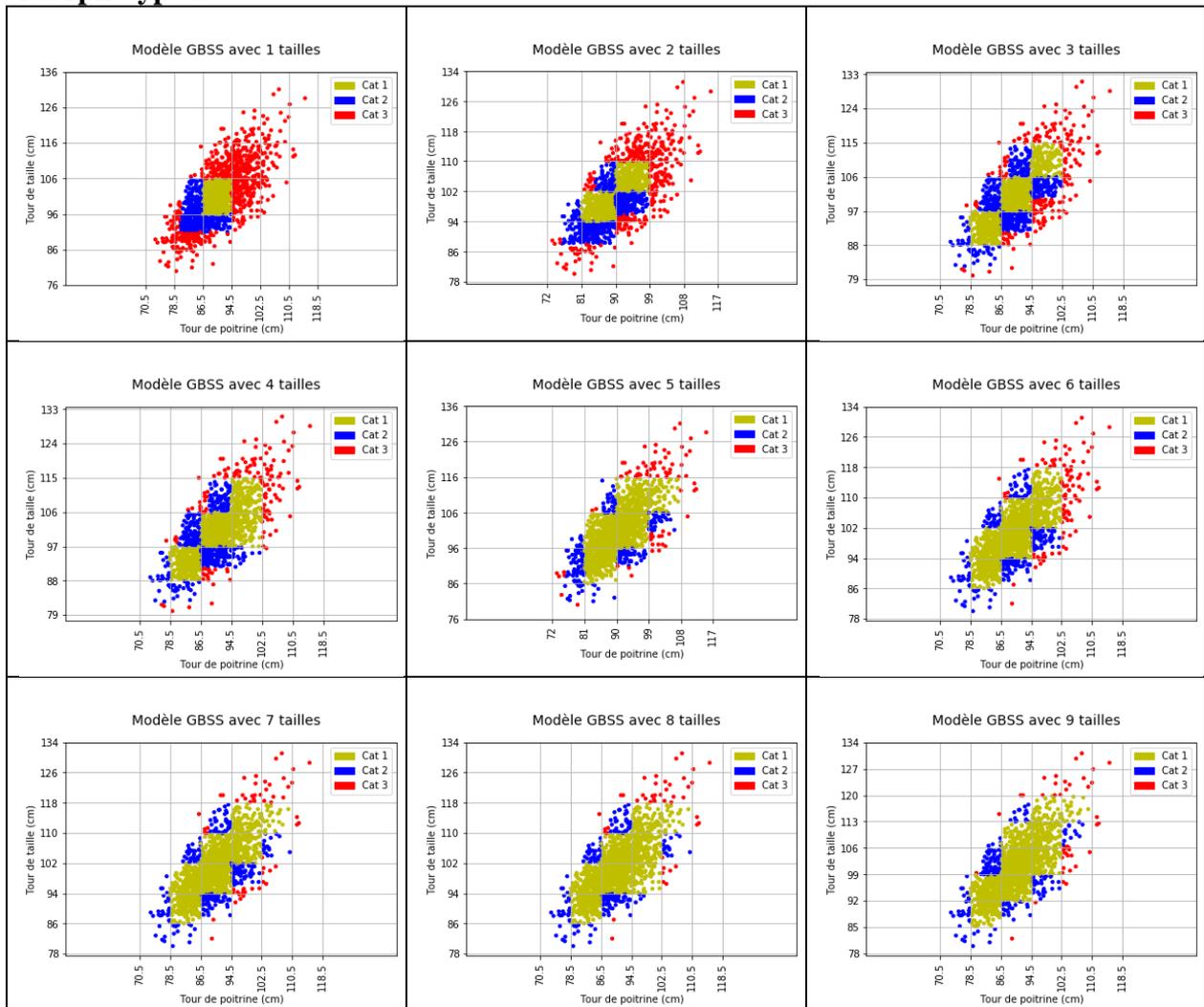
Poitrine/Hanche par morphotype

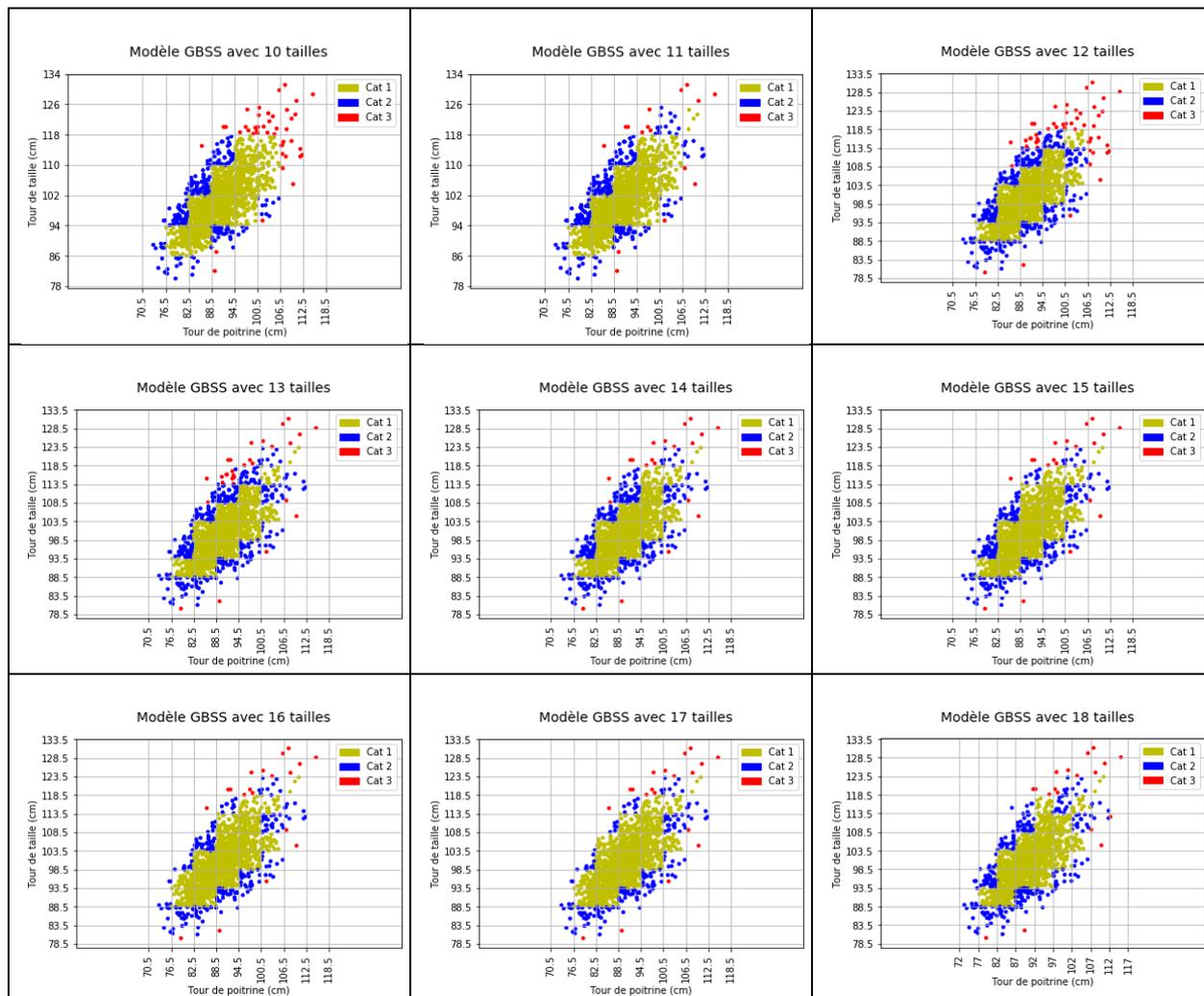
GBSS Morphotype 1



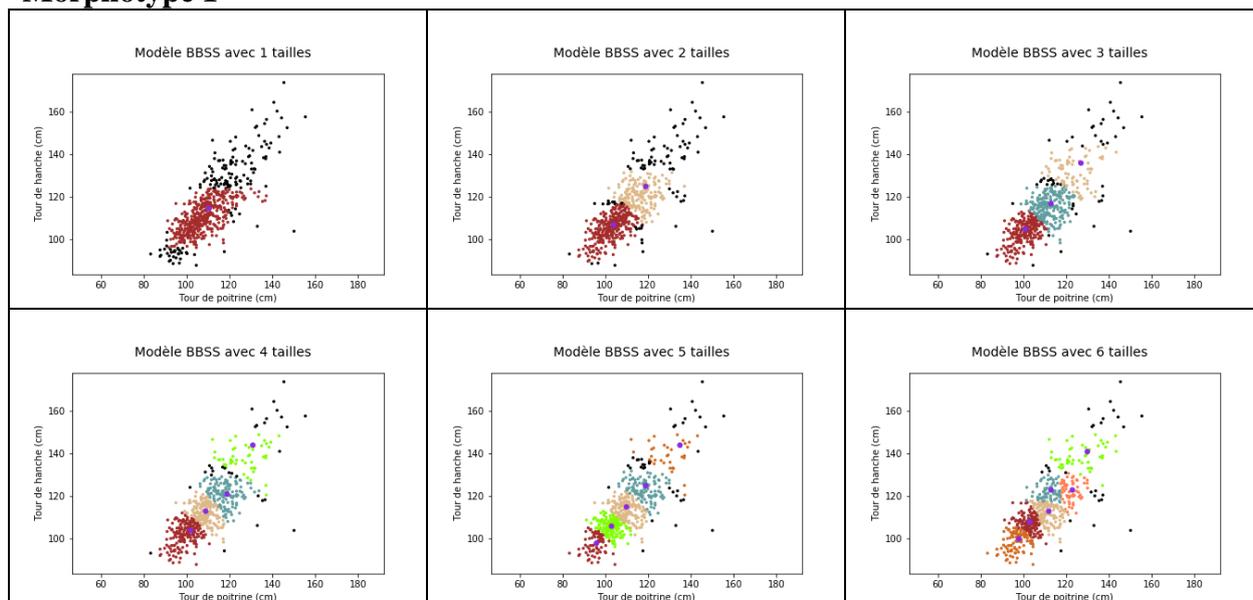


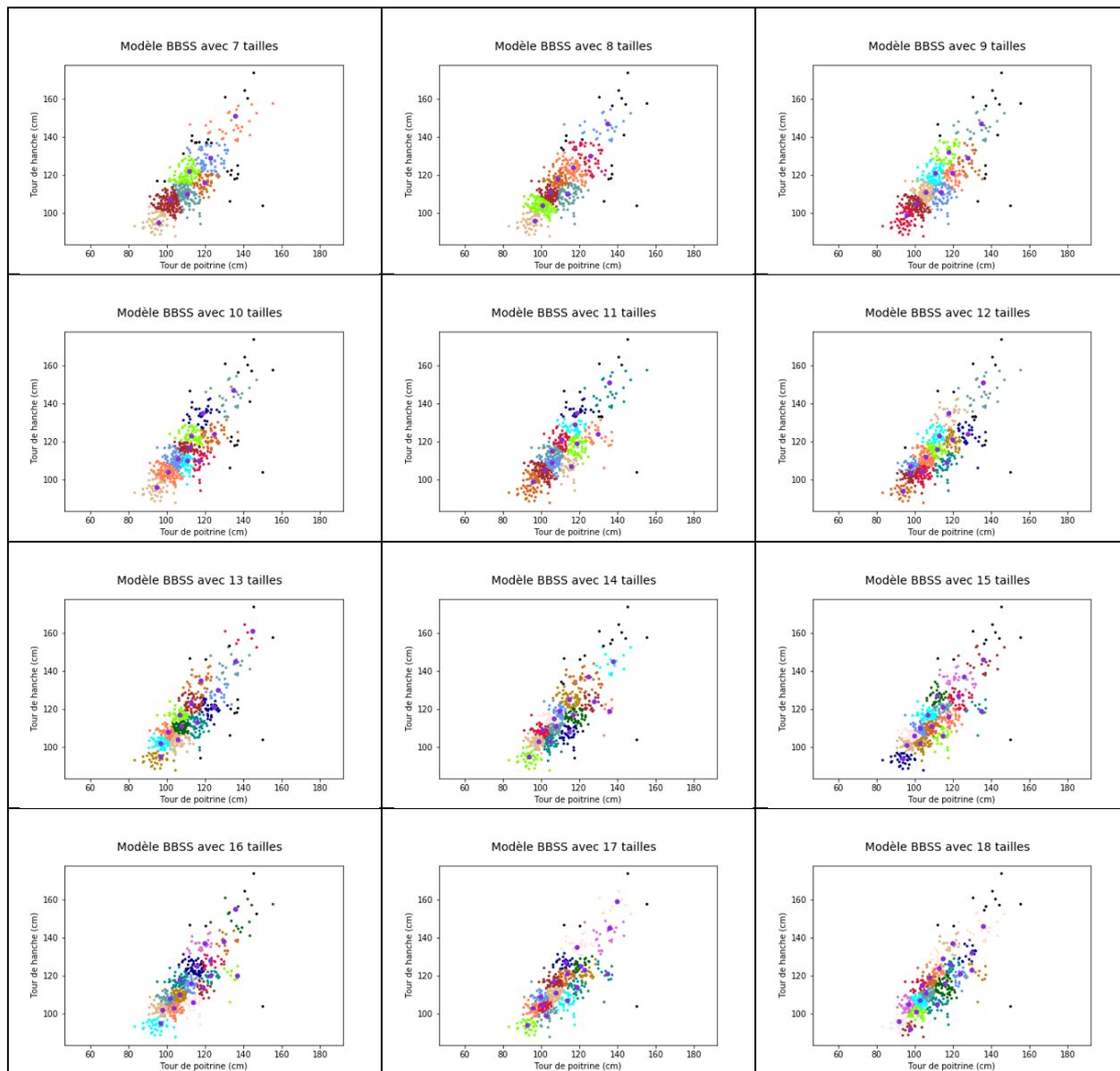
Morphotype 2



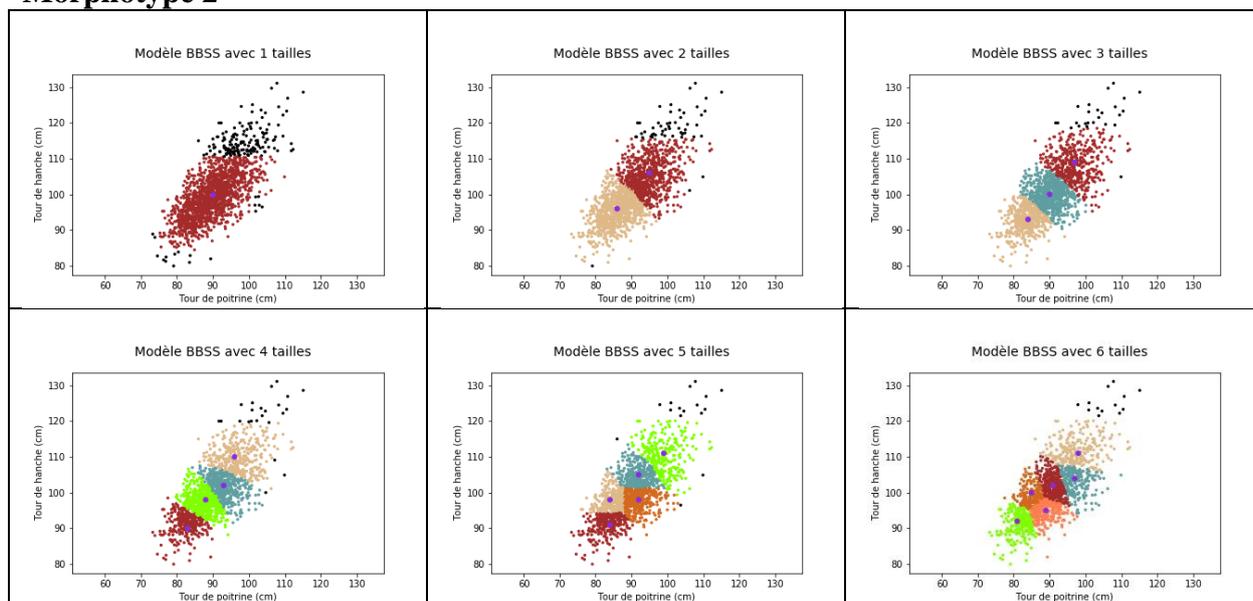


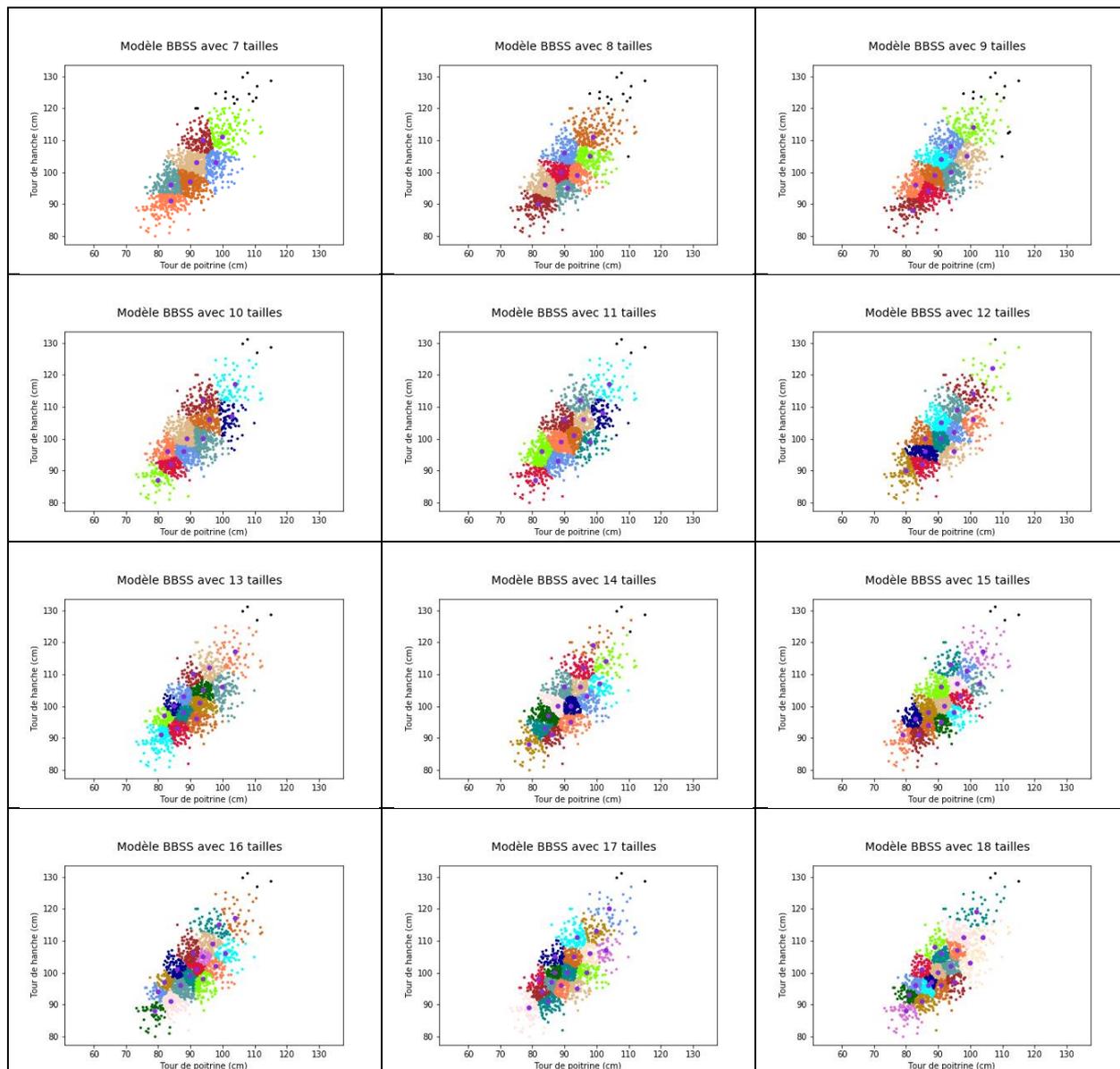
BBSS Morphotype 1



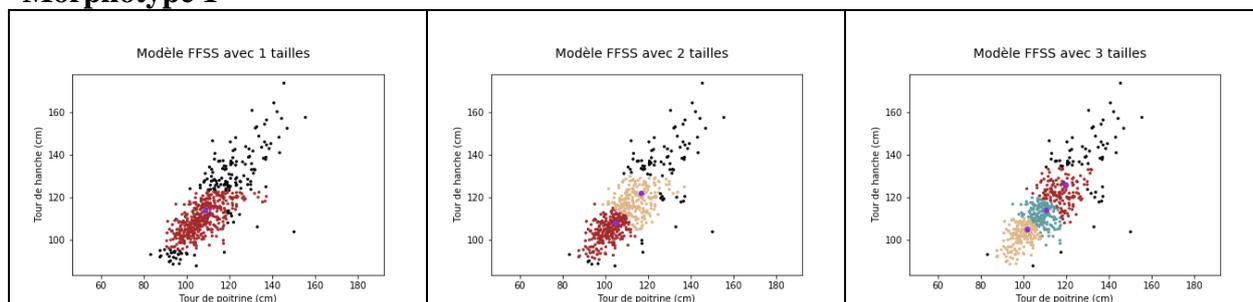


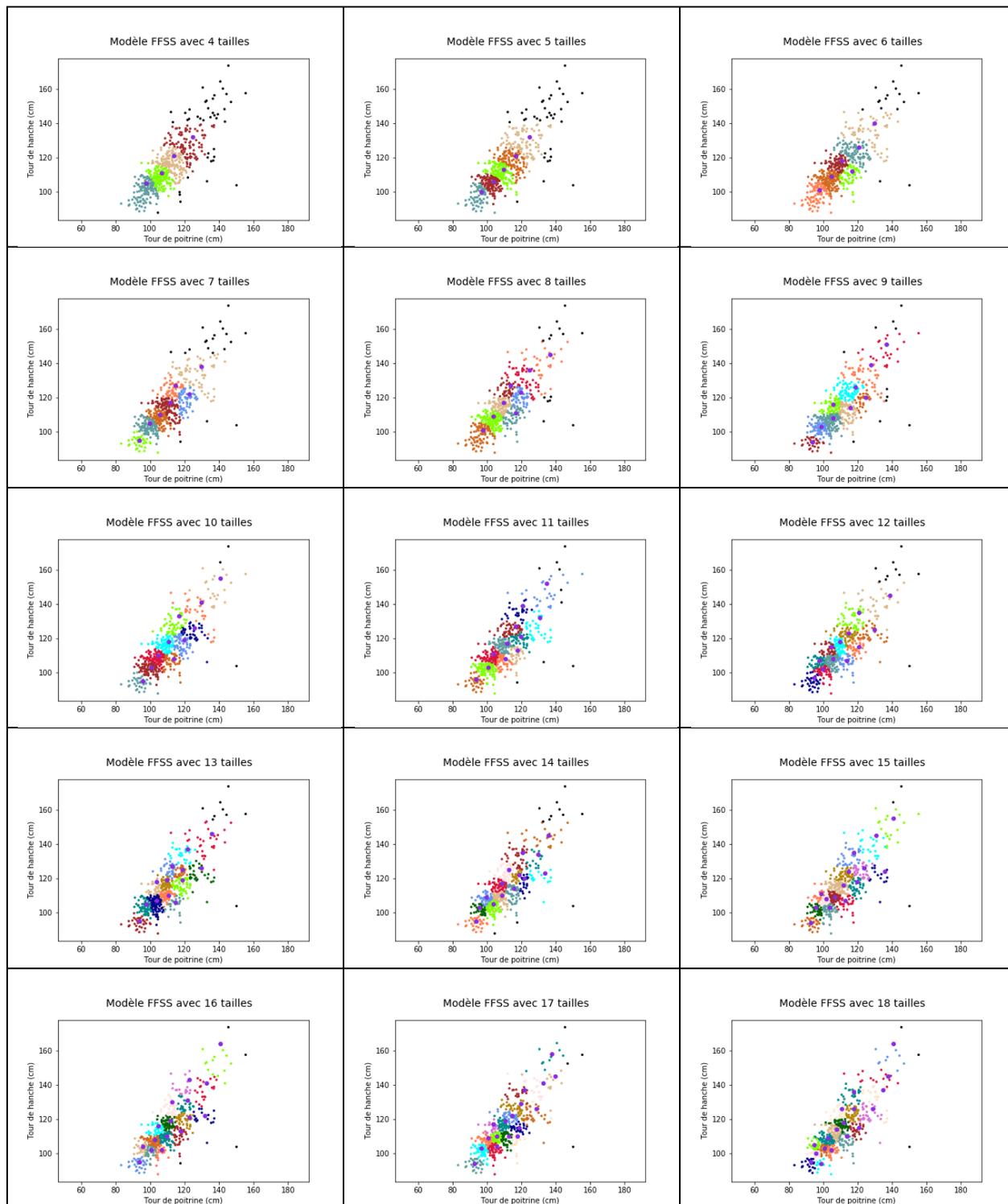
Morphotype 2





FFSS Morphotype 1





Morphotype 2

