

École doctorale n° 432 : Sciences des Métiers de l'Ingénieur

Doctorat ParisTech

THÈSE

pour obtenir le grade de docteur délivré par

l'École Nationale Supérieure d'Arts et Métiers

Spécialité doctorale "Science et Mécanique"

présentée et soutenue publiquement par

Florence DANGLADE

le 07 Décembre 2015

**Traitement de maquettes numériques pour la
préparation de modèles de simulation en conception de
produits à l'aide de techniques d'intelligence artificielle**

Directeur de thèse : **Philippe VERON**

Co-encadrant de thèse : **Jean-Philippe PERNOT**

Jury

Améziane AOUSSAT, Arts et Métiers ParisTech,	Professeur	Président
Benoit EYNARD, Université de Technologie de Compiègne,	Professeur	Rapporteur
Claire LARTIGUE, Université Paris-Sud,	Professeur	Rapporteur
Christian MASCLE, Ecole Polytechnique de Montréal,	Professeur	Examineur
Philippe VERON, Arts et Métiers ParisTech,	Professeur	Examineur
Jean-Philippe PERNOT, Arts et Métiers ParisTech,	Professeur	Examineur
Lionel FINE, Airbus Groups Innovation,	Docteur	Examineur
Arts et Métiers ParisTech		

Laboratoire : UMR 7296 - LSIS - Laboratoire des Sciences de l'Information et des Systèmes

Equipe : INSM - Ingénierie Numérique des Systèmes Mécaniques, Aix en Provence, France

Table des matières

Remerciements	2
Introduction générale	3
I Contexte, état de l’art et verrous scientifiques	6
1 Les représentations intermédiaires du produit pour la simulation numérique	7
1.1 Les représentations multiples de la maquette numérique	8
1.2 La simulation numérique pour la réduction du cycle de développement d’un produit	10
1.3 La préparation des modèles CAO pour la simulation numérique	11
1.3.1 Principales phases de la préparation des modèles intermédiaires . .	11
1.3.2 Modèles intermédiaires pour la simulation numérique	13
1.3.3 Une multitude de processus pour un objectif de simulation	17
1.3.4 Des connaissances en préparation de modèles peu exploitées	20
1.4 Problématique générale et questions soulevées	22
2 Techniques de simplification de modèles CAO pour la simulation numérique	25
2.1 Sélection et classification des techniques de simplification	26
2.1.1 Choix d’une technique de simplification pour l’analyse numérique .	26
2.1.2 Principales familles de méthodes de simplification	27
2.1.3 Pratiques industrielles	28
2.2 Techniques basées sur la suppression de détails	29
2.3 Techniques de simplification basées sur les volumes	32
2.3.1 Filtrage de pièces	32
2.3.2 Substitution	33
2.3.3 Voxélisation	34
2.4 Techniques basées sur les surfaces	35
2.4.1 Décimation et face clustering	35
2.4.2 Construction d’enveloppes	37
2.4.3 Agrégation	38
2.5 Comparaison de modèles géométriques	40

2.5.1	Mesures de similarité	TABLE DES MATIÈRES	46
2.5.2	Identification de caractéristiques globales		41
2.6	Synthèse et verrous relatifs aux méthodes de simplification des modèles CAO		43
3	Techniques d'évaluation de l'impact de la préparation sur le résultat de l'analyse		45
3.1	Écarts dus à une mauvaise définition du modèle géométrique simplifié . . .		46
3.1.1	Méthodes basées sur des critères subjectifs		47
3.1.2	Méthodes basées sur des critères géométriques		48
3.1.3	Méthodes basées sur des critères physiques		49
3.2	Écarts dus au maillage		52
3.2.1	Critères de forme.		52
3.2.2	Critères de taille.		52
3.3	Synthèse et verrous relatifs aux techniques d'évaluation de l'impact de la préparation sur le résultat de simulation.		54
4	Techniques d'intelligence artificielle		56
4.1	Les systèmes experts		57
4.1.1	Les systèmes à base de connaissances		57
4.1.2	Les systèmes à base de cas		57
4.2	Principes de l'apprentissage artificiel		58
4.2.1	Apprentissage supervisé		59
4.2.2	Apprentissage non supervisé		60
4.2.3	Logiciels de création d'algorithmes d'apprentissage		62
4.3	Techniques d'intelligence artificielle utilisées en ingénierie mécanique . . .		63
4.3.1	Objectifs d'apprentissage		63
4.3.2	Variables d'entrée		64
4.3.3	Techniques utilisées		65
4.4	Les principaux classifieurs		65
4.4.1	Réseaux de neurones		65
4.4.2	Machines à vecteur de support		67
4.4.3	Arbres de décision		68
4.5	Méthodologie de conception d'un modèle d'apprentissage		70
4.5.1	Sélection des données d'entrée		70
4.5.2	Construction des modèles d'apprentissage		71
4.5.3	Sélection et évaluation des modèles d'apprentissage		72
4.6	Positionnement des travaux par rapport à l'utilisation des techniques d'IA .		74
4.7	Verrous relatifs aux techniques d'intelligence artificielle		75

5	Approche proposée pour l'estimation de la qualité d'un processus de préparation de modèles de simulation	77
5.1	Structure de l'approche	78
5.2	Présentation générale de l'approche proposée	81
5.3	Contributions	85
6	Modélisation des données pour l'apprentissage	88
6.1	Représentation des processus de préparation	89
6.1.1	Représentation globale d'un processus de préparation	89
6.1.2	Simplification de la géométrie	92
6.1.3	Adaptation	93
6.1.4	Maillage	94
6.1.5	Séquençage des opérations	96
6.1.6	Conclusion sur la représentation des processus de préparation	96
6.2	Modélisation des données pour la description et l'évaluation d'un processus de préparation	98
6.2.1	Processus global de préparation et sous-processus local de simplification	98
6.2.2	Formalisation de la représentation des données	99
6.2.3	Variables à prédire : critères d'évaluation de la qualité d'un processus de préparation	104
6.2.4	Modélisation d'un processus de préparation	106
6.2.5	Modélisation des données pour l'évaluation d'un processus de préparation	116
6.2.6	Constats et synthèse sur la modélisation des données	126
6.3	Identification des variables explicatives	130
6.3.1	Suppression de variables corrélées	131
6.3.2	Sélection des variables explicatives	133
6.3.3	Conclusion sur la sélection des variables explicatives	135
6.4	Construction d'exemples pour l'apprentissage	136
6.4.1	Espace "erreur sur le résultat de l'analyse / niveau de simplification"	136
6.4.2	Contenu de la base d'exemples	137
6.4.3	Méthode pour la construction d'exemples	139
6.4.4	Conclusion sur la construction d'exemples pour l'apprentissage . . .	141
6.5	Synthèse sur la modélisation des données pour l'apprentissage	143
7	Apprentissage et prédictions sur des nouveaux cas	145
7.1	Construction des modèles d'apprentissage	146
7.1.1	Partitionnement des exemples	148

7.1.2	Implémentation et configuration des variables pour l'apprentissage	159
7.1.3	Configuration des classifieurs	157
7.1.4	Évaluation de la classification	159
7.2	Utilisation sur un nouveau cas	161
7.2.1	Approche générale pour l'évaluation de la qualité d'un processus de préparation	161
7.2.2	Prédiction de la qualité de la simplification d'un sous-ensemble	163
7.2.3	Prédiction de la qualité globale d'un processus de préparation	165
7.3	Évaluation des processus de préparation : aide à la décision	168
7.3.1	Indice de performance d'un processus de préparation	168
7.3.2	Identification du meilleur processus de préparation pour le simulation	168
7.4	Conclusion sur l'apprentissage et les prédictions sur un nouveaux cas	171
8	Expérimentations et résultats	173
8.1	Contexte des expérimentations	174
8.1.1	Présentation du cas de simulation	174
8.1.2	Modèles CAO à préparer	175
8.1.3	Simplification des modèles CAO	178
8.1.4	Adaptation des modèles CAO	184
8.1.5	Maillage des modèles	187
8.1.6	Simulation	190
8.1.7	Apprentissage et tests sur de nouveaux cas	191
8.1.8	Synthèse sur la construction d'exemples de processus de préparation de modèles	191
8.2	Configuration des modèles d'apprentissage	195
8.2.1	Modélisation d'exemples pour l'apprentissage et les tests	195
8.2.2	Sélection des variables explicatives	200
8.2.3	Configuration des modèles d'apprentissage	203
8.3	Validation de l'estimateur de qualité d'un processus de préparation	208
8.3.1	Description des nouveau cas	208
8.3.2	Estimation de l'impact de la simplification sur un sous-ensemble (variables y_1 et y_2)	208
8.3.3	Estimation de l'impact de la préparation sur le modèle global (variables y_3 , y_4 et y_5)	210
8.4	Synthèse sur l'estimation de la qualité d'un processus de préparation	212
8.4.1	Domaine de validité du point de vue de l'objectif de préparation	212
8.4.2	Domaine de validité du point de vue des opérations de préparation	213
8.4.3	Domaine de validité du point de vue de l'estimation de la qualité d'un processus de préparation à l'aide de classifieurs	213
	Conclusion et perspectives	214

TABLE DES MATIÈRES	I
Bibliographie	
A Séquençages d'opérations de simplification	II
B Variables explicatives sélectionnées	IV
Liste des abréviations et notations	VIII
Liste des figures	XVI
Liste des tableaux	XXI

Remerciements

Mes remerciements s'adressent tout d'abord à Madame Claire LARTIGUE et Monsieur Benoit EYNARD pour avoir accepté d'être les rapporteurs de mes travaux de thèse.

Je remercie Messieurs Améziane AOUSSAT et Christian MASCLE d'avoir accepté d'évaluer mes travaux de recherche.

Je remercie particulièrement mes deux directeurs de thèse Philippe VERON et Jean-Philippe PERNOT pour avoir encadré mes travaux de recherche durant ces trois années. Merci tout d'abord de m'avoir proposé ce sujet grâce auquel j'ai découvert le domaine de l'intelligence artificielle. Et surtout, merci pour ces heures et souvent ces nuits que vous avez passé à relire mes articles et présentations. Vous m'avez donné le goût de la recherche et l'envie de poursuivre ma carrière dans cette voie.

Je remercie Lionel FINE et toute son équipe d'Airbus Groups Innovation. En plus de me fournir un passionnant support d'étude pour valoriser mes recherches, Lionel a été comme un troisième co-encadrant. Ses conseils toujours très avisés m'ont permis de progresser. Merci Lionel pour ces échanges très riches, j'espère que cette thèse n'est que le début d'une fructueuse collaboration.

Je remercie mon directeur Laurent ARNAUD du campus Arts et Métiers ParisTech de Cluny et mes collègues de l'équipe de construction mécanique pour leur constante bonne humeur et pour leur soutien sans faille tout particulièrement ces derniers mois. Merci donc à Angelo, Frédéric, Michel, Yann et Yoann.

Enfin, merci à Marc de tout cœur pour ta patience, tes encouragements et tes petits plats.

Introduction générale

Objectif et problématique de recherche

Réduire le temps de développement et les délais de mise sur le marché d'un produit tout en améliorant sa qualité est un enjeu industriel majeur. Le cycle de développement d'un produit met en œuvre des activités multiples telles que la simulation en pré dimensionnement ou en dimensionnement, l'optimisation du produit, la simulation des processus de fabrication et le prototypage. A chaque activité correspond une représentation intermédiaire spécifique du produit ayant un niveau de détails plus ou moins élevé et complété par des informations technologiques.

Le processus de préparation d'une représentation initiale vers une représentation dédiée à une activité donnée met en œuvre une chaîne d'opérations obtenues avec différents outils qui doivent prendre en compte un grand nombre de paramètres. Actuellement, les outils pour réaliser ces opérations de préparation existent. Cependant, les processus de préparation sont des tâches complexes qui reposent souvent sur les connaissances des experts et sont peu formalisés. L'impact de la préparation d'une représentation du produit sur la qualité d'une activité n'est pas connue a priori, les experts choisissent un processus de préparation en minimisant les risques. De nombreuses itérations peuvent être nécessaires et la qualité de la représentation après préparation est généralement supérieure au besoin réel. Pouvoir évaluer a priori la qualité d'un processus de préparation avant de le réaliser assurerait une meilleure maîtrise des processus, des coûts de préparation et éviterait la surqualité.

L'objectif de ces travaux est de proposer une aide à la décision durant un processus de préparation d'une représentation d'un produit pour une activité dédiée en utilisant les techniques d'intelligence artificielle.

Les techniques d'intelligence artificielle sont capables d'imiter et de prévoir un comportement à partir d'exemples judicieusement choisis. Cela consiste tout d'abord à capitaliser les connaissances acquises par les experts, à en extraire des données qui caractérisent au mieux le comportement à prédire, puis à les implémenter dans des outils d'intelligence artificielle. Finalement, lorsqu'un nouveau cas se présente, des règles déterminées par

Les techniques d'intelligence artificielle peuvent être exploitées pour estimer la qualité d'un processus de préparation d'une maquette numérique (DMU) de produit pour une activité dédiée. Les connaissances à extraire d'un cas d'exemple sont la description du ou des processus de préparation, les caractéristiques de la maquette numérique originale et préparée, et la description de l'activité. Ces exemples ne sont disponibles qu'en quantité limitée, il est donc nécessaire de proposer une méthode par apprentissage capable de garantir la fiabilité des prédictions malgré un petit nombre d'exemples.

Actuellement, les outils et méthodes de préparation d'une DMU pour une activité dédiée existent. Cependant, les facteurs qui caractérisent les critères d'évaluation d'un processus ne sont pas formalisés. Le principal défi de ces travaux est d'identifier les critères les plus déterminants pour évaluer un processus de préparation d'une DMU.

Les travaux sont appliqués plus particulièrement au transfert d'un modèle CAO vers un modèle pour la simulation numérique qui est une activité qui nécessite de lourdes opérations d'adaptation de la maquette numérique. Au final, il est proposé une méthode pour estimer la qualité de processus de préparation pour différents scénarios dans le but d'aider les analystes dans le choix d'opérations de préparation de modèles CAO pour la simulation. Cela ne dispense pas les analystes de la simulation numérique mais permet d'obtenir plus rapidement un modèle tout en maîtrisant a priori la qualité et les coûts.

La transformation de données extraites des modèles CAO en données utilisables par les techniques d'intelligence artificielle devra être prise en compte. C'est un des verrous scientifiques de cette contribution.

Structuration du manuscrit

La première partie de ce manuscrit présente le contexte général de l'étude, sa problématique et un état de l'art sur les travaux antérieurs.

- Le premier chapitre 1 décrit le positionnement des travaux et la problématique générale relative à l'évaluation a priori de la qualité d'un processus de préparation de modèles CAO pour la simulation numérique.
- Le second chapitre 2 présente les techniques de simplification de modèles CAO pour les objectifs de simulation ciblées par ces travaux et liste les facteurs de comparaison des modèles CAO originaux et simplifiés.
- Le chapitre 3 présente les techniques d'évaluation de l'impact de la préparation d'un modèle CAO sur le résultat de l'analyse numérique en particulier celles qui permettraient d'évaluer a priori la qualité du modèle préparé.
- Le chapitre 4 présente les techniques d'intelligence artificielle et identifie celles qui seraient les plus adaptées pour estimer a priori une évaluation de qualité d'un processus de préparation.

La seconde partie présente les contributions apportées par l'ensemble des travaux.

- Le chapitre 5 propose une approche générale pour l'aide à la décision lors de la préparation de modèles pour la simulation.
- Le chapitre 6 décrit une méthode de modélisation de la base d'exemples pour l'apprentissage.
- Le chapitre 7 décrit une méthode pour configurer un modèle d'apprentissage capable de prédire pour un nouveau cas de processus de préparation ses critères de qualité.
- Le chapitre 8 présente une application des méthodes proposées à la préparation de modèles CAO complexes pour la simulation du transfert thermique dans une enceinte fermée et met en application les méthodes proposées pour construire une base d'exemple, pour configurer un modèle d'apprentissage et finalement estimer un indicateur de performance d'un processus de préparation.

Le dernier chapitre discute de la validation et des perspectives liées à la principale hypothèse des travaux; à savoir que les techniques d'intelligence artificielles peuvent être utilisées pour évaluer a priori un processus de préparation d'un modèle CAO pour la simulation numérique.

Première partie

Contexte, état de l'art et verrous scientifiques

Chapitre 1

Les représentations intermédiaires du produit pour la simulation numérique

Ce chapitre décrit dans une première section 1.1 le processus de développement d'un produit qui donne naissance à un ensemble de modèles intermédiaires qui sont autant de représentations du produit adaptées à des objectifs et des activités multiples.

La section 1.2 montre que les activités de simulations numériques jouent un rôle de plus en plus important dans le cycle de développement du produit. Les simulations sont de plus en plus complexes, réalistes et variées. Afin d'assurer la faisabilité de la simulation et la qualité des résultats de l'analyse, la maquette numérique du produit doit être préparée en fonction de l'objectif de la simulation.

Il sera montré dans une troisième section 1.3 la complexité de la préparation des modèles pour la simulation et comment la réduction des durées et des itérations de ces préparations participe à raccourcir le cycle de développement des produits.

Le positionnement scientifique et la problématique générale relative à la caractérisation et l'évaluation de processus de préparation de modèles de simulation sont synthétisés en fin de chapitre.

1.1 Les représentations multiples de la maquette numérique

SIMULATION NUMÉRIQUE

Le cycle de vie d'un produit manufacturé suit six macro étapes depuis l'étude de marché jusqu'à la fin de vie du produit (figure 1.1). La phase de développement d'un produit passe par un ensemble d'étapes incluant les phases de conception préliminaire et détaillée qui donnent naissance à des représentations géométriques du produit à des stades plus ou moins matures.

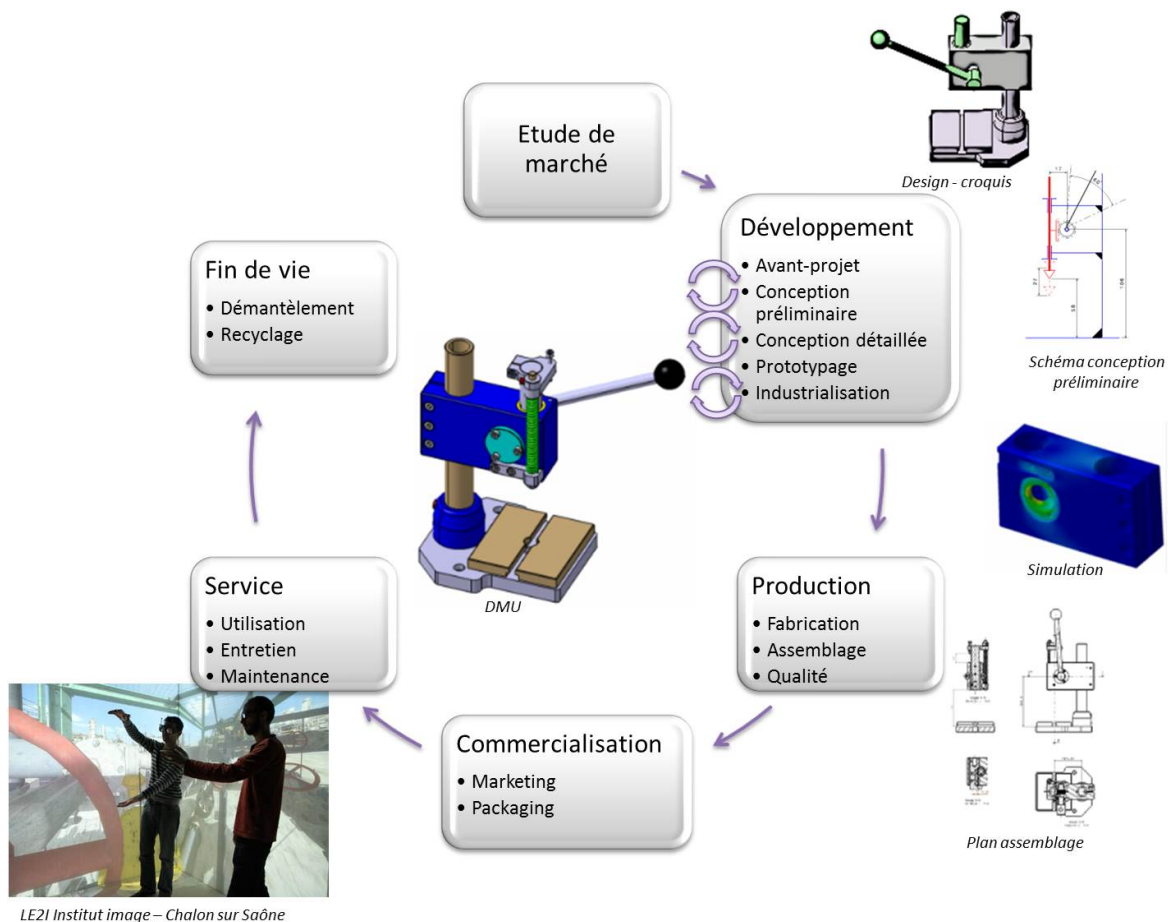


FIGURE 1.1 – Cycle de vie d'un produit et exploitations multiples de la maquette numérique

Une définition de la maquette numérique ou Digital Mock-Up (DMU) a été proposée par le consortium du projet européen AIT – DMU BP (Advanced Information Technology in Design and Manufacturing – Digital Mock-Up Business Process) comme étant : « Une représentation numérique étendue du produit utilisée comme plate-forme de développement produit/processus, de communication et de validation durant toutes les phases de la vie du produit ».

La maquette numérique est utilisée tout au long du cycle de vie du produit (figure 1.1) lors des activités de :

- conception assistée par ordinateur (CAO) ou (CAD : *Computer Aided Design*) pour modéliser la géométrie d'un produit ;

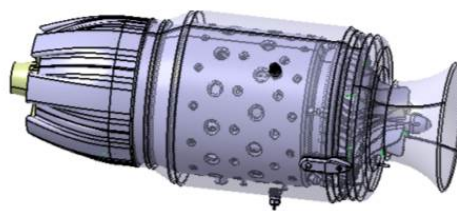
CHAPITRE 1. LES REPRÉSENTATIONS INTERMÉDIAIRES DU PRODUIT POUR LA SIMULATION NUMÉRIQUE

simulation numérique ou (CAE : *Computer Aided Engineering*) pour améliorer la conception des produits en simulant leur performance par le calcul statique ou dynamique par éléments finis, l'analyse thermique et la dynamique des fluides (CFD : *Computational Fluid Dynamics*), l'analyse cinématique ou encore la dynamique de mécanismes ... ;

- simulation de procédés de fabrication (fonderie, moulage, forgeage, enlèvement de matière,...) ;
- simulation de la production pour faire évoluer les procédés, simuler un assemblage, faciliter le suivi de la production ou bien apporter un support à la formation ;
- prototypage virtuel pour la visualisation de processus de maintenance ou de démantèlement ;
- création de documentations techniques (notices de montage/démontage, de maintenance, ou d'utilisation).

Cette multiplicité d'activités met en œuvre un ensemble d'outils et d'acteurs ayant des besoins spécifiques. Cela donne lieu à plusieurs représentations géométriques intermédiaires du produit qui doivent être adaptées à chaque objectif (figure 1.2). Chaque représentation est complétée par des informations technologiques et d'ingénierie propres à chaque activité.

Une représentation géométrique



Des représentations intermédiaires adaptées à des objectifs et activités multiples

Maillage hexaédrique



Elasticité linéaire

Maillage surfacique triangles



Fluides

Images de synthèse



Visualisation

FIGURE 1.2 – Exemples de représentations intermédiaires du produit

CHAPITRE 1 : LES REPRÉSENTATIONS INTERMÉDIAIRES DU PRODUIT POUR LA
SIMULATION NUMÉRIQUE
Les travaux présentés dans cette thèse se positionnent sur le transfert de la maquette numérique définie lors des phases de conception vers des représentations intermédiaires pour la simulation numérique.

1.2 La simulation numérique pour la réduction du cycle de développement d'un produit

La réduction du temps de développement d'un produit est un enjeu majeur pour les industriels. Les entreprises Dassault Systèmes et Sogeti High Tech ont collaborées pour écrire « Virtual Concept - Real Profit » [9] qui explique comment le développement de la simulation numérique et de l'usine virtuelle permet à l'entreprise d'augmenter ses bénéfices. Les auteurs insistent notamment sur l'intérêt de l'utilisation de la simulation numérique au lieu du traditionnel prototypage physique.

Les résultats de la simulation virtuelle sont de plus en plus précis. D'ailleurs, les résultats des simulations sont acceptés par les organismes de certification dans certains cas. De plus, la diminution des prix des outils et les améliorations des interfaces utilisateurs au cours des trente dernières années, ont rendu, aujourd'hui, la simulation plus accessible à tous les ingénieurs et à toutes les entreprises.

Pour innover et in fine augmenter les bénéfices de l'entreprise en utilisant mieux les techniques virtuelles, les auteurs donnent quelques conseils :

- **Il faut accentuer les efforts dès le début de la conception.** Régler des problèmes au niveau de la production entraîne des surcoûts importants et retarde la mise sur le marché du produit. Pouvoir identifier au plus tôt les paramètres importants et ce qui peut être approximé, accélère le développement du produit et permet d'obtenir un produit final de meilleure qualité.
- **Il faut lancer les simulations plus rapidement.** Les simulations de pré dimensionnement ou de dimensionnement et d'optimisation devront être privilégiées aux simulations de validation après conception.

L'idée de lancer au plus tôt l'activité de simulation en parallèle des activités de conception est de plus en plus prise en compte par les entreprises. En effet, la conception séquentielle où chaque phase du cycle de vie du produit est menée l'une après l'autre a été délaissée depuis la fin des années 1990 au profit de l'ingénierie concourante [56]. Ce concept permet aux différents acteurs d'un projet de travailler dès le début du projet sur leurs activités grâce à des échanges simultanés d'informations. La durée du cycle de développement d'un produit est réduite, cependant les itérations sont nombreuses en

La raison pour laquelle l'activité de simulation est décalée par rapport aux activités de conception est souvent liée à la durée de préparation du modèle intermédiaire pour la simulation. Ce délai est dû tout d'abord au fait que les outils utilisés pour la conception et pour la simulation utilisent des modèles de formats différents. Une conversion des données est alors indispensable. La seconde cause est le nombre d'itérations nécessaires avant d'obtenir un modèle de simulation de qualité. En effet, l'analyste ne peut juger la qualité du modèle de simulation seulement lorsque le calcul est réalisé.

1.3 La préparation des modèles CAO pour la simulation numérique

Cette section décrit les principales étapes du processus de préparation d'un modèle pour la simulation ainsi que les caractéristiques des modèles intermédiaires utilisés en conception et en simulation et les freins liés au transfert CAD/CAE. Il peut s'agir de pièces unitaires ou d'assemblages.

1.3.1 Principales phases de la préparation des modèles intermédiaires

Un processus de préparation de modèles intermédiaires pour la simulation numérique prend en entrée un modèle CAO original et fournit un modèle sur lequel sera appliqué le calcul par éléments finis.

Le processus de préparation des modèles pour la simulation numérique consiste à (figure 1.3) :

- A) **Simplifier** le modèle afin d'assurer la faisabilité de la simulation et d'en réduire la durée tout en préservant une qualité de résultat souhaitée. Sans la simplification du modèle original, le maillage et/ou la simulation peuvent être impossibles. Des techniques et outils de simplification appropriés à la préparation de modèles pour la simulation existent et seront analysées dans l'état de l'art du chapitre 2.
- B) **Adapter** le modèle selon l'objectif de la simulation et les conditions limites. Il s'agit généralement de relimiter la géométrie du modèle CAO autour des frontières du maillage et d'identifier ou de créer les surfaces supportant des conditions limites.
- C) **Mailler** le modèle pour permettre une résolution numérique du problème en approximant les domaines par des mailles plus ou moins fines et plus ou moins compliquées (triangles, tétraèdres, hexaèdres, etc.) en fonction des phénomènes physiques simulés, du temps disponible et de la qualité escomptée.

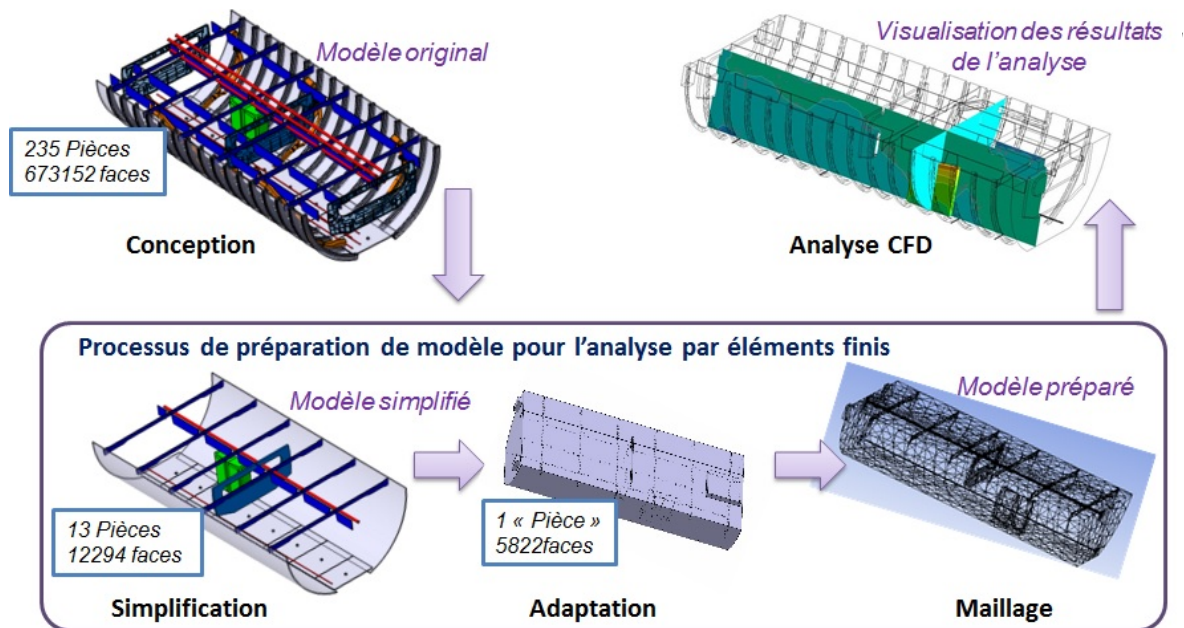


FIGURE 1.3 – Exemple de processus de préparation de modèles CAO pour l'analyse CFD

Lorsque le produit est un ensemble complexe (par exemple un assemblage de plusieurs centaines voir de milliers de pièces) les étapes de maillage et de simulation sont impossibles à obtenir sans simplification.

Malheureusement, le niveau de simplification minimal qui assure la faisabilité du maillage et de la simulation n'est pas connu a priori. Des itérations sont alors nécessaires pour obtenir un modèle préparé de qualité (figure 1.4 a).

Lorsque le modèle a été simulé, les résultats de l'analyse permettent d'identifier les modifications à apporter au produit. Traditionnellement, ces modifications sont apportées sur le modèle CAO original. De nouvelles itérations de préparation puis de calcul sont relancées jusqu'à la validation de la conception.

Les itérations de transfert du modèle de conception vers le modèle de préparation peuvent être réduites en intégrant des processus d'optimisation ou en simplifiant le modèle maillé au lieu du modèle CAO. Cette dernière méthode (figure 1.4 b) a été proposée par Ruding Lou dans sa thèse [37].

Pour réduire la durée de développement, il est fondamental de commencer par diminuer le temps d'obtention des représentations intermédiaires et de limiter le nombre d'itérations (Figure 1.4 c) en obtenant dès le ou les premiers essais la représentation intermédiaire la mieux adaptée à l'objectif.

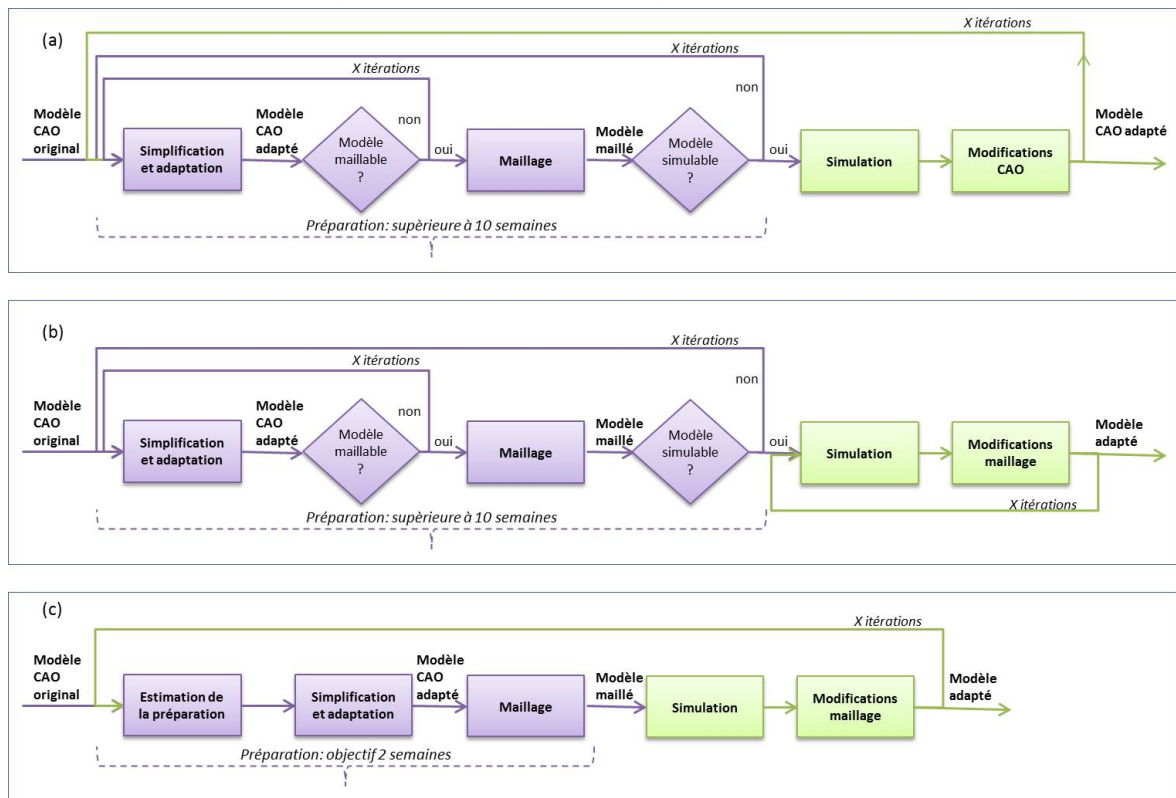


FIGURE 1.4 – Activités de préparation des représentations intermédiaires (RI) et de simulation dans la phase de conception. (a) Méthode traditionnelle. (b) Méthode avec simplification du modèle maillé. (c) Méthode avec estimations a priori du modèle préparé.

1.3.2 Modèles intermédiaires pour la simulation numérique

1.3.2.1 Modèles CAO en phase de conception

La maquette originale est généralement modélisée lors de la phase de conception détaillée à l'aide d'un modelleur 3D. La représentation du modèle original est suffisamment réaliste et contient tous les détails présents sur le produit réel. La modélisation géométrique est basée sur le principe de la représentation B-Rep (Boundary Representation) avec conservation de l'arbre de construction. Une représentation B-Rep est une modélisation par les frontières des solides. Un solide est défini par un domaine intérieur limité par des surfaces, arêtes, et sommets. L'arbre de construction décrit l'historique des opérations utilisées pour créer un solide ou une surface. Il contient les entités qui ont été utilisées pour réaliser ces opérations (points, courbes, esquisses, plans...) et l'ensemble des paramètres qui pilotent ces opérations. Le modèle est alors facilement modifiable.

Le modèle CAO du produit à simuler peut être un ensemble complexe constitué de plusieurs sous-ensembles (figure 1.5 a)). Chaque sous-ensemble est lui-même constitué d'une ou plusieurs pièces de formes et de tailles variées (Figure 1.5 c). Une fonction est attribuée à chaque sous-ensemble qui peut être :

- une condition limite (par exemple une zone où est appliqué un chargement ou un

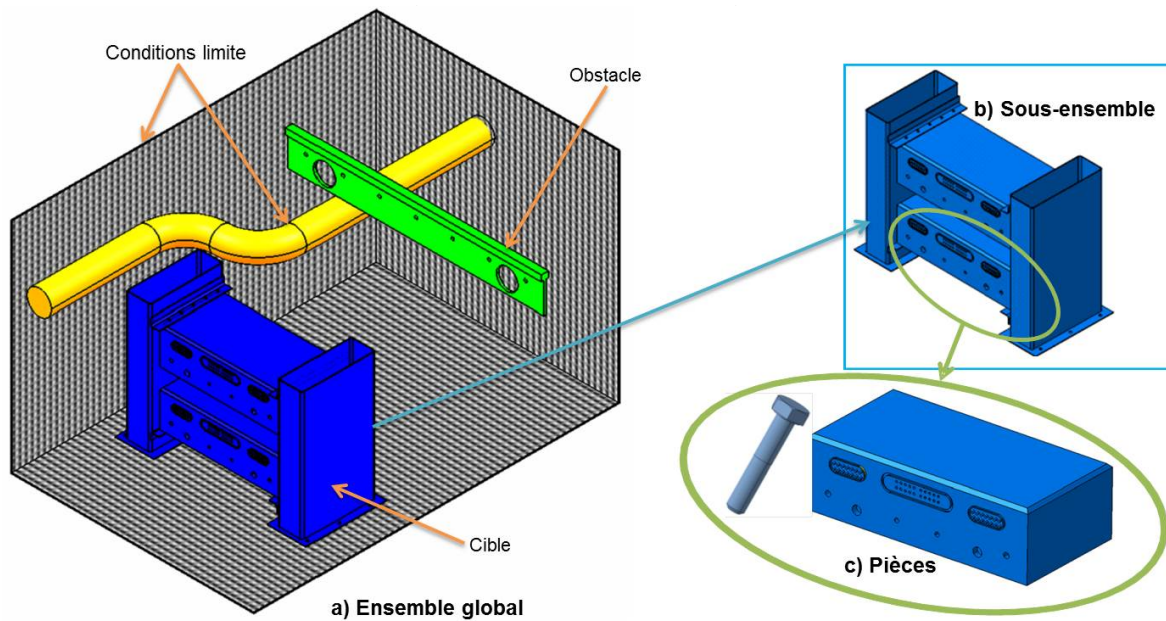


FIGURE 1.5 – Composition du modèle CAO. (a) Ensemble global constitué de conditions limites, de cibles et d'obstacles. (b) Exemple de sous-ensemble. (c) Exemples de pièces

déplacement imposé pour un problème d'élasticité, une paroi ou une source de chaleur pour un problème thermique) ;

- un composant cible sur lequel sera évalué la grandeur physique recherchée ;
- un simple obstacle pour les sous-ensembles qui ne sont pas liés aux conditions limites.

Chaque pièce est modélisée par des formes géométriques et des détails (Figure 1.6) tels que des trous, des bossages, des congés ou des chanfreins. L'ensemble global peut contenir des centaines de sous-ensembles et plusieurs milliers de pièces.

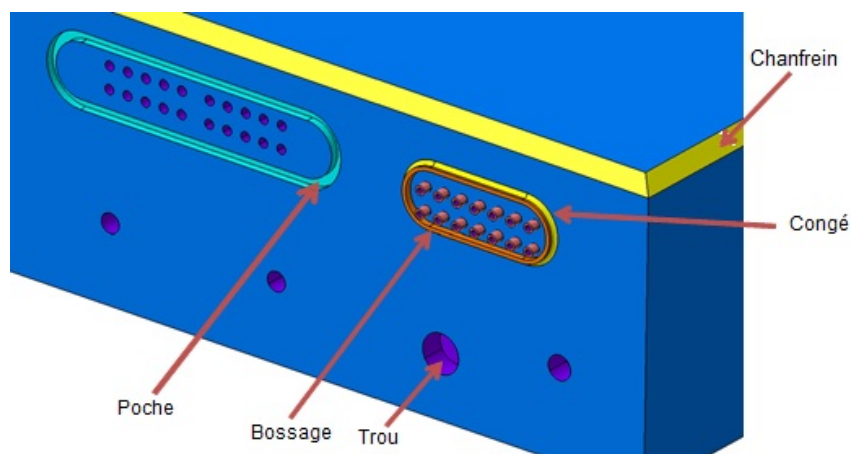


FIGURE 1.6 – Exemples de détails

La maquette numérique originale fournit en plus des données géométriques, des métadonnées qui permettent d'identifier les rôles des composants et leurs caractéristiques

CHAPITRE 1. LES REPRÉSENTATIONS INTERMÉDIAIRES DU PRODUIT POUR LA SIMULATION NUMÉRIQUE
(par exemple la masse volumique, les caractéristiques mécaniques, les caractéristiques thermiques,...). Bien entendu, cette maquette numérique est souvent trop riche pour les différentes phases du cycle de vie du produit. Des adaptations sont donc nécessaires.

1.3.2.2 Modèles mis à disposition des préparateurs pour l'analyse

La maquette mise à disposition par le concepteur pour l'analyste est généralement convertie (Figure 1.7) dans un format de fichier standard (par exemple STEP ou IGES), un format B-Rep polyédrique facétisé (par exemple VRML ou STL) ou un format destiné à la visualisation (par exemple CGR pour CATIA¹) pour des raisons de confidentialité et de rapidité du transfert des données.

Le transfert vers un fichier standard type STEP [1] (Standard for the Exchange of Product Model data, norme ISO 10303) entraîne la perte de l'arbre de construction par contre la maquette est toujours représentée par des solides B-REP et peut être modifiée plus ou moins facilement par les modeleurs 3D et importée directement dans la plupart des mailleurs. La structuration en assemblages et sous-assemblages est préservée.

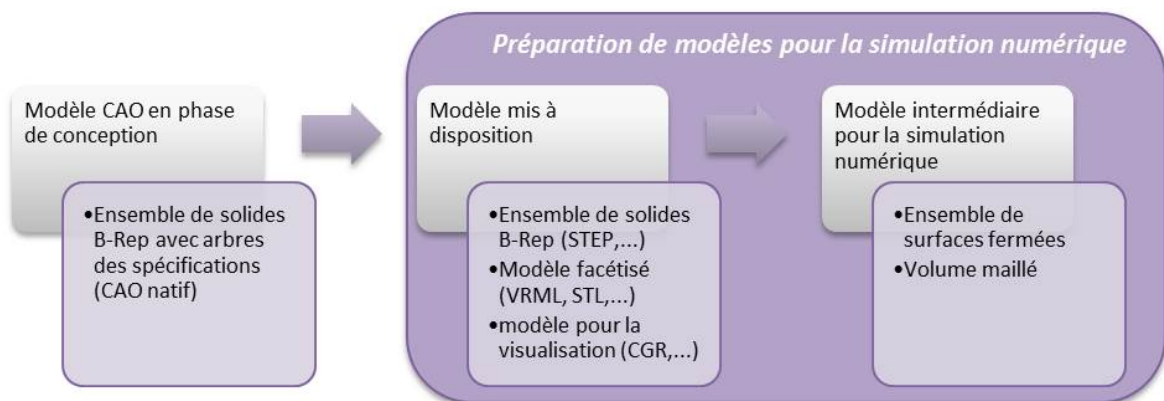


FIGURE 1.7 – Types de formats utilisés du modèle original au modèle intermédiaire pour la simulation numérique

Les formats polyédriques de type VRML (Virtual Reality Markup Language) ou STL (StereoLithography) ne permettent pas les modifications sur les solides. Les modèles destinés à la visualisation permettent d'alléger de façon importante la maquette numérique mais ils ne peuvent être utilisés que par des outils spécifiques les prenant en charge.

Pour un même produit, les sous-ensembles peuvent être transférés dans des formats différents. Il faudra prendre en compte par la suite le fait que le modèle transféré peut contenir une multitude de formats disponibles ce qui peut impacter le coût de la préparation.

1. CATIAV5, Dassault Systems, <http://www.3ds.com/fr/produits-et-services/catia/produits/catia-v5>

CHAPITRE 1 LES REPRÉSENTATIONS INTERMÉDIAIRES DU PRODUIT POUR LA SIMULATION NUMÉRIQUE

1.3.2.3 Modèles préparés pour la simulation numérique

Les modèles préparés sont des représentations intermédiaires du produit sur lesquels les simulations seront appliquées. Il sont obtenus suite aux opérations de simplification, d'adaptation et de maillage telles qu'elles ont été définies dans la section 1.3.1.

Pour définir un modèle intermédiaire pour la simulation, l'analyste commence par répondre à un ensemble de questions sur l'objectif de l'analyse qui vont l'aider à définir les caractéristiques principales du modèle à préparer :

- Quelle grandeur physique sera observée (température, flux, pression, déformation, contrainte,...) ?
- Quel type de simulation doit être mis en œuvre (thermique, fluide, structure, dynamique,...) ?
- Quelle est la zone à observer (point particulier, surface d'un composant, volume limité par les contours d'un composant, ...) ?
- Quel est l'objectif de la simulation. Le dimensionnement ? La validation de tests ? La compréhension d'un phénomène physique ?
- Quel niveau de précision est attendu sur les résultats de la simulation ?
- Quelle méthode de simulation doit être employée :
 - Linéaire ou non-linéaire ?
 - Continue ou discrète ?
 - Modélisation 1D ? 2D ? 3D ?
 - Calcul par éléments finis ? Par volumes finis ? Par différences finies ?
 - Simulation directe ? Simulation grandes échelles ?

Par la suite, les travaux se positionneront sur la préparation de modèles pour l'analyse par éléments ou volumes finis en vue de la validation de tests ou de l'analyse de comportement. La grandeur physique devra être mesurée en tout point d'un volume délimité par le contour intérieur (pour un calcul de structure) ou extérieur (pour un calcul de fluide) du produit (figure 1.8). Le niveau de simplification des surfaces bordant le contour utile sera déterminé afin que l'erreur sur le résultat de l'analyse ne dépasse pas une limite acceptable. Finalement, le modèle préparé sera dans ce cadre un maillage 3D dont les frontières auront été définies par les contours simplifiés du modèle géométrique.

Plus particulièrement, les méthodologies proposées seront appliquées à l'analyse CFD couplée au transfert thermique dans le chapitre 8 résultats.

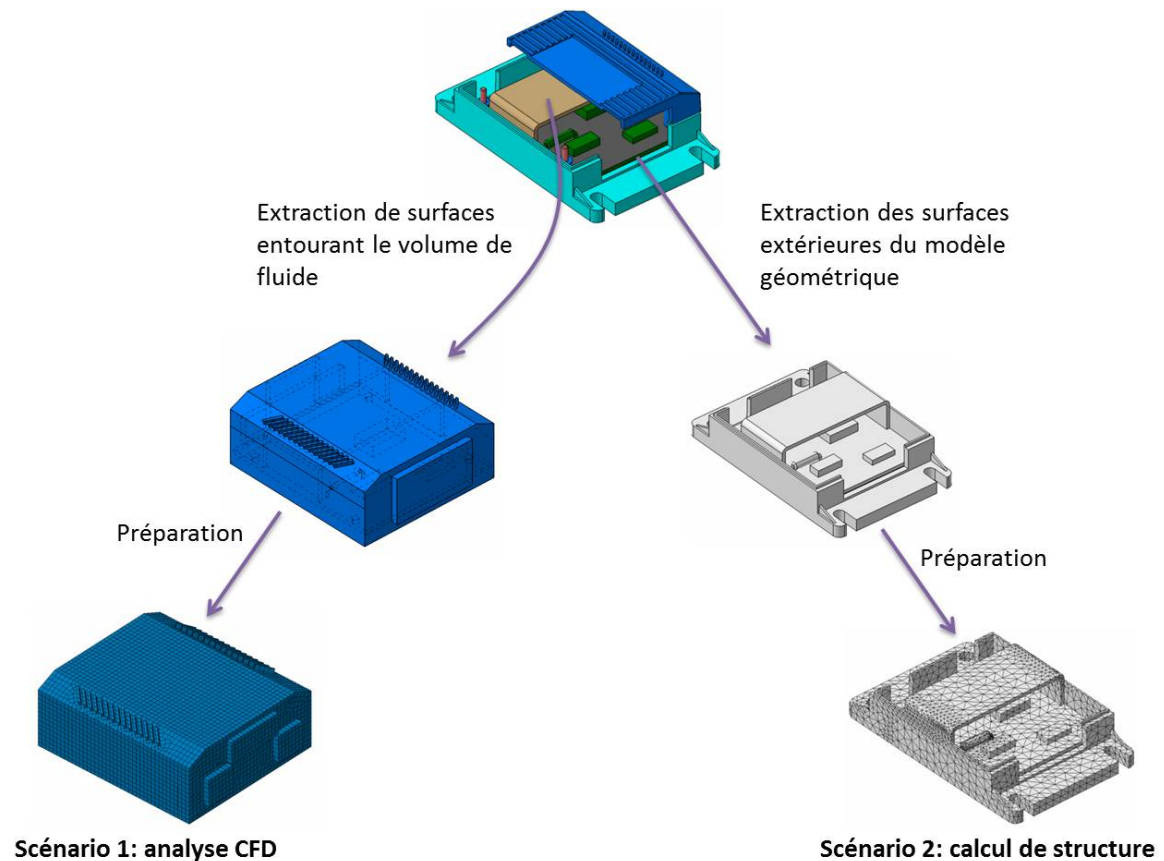


FIGURE 1.8 – Exemples de modèles préparés. Scénario 1 : extraction de surfaces bordant le volume de fluide puis préparation du modèle pour l’analyse CFD. Scénario 2 : extraction du contour du produit puis préparation du modèle pour le calcul de structure.

1.3.3 Une multitude de processus pour un objectif de simulation

1.3.3.1 Travaux antérieurs et pratiques industrielles

L’équipe Ingénierie Numérique des Systèmes Mécaniques du laboratoire LSIS² d’Aix en Provence contribue à la maîtrise des processus de développement de produits et des informations relatives à leur cycle de vie, ceci pour améliorer in-fine les performances et la compétitivité des systèmes complexes. Forte de son expertise sur la manipulation des maquettes numériques et de leur référentiel (simplification, déformation, visualisation,...), l’équipe a mené des études sur les opérateurs de simplification et de maillage des modèles de simulation par éléments finis.

Dans ce contexte, Berthet et al. [8] et Ruf et al. [46] ont constaté que plusieurs opérateurs de simplification, d’adaptation ou de maillage sont utilisés pour un objectif de simulation donné (Figure 1.9). Cependant, les critères de pilotage du processus de préparation des modèles de simulation n’ont pas été formalisés.

2. Laboratoire des Sciences de l’Information et des Systèmes. UMR CNRS N°7296. www.lsis.org

~~CHAPITRE 1. LES REPRÉSENTATIONS INTERMÉDIAIRES DU PRODUIT POUR LA~~
~~Une multitude de processus mettant en œuvre différents séquençages d'opérateurs sont~~
~~possibles. Ces processus dépendent de la nature du modèle à simplifier, du choix des ou-~~
~~tils et du choix des paramètres de contrôle. Plusieurs ingénieurs peuvent proposer des~~
~~processus très différents qui donneront néanmoins des résultats équivalents du point~~
~~de vue de l'analyse par éléments finis. Les pratiques industrielles montrent que lorsque~~
~~les processus sont construits manuellement, le séquençage d'opérations est toujours le~~
~~même, le choix du processus étant généralement basé sur le risque d'échec. Lorsqu'il est~~
~~possible de le créer avec des outils automatisés tels que Model center³, les processus sont~~
~~alors basés sur les hypothèses de la physique, puis sur la précision des résultats et enfin~~
~~l'outillage disponible.~~

3. Phoenix Integration, <http://www.phoenix-int.com/modelcenter/integrate.php>

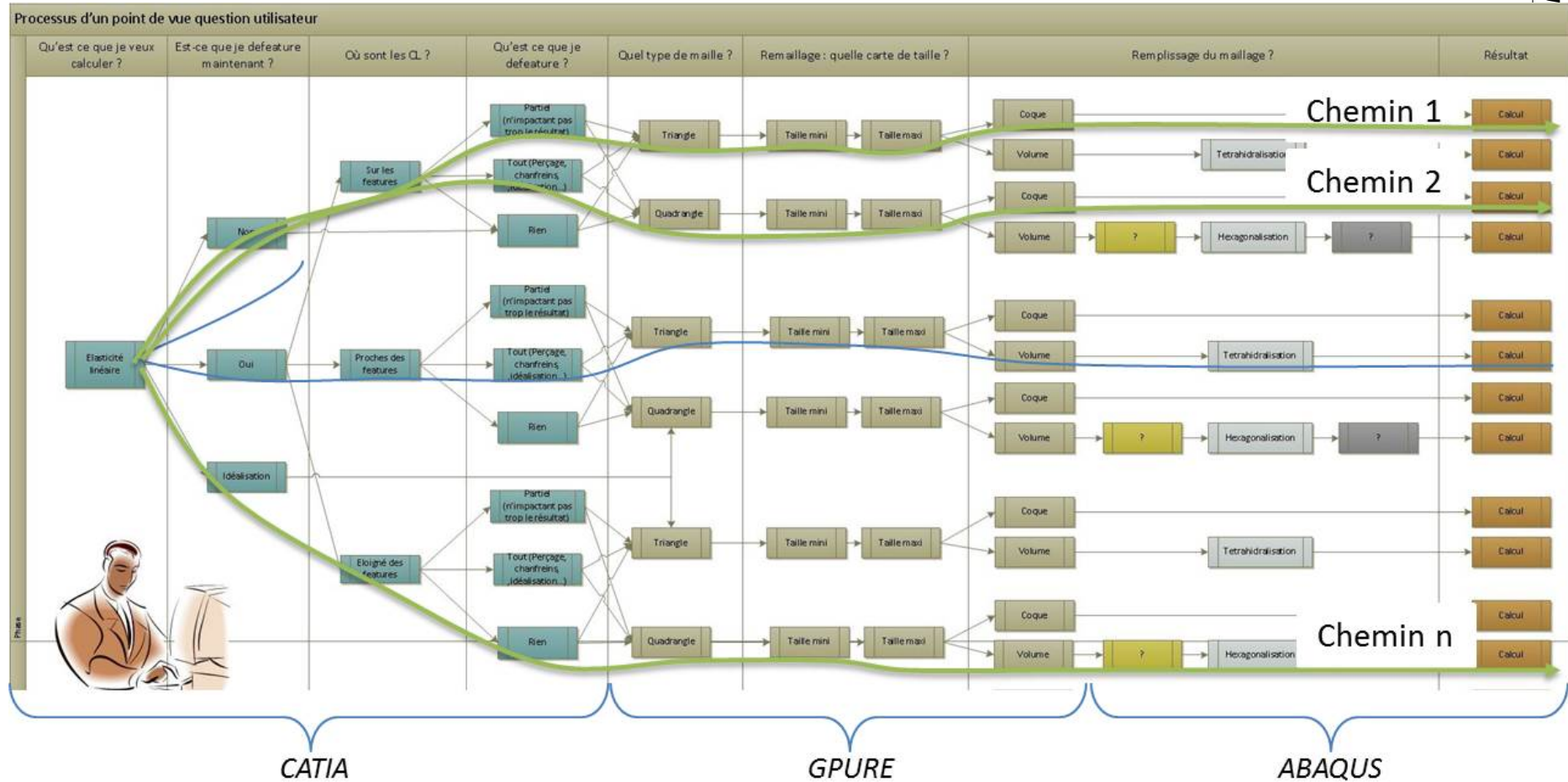


FIGURE 1.9 – Multitude de processus possibles pour la préparation d'un modèle de simulation pour un même objectif de simulation. [8]

Aujourd'hui, il manque une modélisation formalisée des processus de préparation d'un modèle CAO basée sur la précision des résultats des activités en aval de la préparation.

La norme ISO 9000 :2000⁴ (Systèmes de management de la qualité - Principes essentiels et vocabulaire) définit un processus comme étant *"un ensemble d'activités corrélées ou interactives qui transforment des éléments d'entrée en éléments de sortie"*. Un processus se décompose en plusieurs sous-processus qui peuvent eux-même être décomposés en sous-processus de niveaux inférieurs.

Des outils tels que les actigrammes IDEF0/SADT semblent être parfaitement adaptés pour modéliser le processus tel qu'il vient d'être défini. Les actigrammes IDEF0 [41] (*Integration Definition for Function Modeling (Structured Analysis and Design Technique)*) développés aux Etats Unis par Doug Ross en 1977 de la société par Softech⁵ décrivent de façon hiérarchique, du général au particulier, des systèmes complexes en reliant entre elles des activités par différents flux de matière d'œuvre. Dans le cadre d'une étude de faisabilité, ces actigrammes pourraient être une aide à la modélisation des processus de préparation des modèles CAO pour la simulation numérique.

Une représentation formalisée des processus de préparation d'un modèle CAO pour la simulation est proposée dans la section 6.1.

1.3.4 Des connaissances en préparation de modèles peu exploitées

Si l'activité de simulation existe depuis des décennies, il en est de même pour la préparation de modèles intermédiaires pour la simulation. Le chapitre 3, consacrée à l'évaluation de l'impact de la simplification sur le résultat de l'analyse, montrera qu'il existe peu de règles pour identifier le modèle intermédiaire le mieux adapté à une simulation numérique donnée. Au cours de ces années, les analystes ont traité un grand nombre de cas d'analyses. Un expert est en mesure de juger la qualité de ces préparations vis-à-vis de l'objectif de la simulation et des résultats obtenus. Pour cela, il se base sur des critères qu'il a préalablement définis tels que :

- l'objectif de la simulation ;
- la qualité attendue sur les résultats de simulation ;
- le coût des opérations de traitement et de simulation.

4. <http://www.iso.org>

5. Softech Inc., <http://www.softech.com/>

CHAPITRE 1. LES REPRÉSENTATIONS INTERMÉDIAIRES DU PRODUIT POUR LA SIMULATION NUMÉRIQUE

Les connaissances acquises au cours des expériences antérieures facilitent aujourd'hui les prises de décision pour les experts à chaque étape du processus de préparation.

Là où les prises de décision nécessitent de prendre en compte un grand nombre de données, les systèmes experts et les techniques d'intelligence artificielles [39] sont une opportunité pour exploiter ces connaissances et en déduire des règles.

Les systèmes experts à base de connaissance [26] exploitent les connaissances et les raisonnements acquis par les experts. Ces techniques sont rapides et relativement faciles à mettre en œuvre lorsque les connaissances sont formalisées. Cependant, l'extraction de ces connaissances est souvent difficile.

Le raisonnement à partir de cas [45] (CBR : *Case Based Reasoning*) propose alors une alternative aux systèmes experts à base de connaissance.

Les algorithmes d'apprentissage artificiel de type réseaux bayesiens, réseaux de neurones, arbres de décision ou machine à vecteur de support permettent d'identifier des règles et d'expliquer des observations à partir de cas dont la connaissance n'est pas formalisée et dont les critères ne sont pas clairement identifiés.

Ces techniques permettent la recherche de similarités à partir de cas afin de prédire ou de comprendre un comportement.

Ces techniques d'intelligence artificielle peuvent être utilisées pour exploiter des cas de préparation de modèles pour la simulation en vue d'identifier les critères utiles au choix d'un modèle, de qualifier un processus de préparation et finalement d'identifier le meilleur processus pour un objectif de simulation. C'est tout l'objet des travaux présentés dans ce manuscrit.

Un état de l'art sur ces techniques d'intelligence artificielle et leurs applications dans le domaine de l'ingénierie mécanique est proposé dans le chapitre 4.

1.4 Problématique générale et questions soulevées

Synthèse sur le positionnement des travaux :

Les travaux se positionnent au niveau de la préparation de la maquette numérique définie lors des activités de conception vers ses représentations intermédiaires utilisées pour la simulation numérique.

Les méthodologies développées seront appliquées à la transformation de modèles géométriques CAO complexes en modèles de calcul par éléments et volumes finis.

Le modèle intermédiaire pour la simulation est un maillage 3D obtenu à partir de l'extraction de surfaces bordant le modèle géométrique.

Constats :

Les maquettes numériques sont aujourd'hui complexes et très détaillées ce qui rend leur exploitation directe inadaptée pour la simulation. Pour diminuer les durées de développement d'un produit il faut identifier dès le premier essai la représentation intermédiaire la mieux adaptée à l'objectif de la simulation. Cependant, celle-ci n'est pas connue a priori par les concepteurs et les analystes.

Le choix des outils et des opérateurs de préparation et de leurs paramètres de contrôle dépendent du format des maquettes numériques originales et intermédiaires et du type d'analyse mise en œuvre.

Il existe une multitude de processus de préparation d'un modèle intermédiaire pour un objectif de simulation.

Des cas de préparation de modèles intermédiaires existent souvent mais ils ne sont pas exploités dans le cadre de nouvelles simulations.

Face à ces constats, la question de recherche posée est (figure 1.10) :

Problématique générale

Comment identifier a priori le meilleur processus de préparation d'un modèle CAO original pour l'adaptation au besoin de la simulation numérique ?

Pour identifier le meilleur processus de préparation il sera nécessaire de l'évaluer. Pour cela il faut pouvoir caractériser le processus et chaque opération qui le compose. De nouvelles questions se posent alors :

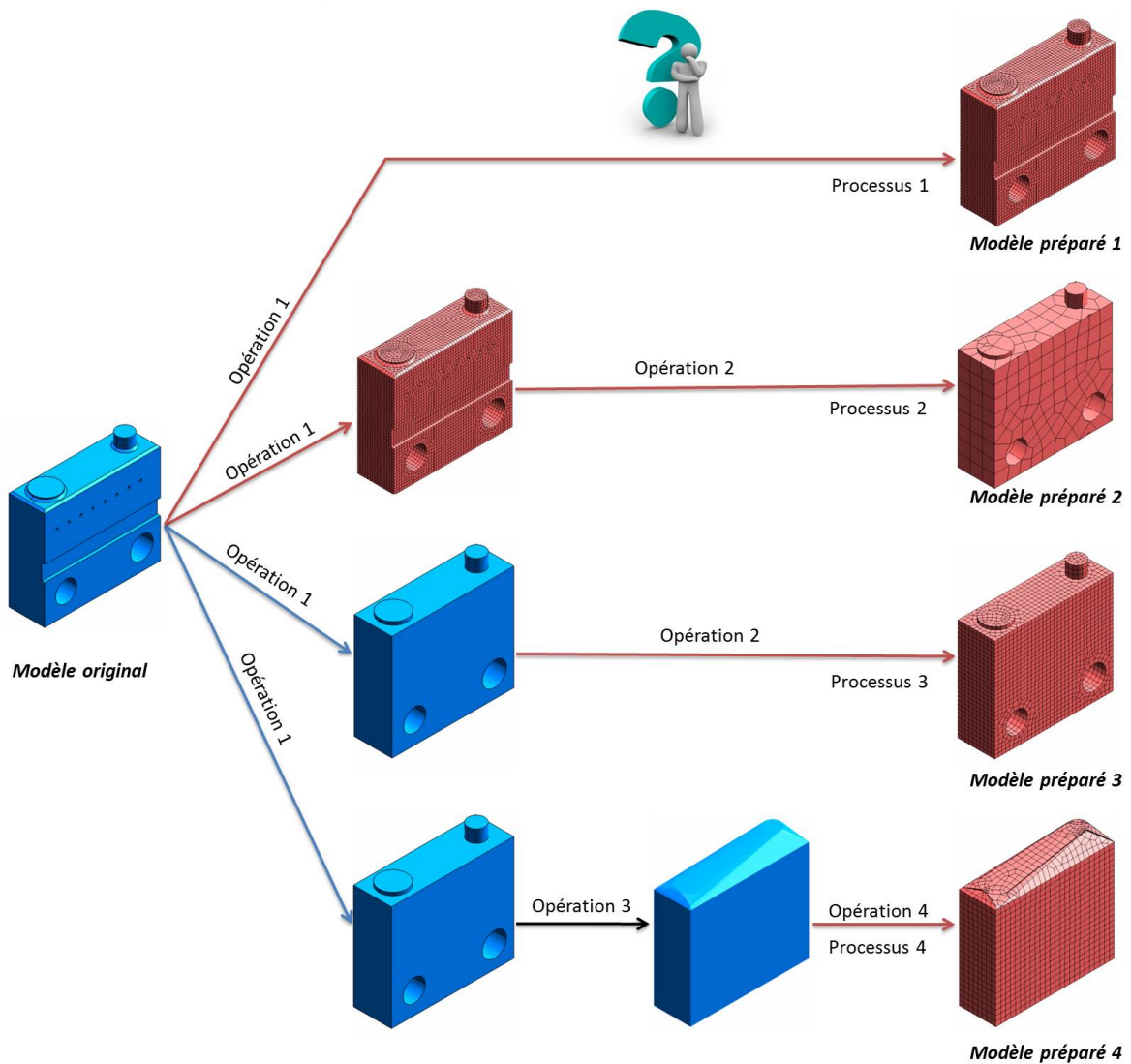


FIGURE 1.10 – Problématique générale : quel est le meilleur processus de préparation ?

Problématiques

- Comment modéliser un processus de préparation d'un modèle pour la simulation numérique ?
- Comment évaluer a priori un processus de préparation ? Avec quels critères ?
- Comment exploiter les cas existants et les connaissances acquises par les experts en préparation de modèles intermédiaires ?

L'identification a priori du processus le plus adapté à la préparation d'un modèle pour la simulation permettra de réduire le nombre d'itérations entre les activités de conception et de simulation. La durée du développement du produit sera ainsi diminuée. In fine, cela permettra de mieux optimiser le produit avec plus d'itérations.

La qualification du processus permettra de connaître a priori son coût et la qualité du résultat de l'analyse.

La prédiction des coûts de préparation du modèle et de la simulation facilitera également l'évaluation de la cotation pour lancer une sous-traitance CAD/CAE.

La prédiction des erreurs sur les résultats de l'analyse assurera une meilleure maîtrise de la précision des calculs tout en limitant la surqualité.

A terme, une formalisation des connaissances sur les processus de préparation des modèles pourra être proposée et permettra de capitaliser les connaissances métier.

Chapitre 2

Techniques de simplification de modèles CAO pour la simulation numérique

L'objectif des travaux présentés dans ce manuscrit est d'identifier a priori le meilleur processus de préparation de modèles intermédiaires pour la simulation numérique. Parmi les trois principales étapes de la préparation d'un modèle, l'étape de simplification est la plus longue, c'est celle qui soulève le plus d'interrogations et qui demande aux ingénieurs de nombreuses prises de décisions. Les méthodes de simplification existantes et les techniques pratiquées dans l'industrie sont synthétisées dans la première section de ce chapitre.

L'état de l'art sur les techniques de simplification est limité aux méthodes utilisables dans le cadre de la préparation de modèles 3D complexes pour le calcul par éléments finis.

Ce chapitre propose tout d'abord des critères de choix des techniques de simplification. Puis, les techniques de simplification des modèles CAO sont présentées selon la classification définie par Thakur et al. [55] . Finalement, les pratiques industrielles de simplification de modèles CAO pour l'analyse numérique par éléments ou volumes finis sont détaillées.

2.1 Sélection et classification des techniques de simplification

LA SIMULATION NUMÉRIQUE

2.1.1 Choix d'une technique de simplification pour l'analyse numérique

Afin de guider l'ingénieur dans le choix d'une technique de simplification, il convient de se poser les questions suivantes :

- Quel(s) niveau(x) de simplification doit-on atteindre ?
- Avec quelle(s) technique(s) de simplification ?
- Avec quel(s) outil(s) ? Quels sont les paramètres de contrôle des opérations ?
- Quelle sera la conséquence de cette simplification sur le résultat de l'analyse ?

Quel(s) niveau(x) de simplification ?

Le premier critère que les ingénieurs doivent prendre en compte avant de lancer une simulation est le niveau de précision attendu sur le résultat de l'analyse. Le niveau de simplification à atteindre (LOS : *Level Of Simplification*) sera faible lorsqu'une précision élevée est recherchée et au contraire plus élevé si l'on veut privilégier des durées de traitement réduites. De manière générale, on cherchera à maximiser le ratio précision/temps. Un LOS est obtenu par comparaison de la géométrie du modèle original avec celle du modèle simplifié. Les critères utilisés pour le définir sont donnés dans les sections suivantes pour chaque opération de simplification.

Avec quelle(s) technique(s) de simplification ?

Le choix d'une technique de simplification dépend essentiellement du type de modèle original (toutes les techniques ne sont pas adaptées à tous les types de modèles), du type de modèle préparé et de la durée acceptable des opérations de préparation (la durée totale de la préparation et de la simulation doit être inférieure à la durée de la simulation faite sur le modèle non simplifié).

Comme cela a été présenté dans la section 1.3.1, la préparation d'un modèle pour la simulation est limitée ici à trois étapes principales : simplification, adaptation et maillage. Cette thèse se limite au choix des opérations et paramètres de simplification.

Avec quel(s) outil(s) ? Quels sont les paramètres de contrôle de l'opération ?

Plusieurs outils existent pour réaliser une opération de simplification. La plupart des outils offrent des options qui permettent de régler le niveau de simplification attendu. Ces paramètres de réglage doivent être identifiés pour chaque type d'opération et chaque outil utilisé. Un choix judicieux des paramètres de contrôle minimisera l'erreur due à la simplification sur le résultat de l'analyse. Pour chaque technique de simplification, quelques

Quelle sera la conséquence de cette simplification sur le résultat de l'analyse ?

L'impact sur le résultat de l'analyse est mesuré à partir de l'erreur sur les grandeurs physiques à calculer. L'impact peut être négligeable (par exemple pour une erreur $< 3\%$) moyenne ou importante (par exemple pour une erreur $> 10\%$). L'impact de la simplification sur le résultat de l'analyse est étudié dans le chapitre 3.

2.1.2 Principales familles de méthodes de simplification

Thakur et al. [55] proposent une classification des techniques de simplification basée sur les surfaces, les volumes, les détails et la réduction de dimension.

Dans le cadre de cette thèse, le modèle préparé devant être un modèle 3D, les méthodes de réduction de dimensions ne peuvent être envisagées. Nous utiliserons ces méthodes uniquement pour extraire un axe ou un plan médian d'un modèle. Pour chaque catégorie de technique de simplification A. Thakur précise son domaine d'application (format d'entrée, types de détails simplifiés, objectif de la simplification), les critères de simplification, ses avantages et ses inconvénients. Cette classification nous a permis d'identifier les techniques utilisables dans notre domaine d'application.

Les techniques retenues qui sont adaptées à la préparation de modèles complexes pour l'analyse numérique par éléments ou volumes finis seront détaillées dans les sous-sections suivantes. Il s'agit de techniques :

- de suppression de détails (defeaturing) (sous-section 2.2) ;
- de filtrage de pièces ou de sous-ensembles, de substitution, ou de voxelisation (sous-section 2.3) ;
- de décimation, de construction d'enveloppes ou d'agrégation (sous-section 2.4).

Pour chacune de ces techniques, il sera précisé :

- le principe de la méthode utilisée ;
- le critère utilisé pour évaluer le niveau de simplification ;
- les outils existants ainsi que le type de données d'entrées qu'ils traitent, leurs intérêts et limites ;
- les paramètres de contrôle qui doivent être réglés par l'opérateur ainsi que les critères de choix de ces réglages.

2.1.3 ~~Pratiques industrielles~~ CHAPITRE 2: TECHNIQUES DE SIMPLIFICATION DE MODÈLES CAO POUR LA SIMULATION NUMÉRIQUE

Les contraintes industrielles relatives au choix d'un outil de simplification sont essentiellement basées sur des critères économiques. Ces critères sont les durées de préparation et de simulation ainsi que le coût d'investissement dans les logiciels et le coût de la formation des ingénieurs.

Des logiciels proposent des outils qui mettent en œuvre ces opérations de simplification, il s'agit par exemple de :

- modeleurs 3D (CATIA¹, Siemens NX5², SolidWorks³, ...) : dans ce cas la simplification est appliquée sur un modèle géométrique B-Rep ;
- outils dédiés à la simplification de modèles 3D (GPURE⁴) : la simplification est appliquée sur un modèle B-Rep polyédrique après conversion du modèle géométrique B-Rep polynomial ;
- mailleurs (AnsysMeshing⁵) ou d'outils de traitement de maillages (MeshLab⁶, outils de la suite OpenFoam⁷) : la simplification est appliquée sur le modèle facettisé ou maillé, la simplification sera réalisée après les opérations d'adaptation et de maillage ;
- outils de simulation (ANSYS⁸, STARCCM+⁹, COMSOL¹⁰, FloEFD¹¹, Abaqus¹²) : ces outils intègrent des fonctions qui facilitent l'adaptation du modèle géométrique pour une application spécifique et/ou qui simplifient le maillage avant simulation.

Les critères proposés par ces outils pour régler les paramètres de contrôle des opérations de simplification sont purement géométriques. Actuellement, les ingénieurs choisissent une opération de simplification, les outils pour la réaliser et les paramètres de réglages afin que l'erreur sur le résultat de l'analyse soit minimale en respectant les contraintes industrielles imposées.

Les extraits d'entretiens ci-dessous avec des ingénieurs chargés de la préparation de modèles pour la simulation montrent que les critères de choix des outils et de leurs paramètres sont subjectifs et qu'ils ne sont pas clairement formalisés :

-
1. CATIAV5, Dassault Systems, <http://www.3ds.com/fr/produits-et-services/catia/produits/catia-v5>
 2. NX5, Siemens PLM Software, <https://www.plm.automation.siemens.com>
 3. SolidWorks, Dassault Systems, <http://www.solidworks.fr>
 4. GPURE, DeltaCad, <http://gpure.net>
 5. AnsysMeshing, ANSYS, <http://www.ansys.com>
 6. MeshLab, <http://meshlab.sourceforge.net>
 7. OpenFoam Mesh Manipulation, OpenFoam foundation, <http://www.openfoam.org>
 8. ANSYS, <http://www.ansys.com>
 9. STARCCM+, CD-Adapco, <http://www.cd-adapco.com>
 10. COMSOL, <http://www.comsol.com>
 11. FLOEFD, Mentor Graphic, <http://www.mentor.com/products/mechanical/floefd>
 12. ABAQUS, Simulia, Dassault Systems, <http://www.3ds.com/products-services/simulia>

- « **Je supprime toutes les petites pièces** », *que veut dire petit ?*
- « **Je supprime toutes pièces qui font moins de 50 mm** », *qu'en est-il si l'ensemble global est de plus petite taille ?*
- « **Je conserve les pièces qui se trouvent à moins de 50 cm d'une source de chaleur** », *comment a été choisi ce « 50 cm » ?*
- « **Les composants positionnés en aval des entrées du fluide et en amont des sorties sont conservés** », *a t-on vérifié cette hypothèse ?*

2.2 Techniques basées sur la suppression de détails

Principes

Les techniques basées sur la suppression de détails (defeaturing) sont les plus utilisées. En comparaison avec les autres méthodes, les travaux de recherches actuels sur les techniques de defeaturing sont les plus riches.

Le defeaturing consiste à supprimer un ou plusieurs détails (trou, poche, bossage, congé, chanfrein) tout en conservant les arêtes extérieures de la pièce. Cette technique est à privilégier lorsque l'on dispose du modèle CAO natif. Le coût de l'opération peut être élevé lorsque l'historique des opérations de construction du modèle n'est pas disponible.

L'influence du defeaturing sur le résultat de l'analyse dépend de la géométrie et de la position du détail par rapport aux conditions limites. Les perturbations causées par le defeaturing peuvent être locales et limitées lorsque les conditions limites sont nulles à la frontière du détail. Les perturbations peuvent se répercuter à l'ensemble du modèle lorsque les conditions limites disparaissent avec le détail.

Des méthodes performantes de suppression de détails et de reconstruction ont été développées. Deux de ces méthodes sont présentées ci-après.

Gao et al. [23] (Figure 2.1) ont développé une méthode pour simplifier des détails sur des modèles maillés. Les détails sont tout d'abord reconnus puis les faces constituant chaque détail sont supprimées, les trous ainsi formés sont ensuite remplis.

Sun et al. [51] proposent une approche pour simplifier un modèle B-Rep (Figure 2.2). La région à simplifier est tout d'abord identifiée puis ses contours sont décomposés en frontières communes avec la pièce et en sous-frontières. Le modèle B-Rep est ensuite reconstruit autour des frontières communes à la région simplifiée et à la pièce.

Le critère utilisé pour évaluer le niveau de simplification est le nombre de détails supprimés. Ce critère est peu significatif. Il est difficile de définir une relation entre le nombre de détails supprimés et l'impact de la simplification sur le résultat de l'analyse.

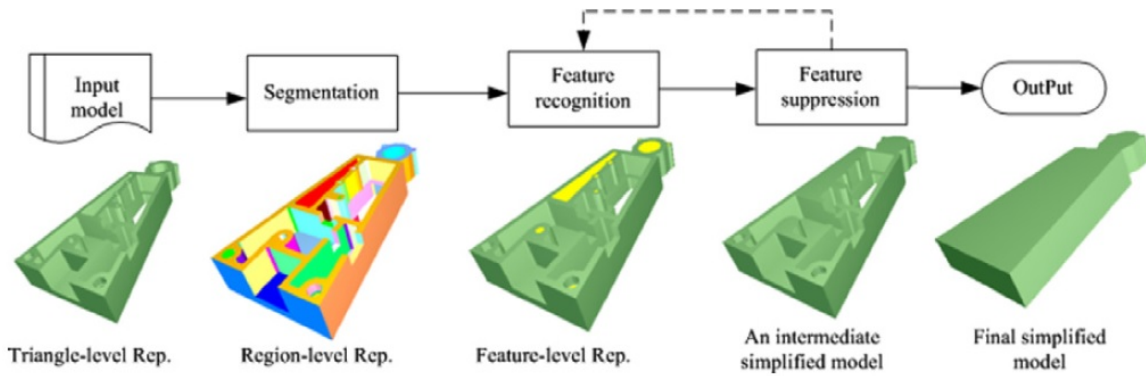


FIGURE 2.1 – Defeaturing. Méthodologie et illustration de l’approche proposée par GAO et al. [23]

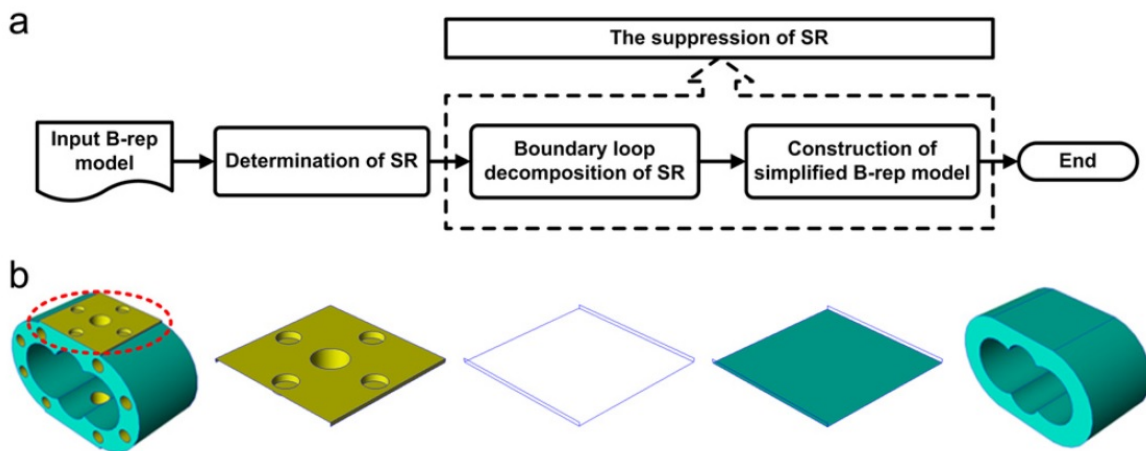


FIGURE 2.2 – (a) Méthodologie de defeaturing. (b) Illustration de l’approche proposée par Sun et al. [51]

Pratiques industrielles

Le paramètre de contrôle de l’opérateur de defeaturing est la liste des détails à supprimer. Afin de minimiser l’impact du defeaturing sur le résultat de l’analyse, les candidats au defeaturing sont sélectionnés en fonction de leur taille ainsi qu’en fonction de leur distance et position par rapport aux conditions limites.

Des exemples de defeaturing sont donnés sur la Figure 2.3.

Certains outils (Tableau 2.1) dédiés à la simplification tels que GPURE proposent des fonctions de defeaturing clé en main basées sur le principe de la suppression suivie de reconstruction. Ces opérateurs suppriment un type de détails avec une méthode spécifique (par exemple en réalisant un remplissage de trous, une découpe des bossages ou la suppression puis la reconstruction de surfaces pour les congés et les chanfreins). L’identification automatique des détails n’est pas toujours satisfaisante et seul le critère «taille» est pris en compte. Une sélection manuelle des détails à supprimer est souvent nécessaire.

Outils	Opérateur	Paramètres	Intérêts	Limites
SIMPLIFICATION NUMÉRIQUE	Suppression	Liste de détails à supprimer	Conservation du modèle CAO natif. Rapidité	Nécessite d'avoir à disposition l'historique des opérations Peut nécessiter une réparation du modèle
GPURE	Fill holes /or Bumps	Diamètre maxi des trous à remplir Longueur maxi des trous borgnes à remplir	Utilisable à partir de fichiers neutres Rapidité	Échecs courants lors du remplissage
GPURE	Remove chamfers and fillets	Dimension maximale des chanfreins/congés à supprimer	Rapidité	Peu de congés et chanfreins identifiés
GPURE	Convert fillets into chamfers	Dimension maximale des congés à simplifier	Rapidité	Peu de congés et chanfreins identifiés
NX5	Simplification détails	Dimensions des détails à supprimer ou sélection manuelle	Possibilité de sélectionner les détails à supprimer. Efficace sur CAO non natif Rapidité	

TABLEAU 2.1 – Outils et techniques de defeaturing

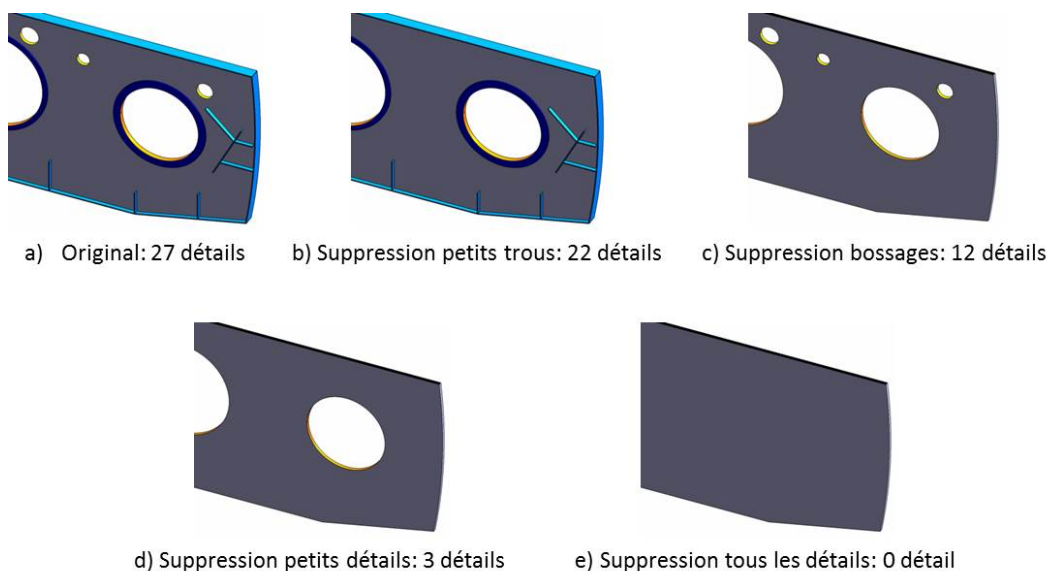


FIGURE 2.3 – Exemples de defeaturing. (a) modèle original d'une pièce. (b) modèle simplifié par suppression des petits trous. (c) modèle simplifié par suppression des bossages. (d) modèle simplifié par suppression des petits détails. (e) modèle simplifié par suppression de tous les détails.

2.3 Techniques de simplification basées sur les volumes

2.3.1 Filtrage de pièces

Principes

Le filtrage (Tableau 2.2) consiste à supprimer des pièces ou des sous-ensembles. Cette opération permet d'alléger la maquette numérique rapidement et sans avoir besoin de disposer d'outil spécifique et quel que soit le format des modèles CAO.

Le critère utilisé pour évaluer le niveau de simplification est le nombre de pièces supprimées. Ce critère ne tient pas compte de la taille des pièces, il est peu significatif en ce qui concerne l'impact de la simplification sur le résultat de l'analyse.

La suppression d'une pièce peut entraîner une perturbation locale lorsque la connexion est rompue entre deux pièces d'un sous-ensemble (Figure 2.4) ou bien une perturbation globale lorsque des conditions limites disparaissant avec la pièce.

Pratiques industrielles

Cette opération est généralement réalisée avant toutes les autres avec un modelleur 3D. Les critères de suppression sont la taille des pièces et leur distance par rapport aux conditions limites les plus proches. Les valeurs limites utilisées restent à l'appréciation de l'opérateur. Pour éviter les perturbations dues à la suppression de pièces dans un sous-ensemble, la modification de pièces voisines peut être nécessaire.

Outils	Opérateur	Paramètres	Intérêts	Limites
CATIA V5	Suppression	Liste de pièces à supprimer	Tous formats	Doit être réalisé en amont par le concepteur
GPURE	Suppression	Liste de pièces à supprimer	Sélection rapide des sous-ensembles à supprimer	Nécessite une conversion des données

TABLEAU 2.2 – Outils et techniques de filtrage

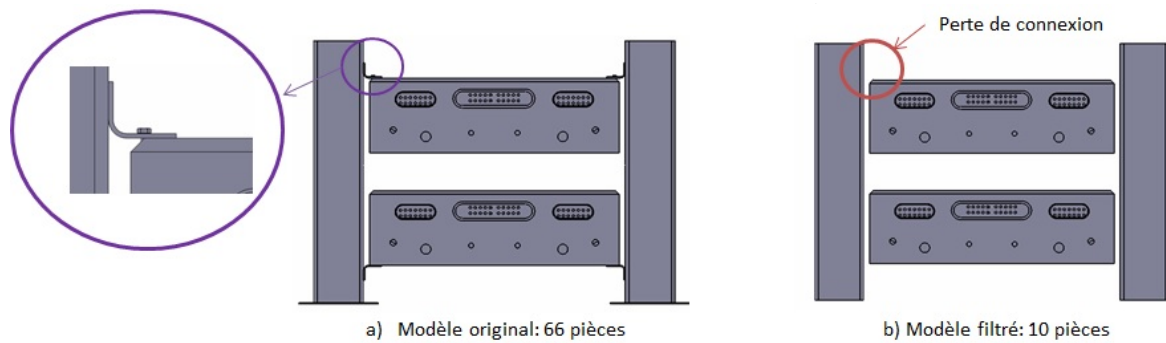


FIGURE 2.4 – Exemples de simplification par filtrage de pièces. (a) Modèle original d'un sous-ensemble. (b) Modèle d'un sous-ensemble filtré. Illustration d'une perte de connexion entre pièces.

2.3.2 Substitution

Principes

Cette opération consiste à supprimer des surfaces et/ou des volumes existants puis à les reconstruire avec une géométrie moins complexe. Les méthodes d'idéalisation par décomposition et réduction [15] sont utilisées pour extraire les axes ou les surfaces médianes du modèle. La géométrie du modèle simplifiée est reconstruite par création de volumes simples tels que des parallélépipèdes (Figure 2.5. b1) ou des cylindres (Figure 2.5. b2), ou bien par ajout d'épaisseurs autour de la surface médiane (Figure 2.5. b3).

Le critère utilisé pour évaluer le niveau de simplification est la distance de Hausdorff, définie dans la section 2.5, mesurée entre le modèle original et le modèle simplifié.

L'impact sur le résultat de l'analyse peut être très important. Dans le cas d'une substitution par parallélépipèdes ou par cylindres, le modèle est simplifié au maximum. Cette technique est plutôt réservée aux essais de faisabilité de la simulation ou pour simplifier un sous-ensemble de type obstacle éloigné des conditions limites.

Pratiques industrielles

Quelques outils (Tableau 2.3) proposent une fonction automatisée pour construire et remplacer un composant de forme élancée par un cylindre après idéalisation de l'axe principal. La reconnaissance de cet axe n'est pas parfaite. En pratique, les opérations de suppression et de reconstruction sont réalisées manuellement à l'aide d'un modèleur 3D.

Les paramètres à prendre en compte (Tableau 2.3) pour une reconstruction sont l'ensemble des dimensions du modèle reconstruit (par exemple qui permettent de conserver la plus grande longueur de la boîte englobante du modèle et son volume).

Outils	Opérateur	Paramètres	Intérêts	Limites
CATIA V5	Suppression puis extrusion	Dimensions du modèle à reconstruire	Conservation du modèle CAO natif	Simulation non automatisée
GPURE	Pipe generation	Aucun	Opération automatisée	Fréquents échecs lors de la détection de l'axe neutre

TABLEAU 2.3 – Outils et techniques de substitution

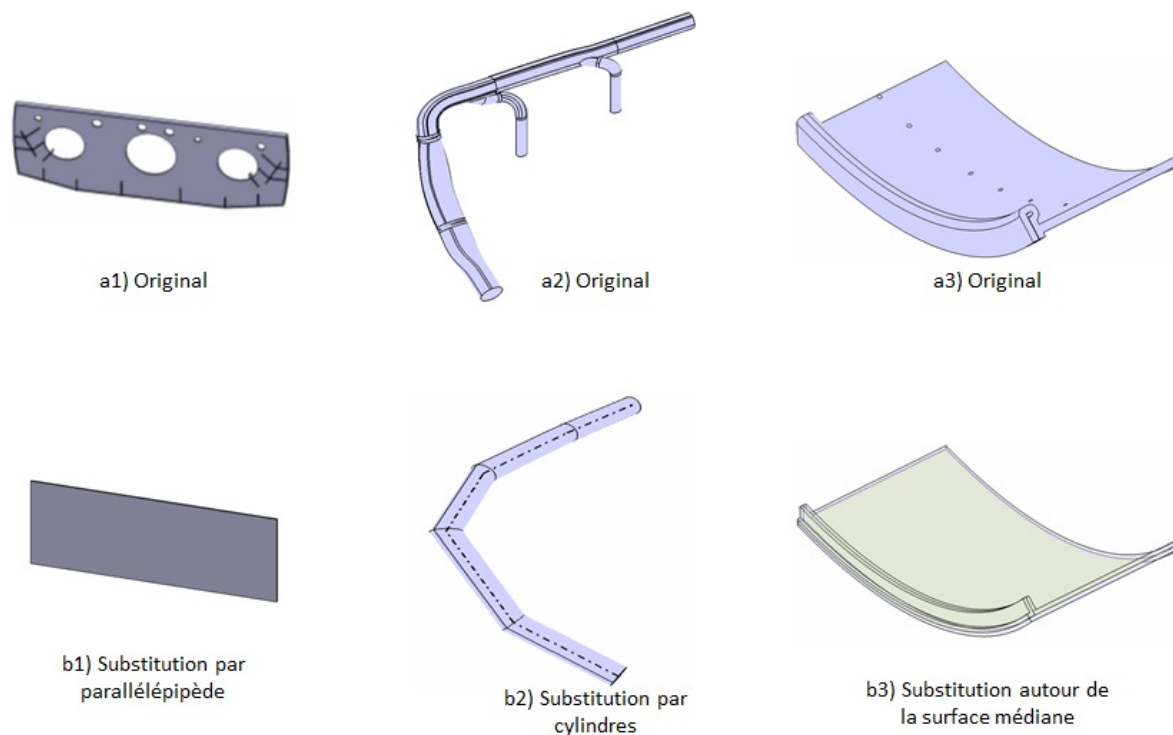


FIGURE 2.5 – Exemples de simplification par substitution. (a1, a2, a3) Modèles originaux. (b1) Modèle substitué par un parallélépipède. (b2) Modèle substitué par des cylindres. (b3) Modèle substitué par un volume autour de la surface médiane.

2.3.3 Voxelisation

Principes

L'opération de voxelisation consiste à construire une surface simplifiée à partir de la discrétisation du modèle CAO [3]. La géométrie du modèle (pièce ou sous-ensemble) est discrétisée en voxels noirs (dans le modèle), blancs (en dehors du modèle) ou gris (sur la frontière du modèle). Les voxels gris sont divisés en voxels plus petits jusqu'à ce qu'il n'y ait plus que des voxels blancs et noirs à une tolérance près. Une surface S est extraite telle que les voxels noirs se trouvent à l'intérieur de S et les voxels blancs à l'extérieur. Cette surface est corrigée par décimation des arêtes.

Là encore, le critère utilisé pour évaluer le niveau de simplification est la distance de

Pratiques industrielles

Les modelleurs facilitent cette opération. Les surfaces extérieures du modèle peuvent être conservées et les détails internes supprimés. Cela peut représenter un intérêt pour l'analyse numérique pour laquelle seules les surfaces extérieures sont maillées. L'impact sur le résultat de l'analyse est alors négligeable. Cependant le modèle voxelisé est « rempli » ce qui alourdi considérablement le modèle.

Le paramètre à régler (Tableau 2.4) est la valeur du grain (ou taille des voxels). Pour régler ce paramètre, les ingénieurs recherchent le meilleur compromis entre une simplification de niveau maximal et une bonne conservation des principales arêtes extérieures.

Outils	Opérateur	Paramètres	Intérêts	Limites
CATIA V5	Enveloppe	Valeur du grain ou taille des voxels	Rapidité	Modèle très lourd peu exploitable (difficile à utiliser)

TABLEAU 2.4 – Simplification par voxelisation

2.4 Techniques basées sur les surfaces

2.4.1 Décimation et face clustering

Principes

Les détails sont simplifiés par décimation ([21] et [59]) des arêtes, faces ou sommets du modèle ou par regroupement de faces (face clustering) ([48], [28]). La technique de décimation sur un modèle polyédrique proposée par Fine et al. (Figure 2.6) est appliquée sur un modèle préalablement transformé dans un format B-Rep polyédrique. Le face clustering proposé par Sheffer [48] puis Inoue [28] consiste à regrouper des faces du modèle B-Rep ayant des caractéristiques de courbure et de dimension proches.

Le critère utilisé pour évaluer le niveau de simplification entre le modèle original et le modèle préparé est la diminution en nombre de faces (ou de clusters) du modèle.

Pratiques industrielles

Des outils tels que GPURE ou QSLIM¹³ proposent des opérations de simplification en partie basée sur ces techniques. La plupart des outils travaillent sur des modèles B-Rep

13. QSLim Simplification Software, <http://www.cs.cmu.edu/~garland/quadrics/qslim.html>

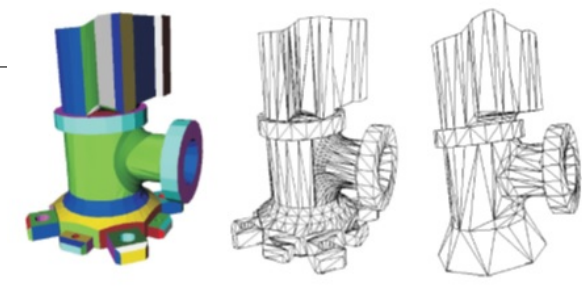


FIGURE 2.6 – Décimation de faces. [21].

polyédriques.

Ce type d'opération permet de simplifier rapidement une pièce ou un sous-ensemble. Lorsque le modèle comporte de nombreux détails de formes et de tailles différentes, le résultat peut s'accompagner d'erreurs (création d'entités non manifold ou création de faces de petites dimensions ou allongées). Dans ce cas, il sera nécessaire d'appliquer une opération de suppression de détails en amont. On peut regretter que les outils commerciaux ne permettent pas de sélectionner une région précise du modèle à simplifier.

Le paramètre à définir (Tableau 2.5) est la précision de la représentation équivalente à l'erreur d'approximation ou l'erreur de corde. La longueur maximale des arêtes créées et l'angle minimal des faces créées peuvent être éventuellement précisées. Certains outils proposent en option la conservation des arêtes extérieures. Pour régler ces paramètres les ingénieurs recherchent un compromis entre un niveau de simplification élevé (Figure 7c) et une bonne conservation des principales arêtes extérieures (Figure 7 b). Ces critères ne sont pas clairement formalisés. L'impact de ces techniques de simplification sur le résultat de l'analyse dépend du niveau de simplification du modèle. Actuellement, il n'existe pas de méthodologie pour savoir a priori si la version proposée sur la figure 7b sera plus ou moins intéressante que celle de la figure 7c du point de vue de l'impact sur le résultat de l'analyse et des coûts induits.

Outils	Opérateur	Paramètres	Intérêts	Limites
CATIA V5	Simplification	Précision	Rapidité	Faible niveau de simplification en cas de conservation des arêtes extérieures
GPURE	Simplify using default tolerance	Erreur de corde	Rapidité	Création d'éléments non-manifold et d'éléments de dimensions irrégulières
GPURE	Simply Tessellation	Aucun	Rapidité	

TABLEAU 2.5 – Simplification par décimation

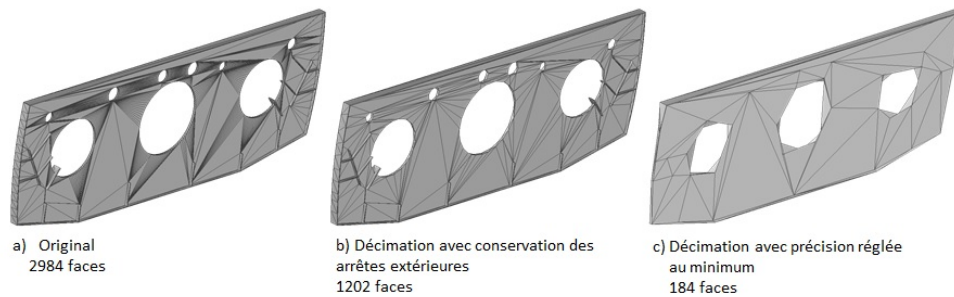


FIGURE 2.7 – Exemples de décimation. (a) modèle original. (b) Modèle décimé par décimation avec conservation des arêtes extérieures. (c) Modèle simplifié avec précision réglée au minimum.

2.4.2 Construction d'enveloppes

Principes

La création d'une enveloppe de type enveloppe convexe [6] ou silhouette permet de conserver l'enveloppe extérieure du modèle. Le modèle d'entrée peut être une pièce ou un assemblage quelconque (format CAO natif, format standard B-Rep, modèle facétisé, modèle pour la visualisation, maillage,...). Pour cela, la représentation initiale est tout d'abord discrétisée, les triangles de la géométrie qui n'appartiennent pas à l'enveloppe sont ensuite filtrés avant reconstruction. Le critère utilisé pour évaluer le niveau de simplification dépend de la méthode de simplification. Il peut s'agir de l'erreur de corde ou de la distance de Hausdorff entre le modèle original et le modèle simplifié.

Pratiques industrielles

La plupart des modeleurs et logiciels de simplification proposent des outils rapides de création d'enveloppes convexes ou de silhouettes. Dans le cas de créations d'enveloppes convexes, la forme extérieure du modèle est fortement modifiée, le niveau de simplification est très élevé (Figure 2.8), particulièrement lorsque le modèle n'est pas convexe comme sur l'exemple a2 et b2. L'impact sur le résultat de l'analyse est alors très important.

Le paramètre à définir (Tableau 2.6) est la précision qui caractérise le niveau de simplification (erreur de corde). Il peut être affiné selon le point de vue utilisé. La précision peut être suffisamment importante pour que les détails appartenant aux surfaces extérieures soient conservés avec très peu de modifications. Les critères sur lesquels s'appuient les ingénieurs pour régler les paramètres sont similaires à décimation (compromis entre une simplification maximale et une bonne conservation des principales arêtes extérieures).

Outils	Opérateur	Paramètres	Intérêts	Limites
CATIA V5	Silhouette	Précision Point de vue	Rapidité Conservation des arêtes extérieures	NUMÉRIQUE échecs lors du maillage des surfaces issues de l'opération
GPURE	Convex hull	Précision	Rapidité Fiabilité	Niveau de simplification très élevé
NX5	Enveloppe	Précision	Rapidité Fiabilité Conservation forme extérieure	

TABLEAU 2.6 – Outils et techniques d'enveloppe convexe

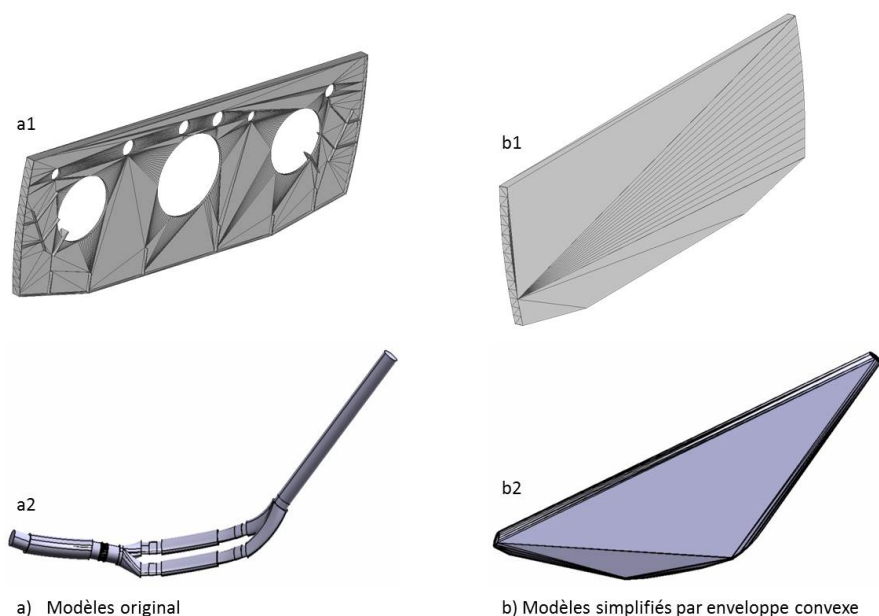


FIGURE 2.8 – Exemples de simplifications par enveloppes convexes. (a) Modèles originaux. (b) Modèles simplifiés par enveloppe convexe.

2.4.3 Agrégation

Principes

L'opération d'agrégation (Tableau 2.7) permet de fusionner plusieurs sous-ensembles en une seule pièce afin de faciliter la manipulation des modèles. Cette opération augmente les chances de faisabilité du maillage et de la simulation.

Le critère utilisé pour évaluer le niveau de simplification est le nombre de pièces fusionnées.

CHAPITRE 2. TECHNIQUES DE SIMPLIFICATION DE MODÈLES CAO POUR LA SIMULATION NUMÉRIQUE

Pratiques industrielles

Des logiciels réalisent cette opération automatiquement. L'utilisation de cette fonction seule impacte très peu le niveau de simplification global et par conséquent le résultat de l'analyse. Cette opération est généralement utilisée en complément d'opérations de décimation ou d'enveloppe. Le choix du séquençage des opérations est important, le résultat final sera très différent si l'agrégation est réalisée après (figure 2.9 b) ou avant (figure 2.9 c) les autres opérations de simplification.

Outils	Opérateur	Paramètres	Intérêts	Limites
CATIA V5	Fusion de parts	Liste des pièces à fusionner	Enregistrement en B-Rep surfacique	Echec en cas de discontinuité entre les pièces
CATIA V5	Fusion des arborescences	Liste de pièces à fusionner	Manipulation aisée des assemblages	Faible niveau de simplification
GPURE	Merge part	Liste de pièces à fusionner	Rapidité de mise en œuvre	Faible niveau de simplification

TABLEAU 2.7 – Outils et techniques d'agrégation

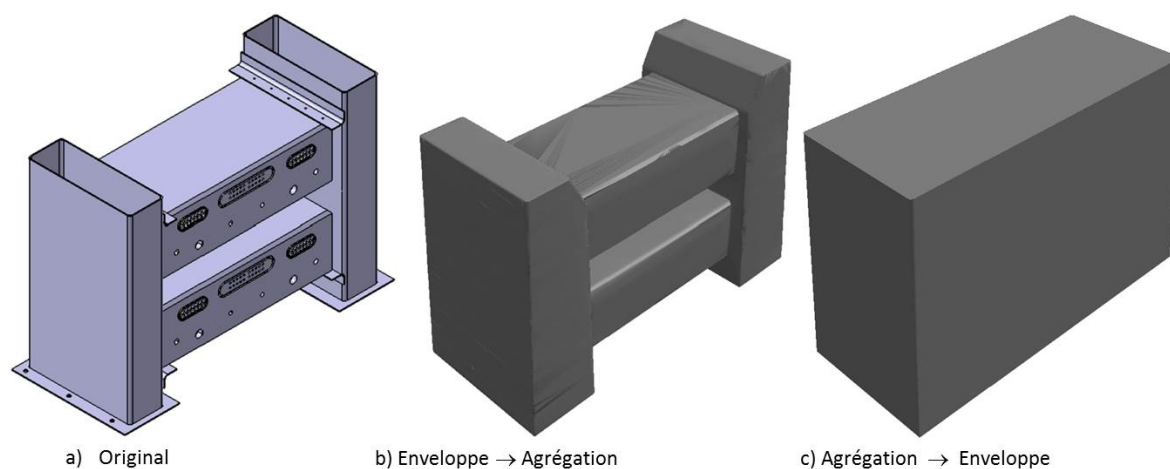


FIGURE 2.9 – Exemples d'agrégations. (a) Modèle original. (b) Modèle agrégé après simplification par enveloppe convexe. (c) Modèle agrégé avant simplification par enveloppe convexe.

2.5 Comparaison de modèles géométriques

Les paramètres de comparaison des géométries du modèle original et du modèle simplifié caractérisent le niveau de simplification d'un modèle qui a subi un traitement à partir d'un ou plusieurs opérateurs. Des critères de comparaison des géométries ont été identifiés lors de travaux sur la simplification de modèles tels que ceux qui ont été présentés précédemment et lors de travaux sur la reconnaissance de formes et de pièces. Iyer et al. [29] proposent un état de l'art détaillé de l'ensemble de ces techniques. Cette thèse se limite aux techniques qui permettent de comparer entre eux différents modèles issus d'un même original. Les techniques de reconnaissance d'objet 3D dans une bibliothèque de modèles ne seront donc pas traitées ici. De même, les paramètres ne serviront pas à définir précisément la forme de l'objet 3D mais simplement de pouvoir différencier les différents modèles.

Les techniques utilisées ici sont la mesure de similarité et l'analyse de caractéristiques globales des modèles.

2.5.1 Mesures de similarité

Soit X un ensemble de N points x_i (par exemple les sommets d'un modèle facétisé) appartenant à un modèle A et Y un ensemble de points y_i appartenant à un modèle B . Les techniques usuelles de mesure de similarité entre les modèles A et B sont :

- La **distance de Minkowski** L_p est définie par l'équation 2.1. La puissance p peut être égale à 1 (distance Manhattan), à 2 (distance euclidienne) ou prendre tout autre valeur réelle.

$$L_p(x, y) = \left[\sum_{i=1}^N |x_i - y_i|^p \right]^{1/p} \quad (2.1)$$

- La **distance de Haudorff** $d_H(X, Y)$ est calculée (équation 2.2) en déterminant pour chaque point x de l'ensemble X le point y le plus proche de l'ensemble Y et inversement. La distance de Hausdorff est le maximum des distances entre les points x et y .

$$d_H(X, Y) = \max \{ \sup_{x \in X} \inf_{y \in Y} d(x, y), \sup_{y \in Y} \inf_{x \in X} d(x, y) \} \quad (2.2)$$

- L'**indice de corrélation** $C(x, y)$ est calculé à partir de l'angle entre deux vecteurs de points définis par leurs moyennes $\bar{x}$ et $\bar{y}$ (équation 2.3).

$$C(x, y) = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}} \quad (2.3)$$

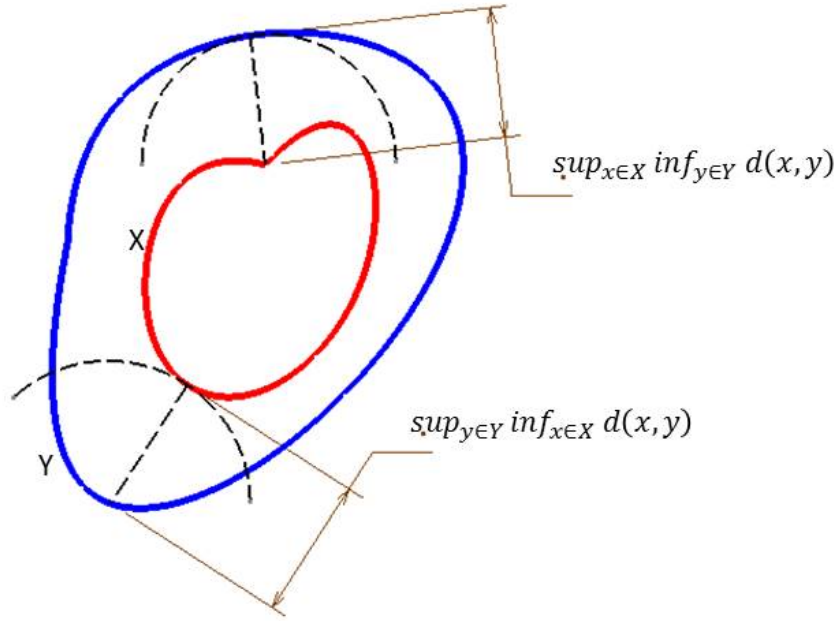


FIGURE 2.10 – Distance de Hausdorff. Distances entre en modèle X et un modèle Y.

La précision de ces paramètres dépend du nombre de points utilisés. Ces points peuvent être obtenus par discrétisation du modèle ou bien ils peuvent correspondre aux sommets du modèle facétisé. Des outils tels que MeshLab proposent des outils de mesures de ces paramètres.

2.5.2 Identification de caractéristiques globales

Il s'agit d'identifier des paramètres qui caractérisent globalement la géométrie d'un modèle et qui seront utilisés comme critères de comparaison des modèles, tels que :

- l'**aire** de la surface du modèle S ;
- le **volume** du modèle V ;
- le **ratio entre l'aire et la surface** $ratio_{S/R}$;
- la **compacité** $Compacite_1$ calculée à partir d'un rapport entre la surface S du modèle et son volume V (équation 2.4) ;

$$Compacite_1 = \frac{V^2}{S^3} \quad (2.4)$$

- la **compacité** $Compacite_2$ (ou *crinkliness*) calculée à partir du rapport entre la surface du sous-ensemble et la surface d'une sphère de même volume (equation 2.5) ;

$$Compacite_2 = \frac{S}{4\pi \left(\sqrt[3]{\frac{3V}{4\pi}} \right)^2} \quad (2.5)$$

- le **rapport entre le volume du modèle et le volume de sa boîte englobante** ;

- le **rapport** entre le volume du modèle et la section de sa boîte englobante (produit des plus petites longueurs de la boîte englobante) ;
- le **moment** d'un modèle 3D $\rho(x_1, x_2, x_3)$ de dimension $p + q + r$ calculé à partir d'intégrales de Riemann (figure 2.11).

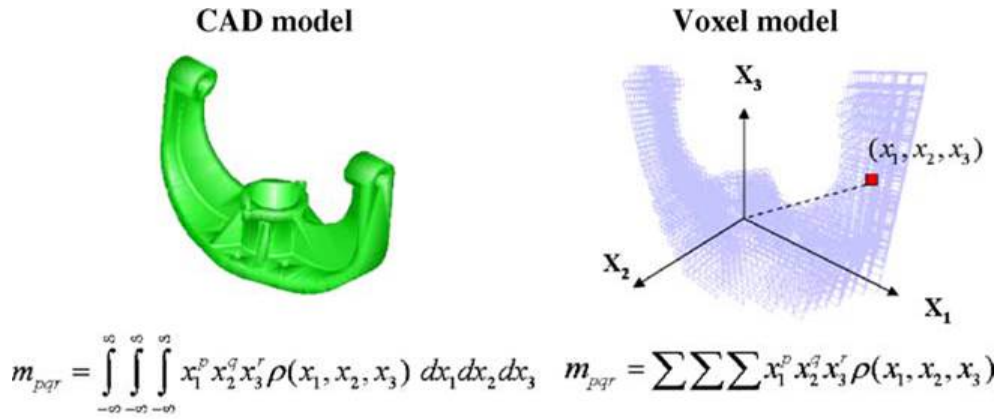


FIGURE 2.11 – Calcul de moments tridimensionnels sur des modèles continus et discrétisés proposés par Iyer et al. [29]

On peut ajouter à cette liste le calcul de courbures discrètes maximales K1 et minimales K2, les courbures de gauss K ou moyennes H. La notion de courbure a été utilisée pour évaluer l'erreur due à la simplification sur des polyèdres [59]. Kim et al. [33] combinent le calcul de courbures et de distances pour améliorer la précision des erreurs géométriques.

Tous ces paramètres peuvent être extraits assez facilement des modèles 3D. Cependant, ils sont assez peu discriminant lorsqu'ils sont utilisés seuls. Il faut donc les combiner pour disposer d'un ensemble de paramètres permettant de caractériser au mieux les modèles géométriques. De plus, ceux qui caractérisent au mieux le niveau de simplification d'un modèle ou l'influence sur le résultat de l'analyse ne sont pas connus.

2.6 Synthèse et verrous relatifs aux méthodes de simplification des modèles CAO

Contributions apportées à l'état de l'art :

- Des familles de techniques de simplification adaptées à la préparation de modèles dans le contexte de cette thèse ont été identifiées. Ces méthodes pourront être utilisées pour simplifier un modèle CAO en vue de l'analyse numérique par éléments ou volumes finis nécessitant un maillage 3D des surfaces extérieures du modèle.
- Pour chaque technique de simplification retenue des méthodes et outils ont été analysés, les paramètres de contrôle ont été identifiés.

Synthèse sur les constats :

Les techniques de simplification sont nombreuses et plusieurs outils sont disponibles pour chaque méthode. Si on combine les possibilités offertes par les outils utilisés, les techniques de simplifications disponibles et leurs paramètres de contrôle on décompte un grand nombre de processus de préparation possibles. Les critères de choix des opérations de simplification et de leurs paramètres ne sont pas formalisés.

Il existe plusieurs critères pour comparer la géométrie d'un modèle original et d'un modèle préparé et évaluer le niveau de simplification du modèle. Or, cette notion est mal définie, chaque technique utilise des critères différents (par exemple nombre d'éléments supprimés, distance de Hausdorff, erreur de corde, différence du nombre de faces,...). Certains critères ne sont pas significatifs en ce qui concerne l'impact de la simplification sur le résultat de l'analyse (nombre de détails ou nombre de pièces supprimés). De plus, plusieurs méthodes de simplification peuvent être utilisées sur un même composant (par exemple defeaturing + décimation + agrégation), ce qui rend impossible l'utilisation de ces critères. Il faudrait donc pouvoir identifier le ou les critères qui caractérisent au mieux le niveau de simplification global d'un modèle.

L'impact de la simplification sur le résultat de la simulation peut être évalué à l'aide de l'erreur sur le résultat de l'analyse. Cet impact dépend du type de simplification mis en œuvre et de ses paramètres de contrôle. Afin d'identifier le meilleur processus de préparation d'un modèle, il faudrait que l'erreur sur le résultat de l'analyse puisse être évaluée a priori.

Les verrous concernant les techniques de simplification sont :

- **Comment identifier tous les processus de préparation possibles ?**
- **Comment évaluer a priori l'impact de la simplification sur le résultat de l'analyse ?**
- **Comment définir un critère générique définissant le niveau de simplification global du modèle ? Comment comparer les modèles intermédiaires ?**

Chapitre 3

Techniques d'évaluation de l'impact de la préparation sur le résultat de l'analyse

La performance d'un processus de préparation peut être évaluée à partir de son coût et de l'impact de la préparation sur le résultat de l'analyse. Ce chapitre décrit les méthodes d'évaluation de l'impact de la préparation du modèle sur le résultat de l'analyse.

Les résultats d'une simulation numérique ne peuvent être strictement égaux aux grandeurs physiques réelles. Les écarts entre la grandeur physique réelle et les résultats de l'analyse peuvent être causés par l'imprécision du solveur, par des données initiales erronées ou manquantes, ou bien par la non qualité du modèle préparé pour la simulation. Ce chapitre traite de cette dernière cause d'erreur. Les écarts peuvent être générés à chaque étape de la préparation du modèle :

- Lors de l'étape de simplification, les écarts entre la grandeur physique réelle et les résultats de l'analyse sont dus à la qualité du modèle géométrique. La section 3.1 traite des écarts dus à une mauvaise définition du modèle géométrique simplifié.*
- Lors de l'étape de maillage, les écarts sont dus à la qualité des éléments et à la définition de la carte de tailles. Les écarts dus au maillage seront détaillés dans la section 3.2.*
- Lors de l'étape d'adaptation et de mise en données dans le solveur, les écarts sont dus par exemple à des données initiales du problème mal définies telles que les propriétés des matériaux, les conditions limites, les principes et les équations mises en œuvre. Ces travaux n'ont pas pour objectif de valider les hypothèses et les cas de calculs proposés par les experts. L'ensemble des données initiales du problème sera considérée dans la suite comme étant connu et ne sera pas remise en cause.*

L'objectif des travaux de thèse est d'identifier des méthodes d'évaluation a priori de la qualité du modèle préparé sans que celui-ci ne soit créé. Cela ne dispensera pas de faire la simulation qui restera nécessaire pour valider l'évaluation.

CHAPITRE 3. TECHNIQUES D'ÉVALUATION DE L'IMPACT DE LA PRÉPARATION SUR LE RÉSULTAT DE L'ANALYSE

3.1 Écart dus à une mauvaise définition du modèle géométrique simplifié

Les différences entre le modèle géométrique original et simplifié sont les premières causes d'écarts entre la réalité et les résultats de l'analyse. Des travaux de recherche antérieurs ont mis en avant trois méthodes pour évaluer les écarts s à la simplification du modèle.

La première méthode est la plus subjective, elle est basée sur la connaissance et l'expérience des analystes [34].

La seconde méthode est basée sur des critères géométriques ([4], [22]).

La troisième méthode est basée sur des critères physiques ([27],[53],[19],[25], [35]).

La plupart de ces techniques sont basées sur des principes généraux d'estimation a priori ou a posteriori des erreurs sur le résultat du calcul par éléments finis ([2], [5], [40]) puis ont été adapté aux erreurs dues aux modifications de la géométrie originale du modèle.

Un écart évalué a priori est calculé sur le modèle CAO original ou le modèle maillé non simplifié. Cela permet de connaître l'erreur sur le résultat de l'analyse due à une modification de la géométrie avant que celle-ci ne soit réalisée.

Calculer a posteriori sur le modèle simplifié, l'écart entre la grandeur réelle et le résultat de l'analyse permet d'estimer par la suite le résultat de l'analyse sur le modèle original.

Pour évaluer une technique d'estimation des écarts s à la définition de la géométrie il est nécessaire de connaître :

- le type de **problème** à résoudre, historiquement les techniques d'estimation d'écarts étaient limitées aux problèmes d'élasticité linéaire, plus récemment les problèmes non linéaires ou relatifs à la thermique ou aux fluides ont été traités ;
- la **technique d'évaluation des écarts** qui peut être très performante mais difficile à mettre en œuvre dans un contexte industriel lorsque le modèle à analyser est riche ;
- si l'**écart** entre la réalité et le calcul numérique est évalué ou non ;
- le **type de modèle** sur lequel est évaluée l'erreur (original pour une évaluation a priori , ou simplifié pour une évaluation a posteriori) ;
- les **techniques de simplifications applicables**, de nombreuses méthodes ne traitent que des écarts s à la suppression de détails avec parfois des restrictions (méthode limitée aux détails éloignés des frontières ou aux détails négatifs de types trous ou poches) ;

Les principaux travaux sur l'estimation de l'écart du à la définition du modèle géométrique et leurs critères d'évaluation sont donnés dans le tableau 3.1.

Référence	Principe	Problème	Limites
Armstrong 1994 [4]	Suppression de détails basée sur le principe de Barré de Saint Venant	Calcul de structure linéaire	Pas d'évaluation des écarts entre la réalité et la simulation.
Fine 2001 [21]	Décimation	Tous	Pas d'évaluation des écarts entre la réalité et la simulation.
Inoue 2001 [28] et Sheffer 2001 [48]	Face clustering	Tous	Pas d'évaluation des écarts entre la réalité et la simulation.
Gao 2010 [23]	Segmentation du modèle puis reconnaissance et suppression des détails	Tous	Pas d'évaluation des écarts entre la réalité et la simulation. Limitée aux simplifications locales.
Sun 2010 [51]	Identification des régions à simplifier puis décomposition et simplification	Tous	Pas d'évaluation des écarts entre la réalité et la simulation.
Foucault 2011 [22]	Critères de suppression des détails basés sur les caractéristiques géométriques du modèle	Tous	Pas d'évaluation des écarts entre la réalité et la simulation. Limitée aux simplifications locales.
Lee 2003 [34]	Minimisation des erreurs par programmation par but.	Tous	Critères déterminants nécessaires.
Gopalakrishnan 2007 [25]	Bornage de la grandeur à calculer à partir des équations de Poisson.	Tout problème gouverné par des équations différentielles partielles du second ordre	Limitée aux simplifications locales. Évaluation sur le modèle simplifié.
Ferrandes 2008 [19]	Évaluation de l'erreur à partir de l'énergie de déformation.	Calcul de structure et thermique	Limitée aux simplifications locales. Évaluation sur le modèle simplifié.
Huang 2013 [27]	Ré-analyse des matrices de rigidité sur les degrés de liberté influencés par les modifications	Calcul de structure	Limitée aux simplifications locales. Évaluation sur le modèle simplifié.
Li 2013 [35])	Intégration d'erreur locale en tant que résultat intermédiaire dans la résolution des équations du problème	Tout problème gouverné par des équations différentielles partielles du second ordre	Limitée aux simplifications locales. Évaluation sur le modèle simplifié.
Tang 2013 [53]	Estimation d'un indice basé sur le principe de réciprocité des actions mécaniques	Calcul de structure	Limitée aux simplifications locales. Évaluation sur le modèle simplifié.

TABLEAU 3.1 – Principales techniques d'évaluation des écarts entre les grandeurs physiques réelles et le résultat de l'analyse dus à des modifications sur la géométrie.

3.1.1 Méthodes basées sur des critères subjectifs

Majoritairement appliquées dans l'entreprise, les méthodes subjectives sont basées sur les connaissances acquises par les ingénieurs. Les analystes recherchent le modèle simplifié dont l'erreur sur le résultat de l'analyse sera la plus faible possible pour un coût minimal de préparation et d'analyse.

Les experts ont identifié des critères qu'ils jugent déterminants pour définir un modèle

CHAPITRE 3. TECHNIQUES D'ÉVALUATION DE L'IMPACT DE LA PRÉPARATION SUR LE RÉSULTAT DE L'ANALYSE
intermédiaire pour un objectif de simulation adapté à leurs buts et à leurs contraintes. Ces critères sont propres à chaque expert. Une même personne peut proposer plusieurs « bons » processus de simplification. Différents experts peuvent proposer des processus de simplification très différents les uns des autres, pourtant les erreurs induites par ces processus seront très semblables. Il est alors difficile d'identifier la meilleure pratique.

Les systèmes experts à base de connaissances pourraient aider les ingénieurs à prendre des décisions lors de la simplification des modèles en vue de minimiser les écarts entre la réalité et les résultats de l'analyse. Cependant, nous verrons dans la section 4.1 qu'il est difficile d'extraire ces connaissances.

Des méthodes rigoureuses, telles que la programmation par but, combinent des principes d'optimisation et de prise de décision.

La programmation par but [34] est une méthode de résolution de problèmes d'optimisation multi-objectifs basée sur les travaux de Edgeworth et Pareto. Un compromis est recherché entre plusieurs critères pour approcher les meilleures solutions. Les décideurs définissent et pondèrent des buts (ou vecteurs de référence) pour chaque objectif. Le problème multi-critère est alors transformé en problème uni-objectif. Les solutions proposées sont obtenues en minimisant la distance entre ces solutions et les vecteurs de référence. Lee et al. [34] ont mis en œuvre cette méthode pour minimiser des erreurs sur la masse, le déplacement ou les fréquences entre un modèle réel et un modèle simplifié.

Cette méthode est performante et intéressante car elle prend en compte des multi-critères. Cependant, si ces critères ne sont pas judicieusement choisis, nous risquons de ne pas trouver de solution Pareto-optimale.

3.1.2 Méthodes basées sur des critères géométriques

Partant du principe de Barré Saint-Venant, Armstrong [4] propose de simplifier un modèle en supprimant les détails qui n'ont pas d'influence sur le résultat de l'analyse par la méthode dite Medial Axis Transform. Pour cela (figure 3.1), un squelette appelé axe médian est créé (ensemble des centres des cercles inscrits dans les frontières du modèle). Les détails (congé, encoche et trous sur la figure 3.1) sont réduits à un point ou à un segment de droite, ils pourront être supprimés par la suite.

En utilisant cette technique, le choix des détails à supprimer ne dépend que de leur taille et de celle de la pièce, l'erreur de la simplification sur le résultat de l'analyse n'est pas quantifiée.

La plupart des techniques de simplification ([21], [23], [28], [48], [51]) présentées dans la section 2 ont été développées dans le but de faciliter le transfert CAD/CAE. Les critères proposés par ces techniques (nombre d'éléments supprimés, distance de Hausdorff, écart

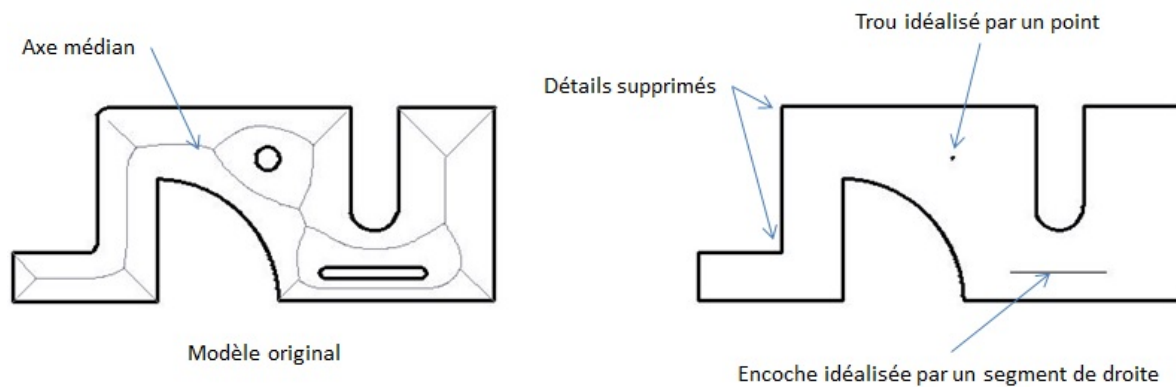


FIGURE 3.1 – Méthode "Medial Axis Transform" et suppression de détails, par Armstrong et al. [4]

de corde, différence du nombre de faces,...) sont purement géométriques. Ils permettent de définir un niveau de simplification en comparant la géométrie du modèle original et du modèle simplifié pour la simulation.

Foucault [22] propose de nouveaux critères pour évaluer les écarts entre la réalité et le résultat d'un calcul par éléments finis. Il s'agit des variations de masse, d'aire locale, de section, de position de centre de gravité et de moment d'inertie entre le modèle original et le modèle simplifié.

Toutes ces méthodes ne donnent pas d'indicateur réel sur les grandeurs physiques des résultats des calculs par éléments finis. De plus, il est possible que d'autres critères qui n'ont pas été énoncés ici puissent être déterminants pour évaluer l'erreur sur le résultat de l'analyse.

3.1.3 Méthodes basées sur des critères physiques

Trois principales méthodes sont basées sur des critères physiques : la ré-analyse, la détermination d'un indice calculé à partir de travaux ou d'énergie et le calcul d'un écart utilisant les équations différentielles partielles du problème.

3.1.3.1 Ré-analyse

L'une des premières méthodes utilisée pour évaluer les écarts entre la réalité et le calcul par élément finis sur des critères physiques s'appuie sur la méthode d'inversion de matrice de rigidité.

Huang et al. [27] ont développé une méthode rapide de ré-analyse. L'analyse est appliquée sur le modèle simplifié, puis, le résultat est estimé sur le modèle original après inversion de la matrice de rigidité. Il est ainsi possible de comparer rapidement deux modèles à condition que les maillages soient très proches. Cette méthode est très performante pour

~~CHAPITRE 3: TECHNIQUES D'ÉVALUATION DE L'IMPACT DE LA PRÉPARATION SUR~~
~~les faibles modifications de la géométrie du modèle. Or, lors du transfert du modèle CAO~~
~~vers le modèle intermédiaire sur des modèles complexes, la structure du modèle et les~~
~~matrices de rigidités sont fortement impactées ce qui rend cette méthode difficile à mettre~~
~~en œuvre.~~

3.1.3.2 Indice calculé à partir de travaux ou d'énergies.

Des travaux très récents sont riches en méthodes basées sur des critères physiques pour évaluer a priori ou a posteriori l'impact de la modification d'un modèle CAO sur le résultat de la simulation numérique. La plupart des études s'intéressent à des problèmes de préparation de modèles pour le calcul de structure par éléments finis.

Ces méthodes mettent en œuvre des principes propres au calcul de structure qui pourront difficilement être transférés à d'autres problèmes tels que le transfert thermique et la mécanique des fluides.

Ferrandes et al. [19] proposent un indicateur a posteriori d'influence de la suppression d'un détail sur le résultat d'un calcul d'élasticité par éléments finis. Cet indicateur d'influence relatif est estimé à partir de l'énergie de déformation sur des domaines dont des détails sont supprimés du modèle original selon des critères a priori. Si la valeur de l'indicateur d'influence est inférieure à une valeur limite alors le détail sera supprimé. Dans le cas contraire le détail sera réintégré dans le domaine. Lorsque tous les détails ont été analysés, on obtient un modèle qui contient uniquement les détails qui ont un impact non négligeable sur le résultat de l'analyse.

L'intérêt majeur de cette méthode porte sur le fait que l'indicateur d'influence ne donne pas directement pour chaque détail l'erreur due à sa suppression, mais apporte une aide à la décision en précisant si le détail doit être supprimé ou conservé. Finalement, un modèle idéal devrait être identifié. Le fait que chaque détail est testé relativement à un modèle simplifié au maximum facilite les analyses. Cependant, l'influence relative entre les différents détails n'est pas prise en compte. De plus, cette méthode est limitée à la simplification par suppression de détails locaux. Les frontières des détails doivent être clairement identifiées. L'indicateur d'influence est plus difficile à estimer dans le cas où la frontière du détail n'est pas clairement délimitée ou dans le cas de simplification globale du produit.

3.1.3.3 Écart calculé à partir d'équations aux dérivées partielles.

D'autres études proposent des méthodes basées sur des critères physiques dont la démarche globale peut être retenue et transposée à un problème autre que le calcul de structure.

CHAPITRE 3. TECHNIQUES D'ÉVALUATION DE L'IMPACT DE LA PRÉPARATION SUR LE RÉSULTAT DE L'ANALYSE

L'estimation de l'impact de la suppression de détail sur un modèle destiné au calcul de transfert thermique a été traitée par Gopalakrishnan et al. [25]. La grandeur physique à calculer est la température d'un système de type semi-conducteur Ω_{device} soumis à une source de chaleur Ω_{coil} placée à l'intérieur d'un ensemble dont un détail Ω_{slot} aura été supprimé (figure 3.2).

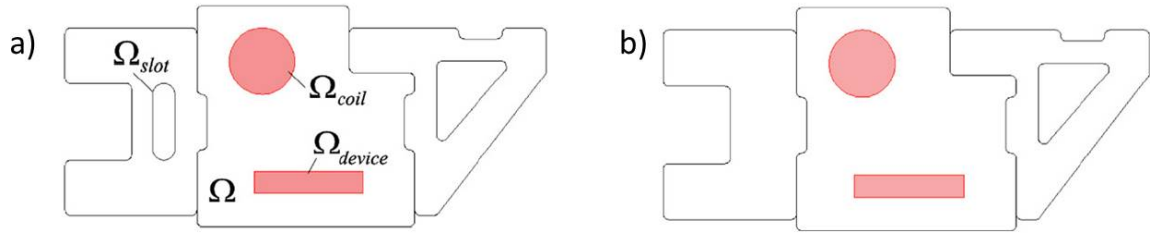


FIGURE 3.2 – (a) A 2-D heat block assembly. (b) A defeatured 2-D block assembly. Par Gopalakrishnan et al. [25]

La température à déterminer T_{device} est bornée par des valeurs inférieures T_{device}^{lower} et supérieures T_{device}^{upper} estimées à partir d'équations aux dérivées partielles du second ordre de Poisson pour deux cas de conditions limites. La condition limite de Neumann impose un flux de chaleur fixe à l'intérieur du domaine. La condition de Dirichlet impose une température fixe sur le bord du domaine supprimé. Cela permet de traiter à la fois les détails situés à l'intérieur et à la frontière du modèle. Cependant, cette méthode se limite aux modèles dont les frontières des détails sont parfaitement définies et ne peuvent facilement s'appliquer aux modèles simplifiés globalement par exemple par décimation ou par enveloppe convexe. Ce constat peut être généralisé à toutes les méthodes basées sur la résolution d'équations aux dérivées partielles.

3.1.3.4 Conclusion sur les méthodes basées sur des critères physiques.

Les études antérieures relatives au transfert CAO/CAE se focalisent essentiellement sur l'opération de simplification du modèle CAO.

De telles méthodes sont performantes pour évaluer la suppression de détails, mais elles nécessitent de définir précisément le domaine du détail à supprimer et le domaine global du modèle. Ces méthodes ne peuvent donc pas être appliquées à des problèmes de simplification globale (décimation globale ou création d'enveloppes par exemple).

De plus, la plupart de ces méthodes ne sont appliquées sur la géométrie du modèle original ce qui rend impossible l'évaluation a priori de l'impact de la préparation sur le résultat de l'analyse .

Lorsque la géométrie n'a pas été simplifiée et que le modèle numérique est très proche du modèle réel, il est malgré tout possible d'obtenir un écart à cause d'un maillage de mauvaise qualité. Un maillage peut être de mauvaise qualité si ses éléments sont dégénérés ou irréguliers ou bien si sa taille n'est pas adaptée aux conditions limites et à la géométrie de la pièce. Afin d'évaluer la qualité du maillage, plusieurs critères peuvent être utilisés.

3.2.1 Critères de forme.

Dompierre [17] propose des critères de formes qui permettent d'identifier les éléments dégénérés ou irréguliers. Un élément est dégénéré si son aire (pour un élément 2D) ou son volume (pour un élément 3D) est proche de zéro. Pour qu'un élément de type triangle ou tétraèdre soit régulier, ses cotés doivent tous être de la même longueur. Il existe un grand nombre de critères de formes pour évaluer la qualité d'un élément K , les trois principaux critères sont :

- le rapport des rayons des cercles inscrit et circonscrit (figure 3.3) ;
- le rapport de la moyenne géométrique sur la moyenne algébrique des valeurs propres de la matrice N de transformation affine d'un élément régulier R vers l'élément K (figure 3.4) ;
- le minimum des angles solides θ_{min} (figure 3.5) de l'élément (aire du secteur sphérique obtenue en projetant la face opposée à P_i sur la sphère unitaire centrée en P_i).

La qualité de l'élément est évaluée en comparant ces critères par rapport une valeur seuil.

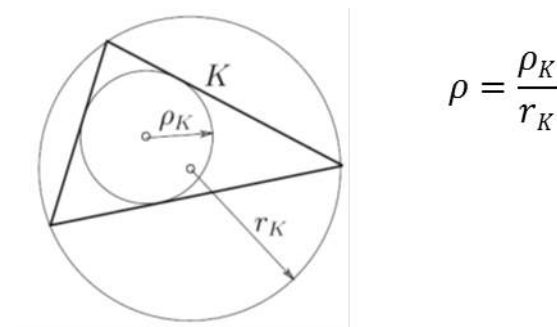


FIGURE 3.3 – Rapport des rayons inscrit et circonscrit. Par Dompierre [17].

3.2.2 Critères de taille.

Des éléments peuvent avoir des formes de bonne qualité sans pour autant être de la bonne taille.

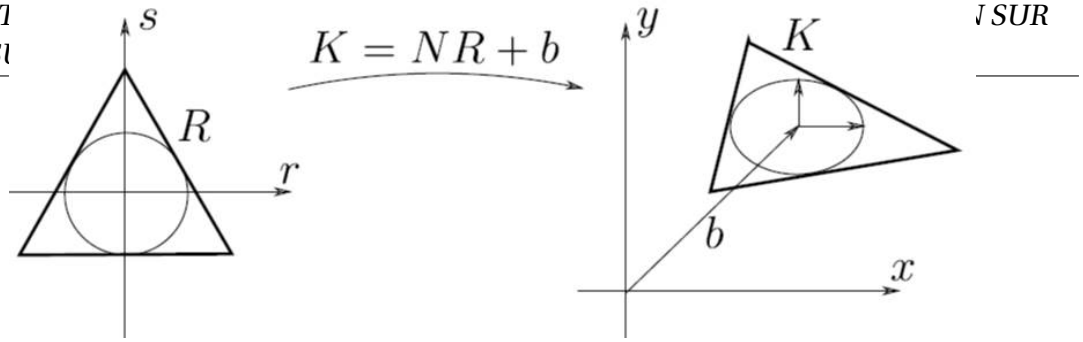
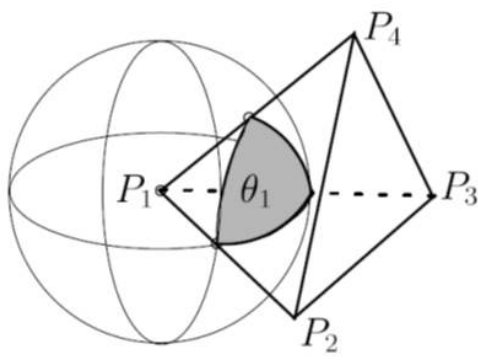


FIGURE 3.4 – Transformation d'un élément régulier en élément réel. Par Dompierre [17].



$$\theta_{min} = \alpha^{-1} \min_{1 \leq i \leq d+1} \theta_i$$

avec $\alpha = \frac{\pi}{3}$ pour un élément 2D
et $\alpha = 6 \arcsin(\frac{\sqrt{3}}{3})$ pour un élément 3D

FIGURE 3.5 – Définition de l'angle solide θ_i . Par Dompierre [17].

Pour définir la taille des éléments, l'analyste impose une carte de maillage définie à partir des conditions limites et de la géométrie du produit.

G. Foucault [22] propose deux indices pour évaluer la qualité de la taille d'un maillage (figure 3.6). La première contrainte J appelée seuil de sur-intensité est un ratio entre la taille souhaitée H et la taille h de l'élément. La valeur de J doit être la plus proche possible de 1. Le second indice est la précision géométrique ϵ ou erreur de discrétisation relative (rapport entre l'erreur de discrétisation absolue δ et la carte de tailles H). ϵ doit être inférieure à un seuil maximal.

Ces indices permettent de s'assurer en cours de maillage que sa qualité est conservée. Cependant, la relation entre ces critères et l'erreur sur le résultat de l'analyse n'est pas établie. Les valeurs choisies pour les seuils J_{max} et ϵ_{max} restent à déterminer.

Tassin et Pernot [54] ont proposé une méthode utilisant des classifieurs de type arbre de décision et réseaux de neurones pour prédire a priori la qualité d'un maillage et proposer des paramètres de maillage optimum à partir d'un modèle géométrique avant même de mailler la pièce. Cette technique s'applique sur un modèle géométrique simplifié et adapté au maillage (l'ensemble des surfaces et/ou volumes à mailler a été créé).

Indice de qualité	Contrainte	Satisfaction de la contrainte
Taille	$J_{\max}$	La taille des ÉF générés doit être supérieure à la taille de maille minimale $h > H_{\min}$, où $H_{\min}$ est définie par le seuil maximal du coefficient de sur-densité : $H_{\min} = \frac{H}{J_{\max}}$
Précision géométrique	$\varepsilon_{\max}$	l'erreur de discrétisation relative $\varepsilon = \delta/H$ doit être inférieure au seuil maximal $\varepsilon_{\max}$: $\varepsilon < \varepsilon_{\max}$

FIGURE 3.6 – Critères de qualité d'un maillage éléments finis. Par Gilles Foucault [22]

3.3 Synthèse et verrous relatifs aux techniques d'évaluation de l'impact de la préparation sur le résultat de simulation.

Constats :

L'impact de la préparation sur le résultat de l'analyse peut être mesuré par les écarts entre la grandeur physique réelle et le résultat de la simulation. Ces écarts peuvent être dus à la qualité du maillage ou à la définition du modèle.

Les erreurs dues au maillage peuvent être identifiées et qualifiées. Les techniques d'intelligence artificielle peuvent être utilisées pour estimer la qualité d'un maillage sans avoir à le réaliser. Cependant, cela nécessite de disposer du modèle géométrique simplifié et adapté.

Les écarts dus à la définition du modèle sont évalués à partir de critères différents. Les méthodes subjectives nécessitent de connaître et de formaliser de manière exhaustive les critères qui influencent l'écart entre la réalité et la simulation. Les connaissances acquises par les experts sont difficiles à extraire et les critères ne sont pas formalisés. Par contre, la construction de cas de simplification pour lesquels l'écart est connu pourrait être envisagée.

Les méthodes basées sur des critères géométriques ne proposent pas d'indicateur précis sur l'erreur d'analyse. L'exhaustivité des critères géométriques n'est pas assurée. La simplicité de mise en œuvre de ces méthodes est néanmoins intéressante. Elles peuvent être suffisantes pour évaluer la qualité d'une simplification sans chercher à évaluer précisément le résultat de l'analyse.

CHAPITRE 3. TECHNIQUES D'ÉVALUATION DE L'IMPACT DE LA PRÉPARATION SUR LE RÉSULTAT DE L'ANALYSE

Pour de nombreuses méthodes d'évaluation a priori (avant la réalisation de la simplification et du maillage) les erreurs sur les résultats de la simulation. Ces méthodes basées sur des critères physiques sont souvent limitées aux techniques de simplification de type defeaturing et sont difficilement transposables dans un contexte industriel où le modèle est complexe et lorsque la simplification est globale. De plus, la plupart ne peuvent s'appliquer qu'à des objectifs de simulation précis tel que le calcul de structure.

Positionnement des travaux :

Les travaux de cette thèse se positionnent sur l'évaluation a priori d'un processus de préparation des modèles CAO pour l'analyse numérique en se basant sur des critères géométriques et sur la qualité du maillage en utilisant des cas connus de préparation de modèles.

Verrous :

A la vue de ces constats et de notre positionnement, des questions demeurent quant à l'évaluation des écarts entre la réalité et la simulation numérique dus à la simplification du modèle.

Verrous relatifs à l'évaluation des écarts entre la réalité et les résultats de l'analyse

- **Comment évaluer l'écart entre la réalité et la simulation pour un modèle complexe simplifié globalement ?**
- **Quels sont les critères géométriques qui influencent le résultat de l'analyse numérique et donc les écarts entre une grandeur réelle et la simulation ?**
- **Comment évaluer a priori cet écart à partir de cas connus ?**

Chapitre 4

Techniques d'intelligence artificielle

L'idée principale de cette thèse consiste à évaluer a priori un processus de préparation de modèle pour la simulation en exploitant des cas existants et les connaissances acquises par les experts sur des cas de simulation. Pour cela, nous proposons d'utiliser les techniques d'intelligence artificielle. Ce chapitre a pour objectif d'identifier les outils qui pourront être utilisés et les types de données manipulées.

John McCarthy de l'université de Stanford a proposé l'appellation "intelligence artificielle" (IA) dans les années 1950 en partant du postulat que toute activité intelligente peut être modélisée et imitable par une machine. L'IA peut être définie comme étant la réalisation de programmes visant à imiter les comportements humains. Les algorithmes construits sont capables de reproduire une tâche à partir de l'analyse d'exemples connus.

Ce chapitre présente les techniques qui permettront d'évaluer et d'identifier un processus de préparation de modèles pour la simulation à partir de cas existants.

Les systèmes experts à base de connaissance et à base de cas semblent pouvoir répondre à cette problématique. Mais la section 4.1 en montrera les limites vis à vis de l'objectif des travaux. La section 4.2 présente les principes de l'apprentissage artificiel, branche de l'IA à laquelle nous nous intéresserons plus particulièrement. La section 4.3 propose une synthèse des applications des techniques d'IA dans le domaine de l'ingénierie mécanique. Puis, chaque technique retenue est décrite dans la section 4.4 en précisant ses principes, paramètres de configuration, intérêts et limites. Une méthodologie usuelle de conception d'un modèle d'apprentissage est enfin décrite dans la section 4.5.

4.1.1 Les systèmes à base de connaissances

Edward Feigenbaum de l'université de Stanford, reconnu comme étant le père des systèmes experts et de l'intelligence artificielle, a donné la définition suivante : *“An expert system is an intelligent computer program that uses knowledge and inference procedures to solve problems that are difficult enough to require significant human expertise for their solution. Knowledge necessary to perform at such level, plus the inference procedures used, can be thought of as a model of the expertise of the best practitioners in the field”*[7].

Un système expert est composé :

- d'une base de règles, souvent de type « si conditions, alors conclusions » élaborée à partir de l'expertise des spécialistes ;
- d'un moteur d'inférences et d'interfaces pour gérer et communiquer autour de ces connaissances.

L'intérêt des systèmes experts est leur rapidité d'exploitation grâce à la mise à disposition de la connaissance formalisée.

La principale limite des systèmes experts concerne l'extraction et la représentation des connaissances. Les connaissances doivent être formalisées pour être traduites en règles. Elles peuvent être contradictoires entre deux experts du domaine. De plus, les experts peuvent être réticents à l'idée de transmettre leur compétence.

Concernant notre étude, les connaissances et les raisonnements qui amènent l'analyste à proposer un modèle intermédiaire pour la simulation ne sont pas formalisés. Il sera donc difficile de retenir cette approche.

4.1.2 Les systèmes à base de cas

Le raisonnement à partir de cas [45] ou case based reasoning (CBR) propose une alternative aux systèmes experts lorsque les connaissances ne peuvent pas être extraites. Le CBR utilise des exemples de cas sources représentatifs d'un problème à résoudre. Un cas source est composé d'un problème et de sa solution. Il s'agirait pour notre étude d'un cas de simulation pour un modèle CAO complexe (problème) et de son modèle CAO préparé (solution).

Le principe d'un CBR est résumé sur la figure 4.1. L'ensemble des cas sources sont indexés en fonction de critères de similarité. Lorsqu'un nouveau problème se présente les cas similaires sont identifiés. L'analyse de leur similarité permet de proposer une solution au nouveau problème. Cette solution est évaluée et éventuellement révisée. La base de données est finalement mise à jour avec le nouveau cas.

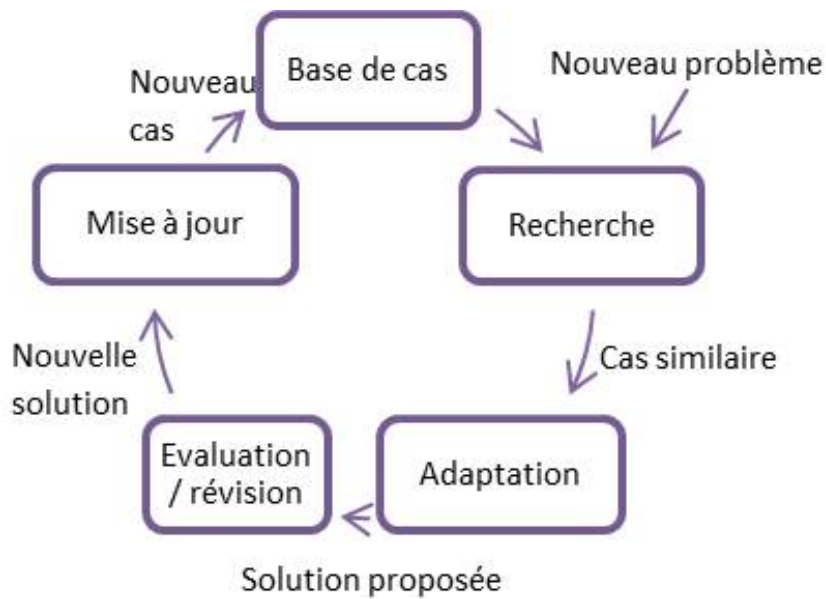


FIGURE 4.1 – Principe d'un système à base de cas (CBR)

L'intérêt des CBR résulte dans le fait qu'ils ne nécessitent pas d'extraction de connaissance ou de raisonnement et que la base est évolutive.

La qualité des résultats dépend de la qualité et de la quantité des cas présents dans la base.

Concernant nos travaux, les cas de préparation de modèles pour la simulation disponibles ne sont pas forcément représentatifs de l'ensemble des cas qui pourraient exister, les critères de recherche de similarité ne sont pas connus, des incertitudes subsistent au niveau de l'évaluation de la qualité des modèles.

4.2 Principes de l'apprentissage artificiel

Pour Cornuéjols et al. [16] l'apprentissage artificiel (AA) *"consiste à construire une représentation du monde en adéquation avec les observations disponibles et permettant de prendre des décisions."*

L'apprentissage artificiel peut avoir comme objectif la prédiction ou la compréhension d'un comportement (dans une moindre mesure).

Les techniques d'AA sont particulièrement intéressantes lorsqu'un grand nombre de données (fouilles de données) sont à analyser ou bien lorsque les algorithmes ne peuvent être créés manuellement.

Autre technique, la fouille de données (*Data Mining*) [39] trouve des applications comme par exemple la recherche d'informations sur internet, la recherche d'informations bancaires ou à des fins commerciales, l'aide au diagnostic médical, la reconnaissance de formes ou d'écriture, ainsi que la recherche en biologie ou en ingénierie.

Concernant l'aide à la création d'algorithmes, on retrouve des applications dans de nombreux domaines tels que par exemple : le traitement du langage, la commande de robots ou d'engins de transport ainsi que la vision par ordinateur.

L'apprentissage artificiel peut être supervisé lorsque l'on cherche à prédire la valeur d'une variable de sortie, ou non supervisé. Dans ce dernier cas, les exemples sont regroupés selon leur similarité formant des groupes homogènes. La prédiction consiste alors à identifier le groupe auquel appartient le nouveau cas.

4.2.1 Apprentissage supervisé

Dans le cas de l'apprentissage supervisé, on dispose de N instances i constituées d'un même nombre de variables explicatives en entrée (généralement modélisées par un vecteur $x^{(i)}$) associées à une ou plusieurs variables de sortie $y^{(i)}$. L'apprentissage supervisé recherche une hypothèse h (fonction ou distribution de probabilité) entre les variables explicatives et de sortie dans le but de prédire pour une nouvelle instance $x^{(k)}$ la valeur de sortie $y^{(k)}$ (figure 4.2). Cette hypothèse h permet de généraliser une fonction à partir des données d'apprentissage.

Les objectifs de l'apprentissage supervisé peuvent être la classification ou la régression.

Dans le cas de la **classification**, les variables de sorties se répartissent dans un nombre fini de classes. Les principaux algorithmes de classification sont :

- les **arbres de décision** (figure 4.3 (a)) qui représentent le problème par un arbre modélisé par la hiérarchisation de tests sur les variables explicatives et dont les extrémités (feuilles) donnent les valeurs de la variable de sortie y_1 , cette technique est présentée dans la section 4.4.3 ;
- les **réseaux de neurones** (figure 4.3 (b)) constitués de neurones formels organisés en couches, les valeurs d'entrée de la première couche correspondent aux valeurs des variables explicatives, la valeur de la variable à prédire est donnée en sortie du neurone de la dernière couche, cette technique est présentée dans la section 4.4.1 ;
- les **machines à vecteurs de support** (ou Support Vector Machine "SVM") qui séparent l'espace des instances par des hyperplans (figure 4.3 (c)) en optimisant les marges (distance entre les instances les plus proches appelées vecteurs supports et l'hyperplan), cette technique est présentée dans la section 4.4.2 ;

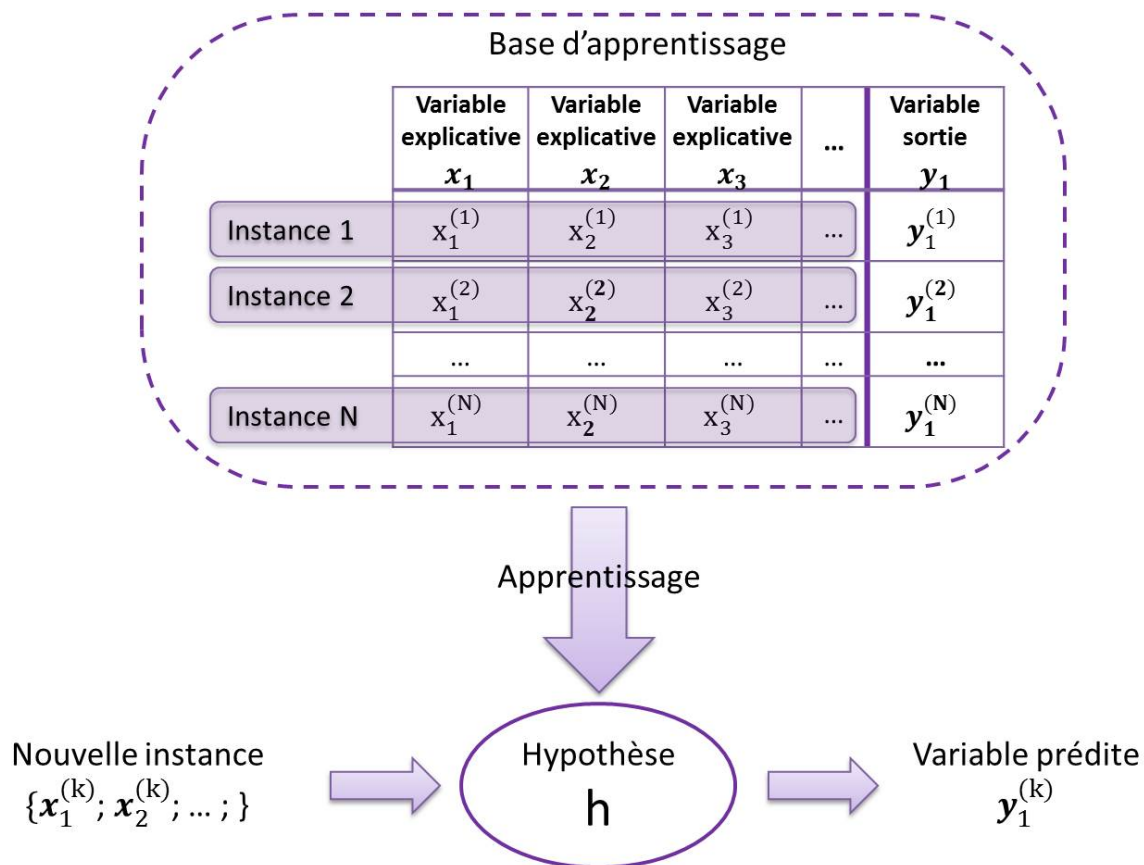


FIGURE 4.2 – Apprentissage supervisé. Prédiction d’une variable de sortie pour une nouvelle instance à l’aide d’une hypothèse obtenue par apprentissage.

- la **classification naïve bayésienne** qui s’appuie sur la théorie de Bayes pour déterminer la probabilité d’appartenance d’une instance à une classe, l’instance testée sera classée dans le groupe dont la probabilité est maximale, cette technique traite les différentes variables indépendamment les unes des autres.

Dans le cas de la **régression**, les variables de sorties peuvent prendre n’importe quelle valeur réelle. Les principaux algorithmes de régression sont basés sur les méthodes statistiques (régression linéaire, polynomiale ou logistique). Les machines à vecteurs de support peuvent également résoudre un problème de régression.

4.2.2 Apprentissage non supervisé

Dans le cas de l’apprentissage non supervisé (ou Clustering), les instances les plus similaires entre elles sont associées dans des groupes homogènes (figure 4.4). Il n’y a pas de variables de sortie imposées. Cette technique est utilisée principalement en fouille de données. La prédiction sur une nouvelle instance consiste à identifier le groupe auquel elle appartient. Contrairement à l’apprentissage supervisé, l’ensemble des données d’apprentissage est nécessaire lors du test sur une nouvelle instance.

Les principaux algorithmes d’apprentissage non supervisés sont :

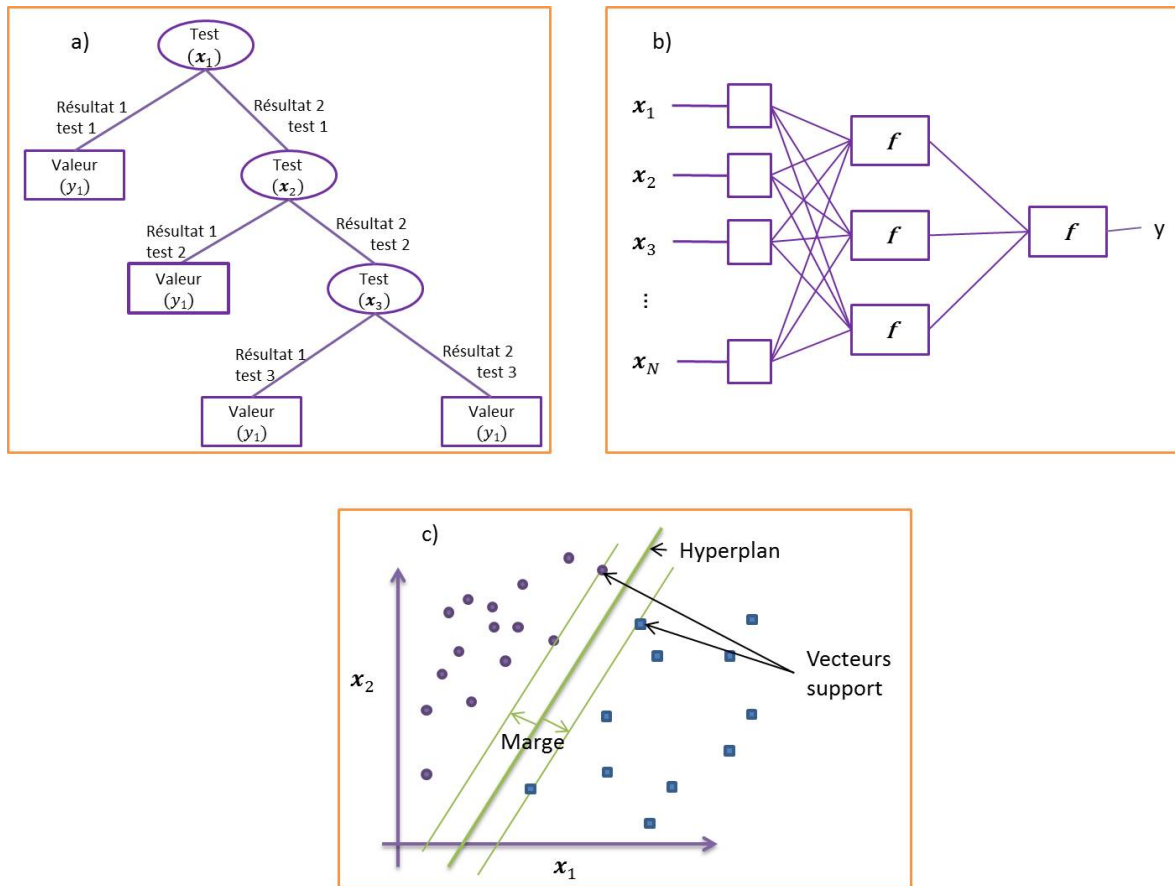


FIGURE 4.3 – Algorithmes de classifications. (a) Arbre de décision. (b) Réseau de neurones. (c) Machine à support de vecteurs.

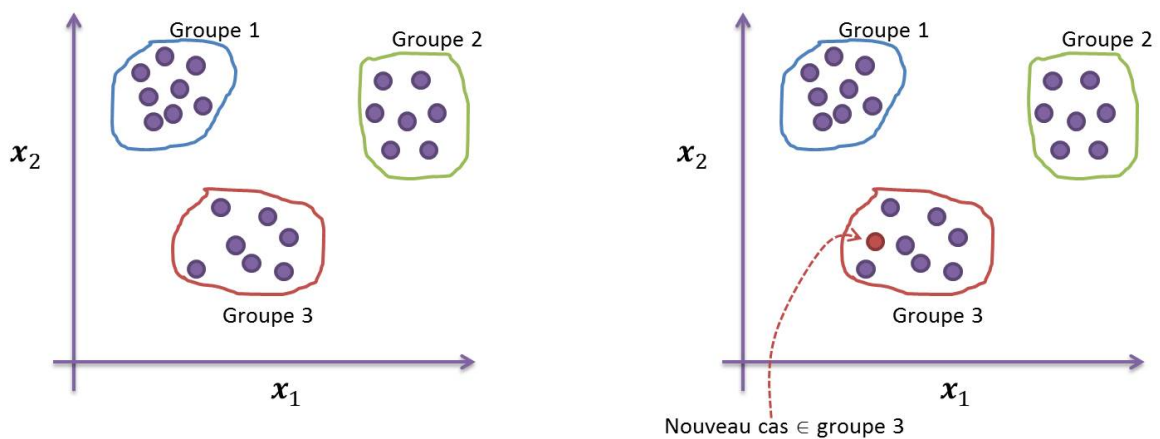


FIGURE 4.4 – Apprentissage non supervisé. Regroupement d'instances proches dans un cluster.

- l'algorithme "**espérance-maximisation**" qui assigne une distribution de probabilité pour chaque instance indiquant la probabilité d'appartenance à chacun des groupes ;
- l'algorithme des "**K-moyennes**" qui partitionne les instances dans k groupes considérant la distance d'une instance à la moyenne des instances d'un même groupe ;
- les algorithmes "**DBSCAN**" et "**OPTICS**" qui regroupent des instances en fonction

4.2.3 Logiciels de création d'algorithmes d'apprentissage

Les logiciels d'aide à la création d'algorithmes d'apprentissage artificiels sont nombreux. Les logiciels commerciaux et libres de droit les plus utilisés sont listés ci-dessous.

les logiciels commerciaux tels que SPSS¹ (Statistical Package for the Social Sciences) sont souvent destinés à l'analyse statistique ou à une application particulière (par exemple la fouille de données pour l'analyse du web). Quelques logiciels commerciaux proposent des algorithmes de classification (Statistica Data Miner²). Ces logiciels ont une interface facile à prendre en main et proposent un rendu esthétique.

Des langages de programmation destinés au calcul numérique tels que Matlab³ intègrent des algorithmes d'apprentissage supervisé et non supervisé.

Les logiciels libres sous licence GNU sont nombreux. Certains tels que R⁴ sont des langages associés à un environnement. Leur principe de fonctionnement à base d'interpréteur de commandes les rendent peu pratiques. Par contre, la qualité du graphisme facilite l'analyse des variables et les rendent intéressants pour présenter des données et des résultats.

Plusieurs logiciels libres proposent les principaux algorithmes d'apprentissage supervisé et non supervisé tels que Rapidminer⁵ ou Tanagra⁶ qui est destiné à l'enseignement et à la recherche.

Weka⁷ (Waikato Environment for Knowledge Analysis) est développé par l'université de Waikato en Nouvelle-Zélande. Il intègre des outils de visualisation de données, de traitement, des classifieurs et du clustering. Weka dispose d'une interface aisée et d'un grand nombre d'algorithmes ce qui rend son utilisation intéressante pour la sélection et la configuration d'algorithmes. C'est ce dernier qui a été utilisé dans le cadre de cette thèse.

1. SPSS, SPSS Inc, <http://www-01.ibm.com/software/analytics/spss/products/statistics/index.html>

2. Statistica, StatSoft Inc., <http://www.statsoft.fr>

3. Matlab, MathWorks, <http://fr.mathworks.com/products/matlab/>

4. R, <http://www.r-project.org>

5. Rapidminer, <https://rapidminer.com>

6. Tanagra, <http://eric.univ-lyon2.fr/~ricco/tanagra/fr/tanagra.html>

7. WEKA, University of Waikato, <http://www.cs.waikato.ac.nz/ml/weka/>

4.3 Techniques d'intelligence artificielle utilisées en ingénierie mécanique

Dans le domaine de l'ingénierie numérique, les techniques d'intelligence artificielle trouvent de nombreuses applications telles que la prédiction d'un comportement physique, l'aide à la conception, la reconnaissance de forme, le reverse engineering, les sciences des matériaux, ou bien le diagnostic.

Pour caractériser une technique d'IA il est nécessaire de connaître :

- l'objectif de l'apprentissage et le type de variables à prédire dans le cas d'apprentissage supervisé ;
- les caractéristiques des variables d'entrées disponibles ;
- les algorithmes utilisés et leurs paramètres.

4.3.1 Objectifs d'apprentissage

Les techniques d'IA sont utilisées dans le domaine de l'ingénierie mécanique pour des objectifs d'apprentissage très variés dont des exemples représentatifs sont donnés ci-dessous.

La prédiction de valeurs géométriques apporte une aide au dimensionnement et à la conception de produits.

Par exemple, Jahangirian et al [30] et Kharal et al. [32] estiment les paramètres géométriques qui définissent un profil d'aile d'aéronef afin d'optimiser sa forme pour un cas aérodynamique ou une pression donnés.

Yuan et al. [60] calculent la coordonnée z de points appartenant à un nuage connaissant les coordonnées x et y pour reconstruire une courbe Nurbs.

Gomes et al. [24] estiment l'épaisseur et l'angle entre les nappes de plaques réalisées en matériaux composites pour un cas de chargement donné.

La prédiction de grandeur physique permet de s'affranchir de la résolution d'équations complexes. Par exemple, Chan et al. [12] estiment l'effort maximal et la contrainte sur des outils d'emboutissage.

Chen et al. [14] estiment la valeur de la pression maximale applicable sur le toit d'un bâtiment.

Mazhar et al. [38] estiment la pression en n'importe quel point d'un aéronef pour le calcul par éléments finis connaissant les résultats du calcul CFD.

Oroumieh et al. [42] estiment les valeurs combinées de la poussée des moteurs, des principales caractéristiques des ailes et du poids d'un aéronef pour des conditions de vol don-

La prédiction de variables statistiques n'estime pas directement la valeur d'une grandeur géométrique ou physique mais donne une information sur leur écart, leur moyenne, leur tendance ou bien leur écart type. Il est également possible d'estimer la probabilité d'apparition d'un effet physique. Uematsu et al. [58] estiment les charges moyennes et les écarts types des pressions dues au vent en n'importe quel point d'un bâtiment. Bucar et al. [11] estiment les tendances et la dispersion sur des courbes de contrainte et de nombre de cycles avant défaillance pour évaluer un risque en fatigue. Lopes et al. [36] estiment la probabilité de défaillance d'un matériau composite.

La classification de formes ([31], [52]) ou de nuages points digitalisés [49] est utilisée pour la reconnaissance de modèles numériques ou pour leur réutilisation.

Les techniques d'IA sont utilisées pour l'automobile [43] et l'aéronautique [20] dans le but de diagnostiquer des causes de pannes ou des risques de défaillance. Pour cela, de nombreuses techniques d'apprentissage supervisé et/ou non supervisé sont mises en œuvre. Les connaissances en matière de diagnostic sont formalisées depuis de nombreuses années dans ces domaines, les techniques à base de systèmes experts peuvent également être utilisées.

On peut constater que les techniques d'intelligence artificielles sont largement exploitées pour des objectifs de classement de formes ou de nuages de points. Leur utilisation est plus confidentielle en ce qui concerne la prédiction de grandeurs physiques et de données géométriques. De plus, dans ces derniers cas d'utilisation, la géométrie du modèle à estimer doit être très proche de la géométrie des cas utilisés pour l'apprentissage. Une alternative, beaucoup répandue dont les résultats sont prometteurs est de prédire non pas la valeur précise d'une grandeur physique ou d'une caractéristique géométrique mais une valeur statistique telle qu'une valeur moyenne, un écart ou une probabilité.

4.3.2 Variables d'entrée

Lors de la mise en œuvre de techniques d'IA il est essentiel de bien définir les variables explicatives pour répondre aux objectifs d'apprentissage. Les variables d'entrée utilisées dans le domaine de l'ingénierie numérique pour répondre aux objectifs cités précédemment peuvent être des grandeurs physiques ([14], [42], [11], [24]) décrivant un problème de mécanique ou des données géométriques.

Les données géométriques peuvent être réduites à de simples coordonnées de points ([58], [38], [49], [60]) ou bien décrites par un vecteur de paramètres ([12], [30], [32]). Dans

Certains cas ([31]), un histogramme peut être utilisé. L'identification de ces variables est un point clé de ce travail de thèse. C'est effectivement au travers de ces variables que les données géométriques (modèles CAO, maillages, ...) seront caractérisées.

4.3.3 Techniques utilisées

Parmi les techniques évoquées dans la section 2.3.1, les principaux algorithmes utilisés en ingénierie mécanique sont des classifieurs de type **réseaux de neurones** pour la prédiction de grandeurs physiques et statistiques ou bien des algorithmes d'apprentissage non-supervisé de type **K-moyenne** pour la classification de formes ou de points digitalisés.

Pour la prédiction de grandeurs géométriques, les réseaux de neurones sont généralement associés à d'autres techniques d'IA telles que la logique floue ou les algorithmes génétiques, ce qui permet d'optimiser les grandeurs recherchées.

4.4 Les principaux classifieurs

Les classifieurs retenus dans cette thèse pour avoir fait leur preuve dans le domaine de l'ingénierie mécanique sont les réseaux de neurones, les arbres de décision et les machines à vecteur de support. Seuls ces classifieurs seront détaillés par la suite.

4.4.1 Réseaux de neurones

Principe

Les réseaux de neurones [18] (figure 4.3 (b)) sont constitués de plusieurs couches de neurones formels.

Un neurone formel (figure 4.5) réalise une fonction de transfert $g(x, w)$ entre un vecteur x de n variables d'entrée x_i et une variable de sortie y . Un vecteur w de n poids w_i est associé à chaque neurone d'entrée i .

Cette fonction $g(x, w)$ paramétrée peut être une fonction linéaire, une fonction à seuil ou une fonction sigmoïde ([18]).

Paramètres de configuration

L'architecture générale d'un réseau de neurones est définie par :

- le nombre de couches cachées de neurones (généralement une ou deux) ;
- le nombre de neurones par couche ;
- le type de fonction de transfert (linéaire, sigmoïde,...).

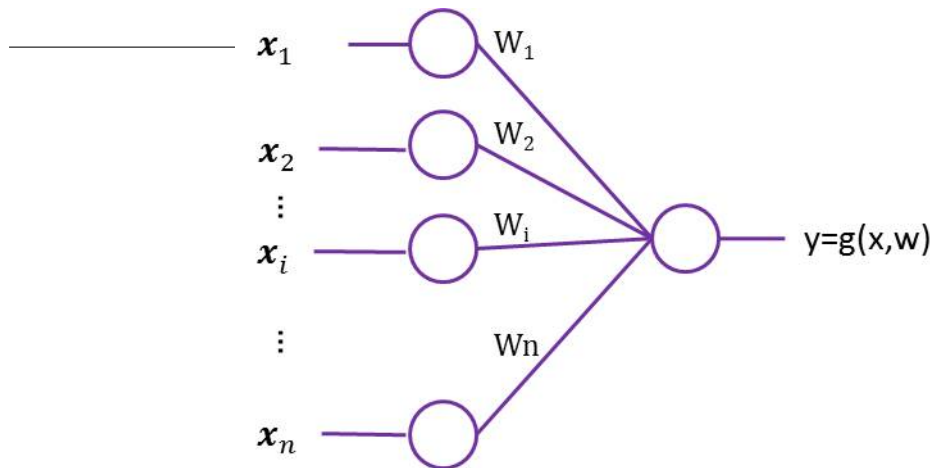


FIGURE 4.5 – Schéma d'un neurone à n entrées x et une sortie y.

Lors de l'apprentissage, les paramètres à définir sont :

- les poids w_i des neurones ;
- le nombre d'itérations avant arrêt de l'apprentissage.

Les poids w_i des neurones sont définis afin que la valeur prédite $g(x^{(k)}, w)$ soit la plus proche de la valeur réelle $y^{(k)}$, c'est à dire pour que la fonction coût $J(w)$ (équation 4.1) soit minimale. $x^{(k)}$ désigne le vecteur des variables d'entrée pour l'instance (ou exemple) k et N le nombre d'instances de la base d'apprentissage.

$$J(w) = \frac{1}{2} \sum_{k=1}^N \left[y^{(k)} - g(x^{(k)}, w) \right]^2 \quad (4.1)$$

Pour cela, la méthode la plus utilisée est l'algorithme de rétro-propagation (*back propagation*) du gradient de la fonction coût qui modifie les poids de manière itérative en sens inverse de la propagation de l'information, c'est à dire de la dernière couche à la première. La rétro-propagation s'arrête lorsque le minimum de la fonction coût est atteint ou lorsqu'un critère arrêt est satisfait [18].

Intérêt et limites

Les réseaux de neurones peuvent être utilisés en association avec d'autres techniques d'intelligence artificielle (systèmes experts, arbres de décision, algorithmes génétiques ou logique floue).

La multitude de choix d'architectures, de paramètres réglables et de méthodes utilisables facilite les chances d'obtenir un algorithme dont la fonction coût est minimale. Cependant, cela rend sa construction difficile et son intégration dans un système d'aide à la décision délicate.

Principe

Les machines à vecteur de support (support vector machine), ou séparateurs à vastes marges (SVM) sont basés sur le principe de séparation des instances par un hyperplan [50].

On appelle marge la distance entre l'hyperplan (figure 4.6) et les instances les plus proches (vecteurs de support).

Lorsque l'ensemble des instances est linéairement séparable, on recherche, parmi l'infinité d'hyperplans valides, à l'aide de la méthode de multiplicateurs de Lagrange, l'hyperplan optimal d'équation $w x + w_0$ pour lequel la marge est maximale. C'est à dire pour lequel $\frac{1}{w}$ est minimal.

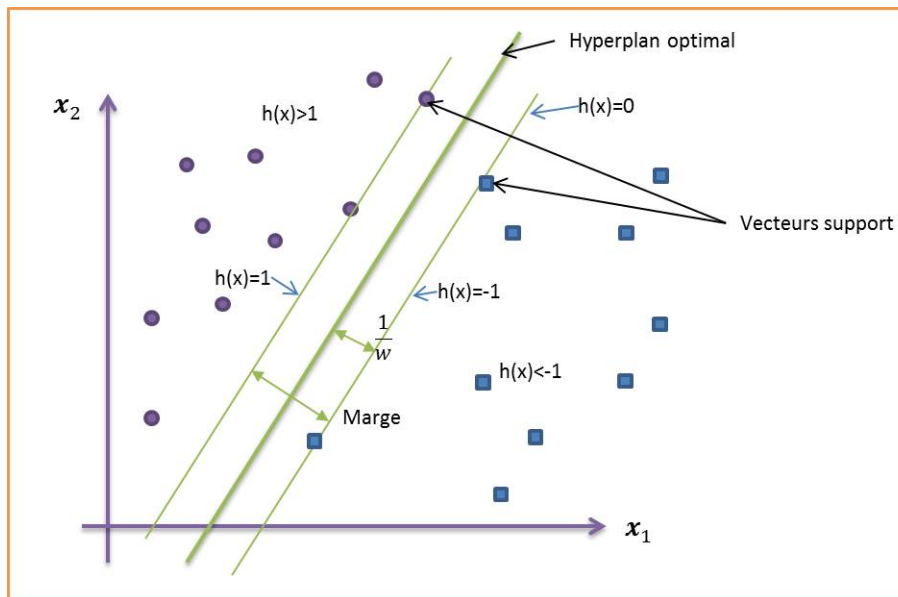


FIGURE 4.6 – Exemple d'un problème de classification à deux classes $y = 1$ ou $y = -1$ avec hyperplan optimal de séparation d'équation $h(x) = wx + w_0$

Dans le cas où l'ensemble des instances n'est pas linéairement séparable, les données sont transformées à l'aide d'une fonction nommée noyau (figure 4.7). Un noyau peut être une fonction polynomiale, gaussienne ou sigmoïde.

Paramètres de configuration

Les paramètres à définir lors de la construction du modèle d'apprentissage SVM sont :

- la fonction noyau k (polynomial ou gaussien),
- le paramètre de tolérance (compromis biais-variance). Plus ce paramètre est grand, plus la marge est grande et la sensibilité aux bruits est faible .

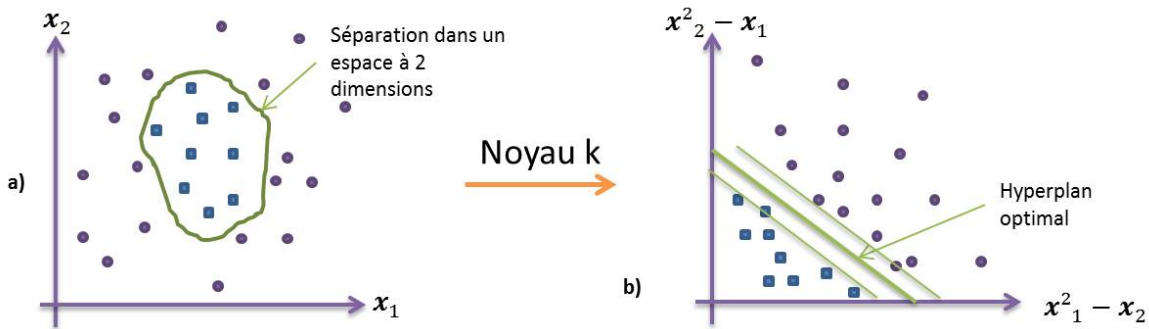


FIGURE 4.7 – Exemple de séparation non linéaire. (a) séparation non linéaire. (b) exemple de linéarisation par fonction noyau polynomiale.

Intérêt et limites

Le principal intérêt des SVM du point de vue de l'utilisateur est le faible nombre de paramètres à choisir lors de la construction du modèle ainsi que sa robustesse.

Néanmoins, les espaces des instances ne sont pas visibles. Ce qui a été appris est donc difficile à interpréter. Un système SVM est perçu comme une boîte noire dans laquelle il est difficile d'intégrer des connaissances a priori.

4.4.3 Arbres de décision

Principe

Un arbre de décision est constitué de nœuds (branches) réalisant des tests sur les variables explicatives x_i et de nœuds terminaux (feuilles) indiquant les valeurs d'une ou plusieurs variables de sortie y_i (figure 4.3 (a)).

Les arbres de décisions utilisés en IA sont de type CART (*Classification and Regression Tree*) issus des travaux de Breiman et al. [10] et les arbres C4.5 issus des travaux de Quinlan [44].

Un arbre de décision est généralement binaire (deux divisions par nœud). La construction d'un arbre consiste tout d'abord à sélectionner à chaque étape le meilleur nœud, puis à définir une condition d'arrêt sur les nœuds terminaux et enfin à affecter une classe ou une valeur à ce nœud terminal.

Les nœuds sont sélectionnés à chaque étape afin que les instances soient divisées de la façon la plus homogène possible. Les critères de sélection des nœuds sont définis à partir d'une mesure d'entropie [44] en minimisant le degré d'hétérogénéité pour les arbres C4.5 ou de l'indice de Gini [10] pour les arbres CART.

Un arbre de grande dimension permettra de classer toutes les instances de la base d'apprentissage mais ne les généralisera pas correctement (on parle alors de surapprentissage

ou d'overfitting). Un arbre de décision doit être alors construit pour que les instances soient correctement classées et pour que les règles obtenues soient généralisables. Il faut alors arrêter l'arbre au bon moment ou l'élaguer (*pruning*).

Paramètres de configuration

Les logiciels offrent des possibilités de construction des arbre de décision à l'aide d'un outil de type "boite noire" qui modélise automatiquement les algorithmes avec peu de paramètres. Seuls le nombre de divisions maximales (de 1 à 5) pour éviter d'avoir un arbre trop long et des paramètres d'élagage sont à renseigner.

Intérêt et limites

Le modèle obtenu est un graphique rapide à exécuter et simple à interpréter. Il permet de formaliser un ensemble de règles. Un grand nombre de variables explicatives peuvent être données en entrée. La sélection des variables déterminantes est intégrée à l'algorithme. Le traitement des variables est facultatif.

Une complexe opération d'élagage est indispensable pour éviter le risque d'over fitting si l'arbre est de grande dimension. Une reconstruction de l'arbre est nécessaire après chaque modification des variables d'entrée. L'arbre peut être profondément modifié par des changements mineurs de variables.

4.5 Méthodologie de conception d'un modèle d'apprentissage

Cornuéjol [16] et Dreyfus [18] ont proposé une méthodologie de conception d'un modèle d'apprentissage basée sur trois phases :

1. sélection des données d'entrée pour se limiter aux variables pertinentes ;
2. construction des modèles d'apprentissage en estimant les paramètres des algorithmes ;
3. sélection du meilleur modèle et estimation de ses performances.

La suite des travaux présentés dans ce manuscrits s'appuient sur les méthodes présentées dans cette section.

4.5.1 Sélection des données d'entrée

L'ensemble des variables d'entrée doit contenir de manière exhaustive les variables explicatives pertinentes et doit chercher à éviter les variables redondantes ou corrélées. Les variables d'entrées qui ont moins d'influence sur la prédiction que le bruit sont des causes d'erreurs sur le résultat la classification.

Limiter le nombre de données explicatives permet également de rendre la construction des algorithmes et leur exploitation plus rapide et plus robuste. Ceci est particulièrement vrai pour les réseaux de neurones dont le nombre de variables d'entrées est égal au nombre de neurones.

Le critère d'Akaike permet de comparer plusieurs modèles ayant des vecteurs de variables explicatives différents. Le meilleur modèle est celui dont le critère AIC "*Akaike Information Criterion*" (équation 4.2) est minimal. N est le nombre total d'instances, q le nombre de variables explicatives, et EQM l'erreur quadratique moyenne sur l'ensemble des données. Si deux modèles ont la même erreur EQM, le modèle retenu sera celui qui aura le moins de variables.

$$AIC = N \log(EQM) + 2(q + 1) \quad (4.2)$$

Il existe plusieurs méthodes de sélection des variables explicatives. Les méthodes les plus simples et largement utilisées dans la pratique sont les méthodes d'élimination (*stepwise backward regression*) et de construction (*stepwise forward regression*).

La méthode **stepwise backward regression** consiste à éliminer une par une les variables les moins pertinentes. On commence par identifier le meilleur modèle parmi tous les modèles ayant $q-1$ variables. Si le modèle à q variables n'est pas meilleur que celui qui vient d'être identifié, on sélectionne le meilleur modèle parmi les modèles de $q-2$ variables auxquels on a retiré une variable. On réitère l'opération jusqu'à ce que le modèle complet soit

La méthode **stepwise forward regression** consiste à ajouter une par une les variables les plus pertinentes. On identifie tout d'abord le meilleur modèle parmi les q modèles comprenant une seule variable. Puis, on crée $q-1$ modèles en ajoutant une variable au modèle retenu précédemment. On identifie le meilleur modèle parmi les $q-1$ modèles comprenant 2 variables. L'opération est répétée jusqu'à ce que la performance des modèles n'évolue plus.

Une troisième méthode, plus performante est la méthode du **descripteur sonde** [13]. Les variables d'entrée sont tout d'abord classées par ordre décroissant d'influence sur la sortie en utilisant une orthogonalisation de Gram-Schmidt. On ajoute aux variables candidates, une variable supplémentaire, appelée descripteur sonde, dont les valeurs sont calculées aléatoirement. On procède au classement des variables en incluant le descripteur sonde et on élimine toutes les variables moins bien classées que lui.

4.5.2 Construction des modèles d'apprentissage

La construction des modèles d'apprentissage consiste à définir leur architecture, à normaliser les données d'entrée et de sortie puis à estimer les paramètres des algorithmes.

L'architecture du modèle d'apprentissage est définie a priori selon le type d'algorithme (réseau de neurones, SVM, arbre de décision,...).

Un modèle d'apprentissage peut être linéaire ou non linéaire. Nous nous intéressons à ce second type d'apprentissage. La normalisation et le centrage des données d'entrée et de sortie est indispensable avant tout apprentissage. En pratique, on cherche à faire en sorte que la moyenne des variables soit égale à 0 et que son écart type soit égal à 1.

Pour un algorithme défini par un vecteur de paramètres w comme un réseau de neurones, l'estimation des paramètres consiste à trouver les composantes de ce vecteur en minimisant la fonction coût $J(w)$ (équation 4.1).

L'apprentissage peut être adaptatif ou non adaptatif et avec ou sans régularisation. G. Dreyfus [18] propose une méthode d'estimation des paramètres des algorithmes en fonction du type de modèle d'apprentissage.

Les méthodes d'apprentissage utilisées pour les modèles linéaires telles que la méthode des moindres carrés ne peuvent être mises en œuvre pour les modèles non linéaires.

Une méthode itérative d'évaluation du gradient de la fonction coût et de modification des paramètres en fonction de ce gradient est nécessaire. La méthode proposée par G.

- initialiser les paramètres ;
- calculer le gradient de la fonction coût (par exemple, avec la méthode de retropagation pour un réseau de neurones [47]) ;
- modifier les paramètres par une méthode de minimisation ;
- arrêter la procédure lorsqu'un nombre maximal d'itération a été atteint ou lorsque la norme du gradient est inférieure à une limite.

Si on souhaite enrichir la base d'apprentissage avec de nouveaux cas, il faut mettre en œuvre un modèle d'apprentissage adaptatif afin de faire évoluer les algorithmes en tenant compte des dérives du processus dues aux nouvelles données. Pour cela, on peut appliquer une technique de modification adaptative des paramètres telle que celle du gradient stochastique [18].

Lorsque le nombre d'exemples est petit, le risque de sur-apprentissage est grand, on peut alors rencontrer des difficultés pour généraliser le problème. Il est alors possible d'intégrer une méthode de régularisation qui permet de limiter l'amplitude des paramètres du modèle. Deux modes de régularisation existent :

- l'**arrêt prématuré** qui consiste à arrêter l'apprentissage avant la convergence ;
- la **pénalisation** qui consiste à ajouter un terme à la fonction de coût ce qui pénalise les algorithmes ayant des poids élevés (méthode de "modération des poids").

4.5.3 Sélection et évaluation des modèles d'apprentissage

La sélection du "meilleur" modèle d'apprentissage consiste tout d'abord à identifier la meilleure architecture puis le meilleur algorithme et ses paramètres pour cette architecture. Le meilleur modèle sera celui dont les performances seront les plus élevées selon les critères définis ci-après (erreur quadratique moyenne, pourcentage d'erreur, surface sous la courbe *Receiver Operating Characteristic Curve* (ROC)).

La construction d'un modèle sur l'ensemble des données d'apprentissage augmente le risque de sur-apprentissage. Pour éviter cela, une méthode de sélection des modèles par validation croisée peut être appliquée. Cette méthode consiste à découper l'ensemble des instances d'apprentissage en deux groupes. L'apprentissage s'effectue sur le premier des deux groupes, puis on teste les résultats des prédictions sur les données du second groupe. On réitère ainsi l'apprentissage sur des groupes différents pour améliorer la généralisation de l'algorithme.

Le choix d'un algorithmes d'apprentissage est basé sur des critères de précision des résultats de prédiction, de robustesse et de coût (durée et facilité de création et d'utilisation des algorithmes).

La précision des résultats de prédiction est évaluée à partir de différents critères :

- l'erreur quadratique moyenne (AQE : *Average Quadratic Error*) d'un problème de régression (équation 4.3), $y^{(k)}$ étant la valeur réelle de la variable de sortie pour l'exemple k parmi l'ensemble des instances apprises N , et $g^{(k)}$ étant la variable prédite ;

$$AQE = \sqrt{\frac{1}{N} \sum_{k=1}^N (y^{(k)} - g^{(k)})^2} \quad (4.3)$$

- le pourcentage d'erreurs de classification et le pourcentage d'erreurs inacceptables. Une erreur sera considérée comme inacceptable lorsqu'elle est optimiste par rapport à la valeur réelle.
- la surface sous la courbe ROC (*Receiver Operating Characteristic Curve*) qui donne le taux de résultats « vrais-positifs » par rapport au taux de « faux-positifs ». La surface sous la courbe ROC doit être la plus grande possible. Si on obtient une courbe bissectrice telle que la courbe 1 de la figure 4.8, les résultats de la prédiction sont équivalents à des résultats choisis au hasard. Si on compare deux modèles d'apprentissage ayant respectivement les courbes ROC 2 et 3, on préférera celui correspondant à la courbe 3 dont l'aire est plus proche de 1.

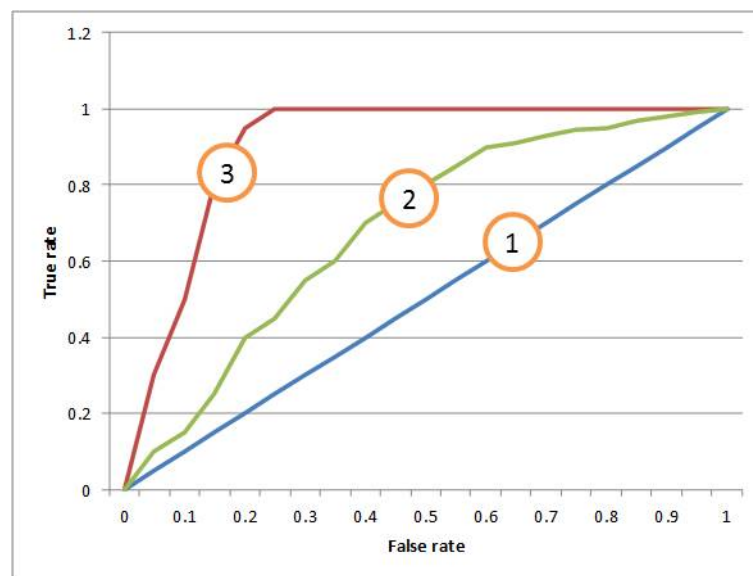


FIGURE 4.8 – exemples de courbes ROC pour l'analyse de "vrais-positifs" / "faux-positifs".

4.6 Positionnement des travaux par rapport à l'utilisation des techniques d'IA

L'objectif de ces travaux est tout d'abord d'estimer a priori l'impact de la préparation d'un modèle CAO sur le résultat de l'analyse ainsi que son coût. In fine, il s'agira d'identifier le meilleur processus de préparation pour un objectif de simulation.

Nous ne cherchons pas à estimer le résultat de l'analyse mais seulement son écart par rapport à la réalité. Un paramètre de type « statistique » semble le plus approprié à notre cas.

Une valeur précise des écarts et des durées n'étant pas nécessaire, les variables de sortie pourront être bornées. Un problème de classification plutôt que de régression doit convenir.

Les techniques d'apprentissage artificiel utilisées seront des classifieurs couplés éventuellement à un algorithme génétique. Des techniques de clustering pourront être mises en œuvre pour associer des variables dans le but de faciliter la compréhension d'un comportement (par exemple un niveau de simplification global du modèle).

Le modèle de simulation peut être décrit par un vecteur de grandeurs physiques comme dans les exemples de la section 4.3 (efforts, pressions, contraintes, déformations, températures, durées,...).

La description des modèles géométriques 3D amène plus d'interrogations. Dans notre cas, la complexité des modèles CAO (assemblages riches) rend difficile l'utilisation de graphes ou d'histogrammes. Le haut niveau de simplification du modèle préparé ne nous permet pas d'utiliser une description basée sur des coordonnées. Un vecteur de paramètres qui restent à définir semble être la meilleure solution.

Le processus de préparation des modèles pourrait être décrit par un arbre, un graphe ou bien un vecteur de paramètres. Le choix du modèle dépend du type de technique d'intelligence artificielle choisie. Si on s'oriente sur des outils de classification on préférera une modélisation par vecteurs de paramètres.

4.7 Verrous relatifs aux techniques d'intelligence artificielle

Constats :

La mise en œuvre de techniques d'apprentissage artificiel nécessite de connaître les variables explicatives qui sont utilisées pour identifier une hypothèse h capable de prédire une variable de sortie. Dans le cas de la préparation de modèles pour la simulation numérique, les variables de sorties sont l'écart entre la grandeur physique réelle et le résultat de l'analyse ainsi que le coût de la préparation. Ces variables de sorties doivent être qualifiées (nombre de classes, bornes,...). Les variables explicatives relatives au processus de préparation, aux modèles 3D originaux, simplifiés et préparés doivent être définies.

Les techniques d'apprentissage artificiel énoncées dans la section précédente nécessitent que l'on dispose d'un grand nombre d'exemples. Or, la préparation de modèles pour la simulation numérique est une activité très longue, le nombre d'exemples est souvent limité.

Trois verrous scientifiques sont associés à ces constats concernant l'utilisation des techniques d'apprentissage artificiel :

Verrous relatifs à la mise en œuvre de techniques d'apprentissage artificiel

- Quelles sont les variables explicatives utilisées pour prédire les variables de sortie ?
- Comment convertir les données géométriques 3D (B-Rep, maillage) en variables explicatives ?
- Comment traduire un processus de préparation de modèle de simulation en variables explicatives ?
- Comment garantir la fiabilité de l'estimation de la qualité d'un processus de préparation à l'aide des techniques d'apprentissage artificiel sur un nombre restreint d'exemples ?

Deuxième partie

Contributions

Chapitre 5

Approche proposée pour l'estimation de la qualité d'un processus de préparation de modèles de simulation

La finalité de l'approche proposée est de guider les analystes dans le choix des opérations de préparation de modèles CAO pour la simulation. Cela ne les dispensera pas de réaliser les opérations de préparation ou de simulation, mais permettra d'obtenir plus rapidement un modèle simplifié de meilleure qualité.

Pour cela, les connaissances acquises par les analystes sur la préparation de modèles pour la simulation peuvent être mobilisées et réutilisées pour analyser un nouveau cas de préparation. La première section de ce chapitre présente différentes approches qui peuvent être envisagées selon l'objectif que l'on se donne à savoir :

- l'identification d'un processus de préparation pour un nouveau cas ;*
- l'évaluation de la qualité d'un processus de préparation donné ;*
- la détermination de règles qui guident l'analyste dans la réalisation des opérations de préparation.*

La seconde section présente l'approche générale élaborée et ses principales phases.

La dernière section de ce chapitre présente les différents éléments qui devront être déterminés pour répondre à l'objectif et les contributions apportées.

5.1 Structure de l'approche

STRUCTURE APPLICATIVE PROPOSÉE POUR L'ESTIMATION DE LA QUALITÉ D'UN PROCESSUS DE PRÉPARATION DE MODÈLES DE SIMULATION

Dans un premier temps, une banque d'exemples de processus de préparation de modèles CAO pour la simulation numérique peut être constituée de cas existants à partir de modèles CAO originaux, de modèles préparés et de rapports d'analyse (figure 5.1) .

L'idée est ensuite de réutiliser cette banque d'exemples existants sur de nouveaux cas. Un "nouveau cas" de préparation de modèle pour la simulation peut être :

- un nouveau produit dont on connaît le modèle CAO original ainsi que les objectifs de la simulation et pour lequel aucune préparation ni simulation n'est disponible ;
- une nouvelle configuration de produit (ajout ou suppression de composants et/ou modification de dimensions) ;
- un nouveau processus de préparation d'un produit ;
- de nouveaux objectifs de simulation d'un produit pour lesquels d'autres objectifs de simulation et processus de préparation ont déjà été testés.

Approche type « systèmes à base de cas » (figure 5.2 (a))

Dans un système à base de cas, lorsqu'un nouveau cas se présente, ce dernier est comparé à l'ensemble des cas présents dans la banque d'exemples. Le processus proposé est alors le processus déjà appliqué à un cas connu et dont les caractéristiques géométriques du modèle original et le cas de simulation sont les plus proches du nouveau cas. Un processus ne peut être proposé que s'il existe un cas similaire présent dans la base, ce qui nécessite d'avoir un très grand nombre d'exemples.

Plusieurs semaines sont nécessaires à un ingénieur pour préparer et simuler une cinquantaine d'exemples de processus de préparation pour un cas complexe. La taille de la base d'exemples sera forcément limitée ce qui rend cette approche limitée. De plus, comme cela a été évoqué dans la section 4.1 présentant les principes des systèmes à base de connaissance, ce système peut ne pas conduire à de nouveaux processus ce qui le rend peu évolutif.

Approche par évaluation et identification d'un processus (figure 5.2 (b))

Industriellement, il est envisageable de disposer d'une banque restreinte d'exemples judicieusement choisis. Lorsqu'un nouveau cas se présente (ici un nouveau processus), un ensemble de processus de préparation doit pouvoir être proposé par l'analyste. Chaque processus est alors évalué à l'aide d'un indicateur de performance (niveau 1 de la figure 5.2 (b)). Le meilleur processus pour ce nouveau cas peut être identifié parmi l'ensemble des processus évalués. Cette proposition est intéressante si ces indicateurs de performance peuvent être prédits a priori sans qu'il soit nécessaire d'exécuter les opérations

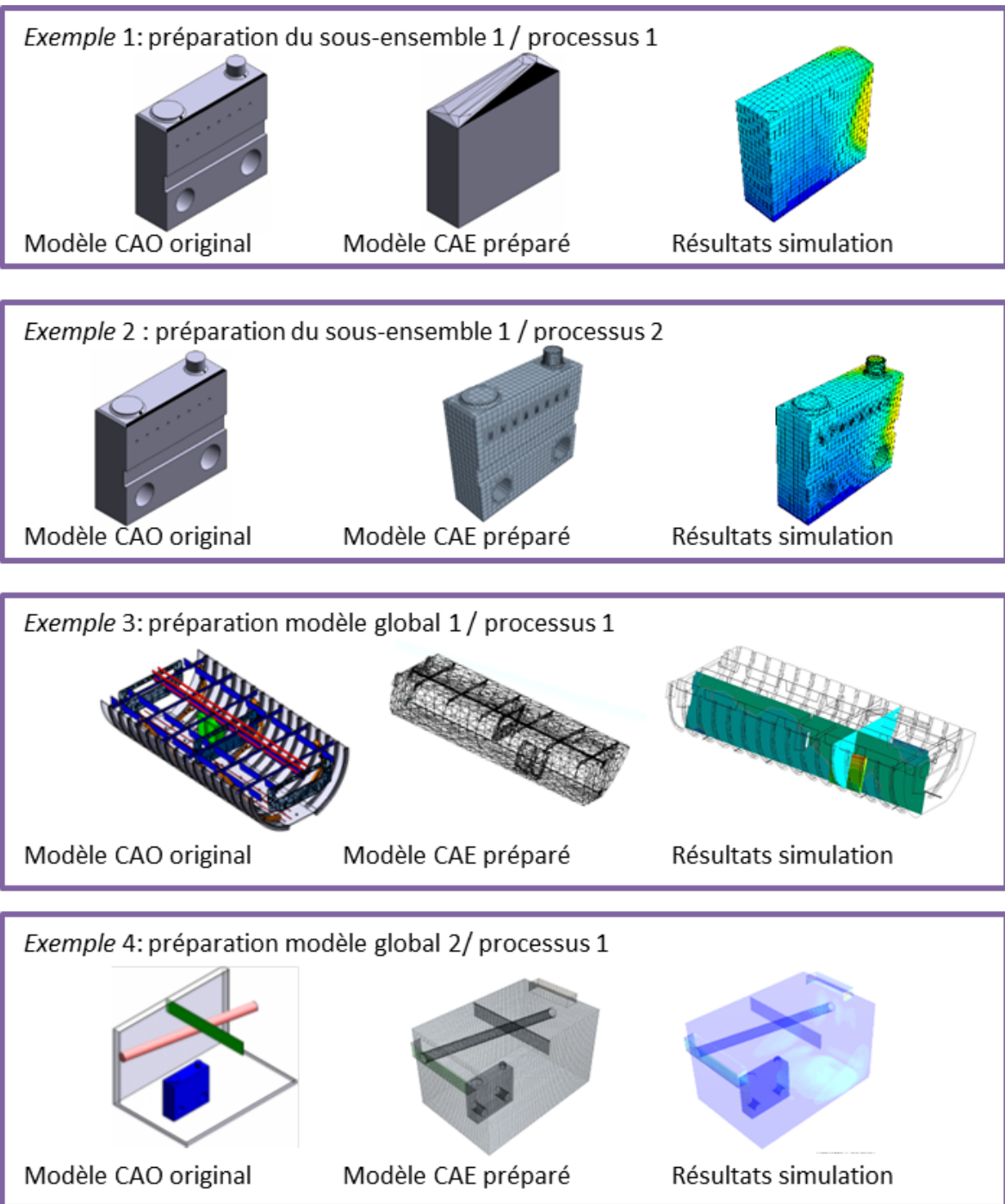


FIGURE 5.1 – Exemples de cas existants de préparation de modèles pour la simulation.

de préparation du modèle.

Ces travaux de thèse s'inscrivent dans cette logique. Une méthode basée sur cette approche d'évaluation a priori d'un processus de préparation de modèle pour la simulation est ainsi présentée dans la section suivante.

Dans la perspective de travaux futurs, l'approche précédente pourra être améliorée en proposant des règles qui donneraient les paramètres de réglage de chaque opération dans

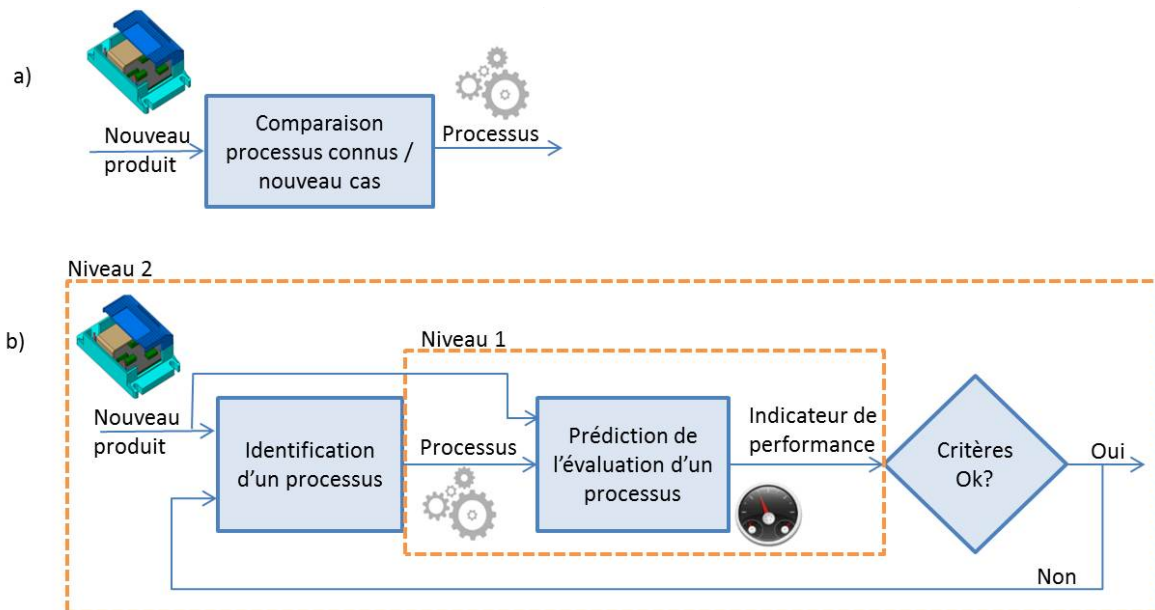


FIGURE 5.2 – Structuration de l'approche. (a) Approche type « systèmes à base de cas ». (b) Approche par identification d'un processus optimal.

le but de construire le processus ayant la meilleure performance possible. Puis, le processus optimal pour un nouveau cas pourrait être identifié (niveau 2 de la figure 5.2 (b)).

Dans la suite, les travaux se concentrent sur le niveau 1 de la figure 5.2 (b) soit sur l'évaluation a priori de la qualité d'un processus de préparation.

5.2 Présentation générale de l'approche proposée

L'approche retenue (figure 5.3) consiste tout d'abord à capitaliser les connaissances des experts dans une base d'exemples de processus de préparation pour ensuite construire un modèle capable de prédire la qualité d'un processus pour un nouveau cas.

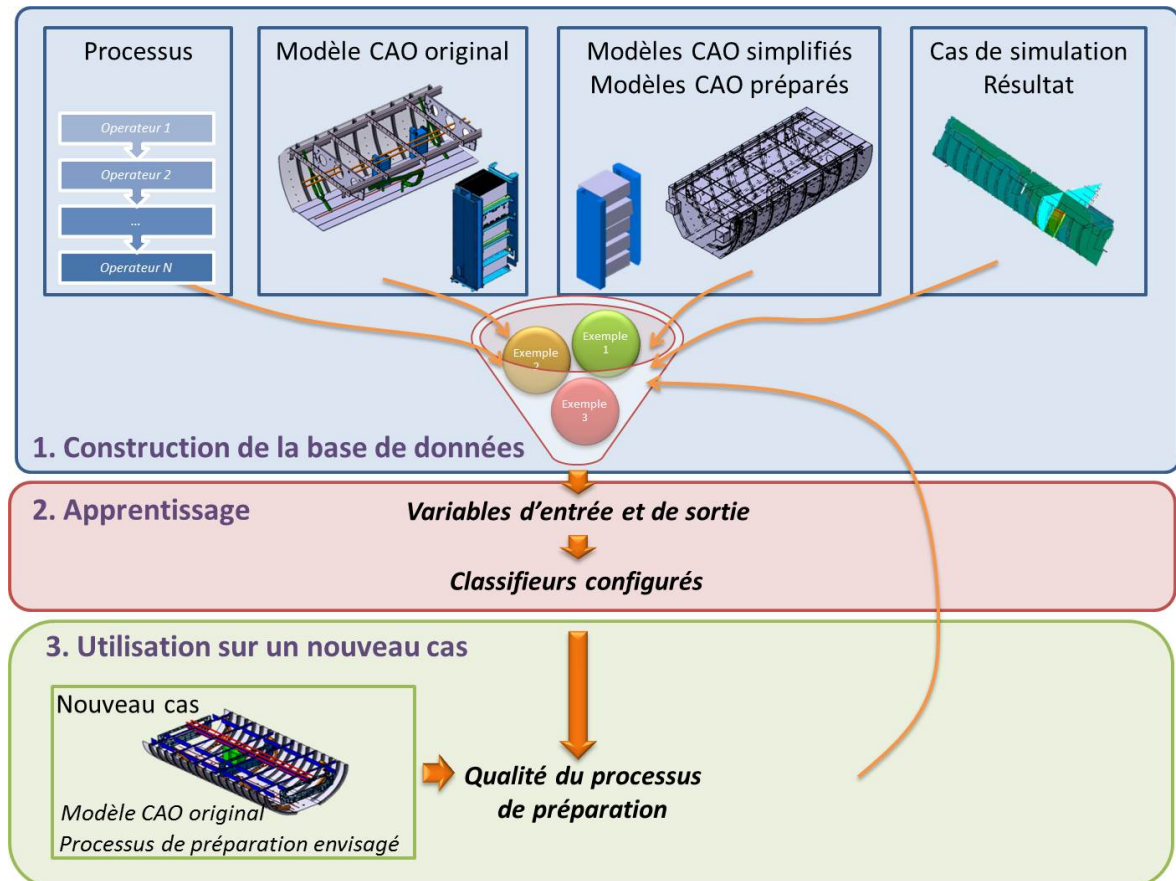


FIGURE 5.3 – Approche générale proposée (niveau 1 de la figure 5.2).

Première étape : construction de la base de données

L'objectif de cette première étape est de construire d'une base de données pour l'apprentissage. Pour cela, les connaissances acquises par les experts en matière de préparation de modèles pour la simulation sont formalisées puis implémentées dans un modèle d'apprentissage (figure 5.4).

La base de données d'exemples de processus de préparation est construite à partir de la représentation formalisée des connaissances extraites :

- d'un modèle de processus de préparation des modèles CAO ;
- de la description des modèles CAO originaux et simplifiés ;
- des informations pour la comparaison des modèles originaux et simplifiés ;

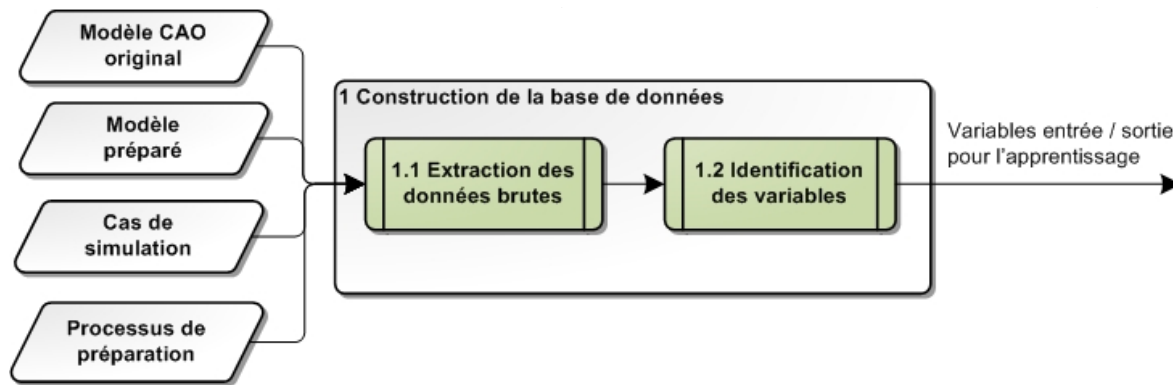


FIGURE 5.4 – Etape 1 de l'approche proposée. Construction de la base de données pour l'apprentissage.

- des informations sur les caractéristiques du maillage appliqué, sa qualité et la durée du processus de maillage ;
- des informations sur le cas de simulation (objectif de la simulation, conditions limites, grandeurs physiques, équations,...), la durée de la simulation et les résultats de l'analyse.

Les données extraites de supports de natures diverses (modèles CAO, maillages, rapports) devront être transformées avant d'être utilisées par un outil d'intelligence artificielle de type classifieur. Une représentation matricielle des données facilitera leur implémentation. Chaque ligne correspond à un exemple de préparation de modèle pour la simulation et chaque colonne à une variable d'entrée (explicative) ou de sortie (à prédire). Les variables à prédire sont les critères d'évaluation d'un processus de préparation d'un modèle pour la simulation. Les variables explicatives décrivent les processus de préparation et caractérisent les variables à prédire.

Afin de caractériser au mieux les variables à prédire, la base de donnée doit contenir toutes les variables explicatives capables de décrire un processus de préparation ou ses critères d'évaluation. Cependant, ces variables explicatives ne sont pas connues. La base de données contient donc dans un premier temps un très grand nombre de variables. Les variables explicatives les plus déterminantes seront ensuite sélectionnées.

Deuxième étape : apprentissage

La seconde étape a pour objectif de configurer un classifieur pour chaque variable à prédire (figure 5.5), en :

- choisissant un type d'apprentissage supervisé ou non supervisé et un type d'algorithme (arbre de décision, réseau de neurones, machine à vecteur de support, classification naïve bayésienne, algorithme des K-moyennes,...). Les principaux classifieurs ont été définis dans la section 4.2 "Principes de l'apprentissage artificiel" ;

CHAPITRE 5. APPROCHE PROPOSÉE POUR L'ESTIMATION DE LA QUALITÉ D'UN PROCESSUS DE PRÉPARATION DE MODÈLES DE SIMULATION

- adaptant les variables au classifieur choisi ;
- identifiant les paramètres du classifieur capable de prédire une variable de sortie à partir d'un vecteur de variables explicatives ;
- évaluant la qualité du modèle d'apprentissage par la validation du choix des variables explicatives et de la configuration du classifieur.

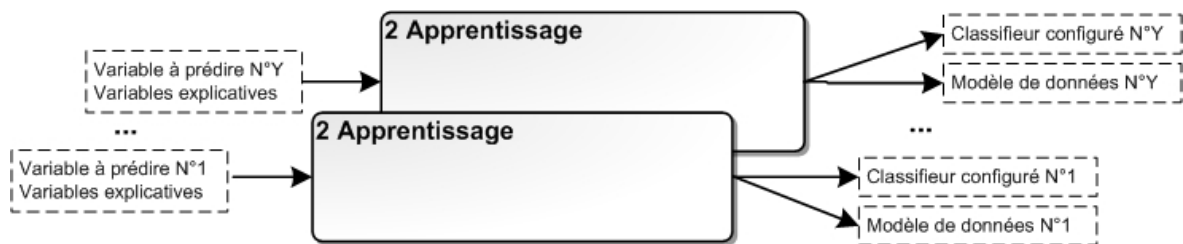


FIGURE 5.5 – Etape 2 de l'approche proposée. Sélection et configuration de classifieurs par apprentissage.

Troisième étape : utilisation sur un nouveau cas

La troisième étape (figure 5.6) consiste à évaluer un processus proposé par l'utilisateur pour un nouveau cas à partir des données disponibles pour ce nouveau cas (modèle CAO original, objectif de simulation, conditions limites et processus de préparation) et des classifieurs précédemment configurés .

Parmi les variables explicatives sélectionnées pour la prédiction des variables à prédire, certaines ne sont pas connues pour un nouveau cas (modèle CAO simplifié, modèle préparé, résultat de la simulation,...). Nous nommerons "variables intermédiaires" ces variables explicatives inconnues pour un nouveau cas. Les premières variables intermédiaires à estimer sont celles qui décrivent le modèle simplifié pour chaque sous-ensemble, à partir des données extraites des modèles originaux des sous-ensembles et des sous-processus de simplification proposés. Connaissant les caractéristiques des modèles simplifiés, on peut ensuite estimer le coût de la simplification, l'influence de la simplification sur chaque sous-ensemble et les caractéristiques du modèle préparé. Les principales variables à prédire (coût de la préparation, erreur sur le résultat de l'analyse et coût de la simulation) seront estimées à leur tour à partir de l'ensemble des variables explicatives connues et intermédiaires.

L'analyste pourra juger la qualité du processus de préparation proposé à partir des valeurs estimées pour ces trois principales variables. Il n'a donc pas besoin de réaliser les simplifications lors de cette estimation a priori.

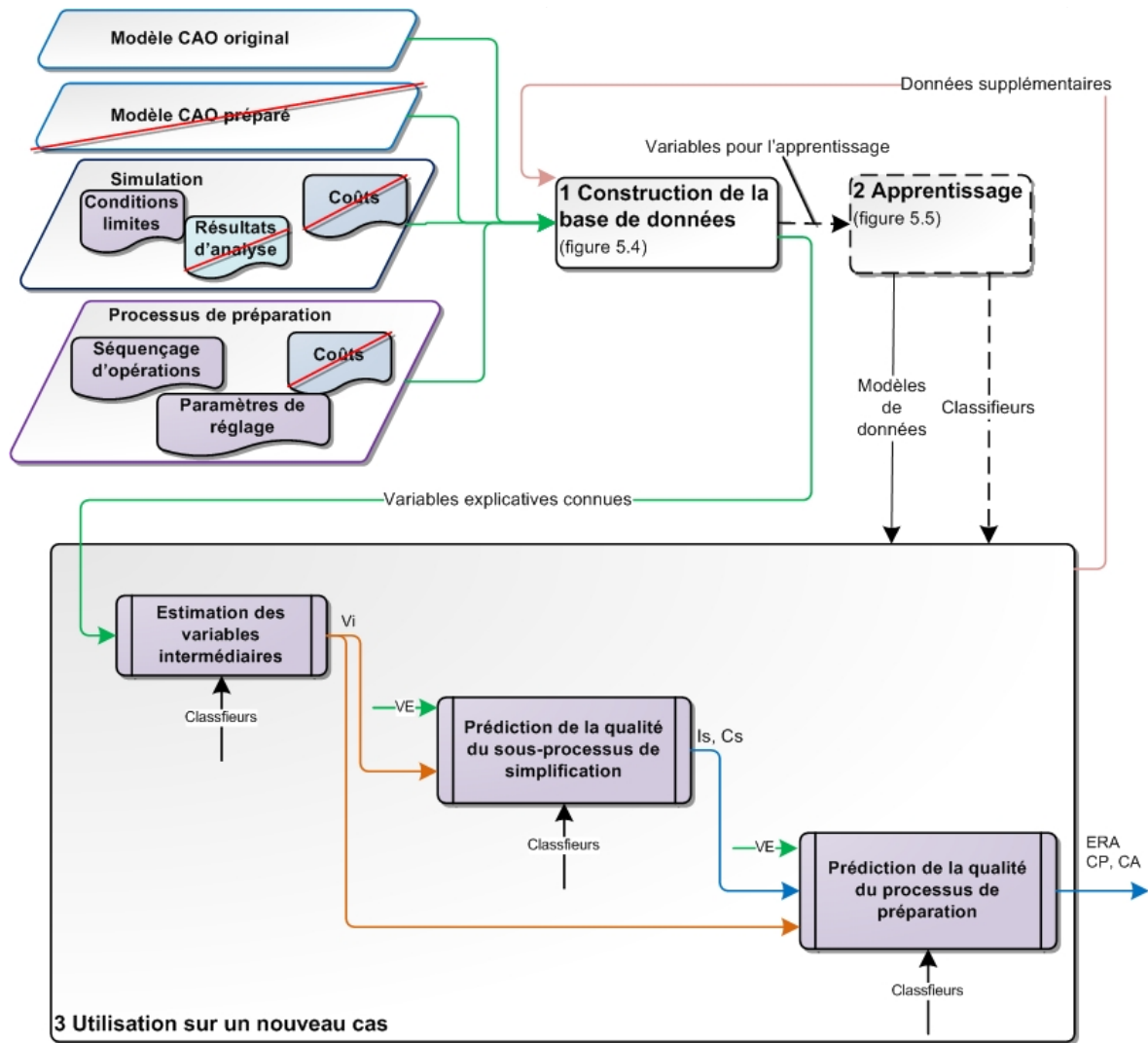


FIGURE 5.6 – Etape 3 de l'approche proposée. Utilisation sur un nouveau cas. Prédiction des critères d'évaluation d'un processus de préparation pour la simulation. VE = variables explicatives connues. VI = variables intermédiaires. CS = coût de la simplification. CP = coût de la préparation. CA = coût de l'analyse. IS = influence de la simplification d'un sous-ensemble sur le résultat. ERA = erreur sur le résultat de l'analyse.

Finalement, les données extraites et estimées sur les nouveaux cas viendront enrichir la base de données d'apprentissage.

5.3 Contributions

CHAPITRE 5.4 CONTRIBUTION PROPOSÉE POUR L'ESTIMATION DE LA QUALITÉ D'UN PROCESSUS DE PRÉPARATION DE MODÈLES DE SIMULATION

Pour définir un classifieur capable d'estimer la qualité d'un processus de préparation à partir d'exemples de processus, il est nécessaire de déterminer cinq éléments (figure 5.7) : (1) les exemples apprentissage, (2) les variables à prédire, (3) les variables explicatives, (4) le type de classifieur et (5) ses paramètres. Pour chacun de ces cinq éléments, il est rappelé ci-dessous les verrous à lever qui ont été identifiés dans les chapitres précédents et les méthodes proposées qui seront détaillées dans les chapitres suivants 6 et 7 .

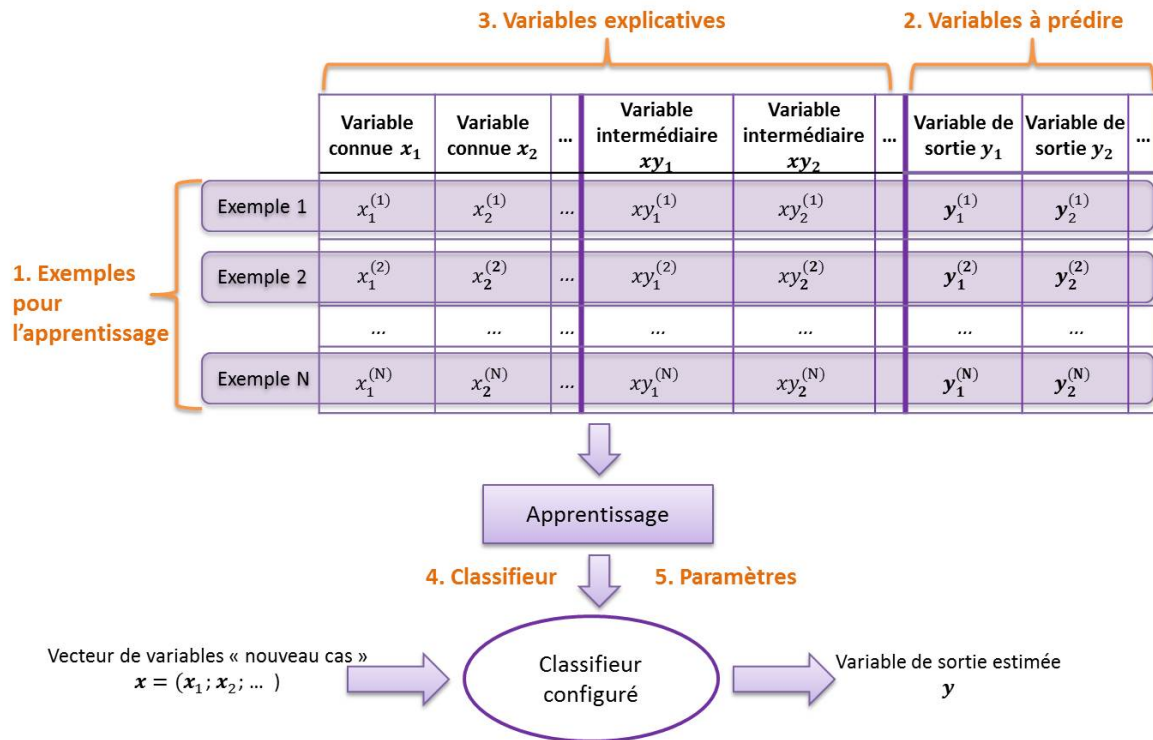


FIGURE 5.7 – Représentation matricielle des données et éléments à identifier pour prédire une variable de sortie par apprentissage.

1) Exemples pour l'apprentissage.

La construction de la base d'exemples pour l'apprentissage a pour objet l'acquisition des connaissances des experts, leur formalisation et leur implémentation dans les outils d'intelligence artificielle (figure 5.4).

L'acquisition de connaissance consiste à identifier des cas existants ou à en construire de nouveaux. La technique d'apprentissage choisie est basée sur la classification d'exemples de processus de préparation pour laquelle il est nécessaire d'avoir un nombre significatif d'exemples dans chaque classe. La préparation et la simulation d'un cas sont des opérations de longue durée, le nombre d'exemples de processus de préparation sera forcément restreint. Une méthode de construction d'une base d'exemples représentatif est proposée dans la section 6.4 du chapitre suivant.

*CHAPITRE 5. APPROCHE PROPOSÉE POUR L'ESTIMATION DE LA QUALITÉ D'UN
PROCESSUS DE PRÉPARATION DE MODÈLES DE SIMULATION*

Le principal frein concernant la formalisation des connaissances est le manque de modélisation du processus de préparation d'un modèle pour la simulation. Une représentation hiérarchique des processus de préparation est proposée dans la section 6.1.

Les facteurs utilisés pour évaluer un processus de préparation peuvent être extraits de supports disponibles (modèles CAO, maillages, rapports). Par contre, ces connaissances ne sont pas formalisées, les facteurs déterminants ne sont pas identifiés (par exemple, les facteurs géométriques qui influencent le résultat de l'analyse ne sont pas complètement connus). Une représentation de l'ensemble des données pouvant décrire et évaluer un processus de préparation est proposée dans la section 6.2.

Les formats des variables sont variés (valeurs discrètes ou continues, bornées ou non bornées). L'implémentation des données nécessite la transformation des données brutes en variables exploitables par les outils d'intelligence artificielle. Les difficultés concernent principalement les données de nature géométrique ainsi que la description des processus. Pour cela, un ensemble d'opérations de traitement des données est proposé dans la section 7.1.2.

2) Variables à prédire.

Les principales variables de sortie à prédire sont les critères d'évaluation d'un processus de préparation. Ces variables sont définies dans la section 6.2.3 au niveau local (coût et influence de la simplification sur un sous-ensemble) et au niveau global (coûts de la préparation, coût de la simulation et erreur sur le résultat de l'analyse sur l'ensemble complet). Là encore, la difficulté réside dans l'identification des bonnes variables permettant de caractériser au mieux la qualité des processus à évaluer.

3) Variables explicatives.

Les variables déterminantes pour la caractérisation des variables à prédire ne sont pas clairement identifiées. La base de données comprend un grand nombre de variables dont certaines sont connues et d'autres inconnues pour un nouveau cas qui devront être estimées par apprentissage. Il s'agit de trouver un compromis entre trop peu de variables et un trop grand nombre. Une méthode de sélection des variables explicatives les plus déterminantes parmi l'ensemble des variables de la base de données est proposée dans la section 6.3.

4) Type de classifieur et architecture générale.

Les classifieurs et leur architecture ont été présentés dans la section 4.4. Deux critères d'évaluation des classifieurs ont été définis dans la section 4.5.2 (pourcentage d'erreur

Une méthode d'aide au choix d'un classifieur, basée sur ces critères, est proposée dans la section 7.1.3.

5) Paramètres des classifieurs.

La nature des paramètres des classifieurs et la méthode pour les obtenir dépendent du type de classifieur. Les principales techniques d'apprentissage ont été présentées dans la section "Méthodologie de conception d'un modèle d'apprentissage 4.5.

Enfin, l'hypothèse générale de la thèse "on peut utiliser les techniques d'intelligence artificielle pour évaluer un processus de préparation d'un modèle pour la simulation" devra être vérifiée. Des critères d'évaluation basés sur la qualité de la configuration des classifieurs et les résultats des prédictions devront être proposés.

La prédiction de la performance d'un processus de préparation sur un modèle complexe impose des contraintes à prendre en compte lors du développement d'une méthode d'apprentissage :

- Le nombre d'exemples de la base de données sera restreint, la méthode d'apprentissage devra être adaptée à cette contrainte.
- Les erreurs ou imprécisions de prédiction des variables intermédiaires risquent d'augmenter l'erreur sur le résultat final de la prédiction. Cette erreur induite devra être prise en compte lors de l'évaluation du classifieur.

Une méthode de configuration et d'évaluation des classifieurs basée sur les pratiques courantes en apprentissage est ainsi proposée dans la section 7.1.

Chapitre 6

Modélisation des données pour l'apprentissage

La première étape de l'approche proposée a pour objectif de construire une base de données pour l'apprentissage à partir de connaissances basées sur les processus de préparation de modèles pour la simulation. L'exploitation de connaissances pour leur ré-utilisation consiste tout d'abord à acquérir les connaissances puis à les formaliser et enfin à les implémenter dans des outils capables de les exploiter.

L'acquisition des connaissances consiste à identifier ou créer des exemples représentatifs de processus de préparation à partir de supports tels que des modèles CAO, des maillages, des rapports et des modèles de processus. Aujourd'hui, il n'existe pas de modélisation du processus de préparation d'un modèle pour la simulation tel qu'il a été défini dans la section 1.3. La première section 6.1 de ce chapitre propose donc une modélisation hiérarchique des processus de préparation sur plusieurs niveaux à l'aide de l'outil IDEF0.

La formalisation des données nécessaires à la description et à l'évaluation d'un processus de préparation est traitée dans la section 6.2 de ce chapitre. Il s'agira de définir la base de données comprenant les variables à prédire et l'ensemble des variables pouvant les caractériser.

Les variables explicatives, les plus déterminantes pour caractériser les variables à prédire sont sélectionnées selon une méthode proposée dans la section 6.3. Afin que tous les cas de préparation soient équitablement représentés dans la base d'exemples, une méthode de construction d'exemples pour l'apprentissage est proposée dans la dernière section 6.4.

CHAPITRE 6.1 Représentation des processus de préparation

6.1.1 Représentation globale d'un processus de préparation

La maquette numérique globale du modèle à préparer est constituée de N sous-ensembles (figure 1.5). Un processus de préparation de modèles pour la simulation numérique devra donc être décrit sur deux niveaux :

1. au niveau du modèle général M_i pour décrire l'ensemble du processus de préparation selon une configuration globale i ;
2. au niveau de chaque sous-ensemble SE_j^n pour décrire les sous-processus de simplification du composant n selon une configuration j .

Un processus (figure 6.1) devra être modélisé par des données d'entrée (modèle original), des données de sortie (modèle préparé) et des activités réalisées par des opérateurs de traitement définis par des outils et pilotées par des paramètres de réglage.

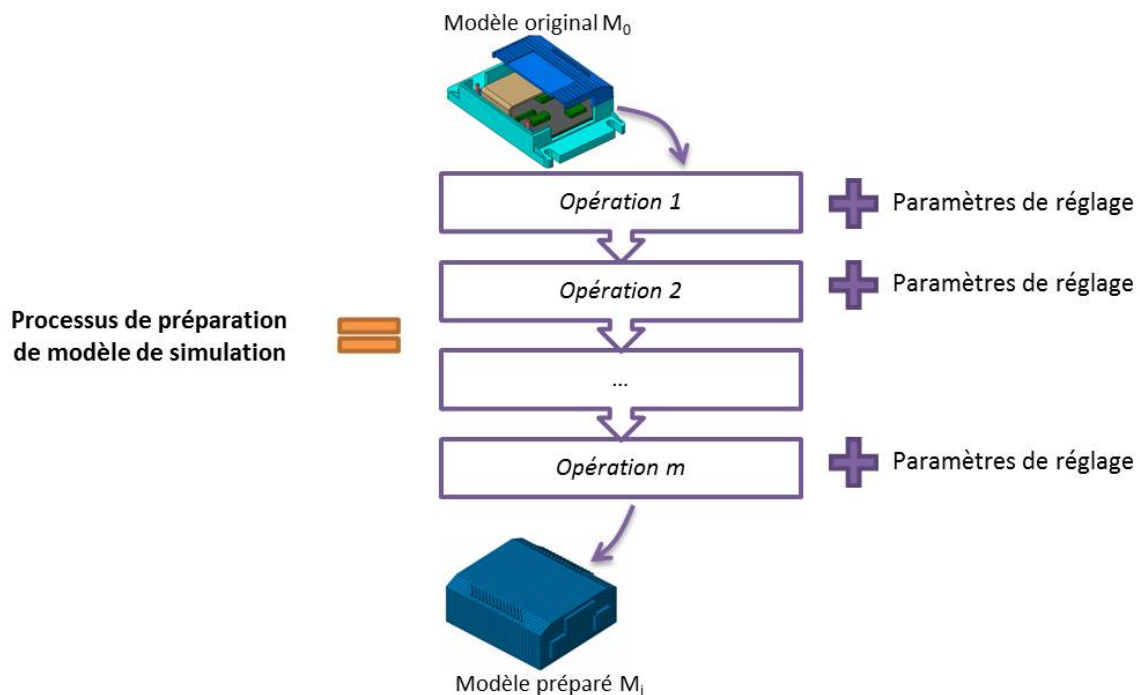


FIGURE 6.1 – Définition d'un processus de préparation de modèles pour la simulation

Nous proposons de caractériser le processus global de préparation par un actigramme de type IDEF0 de niveau A-0 (figure 6.2) dont les données d'entrée sont le modèle CAO original M_0 et les données de sortie une description du modèle M_i préparé pour la simulation.

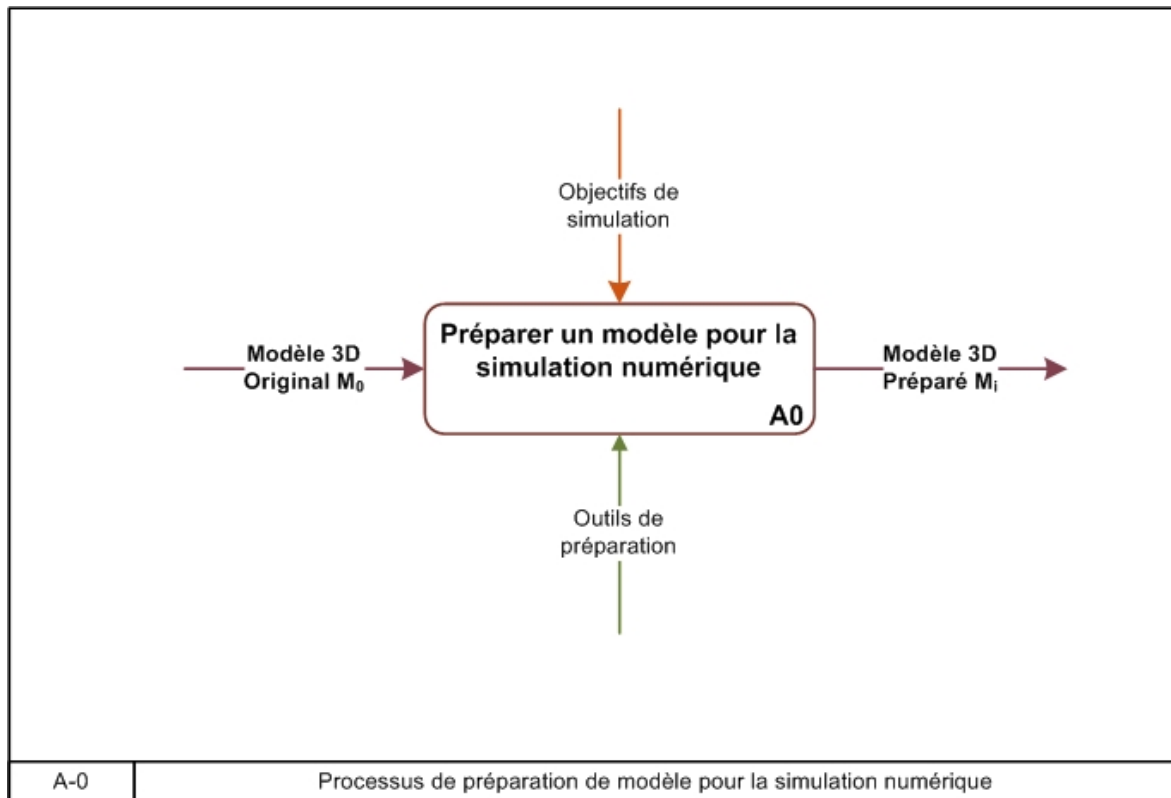


FIGURE 6.2 – Actigramme A-0. Processus de préparation d’un modèle pour la simulation numérique.

6.1.1.1 Définitions d’un modèle original et d’un modèle préparé

Modèle original

Le modèle CAO original M_0 (équation 6.1) est constitué des N sous-ensembles originaux SE_0^n nécessaires à la simulation et extraits de la maquette numérique complète tels qu’ils ont été définis dans la section 1.3.2, où f représente et caractérise la fonction d’assemblage qui lie ces sous-ensembles, il s’agit :

- des composants placés à la frontière du produit à simuler avec l’extérieur ;
- des sous-ensembles liés à une condition limite ;
- des sous-ensembles liés à une zone à observer ;
- des obstacles placés entre les sous-ensembles définis précédemment.

$$M_0 = f(SE_0^1, ..., SE_0^n, ..., SE_0^{Nse}) \quad (6.1)$$

La figure 6.3 montre un exemple de modèle original dont les quatre sous-ensembles ne sont pas simplifiés.

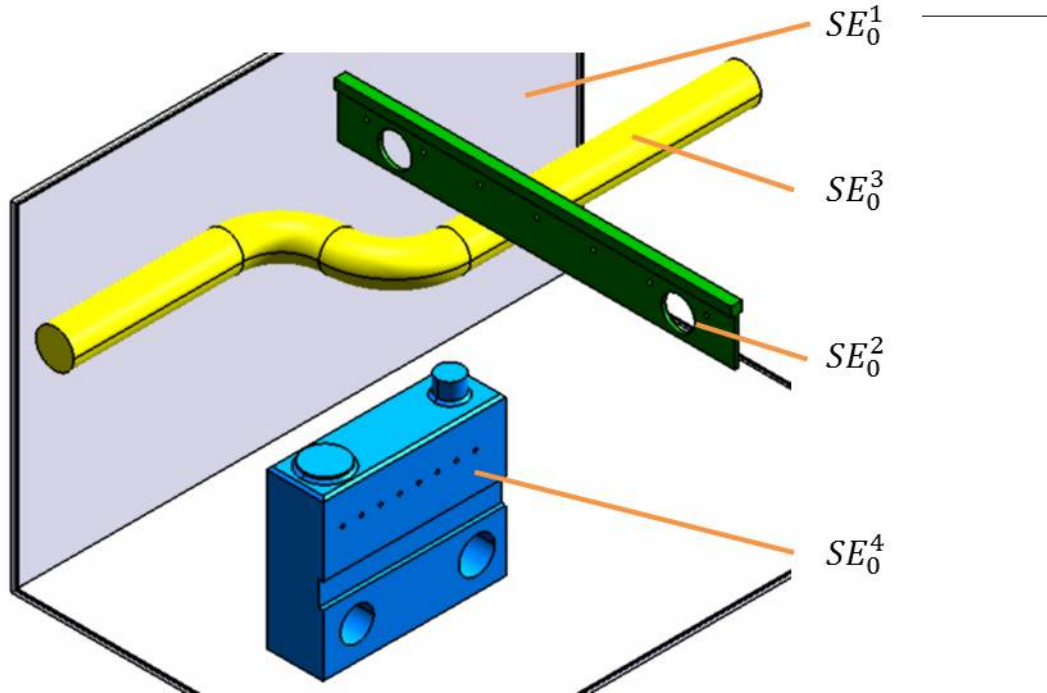


FIGURE 6.3 – Exemple de modèle original constitué de quatre sous-ensembles non simplifiés.

Modèle préparé

Le modèle préparé M_i selon la configuration globale i est le modèle de simulation. Il est constitué des Nse sous-ensembles SE_j^n simplifiés selon une configuration d'indice j (équation 6.11).

$$M_i = f \left(SE_{j1}^1, \dots, SE_{jn}^n, \dots, SE_{jNse}^{Nse} \right) \quad (6.2)$$

6.1.1.2 Modélisation hiérarchique d'un processus de préparation

L'actigramme A-0 est détaillé par un actigramme A0 (figure 6.4) qui comprend les trois activités principales de simplification (A1), d'adaptation (A2) et de maillage (A3) qui seront détaillées dans les sections suivantes.

Chaque activité est décrite par des données d'entrée, des données de sortie, des données de contrôle et des moyens.

Les données d'entrée et de sortie des activités sont les états du modèle 3D avant et après les opérations de traitement, complétés par les informations extraites de la maquette numérique (matériaux, caractéristiques physiques, fonction d'un composant...).

Les données de contrôle peuvent être :

- des informations relatives à l'objectif de la simulation (grandeur mesurée, position et caractéristiques des conditions limites, niveau de simplification souhaité,...) ;

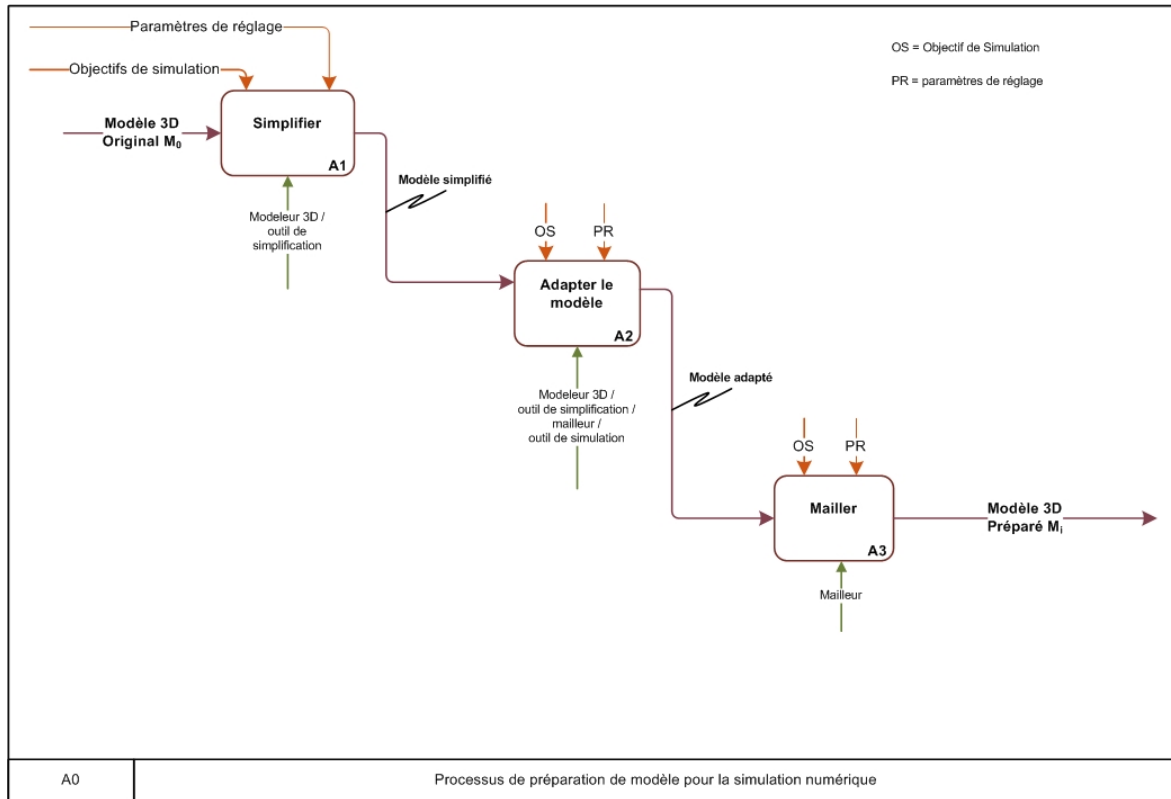


FIGURE 6.4 – Actigramme A0. Processus détaillé de préparation d'un modèle pour la simulation numérique.

— les paramètres de réglage de l'opération.

Les moyens utilisés pour réaliser les opérations sont le type d'outils mis en œuvre (modeleur 3D, logiciel de simplification, mailleur ou logiciel de simulation).

6.1.2 Simplification de la géométrie

Dans la cadre de ces travaux de thèse, six opérations de simplification ont été retenues (le filtrage, le defeaturing, la création d'enveloppe, la décimation, la substitution et l'agrégation) pour préparer un modèle CAO pour la simulation numérique. Ces opérations ont été décrites dans le chapitre 2.

Elles sont appliquées indépendamment sur chacun des N_{se} sous-ensembles constituant le modèle global.

Chaque opération de simplification op sur le sous-ensemble SE_{jn}^n est décrite par une activité $A1.op.Nn$ (figure 6.5). On appellera "sous-processus de simplification" l'ensemble des activités $A1.Nn$ (équation 6.3) sur un sous-ensemble SE_{jn}^n , où Nop représente le nombre d'opérations pour le sous-ensemble n .

$$\{A1.Nn\} = \{A1.1.Nn, A1.2.Nn, ..., A1.Nop.Nn\} \quad (6.3)$$

CHAPITRE 6. MODÉLISATION DES DONNÉES POUR L'APPRENTISSAGE

La première sous-étape (figure 6.6) consiste à identifier les paramètres de réglage tels qu'ils ont été définis dans le chapitre 2 à partir des caractéristiques du modèle CAO avant la simplification et des objectifs de la simulation.

Le modèle est ensuite simplifié puis reconstruit si la simplification a engendré des défauts dans la géométrie du modèle (par exemple la rupture de connexion entre des pièces ou des surfaces, ou la suppression d'un composant parent,...).

L'ordre des opérations peut varier d'un sous-processus de simplification à un autre. Des exemples de séquençages possibles obtenus à partir des six opérations de simplification identifiées dans le chapitre 2 sont décrits dans la section 6.1.5.

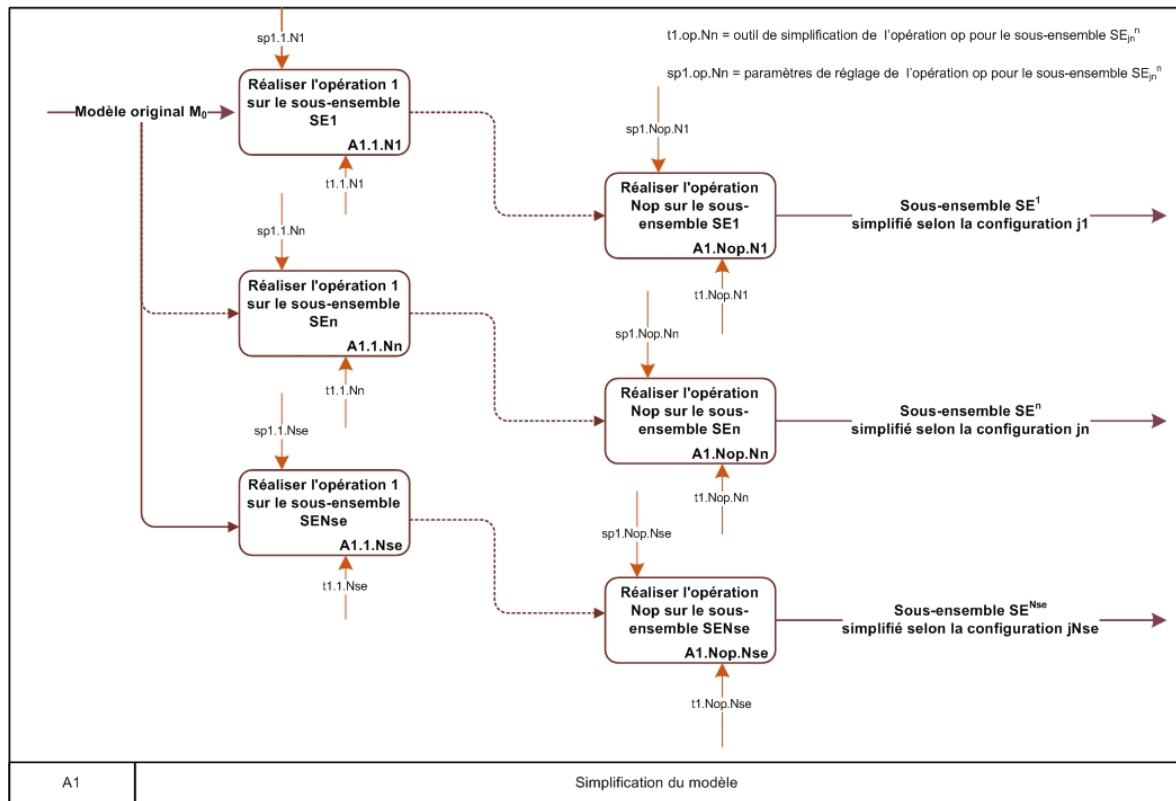


FIGURE 6.5 – Actigramme A1. Simplification du modèle pour la simulation numérique.

6.1.3 Adaptation

Cette étape (figure 6.7) consiste à adapter le modèle géométrique simplifié au modèle à mailler. La première activité (A21) définit les éléments (vertex, arrêtes, surfaces ou volumes) qui supportent les conditions limites à partir des objectifs de la simulation. Ces éléments peuvent être simplement identifiés à l'aide des méta-données extraites de la maquette numérique ou bien créés à l'aide d'un modeleur 3D ou d'un logiciel de simulation.

Dans certains cas, le maillage est appliqué sur un ensemble de surfaces bordant le modèle géométrique global (figure 1.8). La faisabilité du maillage et de la simulation dépendent

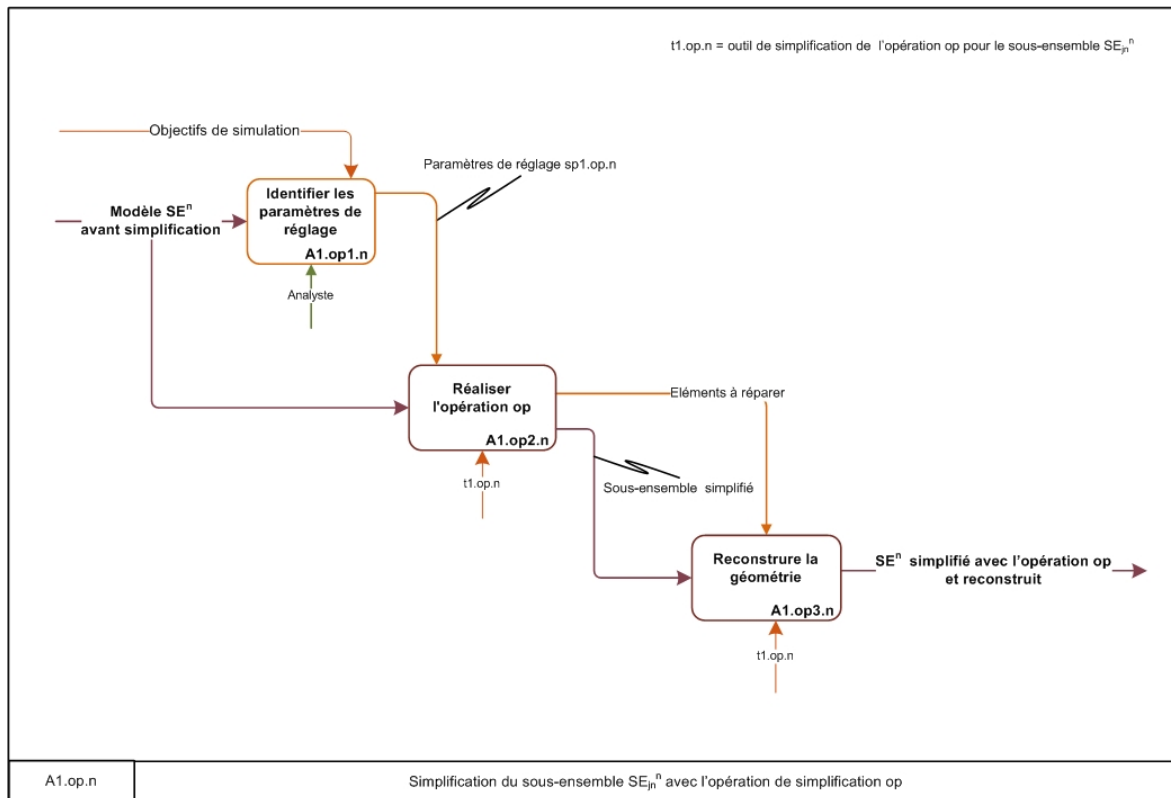


FIGURE 6.6 – Actigramme A1.m.n. Simplification d'un sous-ensemble SE^n par l'opération m .

de la qualité de cet ensemble de surfaces. Il est alors nécessaire d'extraire (activité A22) ces surfaces puis éventuellement de reconstruire le modèle (activité A33).

Le coût de cette opération peut varier considérablement en fonction des outils utilisés, aujourd'hui certains logiciels de simulation réalisent l'ensemble de ces opérations rapidement et de façon automatisée.

6.1.4 Maillage

La première activité de l'étape de maillage A31 (figure 6.8) consiste à définir le type d'éléments maillés et la carte de maillage (taille et forme des éléments, zones de raffinement, orientation des éléments,...) à partir des objectifs de la simulation afin que la qualité du maillage soit satisfaisante selon les critères énoncés dans la section 3.2, et que l'erreur sur le résultat de l'analyse soit minimale pour des durées optimisées de maillage et de simulation.

Le maillage est alors appliqué à l'ensemble des surfaces ou volumes définis dans l'étape précédente (activité A32).

Après évaluation de la qualité et du coût du maillage (activité A33), des boucles d'amélioration peuvent être nécessaires pour obtenir un maillage satisfaisant.

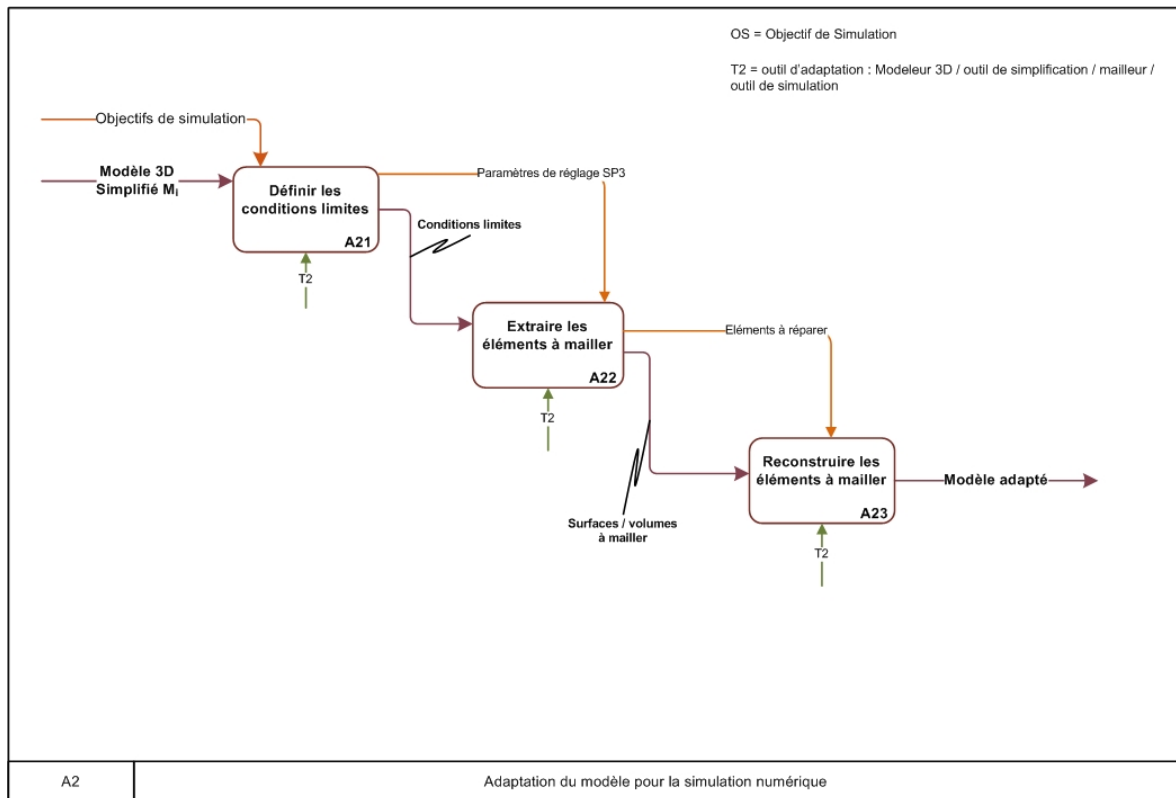


FIGURE 6.7 – Actigramme A2. Adaptation du modèle simplifié au maillage.

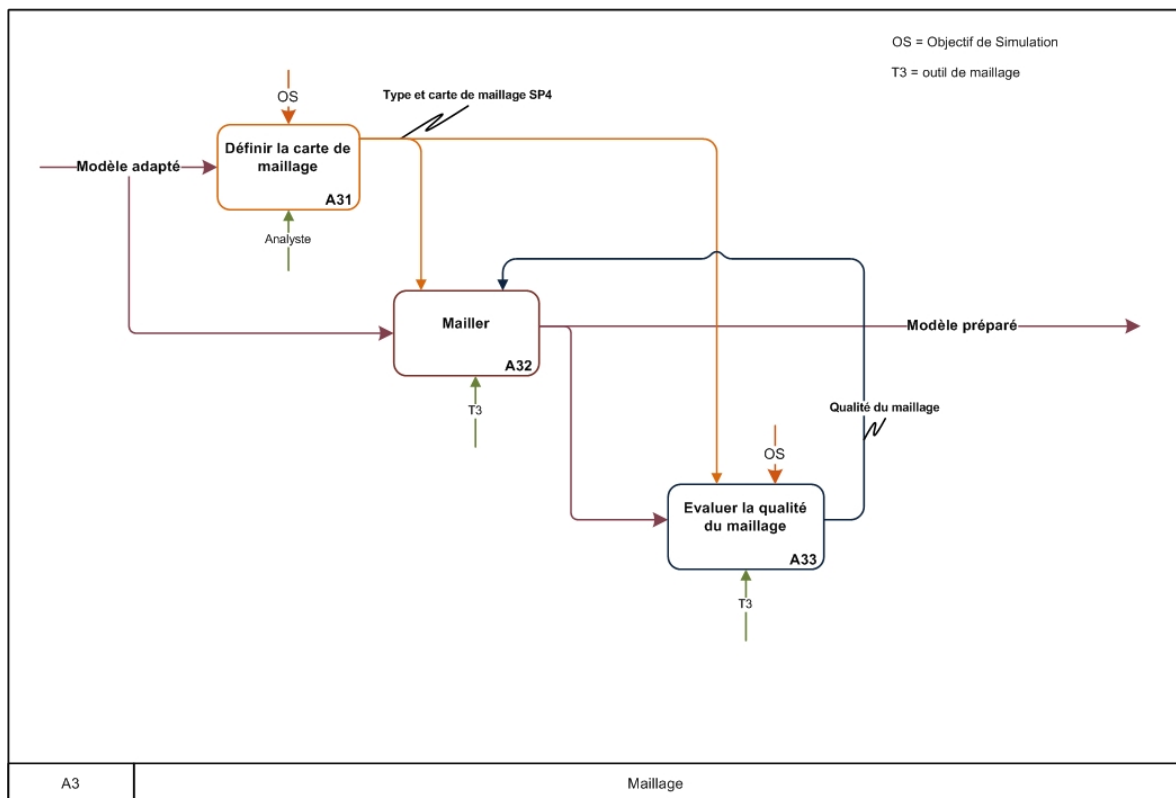


FIGURE 6.8 – Actigramme A3. Maillage.

6.1.5 Séquençage des opérations

Les Nn opérations de simplifications ne sont pas toutes utilisées pour simplifier un sous-ensemble SE_j^n . De plus, ces opérations peuvent être utilisées dans des ordres différents. En théorie, il existe alors un grand nombre de sous-processus de simplification possibles pour un sous-ensemble.

Dans la pratique, tous les arrangements ne sont pas réalistes. Par exemple, le tableau 6.1 donne les antériorités possibles entre les six opérations identifiées dans le chapitre 2. Certaines opérations doivent être menées après d'autres pour faciliter leur réalisation (par exemple le defeaturing doit être réalisé avant la décimation mais après le filtrage). Certaines opérations peuvent être inutiles si elles sont réalisées en complément d'une autre (par exemple le defeaturing est inutile si le modèle est reconstruit par substitution). Si on analyse à partir de cette matrice l'ensemble des combinaisons réalistes d'un point de vue industriel, on peut dénombrer jusqu'à 4 enchainements d'opérations et 40 séquençages différents (tableau A.1 en annexe) pour chaque sous-ensemble SE^n .

↗	Filtrage	Defeating	Décimation	Enveloppe	Substitution	Agrégation
Filtrage		Avant	Avant	Avant	X	Avant
Defeating	Après		Avant	Avant	X	Av/Ap
Décimation	Après	Après		Avant	X	Av/Ap
Enveloppe	Après	Après	Après		X	Av/Ap
Substitution	X	X	X	X		Av/Ap
Agrégation	Après	Av/Ap	Av/Ap	Av/Ap	Av/Ap	

TABLEAU 6.1 – Matrice d'antériorité des opérations de simplification. **Avant** : l'opération (ligne) doit être réalisée avant l'opération (colonne). **Après** : l'opération (ligne) doit être réalisée après l'opération (colonne). **Av/Ap** : l'opération (ligne) peut être réalisée indifféremment avant ou après l'opération (colonne). **X** : les deux opérations ne sont pas utilisées sur un même composant.

6.1.6 Conclusion sur la représentation des processus de préparation

Une représentation hiérarchique des processus de préparation de modèles a été proposée dont les données d'entrée sont les conditions de simulation et le modèle CAO original, et les données de sortie le modèle préparé pour la simulation complété par des critères d'évaluation du processus.

Au niveau de la préparation globale d'un ensemble, le processus de préparation peut être représenté par des macro activités hiérarchisées. Le séquençage global proposé "simplification puis adaptation puis maillage" est celui qui est le plus logique actuellement selon les possibilités offertes par les outils de simplification, de maillage et de simulation.

Au niveau local de la simplification d'un sous-ensemble, le nombre, le type et l'ordre des opérations réalisées varient d'un sous-ensemble à un autre. Lorsque l'ensemble global

~~CHAPITRE 6. MODÉLISATION DES DONNÉES POUR L'APPRENTISSAGE~~
est riche en sous-ensembles, une représentation hiérarchisée du sous-processus de simplification n'est plus adaptée. Il sera proposé dans la section suivante une modélisation vectorielle des processus de préparation et sous-processus de simplification. En plus de résoudre le problème de la représentation d'une multitude de processus, cela facilitera l'implémentation des données descriptives des processus dans les outils d'apprentissage artificiel.

6.2 Modélisation des données pour la description et l'évaluation d'un processus de préparation

Cette section a pour objectif de définir les notions de processus de préparation d'un modèle au niveau global de l'ensemble complet et au niveau local des sous-ensembles, puis d'en proposer une modélisation.

6.2.1 Processus global de préparation et sous-processus local de simplification

Au niveau global (figure 6.9 (a)), il s'agit de décrire les processus de préparation ainsi que d'identifier les facteurs qui caractérisent les critères d'évaluation du processus de préparation. Les données décrivent alors :

- la géométrie du modèle complet (modèles CAO originaux, simplifiés, adaptés et maillés),
- les processus de simplification, d'adaptation et de maillage du modèle complet,
- l'impact global de la simplification,
- la qualité du processus de préparation.

Au niveau local (figure 6.9 (b)), il s'agit de décrire les sous-processus de simplification d'un sous-ensemble ainsi que les facteurs qui caractérisent l'impact de la simplification sur le résultat et le coût de l'analyse. Les données décrivent alors :

- la géométrie du sous-ensemble (modèles CAO originaux, simplifiés),
- le sous-processus de simplification,
- l'influence de la simplification sur l'analyse
- la qualité du sous-processus.

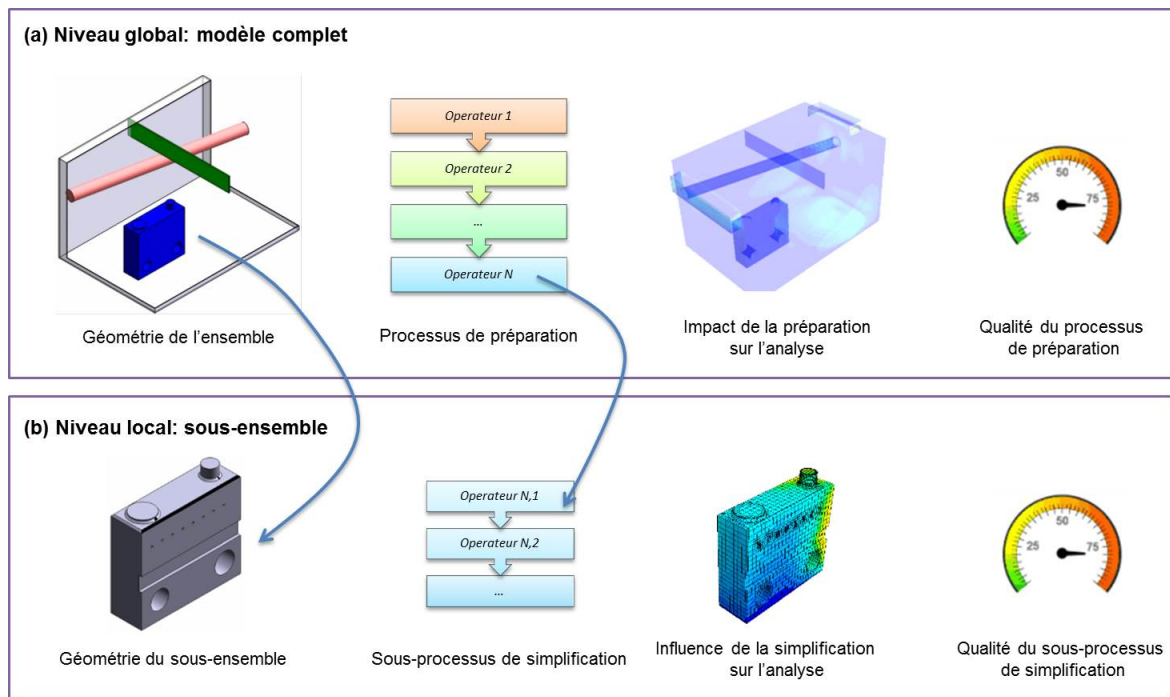


FIGURE 6.9 – Niveaux de représentation. (a) Niveau global pour la description du modèle complet et pour l'évaluation du processus de préparation. (b) Niveau local pour la description du sous-ensemble et pour l'évaluation du sous-processus de simplification.

6.2.2 Formalisation de la représentation des données

La base de données construite dans la première phase de l'approche proposée (figure 5.4) doit contenir l'ensemble des facteurs qui décrivent un processus de préparation et qui caractérisent ses critères d'évaluation. L'objectif de cette section est de proposer une représentation formalisée des données extraites des supports de connaissance (figure 6.10).

La modélisation proposée devra respecter un ensemble de contraintes.

- *Contrainte 1.* La représentation des facteurs décrivant la géométrie des modèles devra être unique quelle que soit la géométrie du modèle, le nombre de sous-ensembles, le nombre de pièces ou le nombre de détails du modèle.
- *Contrainte 2.* Les données seront implémentées ultérieurement dans des outils d'intelligence artificielle de type classifieurs pour lesquels chaque exemple sera décrit par un vecteur de variables.
- *Contrainte 3.* Pour un objectif de préparation donné, les vecteurs qui décrivent le processus de préparation et qui caractérisent ses critères d'évaluation doivent être de dimension finie quel que soit le processus choisi, l'outillage et le cas de simulation.
- *Contrainte 4.* Les variables utilisées pour définir les paramètres de réglage des opérations ne doivent pas dépendre du type d'outil utilisé.

Afin de respecter la contrainte N°1 les données seront représentées sur les deux niveaux

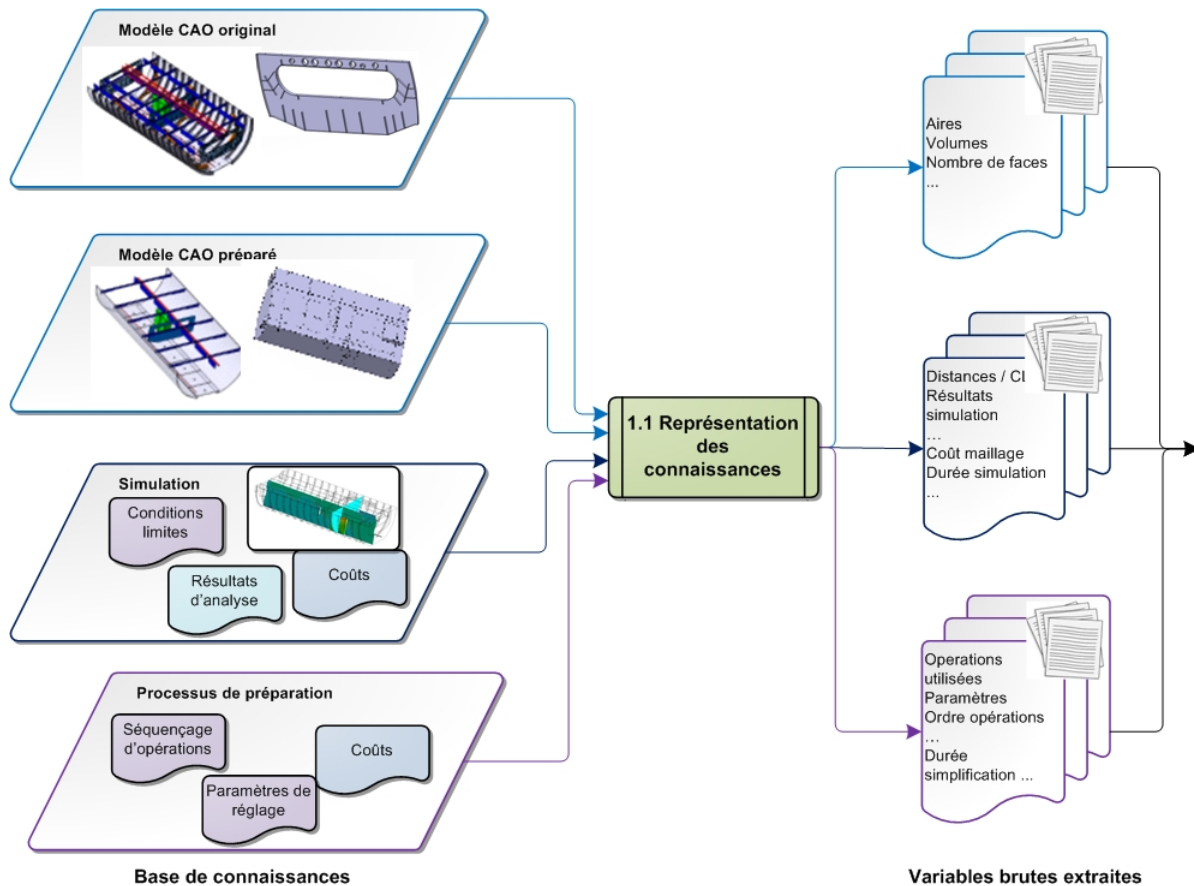


FIGURE 6.10 – Etape 1 construction de la base de données pour l'apprentissage. 1.1 Représentation formalisée des connaissances.

présentés précédemment.

La représentation des données sera matricielle (contrainte N°2). Une ligne correspond soit à un exemple d'ensemble global préparé selon un processus (matrice globale) soit à un exemple de sous-ensemble simplifié (matrice locale). Pour chaque exemple, la matrice contient un même nombre de variables à prédire y (local) ou Y (global) et un même nombre de variables explicatives (contraintes 3 et 4) qui décrivent le processus de préparation ou qui caractérisent ses critères d'évaluation. Ces variables explicatives peuvent être connues pour un nouveau cas (x/X) ou intermédiaires (xy/XY) si elles sont inconnues pour un nouveau cas et devront être estimées.

Les figures 6.11 et 6.12 montrent la structure du contenu des deux matrices qui contiennent :

- **les données décrivant un sous-processus de simplification (local) ou un processus de préparation (global)** en précisant quels sont les opérateurs utilisés, les paramètres de réglages, l'ordonnancement des opérations et les outils utilisés. Ces variables sont connues pour un nouveau cas.
- **les données décrivant les modèles originaux, simplifiés, adaptés et maillés** en précisant les caractéristiques géométriques des modèles (taille et forme) complétés

posants, grandeur physique appliquée, condition limite...). Seule la description du modèle original est connue pour un nouveau cas.

- **les facteurs de comparaison des modèles** simplifiés ou préparés par rapport aux modèles originaux, ces facteurs caractérisent le niveau de simplification du modèle. Ils ne sont pas connus pour un nouveau cas.
- **les facteurs d'influence de la préparation sur le cas de simulation**, ces facteurs quantifient les modifications géométriques dues à la simplification en tenant compte des distances et positions des composants simplifiés par rapports aux conditions limites ou aux cibles de l'analyse (éléments à observer). Ils ne sont pas connus pour un nouveau cas.
- **les critères d'évaluation de la qualité du processus de simplification ou de préparation** (coûts de préparation, coût de simplification et erreur sur le résultat de l'analyse). Ce sont les principales variables à prédire.
Ces variables sont détaillées dans les sections suivantes.

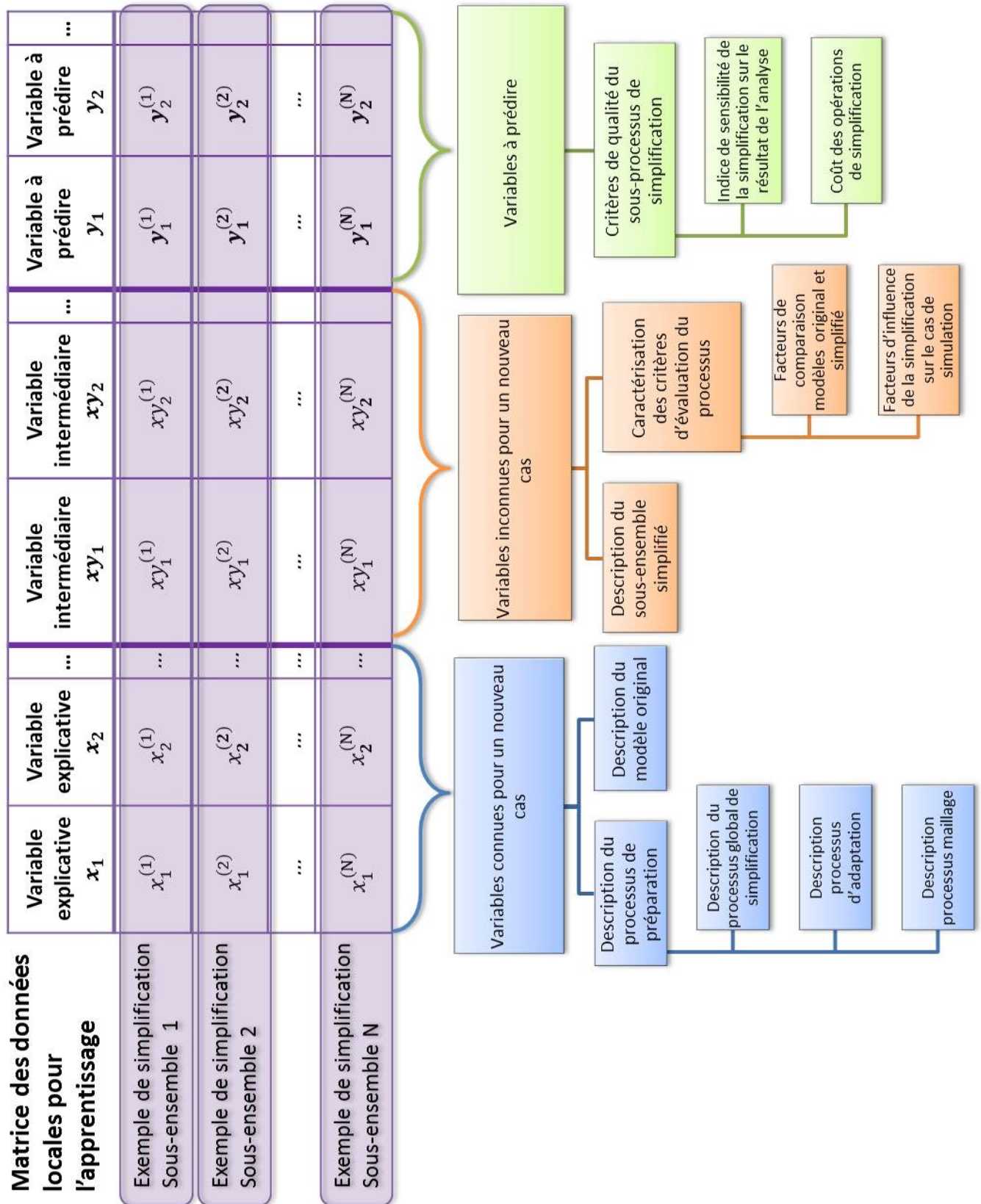


FIGURE 6.11 – Contenu de la matrice de données locale au niveau des sous-ensembles.

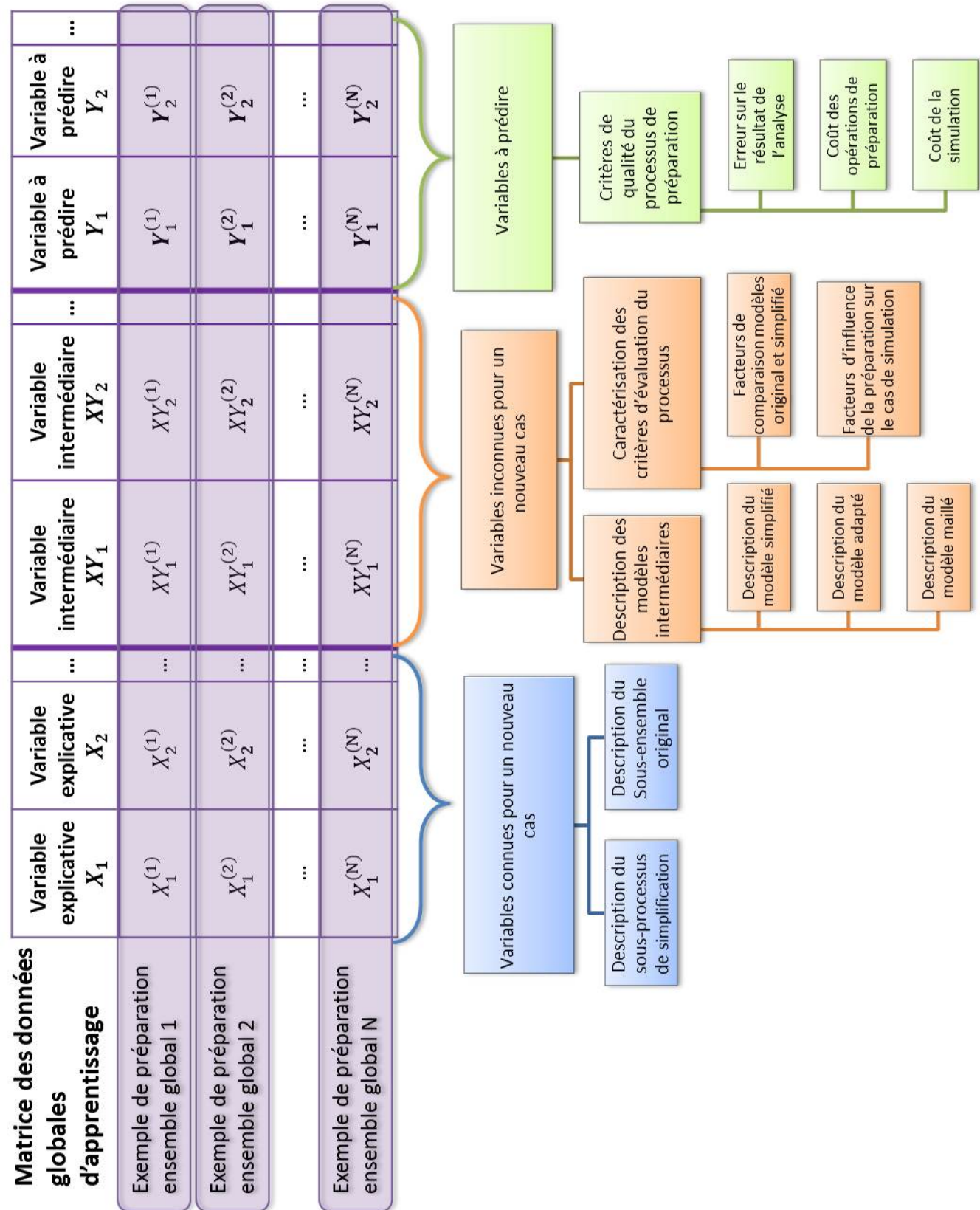


FIGURE 6.12 – Contenu de la matrice de données globale au niveau de l'ensemble complet.

6.2.3 Variables à prédire : critères d'évaluation de la qualité d'un processus de préparation

Les principales variables à prédire sont les valeurs des critères nécessaires à l'évaluation de la qualité d'un processus de préparation d'un modèle pour la simulation. Ces variables sont calculées pour l'apprentissage et estimées pour un nouveau cas.

6.2.3.1 Critères de qualité du sous-processus de simplification

La qualité relative du sous-processus de simplification d'un sous-ensemble est évaluée à l'aide du coût de la simplification et de l'influence de la simplification sur le résultat de l'analyse. Ces deux critères sont des variables intermédiaires qui serviront au final à estimer la qualité du processus global de préparation.

- **Critère 1 : influence de la simplification sur le résultat de l'analyse.** Cet indice IS_j^n est calculé pour l'apprentissage à partir des résultats $R(SE_{simplif.maxi}^n)$ de l'analyse sur les grandeurs calculées lorsque le sous-ensemble n est simplifié au maximum (c'est à dire substitué par un modèle très simple de formes cylindriques ou parallélépipédiques) et des résultats $R(SE_j^n)$ lorsqu'il est simplifié selon la configuration j (équation 6.4).

Ce critère permet d'identifier les processus de simplification qui peuvent mettre en échec la simulation (en rendant impossible les phases d'adaptation ou de maillage) et de comparer plusieurs processus de simplification sur un même sous-ensemble.

$$IS_j^n = \frac{R(SE_j^n) - R(SE_{simplif.maxi}^n)}{R(SE_{simplif.maxi}^n)} \quad (6.4)$$

- **Critère 2 : coût de la simplification.** Le coût CS_j^n (équation 6.5) de la simplification d'un sous-ensemble n est calculé à partir du coût $CS_{op,j}^n$ des Nn opérations de simplification selon des règles proposées par l'utilisateur en prenant en compte par exemple la durée des traitements, les coûts d'investissement et de fonctionnement de l'outillage ou bien le nombre d'actions manuelles.

$$CS_j^n = \sum_1^{Nn} (CS_{op,j}^n) \quad (6.5)$$

La qualité d'un processus global de préparation d'un modèle pour la simulation sera décrit à partir de trois variables à prédire qui sont l'erreur sur le résultat de l'analyse, le coût total de la préparation et le coût de la simulation.

- **Critère 1 : erreur sur le résultat de l'analyse.** Ce critère ERA est calculé pour l'apprentissage (équation 6.6) à partir du résultat $R(M_i)$ de l'analyse obtenu avec le modèle simplifié M_i et du résultat $R(M_{ref})$ obtenu avec le modèle de référence (modèle simulable le moins simplifié possible défini dans la section 6.4.2).

$$ERA_i = \frac{R(M_i) - R(M_{ref})}{R(M_{ref})} \quad (6.6)$$

- **Critère 2 : coût de la préparation.** Ce critère CP (6.7) est calculé pour l'apprentissage à partir du coût des activités de simplification (CP1), d'adaptation (CP2) et de maillage (CP3). Le coût total des activités de simplification CP1 est obtenu à partir du coût pour chacun des N sous-ensembles.

$$CP = CP1 + CP2 + CP3 = \sum_{n=1}^{Nse} (CS_j^n) + CP2 + CP3 \quad (6.7)$$

- **Critère 3 : coût de la simulation.** Ce critère CA est calculé pour l'apprentissage à partir de la durée de l'analyse.

L'objectif de cette section est de proposer une modélisation formalisée des processus de préparation sur un ensemble global et des sous-processus de simplification sur un sous-ensemble. Les notations utilisées dans cette section pour décrire un processus de préparation et un sous-processus de simplification sont données ci-dessous.

Notations pour la modélisation d'un processus de préparation (niveau global) :

- M_i^m : modèle de l'ensemble global m préparé selon la configuration d'indice i
- P_i^m : ensemble de vecteurs décrivant le processus global de préparation de l'ensemble m préparé selon la configuration d'indice i
- Processus de simplification :
 - $SP1_i^m$: ensemble de vecteurs décrivant le processus de simplification de l'ensemble
 - $SP1.x^m$: vecteur décrivant l'opération x de simplification globale
- Processus d'adaptation :
 - $SP2_i^m$: ensemble de vecteurs décrivant le processus d'adaptation de l'ensemble
 - $SP2.x^m$: vecteur décrivant la caractéristique x du processus d'adaptation
 - $T2^m$: vecteur décrivant l'outillage d'adaptation
- Processus de maillage :
 - $SP3_i^m$: ensemble de vecteurs décrivant le processus de maillage de l'ensemble
 - $SP3.x^m$: vecteur décrivant la caractéristique x du processus de maillage
 - $T3^m$: vecteur décrivant l'outillage de maillage

Notations pour la modélisation d'un sous-processus de simplification (niveau local) :

- SE_j^n : modèle du sous-ensemble N^n simplifié selon la configuration d'indice j
- $sp1_j^n$: ensemble de vecteurs décrivant le sous-processus de simplification du sous-ensemble
- $op1_x^n$: vecteur décrivant l'opération de simplification " x " du sous-ensemble
- cs_x^n : vecteur décrivant le coût de la simplification pour l'opération " x " du sous-ensemble
- $t1_x^n$: vecteur décrivant l'outillage de la simplification pour l'opération " x " du sous-ensemble

Le processus de préparation P_i^m d'un modèle M pour un objectif de simulation est représenté par un ensemble de vecteurs qui décrit (figure 6.12) le processus global de simplification $SP1_i^m$, le processus d'adaptation $SP2_i^m$ et le processus de maillage $SP3_i^m$.

$$P_i^m = \{SP1_i^m, SP2_i^m, SP3_i^m\}$$

6.2.4.1 Variables pour la représentation du sous-processus de simplification sur un sous-ensemble

Avant de définir le processus global de simplification $SP1_i^m$ sur l'ensemble du modèle, il est nécessaire de décrire le sous-processus de simplification appliqué à chaque sous-ensemble.

Afin de respecter les contraintes énoncées précédemment, le sous-processus de simplification d'un sous-ensemble n selon une configuration d'indice j est modélisé par un vecteur $sp1_j^n$ (équation 6.8) de dimension finie.

Si on combine les séquençages possibles des opérations de simplification et les paramètres de réglage, il existe un nombre infini de sous-processus de simplification. Ceux-ci devront être ramenés dans un espace fini.

La représentation des processus devra être unique quel que soit le processus choisi parmi les 40 séquençages possibles discutés dans la section précédente.

$$sp1_j^n = \{op1_1^n, \dots, op1_{14}^n, t1_1^n, \dots, t1_{14}^n\} \quad (6.8)$$

Les quatorze premières composantes ($op1_1^n, \dots, op1_{14}^n$), détaillées dans le tableau 6.2 et les paragraphes suivants, décrivent les opérations à réaliser. Afin de respecter la contrainte N°4, ces composantes décrivent les opérations à réaliser sans préciser la fonction utilisée par l'outil et en généralisant les paramètres de réglage.

Les composantes suivantes ($t1_1^n, \dots, t1_{14}^n$) indiquent le type d'outillage $t1_x^n$ utilisé pour chaque opération $op1_x^n$.

La figure 6.13 montre des exemples de simplifications d'un sous-ensemble dont les sous-processus de simplification sont décrits sur le tableau 6.3 et les paragraphes suivants.

Description de l'opération de filtrage

Le paramètre de réglage à définir est la liste des pièces à supprimer. Afin de respecter la contrainte N°1, un nombre ou une liste de pièces ne peuvent être utilisés directement. Pour décider si une pièce doit être supprimée ou conservée, l'analyste

Opération	Paramètres de réglage	GE
Filtrage	Taille : $op1_1^n = Réel \in \{0, 1\}$ Distance CL : $op1_2^n = Réel \in \{0, 1\}$	
Defeaturing	Pour chaque type de détail (trou, poche, bossage, congé/chanfrein) : Taille : $op1_3^n, \dots, op1_6^n = Réel \in \{0, 1\}$ Distance CL : $op1_7^n, \dots, op1_{10}^n = Réel \in \{0, 1\}$	
Décimation	Etat décimation : $op1_{11}^n = \{No, Yes, After - def\}$	
Enveloppe	Etat enveloppe : $op1_{12}^n = \{No, Yes\}$	
Substitution	Etat substitution : $op1_{13}^n = \{No, Cyl, Par\}$	
Agrégation	Etat agrégation : $op1_{14}^n = \{No, After, Before\}$	

TABLEAU 6.2 – Description des sous-processus de simplification d'un sous-ensemble. Composantes du vecteur $sp1_j^n$ (opérations).

Simplification	SE	Seq	op1	op2	op3	op4	op5	op6	op7	op8	op9	op10	op11	op12	op13	op14
Aucune	SE ₀ ²	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Substitution	SE ₁ ²	5	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Enveloppe	SE ₂ ²	4	0	0	0	0	0	0	0	0	0	0	0	1	0	0
Defeaturing petits trous (1)	SE ₃ ²	2	0	0	0,2	0	0	0	1	0	0	0	0	0	0	0
Defeaturing bossages	SE ₄ ²	2	0	0	0	0	1	0	0	0	1	0	0	0	0	0
Defeaturing petits détails	SE ₅ ²	2	0	0	0,2	0	1	0	1	0	1	0	0	0	0	0
Décimation	SE ₆ ²	3	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Defeaturing + décimation	SE ₇ ²	12	0	0	0,2	0	0	0	1	0	0	0	1	0	0	0

TABLEAU 6.3 – Description des J = 8 sous-processus de simplification du sous-ensemble (SE) N°2 (figure 6.13). Seq = référence du séquençage (annexe A)(1) Tous les trous dont la longueur moyenne est inférieure à 0,2 fois la longueur moyenne du modèle global sont supprimés (quelle que soit leur distance par rapport aux conditions limites).

se base sur sa taille et sa distance par rapport aux conditions limites les plus proches (cette distance sera nulle si la pièce est liée à une condition limite).

On notera $Lmoy^m$ la longueur du modèle global m (moyenne des trois longueurs de la boîte englobante du modèle) auquel appartient la pièce.

L'opération de filtrage sera donc définie à l'aide de deux composantes dont les valeurs x sont proportionnelles à $Lmoy^m$:

- composante **taille** $op1_1^n$: les pièces dont la taille est inférieure à $x * Lmoy^m$ sont supprimées ;
- composante **distance par rapport aux conditions limites** $op1_2^n$: les pièces dont la distance par rapport à la condition limite la plus proche est supérieure à $x * Lmoy^m$ sont supprimées.

Les valeurs de ces deux composantes sont des réels compris entre 0 et 1 (la longueur

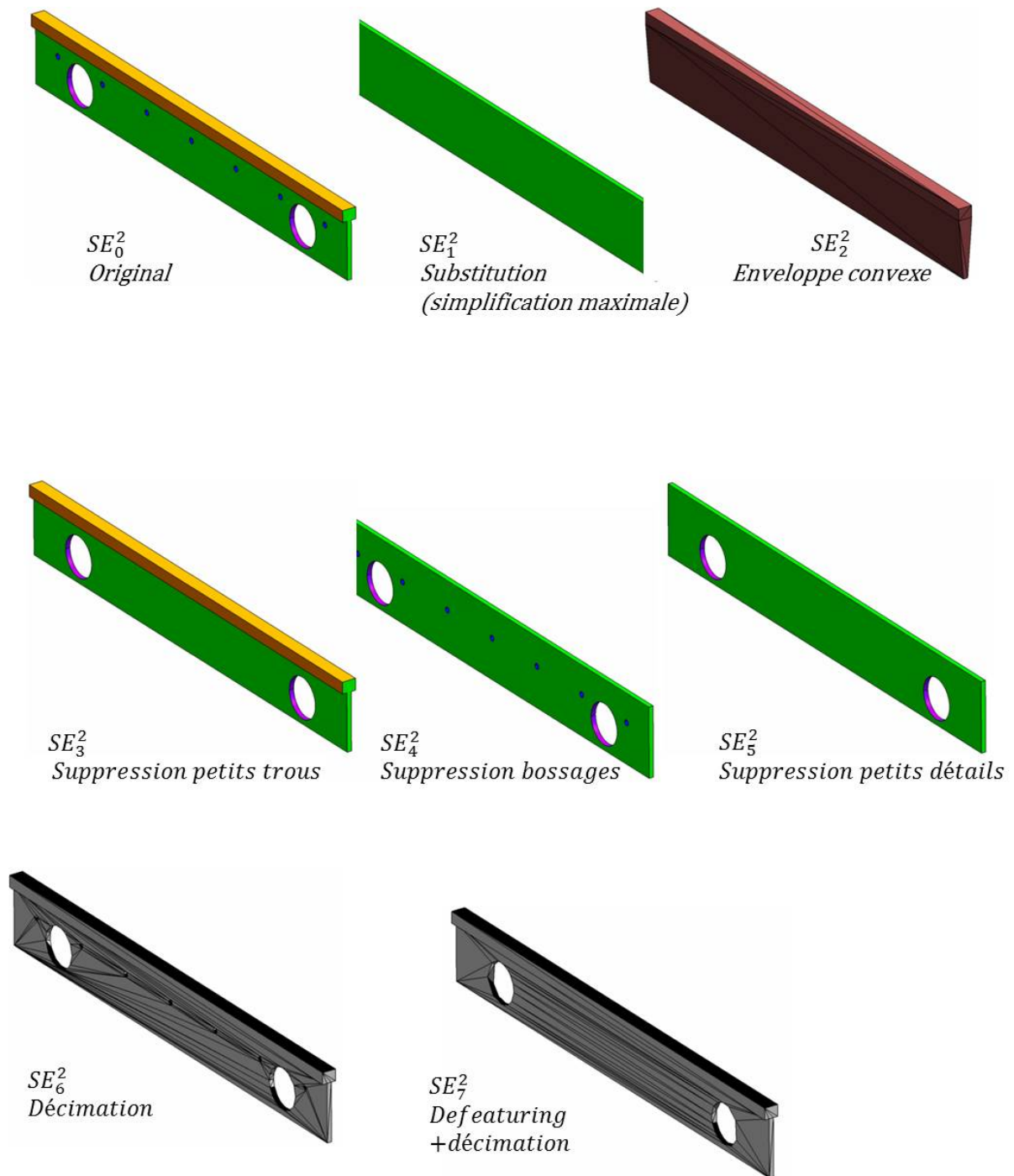


FIGURE 6.13 – Exemples de sous-processus j du sous-ensemble N°2.

moyenne d'une pièce ne peut être supérieure à la longueur moyenne de l'ensemble global), avec :

- $x = 0$ si aucune pièce n'est supprimée ;
- $x = max$ si toutes les pièces sont supprimées.

Description de l'opération de defeaturing

Le paramètre de réglage de l'opération de defeaturing est la liste des détails supprimés. Comme pour le filtrage de pièces, ce paramètre ne peut être utilisé direc-

primé ou non, l'analyste se base sur sa taille, sa distance par rapport aux conditions limites et son type (trou, poche, bossage ou congé/chanfrein).

On notera L_{moy}^n la longueur du sous-ensemble auquel appartient le détail. L'opération de defeaturing sera décrite par deux composantes pour chaque type de détail (tableau 6.4) :

- composantes **taille** $op1_3^n$ à $op1_6^n$: les détails dont la taille est inférieure à $x * L_{moy}^n$ sont supprimés ;
- composantes **distance par rapport aux conditions limites** $op1_7^n$ à $op1_{10}^n$: les détails dont la distance par rapport à la condition limite la plus proche est supérieure à $x * L_{moy}^n$ sont supprimés.

Composante	Trou	Poche	Bossage	Congé/chanfrein
Taille	$op1_3^n$	$op1_4^n$	$op1_5^n$	$op1_6^n$
Distance / CL	$op1_7^n$	$op1_8^n$	$op1_9^n$	$op1_{10}^n$

TABEAU 6.4 – Composantes décrivant l'opération de defeaturing

Les valeurs de ces composantes sont des réels compris entre 0 et 1, avec :

- $x = 0$ si aucun détail n'est supprimé ;
- $x = 1$ si tous les détails sont supprimés quelle que soit leur taille et leur distance par rapport aux conditions limites.

Description de l'opération de décimation

Les outils commerciaux proposent des paramètres de réglage pour l'opération de décimation qui permettent d'ajuster le niveau de simplification. Dans la pratique, les analystes règlent ces paramètres afin d'obtenir un niveau de simplification maximal sans dégénérer le modèle.

Si des opérations de defeaturing précèdent la décimation, le modèle pourra être simplifié à un niveau plus important sans erreur. Il sera donc important de préciser si une opération de defeaturing précède la décimation.

Nous proposons de définir l'opération de décimation d'un sous-ensemble par une composante $op1_{11}^n$ qui peut prendre trois valeurs :

- **"No"** : lorsque le sous-ensemble n'est pas décimé ;
- **"After def"** : lorsque le sous-ensemble est décimé après defeaturing ;
- **Yes** : lorsque le sous-ensemble est décimé sans (ou avant) le defeaturing.

Description de l'opération de création d'enveloppe

Réalisée par un outil commercial, l'opération de création d'enveloppe est souvent de type "boite noire". Le résultat et le niveau de simplification obtenus dépendent de l'outil et de la fonction utilisée.

Cette opération pourrait alors être définie par une composante $op1_{12}^n$ dont la valeur est textuelle :

- **"No"** : lorsque la création d'enveloppe n'est pas réalisée ;
- **"Yes"** : lorsque la création d'enveloppe est réalisée.

Description de l'opération de substitution

L'opération de substitution peut générer deux types de géométrie : cylindrique ou parallélépipédique. Pour définir, la géométrie créée, l'analyste propose une règle qui sera appliquée à tous les éléments du modèle (pièces ou sous-ensembles) substitués. Par exemple, la plus grande longueur de la boite englobante et le volume du composant à substituer sont conservés.

L'opération de substitution sera définie par une composante $op1_{13}^n$ dont les valeurs sont :

- **"No"** : lorsqu'il n'y a pas de substitution ;
- **"Cyl"** : lorsque les éléments sont substitués par un cylindre ;
- **"Par"** : lorsque les éléments sont substitués par un parallélépipède.

Description de l'opération d'agrégation

L'opération d'agrégation n'influence pas directement le niveau de simplification. Par contre, le résultat des opérations de simplification globales telles que la substitution ou la création d'enveloppe est très différent selon que ces opérations sont réalisées avant l'agrégation (chaque pièce est simplifiée indépendamment) ou après (le résultat est un volume unique fortement simplifié, figure 2.9).

L'opération d'agrégation sera définie par une composante $op1_{14}^n$ dont les valeurs textuelles sont :

- **"No"** : lorsqu'il n'y a pas d'agrégation ;
- **"After"** : lorsque l'agrégation est réalisée après les autres opérations de simplification ;
- **"Before"** : lorsque l'agrégation est réalisée avant les autres opérations de simplification.

Des exemples de simplification globale d'un ensemble sont donnés sur la figure 6.14.

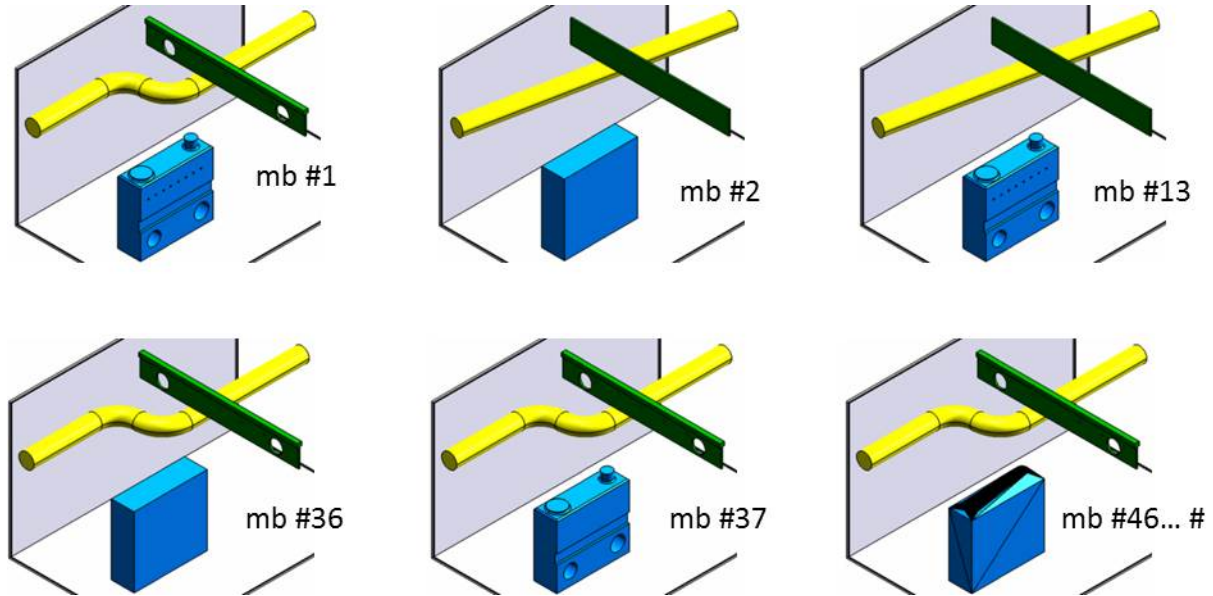


FIGURE 6.14 – Exemples de simplifications du modèle basique.

14 variables ($op1_1, \dots, op1_{14}$) ont été utilisées pour décrire le sous-processus de simplification d'un sous-ensemble. Le processus global de simplification $SP1_i^m$ selon une configuration i pourrait être représenté par une matrice de dimension $Nse * 14$, Nse étant le nombre de sous-ensembles du modèle original. Cela impliquerait que la dimension de la base dépendrait du nombre de sous-ensembles du modèle global original ce qui n'est pas souhaitable (contrainte N°1).

Afin de s'affranchir du nombre de sous-ensembles, ceux-ci sont regroupés dans trois catégories selon leur fonction :

1. "Conditions Limites" *CL* (zone où est appliqué un chargement ou condition de déplacement pour un problème d'élasticité, paroi ou une source de chaleur,...) ;
2. "Cibles" sur lesquels la grandeur physique recherchée sera évaluée ;
3. et "Obstacles" (sous-ensemble qui n'est ni une *CL* ni une cible).

Pour chaque catégorie, le processus sera décrit par six variables (équation 6.9) indiquant le niveau de simplification global pour chaque type d'opération. Ces variables sont estimées à partir de facteurs tels que des ratio des aires ou la distance de Hausdorff (tableau 6.5).

On notera :

- f ($x = 0$ si l'élément est supprimé, $x = 1$ si l'élément est conservé) ;
- $Aire_0^p$ l'aire d'une pièce p (avant toute autre opération de simplification) ;
 - $Aire_0^f$ l'aire d'un détail f (avant toute autre opération de simplification) ;
 - $Aire_0^m$ l'aire totale de l'assemblage global original ;
 - $Aire_j^n$ l'aire d'un sous-ensemble n ;
 - Np le nombre total de pièces du modèle ;
 - Nf le nombre total de détails ;
 - Nse le nombre total de sous-ensembles ;
 - HD_j^n la distance de Hausdorff entre le sous-ensemble simplifié et le sous-ensemble original.

$$SP1_i^m = \overbrace{\{SP1.1^m, ..., SP1.6^m\}}^{CL} \quad \overbrace{\{SP1.7^m, ..., SP1.12^m\}}^{Cibles} \quad \overbrace{\{SP1.13^m, ..., SP1.18^m\}}^{Obstacles} \quad (6.9)$$

	Filtrage	Defeaturing	Décimation	Enveloppe	Substitution	Agrégation
<i>CL</i>	<i>SP1.1</i>	<i>SP1.2</i>	<i>SP1.3</i>	<i>SP1.4</i>	<i>SP1.5</i>	<i>SP1.6</i>
<i>Cible</i>	<i>SP1.7</i>	<i>SP1.8</i>	<i>SP1.9</i>	<i>SP1.10</i>	<i>SP1.11</i>	<i>SP1.12</i>
<i>Obstacle</i>	<i>SP1.13</i>	<i>SP1.14</i>	<i>SP1.15</i>	<i>SP1.16</i>	<i>SP1.17</i>	<i>SP1.18</i>
	$= \frac{\sum_1^{Np}(xp.Aire_0^p)}{\sum_1^{Np}(Aire_0^p)}$	$= \frac{\sum_1^{Nf}(xf.Aire_0^f)}{\sum_1^{Nf}(Aire_0^f)}$	$= \frac{\sum_1^{Nse}(HD_j^n.Aire_j^n)}{Aire_0^m}$			

TABLEAU 6.5 – Paramètres globaux de réglage des opérations de simplification

Des exemples de représentation de processus de préparation sont donnés dans le tableau 6.6.

Op.	Filtrage			Defeaturing			Enveloppe			Substitution			Agrégation		
Cp.	CL	Cible	Obs.	CL	Cible	Obs.	CL	Cible	Obs.	CL	Cible	Obs.	CL	Cible	Obs.
Var.	SP1.1	SP1.7	SP1.13	SP1.2	SP1.8	SP1.14	SP1.4	SP1.10	SP1.16	SP1.5	SP1.11	SP1.17	SP1.6	SP1.12	SP1.18
M_0^b	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M_1^b	0	0	0	0	0	0	0	0	0	0,05	0,08	0,04	0	0	0
M_{13}^b	0	0	0	0	0	0	0	0	0	0,05	0	0,04	0	0	0
M_{36}^b	0	0	0	0	0	0	0	0	0	0	0,08	0	0	0	0
M_{37}^b	0	0	0	0	0,08	0	0	0	0	0	0	0	0	0	0
M_{46}^b	0	0	0	0	0,3	0	0	0,08	0	0	0	0	0	0	0

TABLEAU 6.6 – Exemples de modélisation de processus global de préparation sur le modèle M^b (figure 6.14). **Var.** = niveaux de simplifications (définis dans le tableau 6.5) pour chaque opération (**Op.**) et chaque type de composant (**Cp.**).

Le processus d'adaptation modélisé dans la section 6.1.3 du modèle simplifié pour la simulation consiste à définir les surfaces supportant les conditions limites et à relimiter les frontières du modèle à mailler.

Le nombre et la nature des opérations d'adaptation dépendent de l'objectif de la simulation. Une base d'apprentissage sera constituée d'exemples ayant le même objectif de simulation (calcul de structure, dynamique des fluides, simulation de procédés, visualisation....).

Dans le cadre de cette thèse, le processus d'adaptation est défini par un vecteur $SP2_i^m$ comprenant sept variables.

$$SP2_i^m = \{SP2.1^m, ..., SP2.6^m, T2^m\}$$

- $SP2.1^m$ **dimension**. Cette variable textuelle spécifie la dimension des éléments à adapter pour le maillage parmi trois valeurs : linéaires, surfaciques ou volumiques ;
- $SP2.2^m$ **éléments à extraire**. Cette variable textuelle spécifie le type d'éléments à extraire (éléments qui seront maillés) parmi trois valeurs : tous éléments, enveloppe extérieure, enveloppe intérieure ;
- $SP2.3^m$ à $SP2.6^m$ **éléments à créer**. Ces variables numériques indiquent le nombre d'éléments à créer (supports de conditions limites ou éléments de fermeture) :
 - $SP2.3^m$ nombre d'éléments 1D ,
 - $SP2.4^m$ nombre d'éléments 2D ,
 - $SP2.5^m$ nombre d'éléments surfaciques ,
 - $SP2.6^m$ nombre d'éléments volumiques.
- $T2^m$ **outillage**. Cette variable textuelle indique le type d'outillage utilisé pour l'adaptation (référence du logiciel utilisé et de ces modules). Certains outils réalisent automatiquement toutes les opérations d'adaptation pour un objectif de simulation donné. Cette variable indique donc indirectement le niveau d'automatisation des opérations d'adaptation.

Les activités de maillage ont été modélisées dans la section 6.1.4. Les paramètres de réglage des activités de maillage dépendent essentiellement de l'objectif de la simulation.

$$SP3_i^m = \{SP3.1^m, ..., SP3.6^m, T3^m\}$$

Dans le cadre de cette étude, le processus de maillage sera décrit par un vecteur $SP3_i^m$ ayant sept composantes spécifiques à l'objectif de simulation.

- $SP3.1^m$ **Éléments surfaciques.** Cette variable textuelle spécifie la forme des éléments surfaciques à mailler parmi trois valeurs : aucun (pas d'éléments surfaciques), triangles ou quadrilatères ;
- $SP3.2^m$ **Éléments volumiques.** Cette variable textuelle spécifie la forme des éléments volumiques à mailler parmi trois valeurs : aucun (pas d'éléments volumiques), tétraèdres ou hexaèdres ;
- $SP3.3^m$ **Taille.** Cette variable numérique spécifie la taille des éléments (taille moyenne hors raffinement) ;
- $SP3.4^m$ **Raffinement.** Cette variable numérique spécifie le niveau de raffinement autour des conditions limites, les valeurs de $SP3.4_i^m$ sont comprises entre 0 (pas de raffinement) et 1 (raffinement très fin autour de toutes les conditions limites) ;
- $SP3.5^m$ **Qualité de taille.** Cette variable numérique spécifie le niveau de la qualité de la taille des éléments attendu selon les critères définis dans la section 3.2.
- $SP3.6^m$ **Qualité de forme.** Cette variable numérique spécifie le niveau de la qualité de la forme des éléments attendus (section 3.2).
- $T3^m$ **outillage.** Cette variable textuelle indique le type d'outillage utilisé pour le maillage (référence du logiciel utilisé et de ces modules).

Ces variables sont les principales caractéristiques d'un maillage. Elles pourraient être complétées par les paramètres et options spécifiques à l'outil de maillage.

de préparation

L'objectif de cette section est de proposer une modélisation formalisée des données extraites de modèles 3D au niveau d'un ensemble global et d'un sous-ensemble, des facteurs de comparaison entre les modèles originaux et simplifiés ainsi que des facteurs d'influence de la simplification sur le cas de la simulation. Les notations utilisées dans cette section pour représenter ces données descriptives et facteurs sont données ci-dessous.

Notations pour la modélisation des données pour l'évaluation d'un sous-processus de simplification (niveau local) :

- CAO_0^n : ensemble de vecteur décrivant les caractéristiques géométriques du sous-ensemble n original ;
- CAO_j^n : ensemble de vecteur décrivant les caractéristiques géométriques du sous-ensemble n simplifié selon la configuration j ;
- $CAO.x_j^n$: vecteur décrivant la caractéristique géométrique x ;
- $Comp_j^n$: ensemble de vecteur décrivant les facteurs de comparaison entre le modèle original SE_0^n et le modèle simplifié SE_j^n ;
- $Inter_j^n$: ensemble de vecteur décrivant les facteurs d'influence des composants du sous-ensemble sur les conditions limites ou les objectifs de la simulation.

Notations pour la modélisation des données pour l'évaluation d'un processus de préparation (niveau global) :

- $CAOM1_i^m$: ensemble de vecteur décrivant les caractéristiques géométriques du modèle m simplifié selon la configuration i ;
- $CompM1_i^m$: ensemble de vecteur décrivant les facteurs de comparaison entre le modèle CAO original $M1_0^m$ et le modèle simplifié $M1_i^m$;
- $CAOM2_i^m$: ensemble de vecteur décrivant les caractéristiques géométriques du modèle m adapté ;
- $CompM2_i^m$: ensemble de vecteur décrivant les facteurs de comparaison entre le modèle préparé de référence $M2_{ref}^m$ et le modèle préparé $M2_i^m$ selon la configuration i ;
- $Mail_i^m$: ensemble de vecteurs décrivant les caractéristiques du modèle m maillé ;
- $CompM3_i^m$: ensemble de vecteur décrivant les facteurs de comparaison entre le maillage de référence $M3_{ref}^m$ et le maillage $M3_i^m$ préparé selon la configuration i ;

Il s'agit de proposer une représentation formalisée des facteurs qui caractérisent l'influence de la simplification d'un sous-ensemble sur le résultat de l'analyse et sur le coût de la simplification.

Il peut s'agir :

- de **données géométriques** décrivant la taille et la forme des modèles CAO représentés par les ensembles de vecteurs CAO_0^n (modèle original) et CAO_j^n (modèle simplifié) ;
- de **facteurs de comparaison des modèles** géométriques (ensemble de vecteurs $Comp_j^n$) tels qu'ils ont présentés dans la section 2.5 ;
- de **facteurs d'influence** des composants du sous-ensemble et les conditions limites ou les objectifs de la simulation (ensemble de vecteurs $Inter_j^n$).

Ces facteurs sont définis dans les paragraphes suivants. Parmi toutes les données qui peuvent être extraites d'un modèle géométrique, seules celles qui semblent être les plus logiques pour évaluer un coût de préparation, un coût de simulation et une erreur sur le résultat de l'analyse sont présentées dans cette section. Dans un premier temps la liste est assez exhaustive. La sélection de variables déterminantes se fera par apprentissage dans le chapitre suivant.

Description des modèles CAO.

Les géométries du modèle original et du modèle simplifié sont décrites à l'aide de variables caractéristiques de la taille du modèle et de sa forme listées dans le tableau 6.7). Soit $CAO.x_j^n$ la valeur brute d'une des composantes du vecteur CAO_j^n . Les valeurs $CAO.x_j^n$ sont mesurées sur les modèles CAO du sous-ensemble SE_j^n ou sur sa boîte englobante (parallélépipède minimal contenant le modèle CAO et aligné sur ses axes principaux).

Ces variables serviront essentiellement à calculer les composantes des ensembles de vecteurs $Comp_j^n$ et $Inter_j^n$. Certaines caractéristiques de la complexité d'un modèle seront déterminantes pour le calcul des coûts de préparation et de simulation. Dans une moindre mesure, elles interviendront dans le calcul de l'influence de la simplification sur le résultat de l'analyse.

En plus de la valeur brute $CAO.x_j^n$, la base de données peut contenir des variantes qui permettront de prendre en compte les dispersions entre les différentes pièces d'un sous-ensemble, comme :

- l'écart entre les valeurs minimales et maximales $\overline{\text{L CAO.x}_j^n}$;
- la moyenne des valeurs $\overline{\text{CAO.x}_j^n}$;
- l'écart type des différentes pièces du sous-ensemble $\sigma(\text{CAO.x}_j^n)$;
- un ratio par rapport à une autre grandeur.

La base de donnée descriptive d'un modèle géométrique contient alors plus de 300 variables.

	Vecteur CAO_j^n	Définition
Facteurs de taille	Aire_j^n	Aire du sous-ensemble
	Volume_j^n	Volume du sous-ensemble
	N.pieces_j^n	Nombre de pièces du sous-ensemble
	Triangles_j^n	Nombre de triangles du sous-ensemble
	Patches_j^n	Nombre de patches du sous-ensemble
	Vol.boite_j^n	Volume de la boîte englobante du sous-ensemble
	L.boite_j^n	Longueur (plus grande dimension) de la boîte englobante du sous-ensemble
	Lmoy_j^n	Longueur moyenne (moyenne des trois longueurs) de la boîte englobante du sous-ensemble, caractérise globalement la taille du sous-ensemble
	S.boite_j^n	Section (plus grande dimension) de la boîte englobante du sous-ensemble
Facteurs de forme	Compacité.1_j^n	Compacité du sous-ensemble de type 1 (équation 2.4 de la section 2.5)
	Compacité.2_j^n	Compacité du sous-ensemble de type 2 (équation 2.5 de la section 2.5)
	$\text{Ratio.vol.volboite}_j^n$	Ratio entre le volume du sous-ensemble sur le volume de sa boîte englobante
	$\text{Ratio.section.vol}_j^n$	Ratio entre la section et le volume du sous-ensemble
	N.détails_j^n	Nombre de détails du sous-ensemble
	Aire.détails_j^n	Aire des surfaces des détails du sous-ensemble
	$\text{Triangles.détails}_j^n$	Nombre de triangles des surfaces des détails du sous-ensemble
	$\text{Courbure.moyenne}_j^n$	Courbure moyenne des faces du sous-ensemble

TABLEAU 6.7 – Composantes du vecteur CAO_j^n . Facteurs descriptifs de la géométrie d'un sous-ensemble.

Des exemples de vecteurs descriptifs de la géométrie d'un sous-ensemble sont donnés dans le tableau 6.8.

	Aire	Volume	Nb.pieces	Nb.triangles	Nb.patchs	Compacité	Ratio volume / vol. boîte	Ratio nb. triangles / patchs	Courbure k1	Courbure k2
	CAO.1 ⁿ _j	CAO.2 ⁿ _j	CAO.3 ⁿ _j	CAO.4 ⁿ _j	CAO.5 ⁿ _j	CAO.10 ⁿ _j	CAO.12 ⁿ _j	CAO.14 ⁿ _j	CAO.18 ⁿ _j	CAO.19 ⁿ _j
SE ² ₀	1,46	0,016	1	1304	28	92	1,25	47	-62	-2
SE ² ₁	1,45	0,016	1	548	18	92	1,26	30	-48	2
SE ² ₂	1,29	0,012	1	1288	24	111	0,92	54	-62	-2
SE ² ₃	1,53	0,017	1	784	24	92	1,33	33	-62	-2
SE ² ₄	1,28	0,012	1	532	10	110	0,92	53	-48	2
SE ² ₅	1,36	0,013	1	12	6	108	1	2	-50	0
SE ² ₆	1,46	0,016	1	326	326	93	1,25	1	-144	90
SE ² ₇	1,46	0,016	1	136	136	92	1,26	1	-100	63

TABLEAU 6.8 – Description du modèle CAO du sous-ensemble N°2 (figure 6.13). Exemples de variables explicatives limitées aux variables déterminantes.

Comparaison entre le modèle original et le modèle simplifié.

Des facteurs de comparaison et de similarité entre les modèles géométriques ont été présentés dans la section 2.5. Les critères de comparaison entre le modèle original SE_0^n et le modèle simplifié SE_j^n donnent une indication sur le niveau de simplification du modèle. Les modèles sont comparés à l'aide de deux types de facteurs (tableau 6.9) :

- les bénéfices $Comp.x_j^n$ de la simplification pour chacune des valeurs brutes CAO. x_j^n

$$Comp.x_j^n = \frac{CAO.x_j^n - CAO.x_0^n}{CAO.x_0^n}$$

- et les distances de Hausdorff (figure 2.10) HD_j^n , $HD.moy_j^n$ et $HD.quad_j^n$, respectivement maximales, moyennes et moyennes quadratiques, calculées entre le modèle original SE_0^n et le modèle simplifié SE_j^n .

Bénéfices sur la taille	Bénéfices sur la forme	Distance de Hausdorff
$Comp.Aire_j^n$	$Comp.Compacité1_j^n$	HD_j^n
$Comp.Volume_j^n$	$Comp.Compacité2_j^n$	
$Comp.Npièces_j^n$	$Comp.détails_j^n$	
$Comp.Triangles_j^n$	$Comp.aire.détails_j^n$	
$Comp.Patch_j^n$	$Comp.triangles.détails_j^n$	
	$Comp.courbures_j^n$	

TABLEAU 6.9 – Composantes du vecteur $comp_j^n$. Facteurs de comparaison entre le sous-ensemble simplifié et original.

Pour minimiser l'impact de la simplification sur le résultat de l'analyse, les bénéfices relatifs aux facteurs de taille devraient tendre vers zéro. Au contraire, afin de

minimiser les coûts de préparation et de simulation, les bénéfices relatifs aux facteurs de forme doivent être les plus importants possible. Un compromis entre ce deux types de facteurs doit être recherché.

Des exemples de vecteurs descriptifs des facteurs de comparaison de sous-ensembles original et simplifié sont donnés dans le tableau 6.10.

	Nb. pièces	Nb. triangles	Compacité	Ratio volume / vol. boîte	Courbure k1	Courbure k2	Distance de Hausdorff	Distance / cl	Distance / cible	Moment distance / cl
	$comp.3_j^n$	$comp.4_j^n$	$comp.7_j^n$	$comp.8_j^n$	$comp.13_j^n$	$comp.14_j^n$	$comp.16_j^n$	$inter.1_j^n$	$inter.2_j^n$	$inter.5_j^n$
SE ₀ ²	0	0	0	0	0	0	0	92	514	0
SE ₁ ²	0	0,58	0,006	-0,004	0,238	1,953	0,001	92	514	0,003
SE ₂ ²	0	0,012	-0,084	0,266	0	0	0,004	92	475	0,118
SE ₃ ²	0	0,399	-0,01	-0,061	0	0	0,007	92	475	-0,051
SE ₄ ²	0	0,592	-0,077	0,262	0,238	1,953	0,004	92	475	0,121
SE ₅ ²	0	0,991	-0,08	0,201	0,197	1	0,008	92	475	0,07
SE ₆ ²	0	0,75	-0,002	-0,002	-1,313	39,131	0	92	475	-0,003
SE ₇ ²	0	0,896	0,004	-0,008	-0,6	27,694	0,001	92	475	-0,002

TABLEAU 6.10 – Exemples de comparaison $comp.x_j^n$ de sous-ensembles originaux et simplifiés (figure 6.13) et de facteurs d'influence $inter.x_j^n$ sur le cas de simulation. Variables explicatives limitées aux variables déterminantes.

Interactions avec le cas de simulation.

Il s'agit de définir les interactions entre le sous-ensemble à caractériser et les composants liés aux conditions limites (CL) et aux éléments à observer (cibles) en prenant en compte les positions relatives entre le sous-ensemble et les CL ou cibles.

Les facteurs qui caractérisent ces interactions sont la fonction du sous-ensemble, les distances entre le sous-ensemble et les conditions limites ainsi que la taille des composants qui constituent le sous-ensemble (selon les constats faits dans la section 3.1, un élément de taille importante aura plus d'influence sur le résultat de l'analyse qu'un élément de très petite taille proche d'une condition limite). Les six composantes du vecteur $Inter_j^n$ sont listées dans le tableau 6.11.

Distances aux conditions limites

On notera DCL^n la distance euclidienne entre le sous-ensemble et la condition limite la plus proche, et $DCible^n$ la distance euclidienne entre le sous-ensemble et la cible la plus proche. Ces distances varient peu quel que soit le niveau de simplification du sous-ensemble, elles peuvent être considérées comme constantes donc connues pour un nouveau cas.

Taille des composants.

Un sous-ensemble est constitué de pièces de tailles très variées. Le calcul de la dispersion entre les pièces doit tenir compte à la fois de la taille des pièces et de leurs distances aux conditions limites. Deux facteurs inspirés du théorème de Huygens ont été proposés à partir de l'aire $Aire_j^p$ ou la taille $Lmoy_j^p$ de chaque pièce P_j^p (longueur moyenne des dimensions principales de la boîte englobante de la pièce).

- Facteur relatif à la taille des pièces $Taille.CL_j^n$ du sous-ensemble SE_j^n :

$$Taille.CL_j^n = \frac{\sum_{p=1}^{Np} (DCL_j^p \cdot Lmoy_j^p)}{\sum_{p=1}^{Np} (DCL_0^p \cdot Lmoy_0^p)}$$

- Facteur relatif à l'aire des pièces $Aire.CL_j^n$ du sous-ensemble SE_j^n :

$$Aire.CL_j^n = \frac{\sum_{p=1}^{Np} \left((DCL_j^p)^2 \cdot Lmoy_j^p \right)}{\sum_{p=1}^{Np} \left((DCL_0^p)^2 \cdot Lmoy_0^p \right)}$$

Fonction du sous-ensemble.

Si l'une ou l'autre des variables DCL^n ou $DCible^n$ est nulle, alors le sous-ensemble n est lui même une condition limite ou une cible.

Lorsque le sous-ensemble n est une condition limite, des informations peuvent compléter la description de sa fonction telles que le type de grandeur physique appliquée au sous-ensemble (action mécanique, chaleur spécifique,...) ou l'ordre de grandeur des valeurs.

Distances aux conditions limites	Taille des composants	Fonction
DCL^n	$Taille.CL_j^n$	$Grandeur_j^n$
$DCible^n$	$Aire.CL_j^n$	$Valeur.grandeur_j^n$

TABLEAU 6.11 – Composantes du vecteur $Interactions_j^n$. Facteurs d'interactions entre le sous-ensemble et le cas de simulation.

Des exemples de vecteurs descriptifs des facteurs d'interactions entre le sous-ensemble et le cas de simulation sont donnés dans le tableau 6.10.

6.2.5.2 CHAPITRE 6. MODÉLISATION DES DONNÉES POUR L'APPRENTISSAGE Variables pour la représentation des données au niveau de l'ensemble global

Il s'agit de lister les facteurs qui caractérisent les critères d'évaluation d'un processus de préparation d'un modèle pour la simulation. On notera pour un modèle global m :

- $M1_i^m$ le modèle CAO **simplifié** selon la configuration i ;
- $M2_i^m$ le modèle **adapté** ;
- $M3_i^m$ le modèle **maillé** .

Description de la géométrie des modèles CAO.

L'ensemble global $M1_i^m$ simplifié est un assemblage des N_c sous-ensembles SE_j^n dont les caractéristiques géométriques ont été définies dans la sous-section précédente. Le modèle CAO global est décrit par un vecteur $CAOM1_i^m$ dont les composantes sont des grandeurs géométriques similaires à celles des sous-ensembles (tableau 6.12). Les valeurs de ces composantes seront notées $CAOM1.x_i^m$.

Comparaison des modèles simplifiés.

Les modèles globaux originaux $M1_0^m$ et simplifiés $M1_i^m$ sont comparés en calculant le bénéfice de la simplification entre la valeur $CAOM1.x_i^m$ du modèle simplifié et la valeur $CAOM1.x_0^m$ du modèle original .

$$CompM1.x_i^m = \frac{CAOM1.x_i^m - CAOM1.x_0^m}{CAOM1.x_0^m}$$

Description du modèle adapté.

Le modèle adapté (modèle géométrique à mailler) est décrit par un vecteur $CAOM2_{2i}^m$ dont les composantes sont similaires au modèle simplifié (tableau 6.12).

Des exemples de facteurs descriptifs des modèles CAO simplifiés sont donnés sur le tableau 6.13.

Comparaison des modèles adaptés.

Les opérations d'adaptation sur un modèle CAO complexe sont généralement impossibles sans simplification. Le modèle adapté selon une configuration de simplification i est comparé à un modèle simplifié selon une configuration de référence ref (modèle adaptable ayant le plus faible niveau de simplification) en calculant :

- pour chaque composante du vecteur $CAOM2_i^m$ les bénéfices $CompM2_i^m$ entre le modèle adapté et le modèle de référence (défini dans la section 6.4.2) ;

$$CompM2.x_i^m = \frac{CAOM2.x_i^m - CAOM2.x_{ref}^m}{CAOM2.x_{ref}^m}$$

	Vecteur CAOM1 _i ^m	Vecteur CAOM2 _i ^m	Définition
Facteurs de taille	Aire1 _i ^m	Aire2 _i ^m	Aire du modèle global ou adapté
	Volume1 _i ^m	Volume2 _i ^m	Volume de l'ensemble global ou adapté
	N.pièces1 _i ^m	∅	Nombre de pièces de l'ensemble global
	Triangles1 _i ^m	Triangles2 _i ^m	Nombre de triangles de l'ensemble global ou adapté
	Patches1 _i ^m	Patches2 _i ^m	Nombre de patches de l'ensemble global ou adapté
	Vol.boite1 _i ^m	∅	Volume de la boîte englobante de l'ensemble global
	Lmoy1 _i ^m	∅	Longueur moyenne de la boîte englobante de l'ensemble global
Facteurs de forme	Compacité1.1 _i ^m	Compacité.1 _{2,i} ^m	Compacité de type 1 (tableau 6.7) de l'ensemble global ou adapté
	Compacité1.2 _i ^m	Compacité2.2 _i ^m	Compacité de type 2 (tableau 6.7) de l'ensemble global ou adapté
	N.détails1 _i ^m	∅	Nombre de détails de l'ensemble global
	Aire.détails1 _i ^m	∅	Aire des surfaces des détails de l'ensemble global
	Triangles.détails1 _i ^m	∅	Nombre de triangles des surfaces des détails de l'ensemble global
	Courbure.moy1 _i ^m	Courbure.moy2 _i ^m	Courbure moyenne des faces de l'ensemble global ou adapté

TABEAU 6.12 – Composantes des vecteurs CAOM1_i^m et CAOM2_i^m. Facteurs descriptifs de la géométrie du modèle global et adapté. ∅ : variable non utilisée (la notion de pièces et de détails ne s'applique pas au modèle adapté).

	Nb. patches	Ratio volume / section boîte	Aire	Volume	Nb. pièces	Nb. triangles	Ratio volume / volume boîte	Ratio volume / section boîte	Ratio nb. triangles / patches
	CAOM1.21	CAOM1.28	Comp1.33	CompM1.34	CompM1.35	CompM1.36	CompM1.40	CompM1.41	CompM1.42
M ₀ ^b	102	0,060	0	0	0	0	0	0	0
M ₁ ^b	28	0,063	0,009	-0,052	0,545	0,975	-0,05	-0,05	0,908
M ₃ ^b	74	0,059	0,008	0,019	0,545	0,724	0,019	0,019	0,619
M ₃₆ ^b	56	0,064	0,002	-0,070	0	0,251	-0,07	-0,07	-0,37
M ₃₇ ^b	78	0,060	0	0	0	0,092	0	0	-0,19
M ₄₆ ^b	276	0,063	0,007	-0,048	0	0,230	-0,05	-0,05	0,716

TABEAU 6.13 – Exemples de description CAOM1 et de comparaison CompM1 d'ensembles CAO représentés sur la figure 6.14. Variables explicatives limitées aux variables déterminantes.

- les distances de Hausdorff HD2_{2i}^m, HD.moy2_i^m et HD.quad2_i^m respectivement maximales, moyennes et moyennes quadratiques entre le modèle adapté et le modèle de référence (défini dans la section 6.4.2).

Le modèle préparé est le modèle maillé $M3_i^m$. Les facteurs relatifs au modèle maillé qui influencent la qualité du processus de préparation sont décrits par un vecteur $Mail3_i^m$ dont les composantes sont les caractéristiques du maillage et ses critères de qualité, soit :

- le nombre d'éléments maillés $Eléments3_i^m$;
- le nombre d'arêtes $Arêtes3_i^m$;
- le nombre de sommets $Sommets3_i^m$;
- le critère de qualité de forme des éléments maillés $Qualité.Forme3_i^m$ définis selon les critères présentés dans la section 3.2 ;
- le critère de qualité de taille des éléments maillés $Qualité.Taille3_i^m$ (section 3.2) .

Les paramètres de réglage de l'opération de maillage (type d'éléments, carte de maillage, ...) sont des données connues et définies précédemment (description des processus de préparation).

Comparaison des modèles préparés.

Le modèle non simplifié ne pouvant être maillé, le modèle préparé selon la configuration de simplification i sera comparé au modèle préparé de référence ref (modèle maillable le moins simplifié possible tel qu'il est défini dans la section 6.4.2) en calculant le bénéfice $CompM3.x_i^m$.

$$CompM3.x_i^m = \frac{Mail3.x_i^m - Mail3.x_{ref}^m}{Mail3.x_{ref}^m}$$

Des exemples de variables pour la description et la comparaison des modèles adaptés et maillés sont donnés dans le tableau 6.14.

	Altre	Volume	Nb. triangles	Nb parchs	Volume	Compacité	Ratio_vol_vol_boite	Distance de Hausdorff	Nb. éléments	Durée maillage	Nb. éléments
	CAOM2.1	CAOM2.2	CAOM2.3	CAOM2.4	CompM2.13	CompM2.17	CompM2.18	CompM2.25	Mail3.3	Mail3.4	Comp3.7
M_0^b	34,1	10,3	7760	102	0	0	0	0	99951	36	0
M_1^b	33,8	10,3	296	28	0,003	0,01	0,05	0,05	86478	37	12
M_{13}^b	33,8	10,3	1780	74	-0,001	0,01	-0,02	0,04	90966	40	8
M_{36}^b	34,1	10,3	6276	56	0,004	0	0,07	0,01	93208	36	6
M_{37}^b	34,1	10,3	7248	78	0,000	0	0	0	97745	37	2
M_{46}^b	34,1	10,3	6490	276	0,003	0	0,05	0,01	97752	34	2

TABEAU 6.14 – Exemples de description de modèles adaptés CAOM2, de description de maillages $MailM3$, de comparaison de modèles adaptés $CompM2$, de comparaison de maillages $CompM3$. Variables explicatives limitées aux variables les plus déterminantes

Il s'agit de proposer des facteurs qui caractérisent l'influence de la simplification sur le cas de simulation en décrivant les interactions d'un sous-ensemble avec les conditions limites et les éléments à observer (cibles). Ces facteurs sont ceux qui ont été définis pour décrire les interactions entre le sous-ensemble et les conditions limites généralisées à l'ensemble des Nse sous-ensembles du modèle. Ils sont calculés à partir de la distance de chaque pièce p du sous-ensemble n à la condition limite ou la cible la plus proche (respectivement $DCL^{n,p}$ et $DCible^{n,p}$ et à partir de l'aire $Aire^{n,p}$ ou de la taille $Lmoy^{n,p}$ de la pièce, soit :

- le facteur relatif à la taille des Np pièces et à leur distance aux conditions limites $Taille.CL3_i^m$;

$$Taille.CL3_i^m = \frac{\sum_{n=1}^{Nse} \left(\sum_{p=1}^{Np} \left(DCL^{n,p} \cdot Lmoy_j^{n,p} \right) \right)}{\sum_{n=1}^{Nse} \left(\sum_{p=1}^{Np} \left(DCL^{n,p} \cdot Lmoy_0^{n,p} \right) \right)}$$

- le facteur relatif à l'aire des Np pièces et à leur distance aux conditions limites $Aire.CL_i$ pour tous les sous-ensembles n ;

$$Aire.CL3_i^m = \frac{\sum_{n=1}^{Nse} \left(\sum_{p=1}^{Np} \left((DCL^{n,p})^2 \cdot Aire_j^{n,p} \right) \right)}{\sum_{n=1}^{Nse} \left(\sum_{p=1}^{Np} \left((DCL^{n,p})^2 \cdot Aire_0^{n,p} \right) \right)}$$

- le facteur relatif à la taille des pièces et à leur distance aux cibles $Taille.Cible_i$;

$$Taille.Cible3_i^m = \frac{\sum_{n=1}^{Nse} \left(\sum_{p=1}^{Np} \left(DCible^{n,p} \cdot Lmoy_j^{n,p} \right) \right)}{\sum_{n=1}^{Nse} \left(\sum_{p=1}^{Np} \left(DCible^{n,p} \cdot Lmoy_0^{n,p} \right) \right)}$$

- le facteur relatif à l'aire des pièces et à leur distance aux cibles $Aire.Cible_i$;

$$Aire.Cible3_i^m = \frac{\sum_{n=1}^{Nse} \left(\sum_{p=1}^{Np} \left((DCible^{n,p})^2 \cdot Aire_j^{n,p} \right) \right)}{\sum_{n=1}^{Nse} \left(\sum_{p=1}^{Np} \left((DCible^{n,p})^2 \cdot Aire_0^{n,p} \right) \right)}$$

Une modélisation formalisée des données qui caractérisent les critères d'évaluation d'un processus de préparation pour la simulation a été proposée.

Afin d'implémenter ultérieurement ces données dans des classifieurs et afin de proposer une représentation unique quel que soit le nombre de sous-ensembles et de pièces du modèle, les données sont représentées dans deux matrices.

La matrice globale décrit le processus de préparation du modèle ainsi que les caractéristiques des critères d'évaluation de ce processus.

La matrice locale décrit le sous-processus de simplification d'un sous-ensemble ainsi que les caractéristiques des critères d'évaluation de ce sous-processus. Cette matrice locale sera utilisée pour déterminer les composantes de la matrice globale.

Chaque ligne correspond à un exemple de sous-ensemble simplifié (matrice locale) ou de modèle global simplifié (matrice globale).

Chaque colonne est une variable qui décrit le processus de préparation du modèle ou qui caractérise ses critères d'évaluation.

Les variables relatives au sous-processus de simplification (matrice locale) sur un sous-ensemble sont synthétisées sur la figure 6.15.

Les variables relatives au processus global (matrice globale) sont synthétisées sur la figure 6.16.

Le sous-processus de simplification d'un sous-ensemble est modélisé par un vecteur qui décrit le type d'opération réalisée, ses paramètres de réglage et le type d'outils utilisés. Cette description permet de formaliser simplement les six opérations de simplification identifiées et les 40 processus réalistes (annexe A.1). D'autres opérations de simplification et d'autres sous-processus pourraient être proposées pour d'autres objectifs de préparation (calcul de structure, visualisation,...).

Le processus global de préparation d'un modèle est modélisé par un vecteur P^m indiquant le niveau de simplification du modèle relatif à chaque opération, les paramètres de réglages et les outils d'adaptation et de maillage. Cette description est spécifique à un objectif de préparation précis. Si tous les exemples de la base n'ont pas le même objectif de préparation, une variable supplémentaire devra indiquer l'objectif de la simulation (calcul de structure, thermique, ...).

Dans le cadre de cette étude, les variables à prédire par apprentissage sont les critères d'évaluation d'un sous-processus de simplification y (local) et les critères d'évaluation Y d'un processus de préparation (global) basées sur le calcul des coûts de

La base de données doit contenir tous les facteurs qui caractérisent les six critères d'évaluation d'un processus de préparation. Ces facteurs décrivent les modèles originaux, simplifiés, adaptés ou maillés, les comparent et quantifient les modifications sur les interactions avec le cas de simulation.

Un grand nombre de facteurs (190 variables pour la base locale au niveau des sous-ensembles et plus de 220 variables pour la base globale de l'ensemble complet) a été proposé. Les facteurs les plus déterminants ne sont pas encore connus, par exemple :

- La simplification d'un modèle de grande taille influencera plus les résultats de l'analyse. Est-ce l'aire, le volume du modèle ou bien un autre facteur qui caractérise au mieux la taille d'un sous-ensemble ?
- Doit-on utiliser la valeur totale de la taille d'un sous-ensemble ? Ou bien la taille moyenne des pièces qui le compose ? Ou bien leur écart type ?
- Le coût de la simplification sera plus important si le modèle est complexe. Comment caractériser cette complexité ? Avec un critère de compacité ? En dénombrant le nombre de faces ? Ou bien le nombre de triangles du modèle ?

Afin de répondre à ces questions et afin de limiter le nombre de facteurs, les variables les plus déterminantes doivent être identifiées. Pour cela, une méthode est proposée dans la section suivante. Des critères basés sur les résultats des prédictions devront évaluer la qualité de cette sélection.

D'autres facteurs auraient pu être ajoutés (position du centre de gravité, moments,...), la méthode proposée devra donc également s'assurer de l'exhaustivité des variables pour un objectif d'apprentissage particulier.

Représentation locale des données (sous-ensemble)				
Processus de simplification	Caractéristiques des critères d'évaluation du processus de simplification			Critères d'évaluation d'un sous-processus de simplification
	Valeurs brutes, Ecart mini / maxi, Moyenne par pièce, Ecart type	Comparaison modèle original et simplifié	Interactions avec le cas de simulation	
Vecteurs de variables : $SP1_j^n$	Description des modèles CAO : Vecteurs de variables : $\{CAO_0^n; CAO_j^n\}$	Vecteurs de variables : $Comp_j^n$	Vecteurs de variables : $Inter_j^n$	Vecteurs de variables : y_j^n
20 var. connues	68 var. connues + 68 var. inconnues	48 var. inconnues	4 var. connues + 2 var. inconnues	2 var. à prédire
■ Description des opérations de simplification	■ Taille du sous-ensemble	■ Bénéfices sur les valeurs caractéristiques des modèles CAO	■ Rôle du sous-ensemble	■ Sensibilité de la simplification sur le résultat de l'analyse (IS = y1)
■ Outils utilisés pour la simplification	■ Taille de la boîte englobante du sous-ensemble	■ Distance aux conditions limites et aux cibles	■ Distance aux pièces / distance aux conditions limites et aux cibles	■ Coût de la simplification (CS = y2)
	■ Forme du sous-ensemble	■ Distance de Hausdorff		
	■ Description des détails du sous-ensemble			

FIGURE 6.15 – Contenu de la matrice de données locale au niveau des sous-ensembles.

Représentation globale des données (ensemble complet)								
Processus de préparation		Caractéristiques des critères d'évaluation du processus de préparation					Critères d'évaluation d'un processus de préparation	
		Description des modèles CAO	Description du modèle adapté	Description du modèle maillé	Comparaison modèles de référence et modèles préparés			
		Opérations de simplification	Opérations d'adaptation et de maillage	Valeurs brutes, Ecart mini / maxi, Moyenne par pièce, Ecart type				
Vecteurs de variables : $SP1_i^m$	Vecteurs de variables : $SP2_i^m$ $SP3_i^m$	Vecteurs de variables : $CAOM1_0^m$ $CAOM1_i^m$	Vecteurs de variables : $CAOM2_i^m$	Vecteurs de variables : $Mail3_i^m$	Vecteurs de variables : $CompM1_i^m$ $CompM2_i^m$ $CompM3_i^m$	Vecteurs de variables : $Inter_i^m$	Vecteurs de variables : Y_j^n	
18 var. connues	14 var. connues	52 var. connues + 52 var. inconnues	28 var. inconnues	5 var. inconnues	83 var. inconnues	4 var. inconnues	3 var. à prédire	
■ Description globale des opérations de simplification	■ Description des opérations d'adaptation ■ Description des opérations de maillage	■ Taille de l'ensemble ■ Taille de la boîte englobante de l'ensemble ■ Forme de l'ensemble ■ Description des détails de l'ensemble	■ Taille de l'ensemble adapté ■ Forme de l'ensemble adapté	■ Nombre d'éléments ■ Qualité du maillage	■ Bénéfices sur les valeurs caractéristiques des modèles CAO ■ Bénéfices sur les valeurs caractéristiques des modèles adaptés ■ Distance de Hausdorff entre modèles adaptés de référence et testés ■ Bénéfices sur les valeurs caractéristiques des modèles maillés	■ Moment taille ou aire pièces / distance aux conditions limites et aux cibles	■ Erreur sur le résultat de l'analyse (ERA = Y3) ■ Coût de la préparation (CP = Y4) ■ Coût de la simulation (CA = Y5)	

FIGURE 6.16 – Contenu de la matrice de données globale au niveau de l'ensemble complet.

6.3 Identification des variables explicatives

La base de données brute contient un très grand nombre de variables (plus de 250). L'étape d'identification (Figure 6.17) consiste à sélectionner parmi ces 250 variables celles qui sont les plus déterminantes pour prédire les variables de sortie et intermédiaires qui seront appelées variables explicatives.

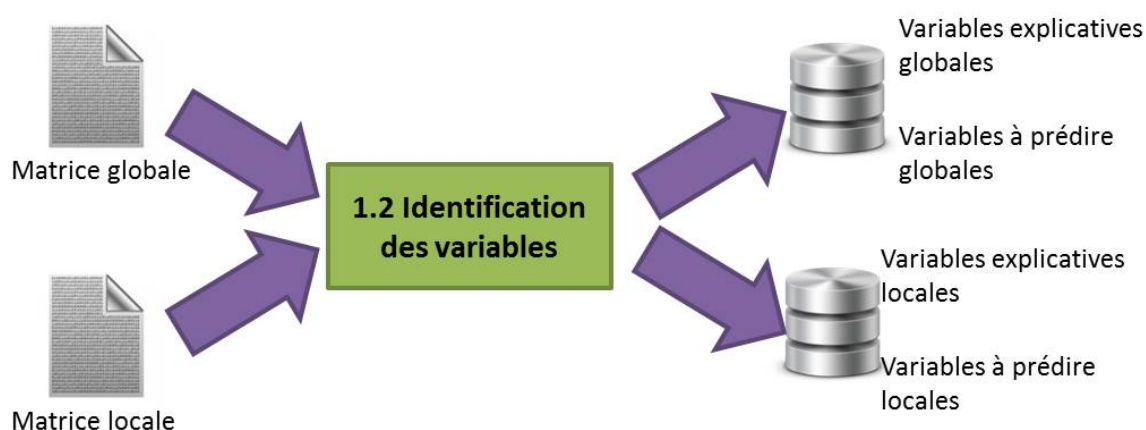


FIGURE 6.17 – Etape 1 construction de la base de données pour l'apprentissage (figure 5.4).1.2 Sélection des variables explicatives et variables à prédire.

La réduction du nombre de variables d'entrée se justifie pour plusieurs raisons.

- *Justification N°1.* Un grand nombre de variables explicatives nécessite d'avoir un très grand nombre d'exemples. Ce qui est difficile dans le cadre de cette étude et dans un contexte industriel .
- *Justification N°2.* Certaines variables d'entrées ne sont pas connues pour un nouveaux cas. Réduire le nombre de variables d'entrée permettra de réduire le nombre de variables intermédiaires à prédire.
- *Justification N°3.* Les variables d'entrées qui ont moins d'influence qu'un bruit sur la prédiction sont des causes d'erreur sur le résultat de la classification.
- *Justification N°4.* La configuration des classifieurs est plus rapide si le nombre de variables explicatives est limité.
- *Justification N°5.* Le facteur le plus déterminant pour décrire une grandeur n'est pas toujours connu a priori. La base de données brute contient donc initialement plusieurs variables pour décrire une même grandeur. Ces différentes variables sont alors fortement corrélées. La plus déterminante devra être identifiée.
- *Justification N°6.* Si on généralise la problématique précédente on peut considérer que l'identification de variables déterminantes non connues a priori permet de formaliser la connaissance.

L'identification des variables explicatives consiste dans un premier temps à supprimer les variables corrélées, puis à sélectionner les variables explicatives pour chaque variable à prédire et à valider la sélection.

6.3.1 Suppression de variables corrélées

Deux variables sont corrélées lorsqu'elles présentent des interdépendances. La suppression de variables corrélées permet d'éliminer les variables redondantes en ne sélectionnant que la variable la plus déterminante.

6.3.1.1 Méthode proposée pour la suppression de variables corrélées

La méthode proposée consiste tout d'abord à identifier des ensembles de variables corrélées entre elles. Les outils mathématiques ne permettant de déceler que les corrélations linéaires, cette tâche reste visuelle. L'opération est facilitée par les outils de data mining qui intègrent des outils de visualisation de données.

Pour chaque groupe de variables corrélées, le coefficient de corrélation $\rho(x, y)$ (équation 6.10) est calculé entre chaque variable à analyser et chaque variable à prédire.

$$\rho(x, y) = \frac{\sum (x - \bar{x}) \cdot (y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2} \cdot \sqrt{\sum (y - \bar{y})^2}} \quad (6.10)$$

Avec :

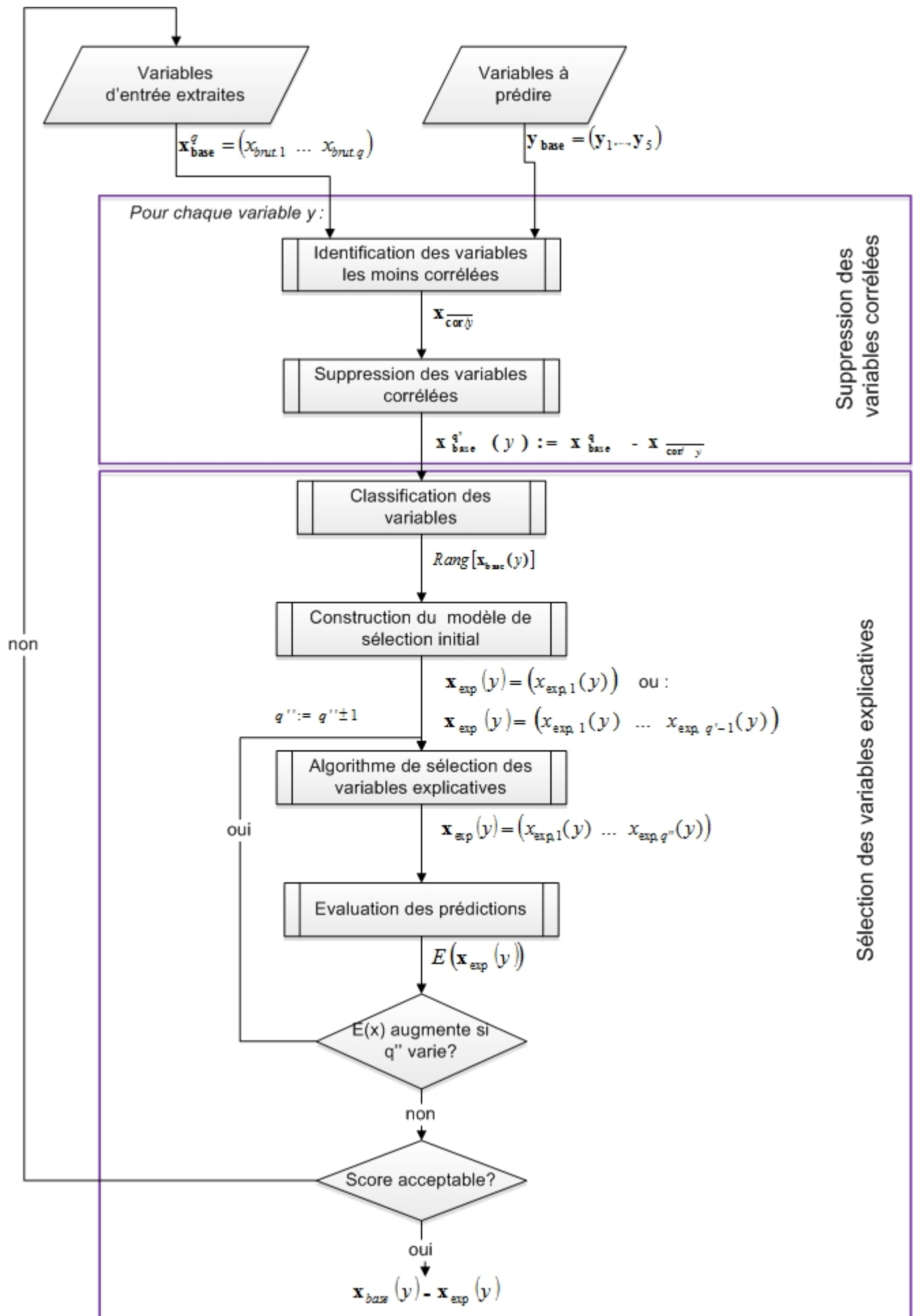
x, y : variables à analyser

$\bar{x}, \bar{y}$: moyennes respectives des variables x et y .

L'un des traitements nécessaires à l'implémentation des données qui sera développé dans le chapitre suivant est la suppression de variables aberrantes. Le coefficient de corrélation est très sensible aux valeurs aberrantes, il est nécessaire de s'assurer que la base de données n'en contient plus.

La variable retenue sera la variable la plus corrélée avec les variables à prédire (variable dont la valeur absolue du coefficient de corrélation est le plus élevé). Si un groupe comprend à la fois des variables connues et inconnues pour un nouveau cas et que leurs coefficients de corrélation sont assez proches, la variable connue sera retenue.

La base X_{base}^q contient initialement q variables. Les variables $\mathbf{x}_{cor,y}$ les moins corrélées



132 FIGURE 6.18 – Méthode pour l'implémentation et la sélection de variables explicatives.

Les variables à prédire y sont supprimées de la base de données qui est alors réduite à q' variables.

6.3.1.2 Exemples de suppression de variables

La figure 6.19 montre un exemple de variables non corrélées (distance de Hausdorff moyenne et bénéfice sur l'aire du modèle préparé) dont les valeurs occupent tout l'espace, de variables faiblement corrélées (distance de Hausdorff moyenne et maximale) et de variables fortement corrélées (distance de Hausdorff moyenne et quadratique) dont les valeurs sont visiblement interdépendantes. Le coefficient de corrélation est proche de 0 dans le premier cas et proche de 1 dans le dernier cas.

Le tableau 6.15 donne les résultats du coefficient de corrélation entre ces variables et les variables principales à prédire. La valeur moyenne quadratique sera retenue pour la prédiction de l'erreur sur le résultat de l'analyse. La valeur moyenne sera retenue pour la prédiction des coûts. Les valeur maximale sera supprimée de la base de données.

Coefficient de corrélation $\rho(x, y)$		Distance de Hausdorff		
x y		Maximale	Moyenne	Quadratique
Erreur sur le résultat de l'analyse		-0,062	-0,343	-0,476
Coût de la simulation		-0,028	0,056	0,030
Coût de la préparation		0,047	-0,363	-0,210

TABEAU 6.15 – Coefficients de corrélation entre les distances de Hausdorff (maximale, moyenne ou quadratique) et les variables à prédire.

6.3.2 Sélection des variables explicatives

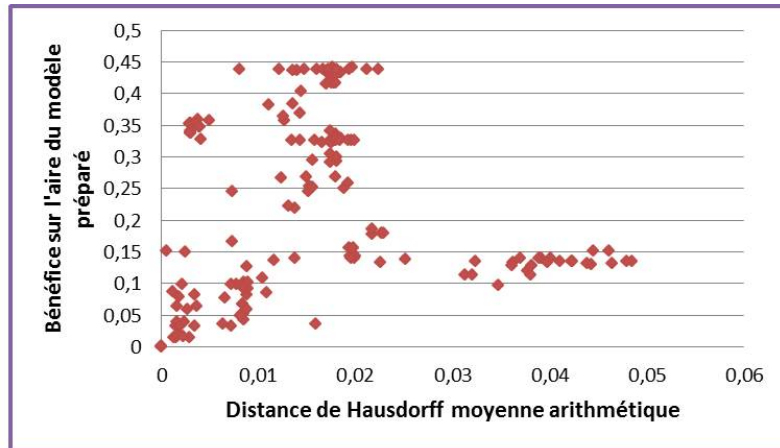
Après traitement et suppression des variables corrélées, la base de donnée $\mathbf{x}_{\text{base}}^{q'}$ contient q' variables peu corrélées entre elles.

Pour chaque variable à prédire y , les variables sont classées en fonction de leur influence sur la variable y .

Les deux algorithmes de sélection des variables explicatives les plus courants (stepwise backward regression et stepwise forward regression), présentés dans la section 4.5, ont été utilisés. Un premier modèle est construit ayant soit $q' - 1$ variables (pour l'algorithme stepwise backward regression) soit une seule variable (pour les algorithmes stepwise forward regression et descripteur sonde).

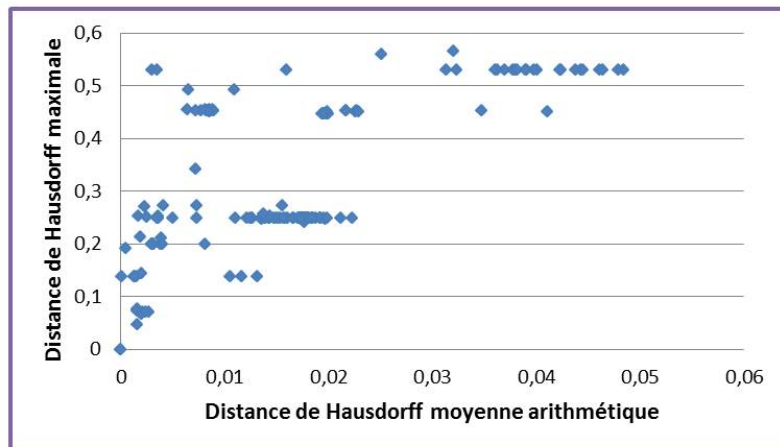
(a) Variables peu corrélées

$\rho(a) = 0,035$



(b) variables faiblement corrélées

$\rho(b) = 0,63$



(c) variables fortement corrélées

$\rho(c) = 0,96$

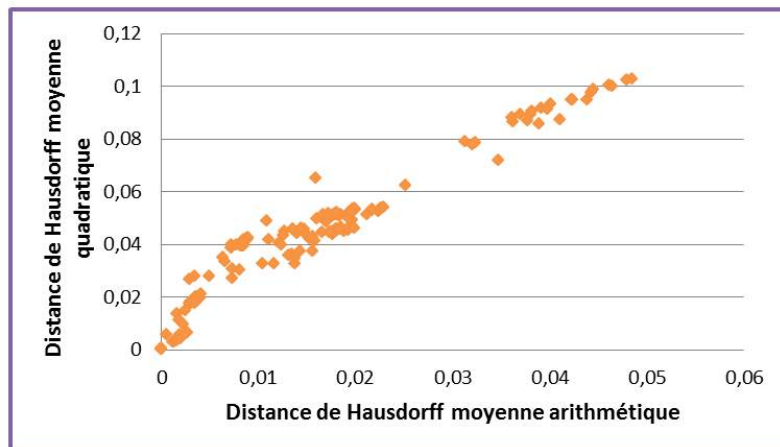


FIGURE 6.19 – Exemples de corrélation. (a) Variables non corrélées. (b) Variables faiblement corrélées. (c) Variables fortement corrélées. $\rho(a), \rho(b), \rho(c)$: coefficients de corrélation.

Des modèles successifs sont créés auxquels une variable est ajoutée ou retirée à chaque itération en fonction de son rang $\text{Rang}[x_{base}(y)]$. Les algorithmes évaluent ces modèles à l'aide de critères propres à chaque algorithme. Les itérations s'arrêtent lorsque le critère d'évaluation a atteint un seuil acceptable et lorsque l'ajout d'une nouvelle variable supplémentaire n'améliore pas le critère d'évaluation.

Si, quel que soit l'algorithme de sélection, le critère d'évaluation n'a pas atteint un seuil acceptable et n'évolue plus, l'exhaustivité des variables initiales est remise en

6.3.3 Conclusion sur la sélection des variables explicatives

La sélection des variables explicatives permet de s'assurer de la qualité de la classification et aide à la formalisation des connaissances.

Des méthodes ont été proposées pour sélectionner les variables corrélées à supprimer et pour sélectionner les variables explicatives à partir d'algorithmes courants de sélection.

Des critères d'évaluation propres aux algorithmes de sélection évaluent la qualité de la sélection. La valeur du seuil de ces critères est définie par l'utilisateur en fonction de la précision attendue sur les prédictions. Le choix et le nombre des variables explicatives est validé lorsque l'erreur sur les prédictions atteint un score acceptable et qu'elle n'évolue plus.

Pour chaque variable à prédire, la sélection des variables explicatives est donc conditionnée par le seuil défini et les algorithmes de sélection utilisés. L'exhaustivité des variables explicatives, devra être confirmée par des tests de classification présentés dans le chapitre suivant.

6.4 Construction d'exemples pour l'apprentissage

6.4.1 Espace "erreur sur le résultat de l'analyse / niveau de simplification"

Un exemple est un cas de préparation pour un modèle CAO et pour un objectif de simulation. Les variables de sortie sont discrétisées et réparties dans plusieurs classes. La base d'exemples pour l'apprentissage doit contenir un nombre significatif de cas pour chaque classe de variables d'entrée (description du processus de préparation et des modèles géométriques) et de sortie (erreur sur le résultat de l'analyse et coûts). Les exemples choisis pour l'apprentissage devront donc couvrir tout le domaine « erreur sur le résultat » pour tous les niveaux de simplification du modèle (figure 6.20).

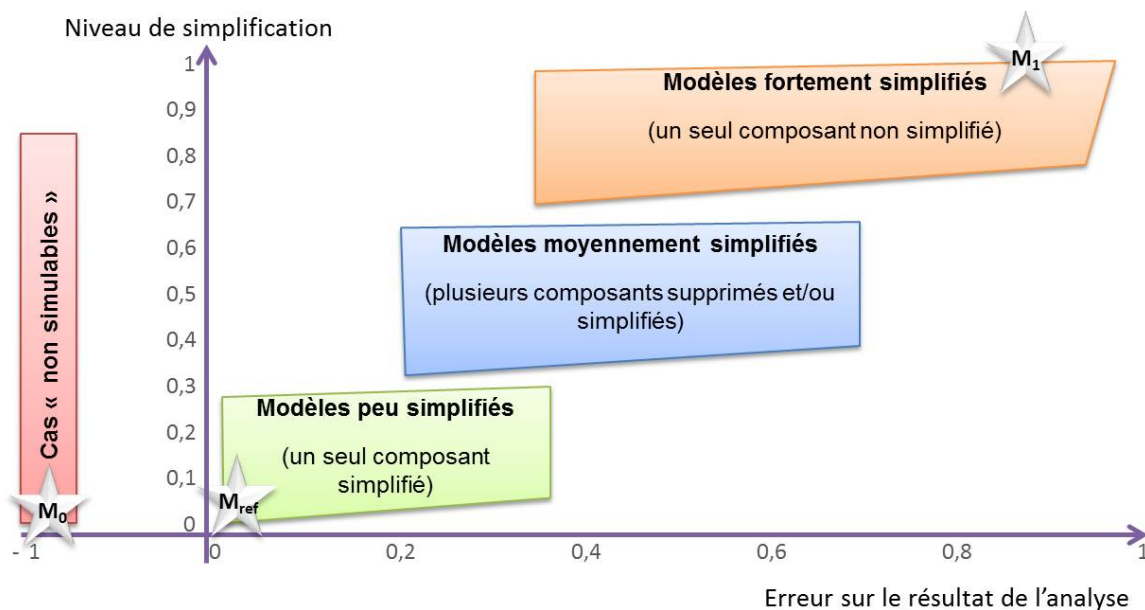


FIGURE 6.20 – Espace "Erreur sur le résultat de l'analyse / Niveau de simplification du modèle"

L'erreur sur le résultat de l'analyse (ERA équation 6.6) est calculée pour l'apprentissage à partir du résultat de la simulation sur le modèle à évaluer et le modèle de référence (modèle simulable le moins simplifié possible).

Le niveau de simplification (LOS) est évalué à l'aide de la moyenne quadratique de la distance de Hausdorff entre le modèle original et le modèle simplifié. Ce critère a été choisi parmi les facteurs de comparaison entre les modèles énoncés dans la section 6.2.5 pour avoir le plus d'influence sur le résultat de l'analyse (coefficient de corrélation le plus élevé).

6.4.2. Contenu de la Base d'exemples

Le modèle géométrique de l'ensemble global M_i (équation 6.11) comprend Nse sous-ensembles SE_j^n liés aux conditions limites (CL) ou hors conditions limites, pour lesquels n est l'indice du sous-ensemble ($1 < n < k$ pour les sous-ensembles liés aux conditions limites ou aux éléments à observer) et $k + 1 < n < N$ pour les sous-ensembles hors conditions limites).

$$M_i = f \left(\overbrace{(SE_{j1}^1, \dots, SE_{jk}^k)}^{CL}, \overbrace{(SE_{jk+1}^{k+1}, \dots, SE_{jN}^N)}^{Hors CL} \right) \quad (6.11)$$

L'indice j donne la référence de la configuration du sous-processus de simplification du sous-ensemble SE_j^n . Ses valeurs particulières sont :

- $j = 0$ si le sous-ensemble n'est pas simplifié,
- $j = 1$ si le sous-ensemble est simplifié au maximum (simplification par agrégation de toutes les pièces du sous-ensemble puis substitution par un parallélépipède ou un cylindre),
- $j = sup$ si le sous-ensemble est supprimé (filtrage de toutes ses pièces),
- $j = best$ correspond à la meilleure configuration du sous-ensemble (meilleur compromis entre l'erreur sur le résultat de l'analyse et du coût de la simplification et de simulation),
- $j = ref$ si le sous-ensemble est simplifié au minimum (bénéfice de l'aire du modèle minimale) et si cette simplification ne cause pas l'échec de la simulation.

L'indice i est la référence de la configuration du processus de préparation du modèle global. Ses valeurs particulières sont :

- $i = 0$ pour le modèle original,
- $i = 1$ pour le modèle simplifié au maximum,
- $i = ref$ pour le modèle de référence (modèle qui peut être adapté, maillé et simulé et dont le bénéfice sur l'aire est minimal),
- $i = best$ pour le meilleur modèle.

Afin de tester l'approche proposée en couvrant tout l'espace "niveau de simplification", la base d'exemples a été construite en contenant pour chaque modèle (figure 6.20) :

- le modèle original M_0 non simplifié ;

$$M_0 = f (SE_0^1, \dots, SE_0^N)$$

- le modèle minimal M_1 dont tous les sous-ensembles hors conditions limites sont supprimés et les sous-ensembles liés aux conditions limites sont simplifiés au maximum,

$$M_1 = f\left(SE_1^1, \dots, SE_1^k, SE_{sup}^{k+1}, \dots, SE_{sup}^N\right)$$

- le modèle de référence M_{ref} dont tous les sous-ensembles sont simplifiés selon leur configuration de référence SE_{ref}^n ,

$$M_{ref} = f\left(SE_{ref}^1, \dots, SE_{ref}^N\right)$$

- des modèles ayant des niveaux de simplification élevés (tous les sous-ensembles sont simplifiés au maximum ou supprimés sauf le sous-ensemble SE_j^{n1}),

$$M_{LOS.Max.} = f\left(SE_{sup ou 1}^1, \dots, SE_j^{n1}, \dots, SE_{sup ou 1}^N\right)$$

- des modèles ayant des niveaux de simplification bas (tous les sous-ensembles sont simplifiés selon leur configuration de référence sauf un qui est filtré SE_{sup}^{n1} ou simplifié au maximum SE_1^{n1} ,

$$M_{LOS.min.} = f\left(SE_{ref}^1, \dots, SE_{sup ou 1}^{n1}, \dots, SE_{ref}^N\right)$$

- des modèles ayant des niveaux de simplification intermédiaires, obtenus par remplacements successifs d'un sous-ensemble par une version plus ou moins simplifiée selon la méthode proposée dans la section suivante.

Afin de couvrir tout l'espace "erreur sur le résultat de l'analyse", la base d'exemples devra contenir un nombre significatif d'exemples :

- dont l'erreur sur le résultat de l'analyse est négligeable (inférieure à une limite fixée par les analystes) ;
- dont l'erreur sur le résultat de l'analyse est inacceptable (supérieure à une limite fixée par les analystes) ;
- pour lesquels la préparation ou la simulation n'est pas réalisable ;
- distribués dans les autres classes (erreurs comprises entre "négligeables" et "inacceptables").

6.4.3. Méthode pour la construction d'exemples

La méthode proposée pour créer les exemples significatifs de cas de préparation consiste tout d'abord à créer des exemples de simplification pour chaque sous-ensemble puis à créer des exemples globaux de préparation.

6.4.3.1 Construction d'exemples de simplification d'un sous-ensemble

Les simulations sont généralement impossibles lorsqu'un ou plusieurs sous-ensemble n'est pas simplifié. Il est souhaitable que la base d'exemples ne contienne pas un trop grand nombre de cas dont la simulation est impossible (le nombre d'exemples doit être régulièrement réparti dans toutes les classes). La méthode proposée schématisée sur la figure 6.21 consiste à identifier pour chaque sous-ensemble les modèles non simulables et les modèles particuliers.

Pour cela, une série de "J" sous-processus de simplification est proposée par l'analyste pour chaque sous-ensemble n (les 40 séquençages ne sont pas tous appropriés à tous les types de sous-ensembles) en précisant pour chaque processus le séquençage des opérations et leurs paramètres de réglages.

Pour chaque sous-processus de simplification " j " proposé, le sous-ensemble SE_j^n est simplifié à partir du modèle original SE_0^n . Les niveaux de simplification $LOS(SE_j^n)$ et coûts de simplifications $CS(SE_j^n)$ sont calculés.

Le sous-ensemble est intégré dans l'ensemble global $M_{LOS.Max.}$ à fort niveau de simplification dont tous les autres sous-ensembles $SE_1^{≠n}$ sont simplifiés au maximum afin d'éviter les risques d'échec. Le modèle global ainsi obtenu à fort niveau de simplification est simulé. Les résultats de la simulation $R(SE_j^n)$ et l'influence $IS(SE_j^n)$ de la simplification sur le sous-ensemble sont relevés. Le modèle de référence SE_{ref}^n (modèle simulable ayant un faible niveau de simplification) et le meilleur modèle SE_{ref}^n (modèle ayant l'erreur la plus faible pour un coût moindre) sont identifiés.

Finalement, les différentes configurations de simplification j du sous-ensemble n sont classées selon la valeur de l'influence $IS(SE_j^n)$.

6.4.3.2 Construction d'exemples de préparation d'ensembles globaux

Il s'agit de construire progressivement des modèles globaux M_i ayant des niveaux de simplification faibles, intermédiaires et élevés.

Les exemples de modèles à fort niveau de simplification comprennent :

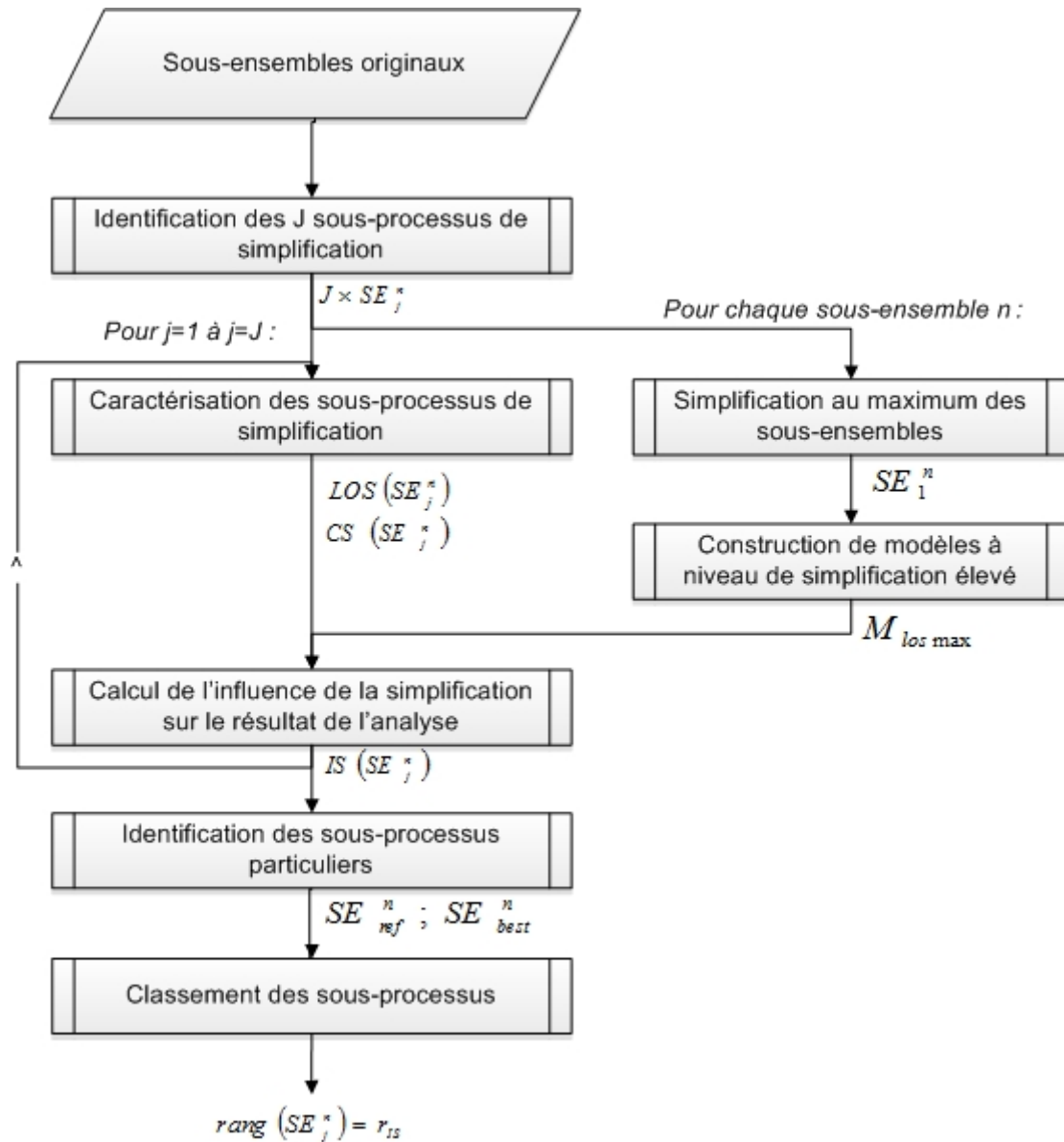


FIGURE 6.21 – Méthode pour la construction d'exemples de simplification de sous-ensembles

- un modèle M_1 dont tous les sous-ensembles sont simplifiés au maximum et conservés ;
- un modèle $M_{LOS.Max}$ dont tous les sous-ensembles sont simplifiés au maximum et dont les sous-ensembles hors condition limites sont supprimés ;
- des modèles dont tous les composants sont simplifiés au maximum sauf un qui est simplifié selon une configuration j ;
- des modèles dont tous les composants sont simplifiés au maximum et dont un ou plusieurs sous-ensembles hors condition limite est supprimé.

Les exemples de modèles à niveau de simplification intermédiaires comprennent :

- des modèles dont les sous-ensembles sont simplifiés au maximum ou selon leur meilleur configuration, obtenus progressivement en remplaçant un par un les sous-ensembles SE_1^n par un sous-ensemble SE_{best}^n en commençant par celui dont l'influence sur le résultat de l'analyse est la plus faible ;
- des modèles dont les sous-ensembles sont simplifiés selon la meilleur configuration et dont un ou plusieurs sous-ensembles est supprimé.

Les exemples de modèles à faible niveau de simplification intermédiaires comprennent :

- le modèle original global M_0 (généralement non simulable) ;
- le modèle de référence M_{ref} ;
- des modèles dont tous les sous-ensembles sont simplifiés selon la configuration de référence SE_{ref}^n sauf un qui est simplifié au maximum SE_1^n ou selon sa meilleure configuration SE_{best}^n ;
- des modèles dont tous les sous-ensembles sont simplifiés selon la configuration de référence SE_{ref}^n et dont un sous-ensemble est supprimé.

6.4.4 Conclusion sur la construction d'exemples pour l'apprentissage

Une méthode pour construire efficacement une base d'exemples de processus préparation a été proposée. Pour s'assurer que la base d'exemples ainsi construite est suffisante, il est nécessaire de vérifier qu'un nombre équivalent d'exemples est présent dans chaque classe de variables à prédire (principales ou intermédiaires). Des expérimentations devront déterminer le nombre minimal d'exemples que doivent contenir chaque classe de variables explicatives.

Hypothèse 1 :

Le modèle global de référence M_{ref} est la concaténation de tous les sous-ensembles de référence SE_{ref}^n . Cette hypothèse se justifie par le fait qu'il ne peut y avoir de modèle simulable ayant un niveau de simplification plus faible. Par contre, cela implique que ce modèle global de référence est simulable. Dans le cas contraire, le modèle M_{ref} devra être identifié parmi les modèles à faible niveau de simplification.

Hypothèse 2 :

L'erreur globale sur le résultat de l'analyse est comparée au résultat d'un modèle de référence dont on considère l'erreur négligeable. Cette hypothèse pourra être vérifiée en réalisant des tests sur un modèle moins complexe dont la configuration non simplifiée est simulable ou bien en procédant à un test physique sur une configuration donnée.

CHAPITRE 6. SYNTHÈSE SUR LA MODÉLISATION DES DONNÉES POUR L'APPRENTISSAGE

6.5 Synthèse sur la modélisation des données pour l'apprentissage

Une modélisation des données descriptives d'un processus de préparation et des données extraites des modèles géométriques (CAO ou maillages) a été proposée. Afin d'être implémentées dans des outils d'apprentissage artificiels de type classifieurs, les données pour l'apprentissage sont représentées dans deux matrices de variables. Une matrice au niveau local (figure 6.15) décrit le sous-processus de simplification d'un sous-ensemble. Une matrice au niveau global (figure 6.16) décrit le processus de préparation de l'ensemble complet. Pour chaque matrice, une ligne correspond à un exemple (configuration de sous-ensemble simplifié ou de modèle global préparé). Chaque colonne est un vecteur de variables.

La taille de la base de données qui contenait initialement un grand nombre de variables (plus de 300) a été réduite grâce à une analyse des variables corrélées et une sélection des variables les plus déterminantes.

Une méthode a ensuite été proposée pour construire une base d'exemples représentative de l'ensemble des cas de processus de préparation.

A la fin de cette phase, deux des cinq éléments nécessaires à la configuration d'un modèle d'apprentissage ont été complètement définis (figure 6.22) : (1) exemples pour l'apprentissage et (2) variables à prédire. Les variables explicatives (3) ont été identifiées, elles devront être configurées pour s'adapter aux classifieurs choisis. L'étape suivante (chapitre 7) consistera alors à choisir (4) puis configurer les paramètres (5) des classifieurs et des variables (3).

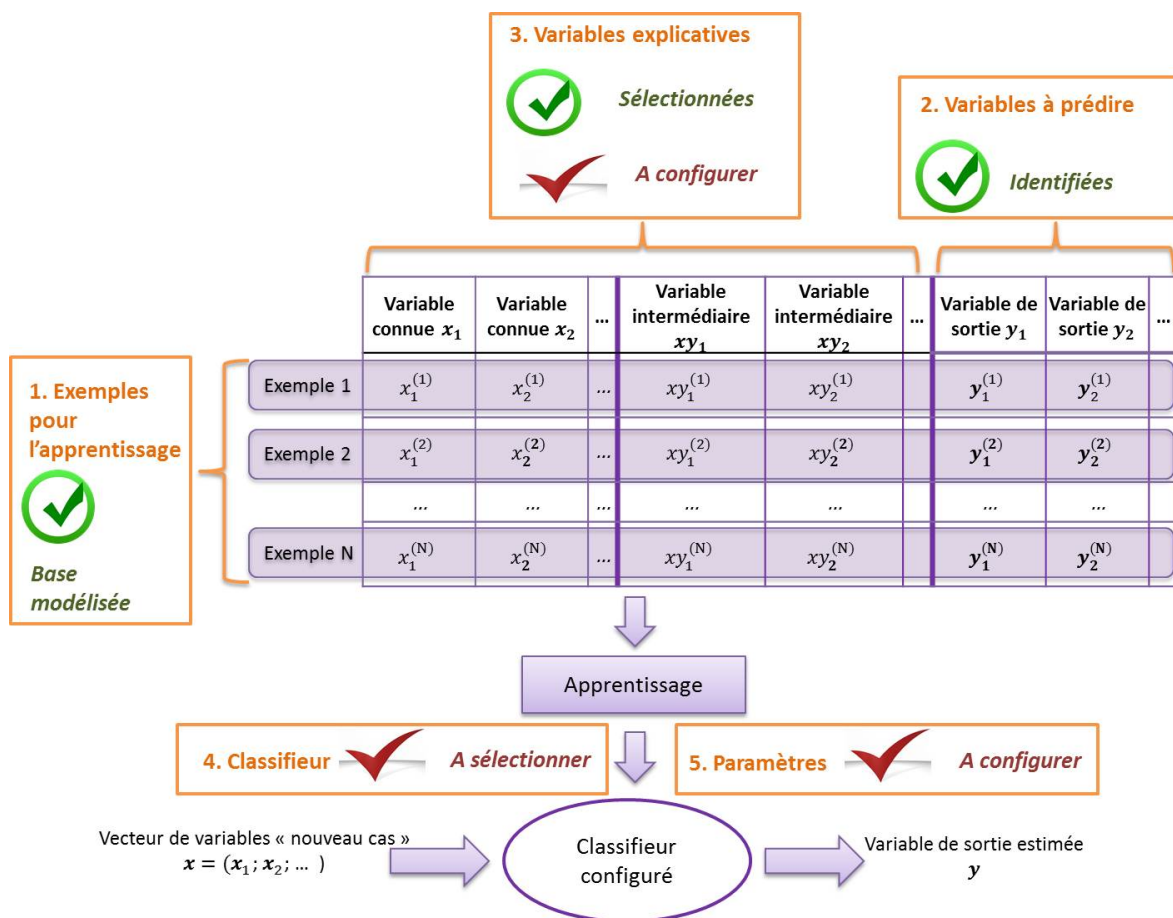


FIGURE 6.22 – Cinq éléments à déterminer pour la configuration d'un modèle d'apprentissage : avancement après la phase de construction de la base de données.

Chapitre 7

Apprentissage et prédictions sur des nouveaux cas

La phase précédente avait pour objectif de construire une base d'exemples de processus de préparation de modèles pour la simulation décrits par des variables explicatives et des variables à prédire. Les variables explicatives qui caractérisent au mieux les variables à prédire ont été sélectionnées puis traitées afin d'assurer leur implémentation dans les outils d'apprentissage artificiel.

Un modèle d'apprentissage est un couple de données et de classifieurs configurés pour prédire le plus précisément possible une variable à partir d'un nouveau vecteur de variables explicatives. La phase d'apprentissage consiste à construire des modèles d'apprentissage pour l'estimation de chaque variable à prédire.

Il a été mis en évidence dans la section 6.2 du chapitre précédent que les variables explicatives ne sont pas toutes connues pour un nouveau cas. Il sera alors nécessaire de définir une stratégie pour évaluer ces variables intermédiaires et prendre en compte l'erreur induite par leur estimation sur le résultat global.

Un évaluateur devra être proposé afin de vérifier la méthode proposée et plus largement l'hypothèse qu'il est envisageable d'utiliser des techniques d'apprentissage artificiels pour l'évaluation a priori de la qualité des processus de préparation pour la simulation.

La section 7.1 propose une méthode pour la sélection et la configuration d'un modèle d'apprentissage ainsi que pour l'évaluation de la qualité des classifications.

La section 7.2 montre comment ces modèles d'apprentissage seront utilisés lorsqu'un nouveau cas se présente pour évaluer un sous-processus de simplification sur un sous-ensemble puis pour évaluer un processus global de préparation.

L'objectif de cette section est de configurer des modèles d'apprentissage capables d'estimer le plus précisément possible une variable à prédire dans un délai acceptable en configurant au mieux des modèles d'apprentissage (couples données / classifieurs).

Notations pour la construction de modèles d'apprentissage :

- N : nombre d'exemples total de la base ;
- N_a : nombre d'exemples pour l'apprentissage ;
- N_t : nombre d'exemples pour les tests ;
- NK : nombre d'exemples pour la validation croisée ;
- $x_{ap}(y)$: vecteur de variables explicatives pour la prédiction de la variable y ;
- $Cl(y)$: classifieur pour la prédiction de la variable y ;
- $\{x_{ap}(y), Cl(y)\}$: modèle d'apprentissage ;
- AQE : erreur quadratique moyenne (Average Quadratique Error), eq. 4.3 ;
- ROC : surface sous la courbe ROC (Receiver Operating Characteristic), définie dans la section 4.5.2 ;
- D_{ap} : durée de mise en œuvre de l'apprentissage ;
- IC : indicateur de confiance du classifieur.

CHAPITRE 7. APPRENTISSAGE ET PRÉDICTIONS SUR DES NOUVEAUX CAS

La méthode proposée pour la construction des modèles d'apprentissage est illustrée sur la figure 7.1 puis détaillée sur la figure 7.2 et les sous-sections suivantes.

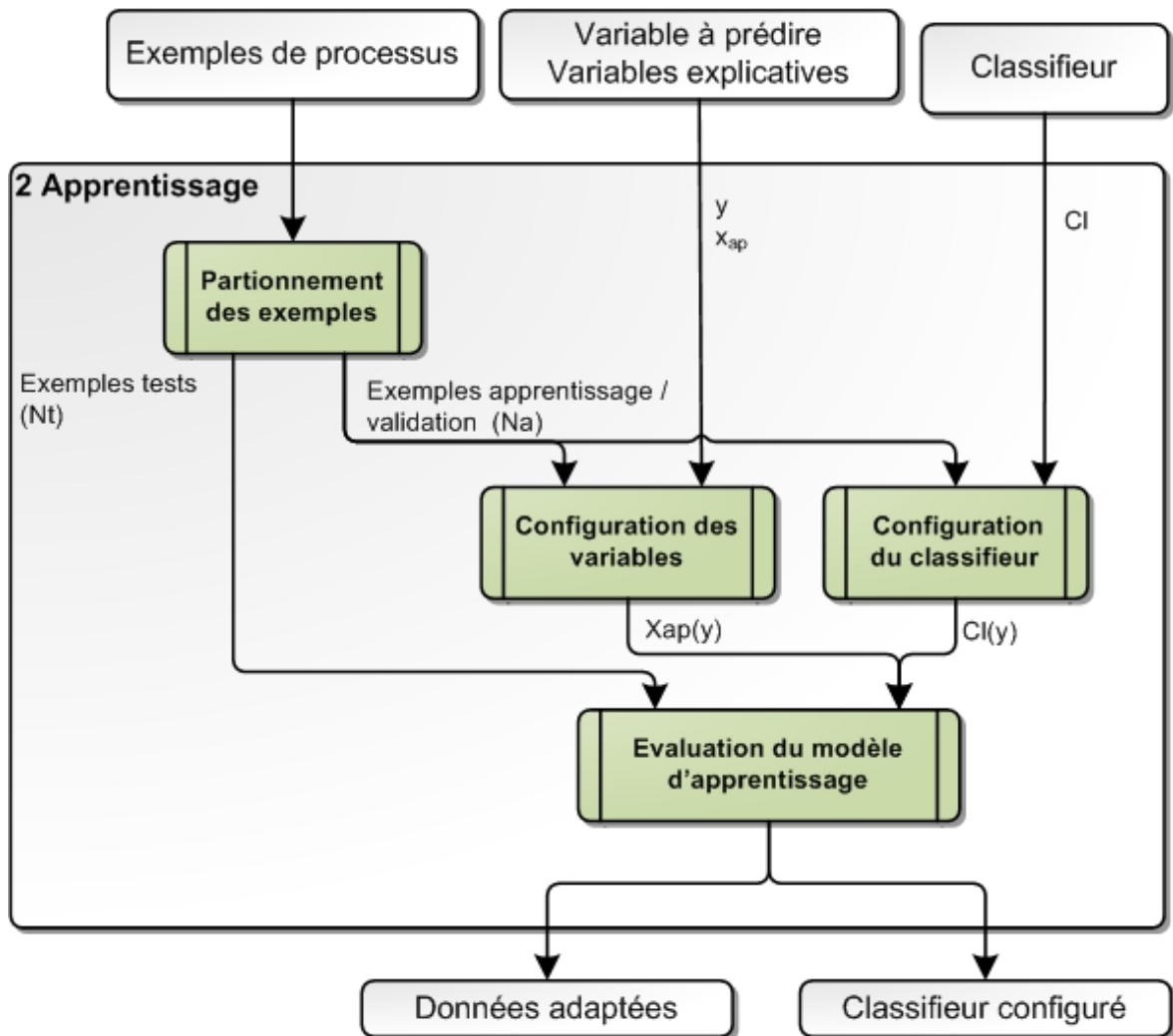


FIGURE 7.1 – Approche générale pour la construction de modèles d'apprentissage.

Les exemples de la base construits précédemment sont partitionnés dans deux principaux ensembles. Les N_a exemples du premier ensemble seront ré-échantillonnés en K groupes pour l'apprentissage et la validation. Les N_t exemples du second ensemble seront utilisés pour les tests de classification.

Afin d'identifier, pour chaque variable y à prédire, le meilleur modèle d'apprentissage $\{x_{ap}(y), Cl(y)\}$, différentes configurations de variables $x_{ap}(y)$ et d'architectures de classifieurs $Cl(y)$ seront évaluées.

Les modèles d'apprentissage dont l'évaluation est satisfaisante sont sélectionnés puis optimisés en identifiant précisément chacun de leurs paramètres.

7.1.1 Partitionnement des exemples

Lorsque les exemples utilisés pour les tests sont les mêmes que ceux utilisés pour l'apprentissage les résultats de prédiction collent aux données d'apprentissage. L'erreur de prédiction est trop optimiste et manque de fiabilité. Pour y remédier, les N exemples stockés dans la base de données sont partitionnés en N_a exemples pour l'apprentissage qui seront utilisés lors de la configuration des classifieurs et en N_t exemples pour les tests (étape 2.1 de la figure 7.2 et figure 7.3).

Comme pour les N exemples initiaux, les exemples des deux groupes pour l'apprentissage et les tests devront couvrir tout l'espace "erreur sur le résultat de l'analyse / niveau de simplification" tel que cela a été défini dans la section 6.4.

Le nombre d'exemples étant restreint, afin d'éviter le risque de sur-apprentissage, une méthode de validation croisée "*k-fold cross-validation*" sera appliquée. Pour cela, les exemples pour l'apprentissage sont divisés en K groupes aléatoires de N_k exemples (étape 2.2 de la figure 7.2). La figure 7.3 donne un exemple de ré-échantillonnage en $K = 5$ groupes.

Lors du partitionnement des exemples, quatre paramètres sont à définir.

- Paramètre "partitionnement" 1 : nombre minimal d'exemples dans chaque classe de variable à prédire.
- Paramètre "partitionnement" 2 : nombre d'exemples pour l'apprentissage N_a .

$$N_a \simeq N \times 75\%$$

- Paramètre "partitionnement" 3 : nombre de groupes K pour la validation croisée.
- Paramètre "partitionnement" 4 : nombre d'exemples pour les tests N_t .

$$N_t = N - N_a$$

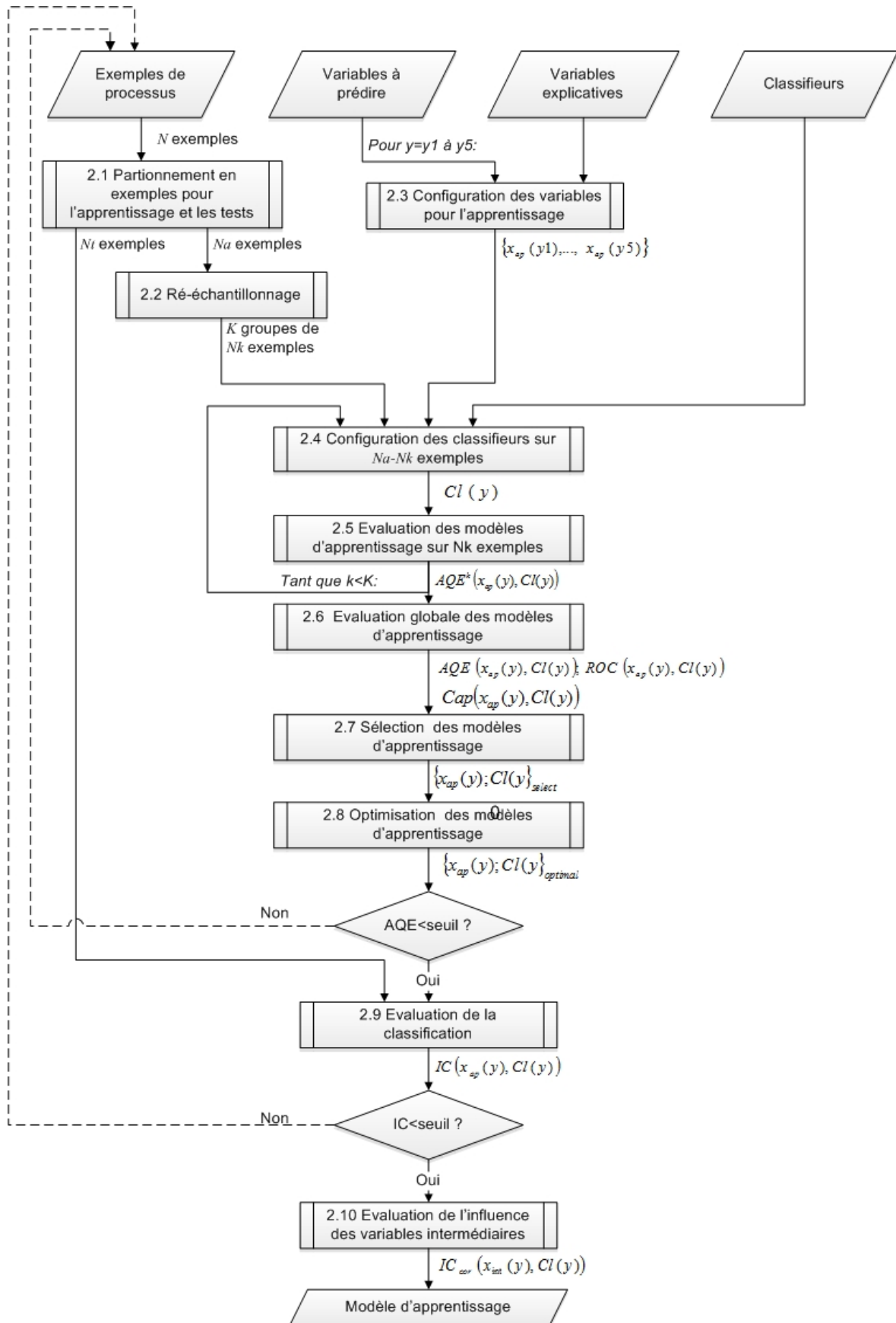


FIGURE 7.2 – Méthode pour la construction de modèles d'apprentissage.

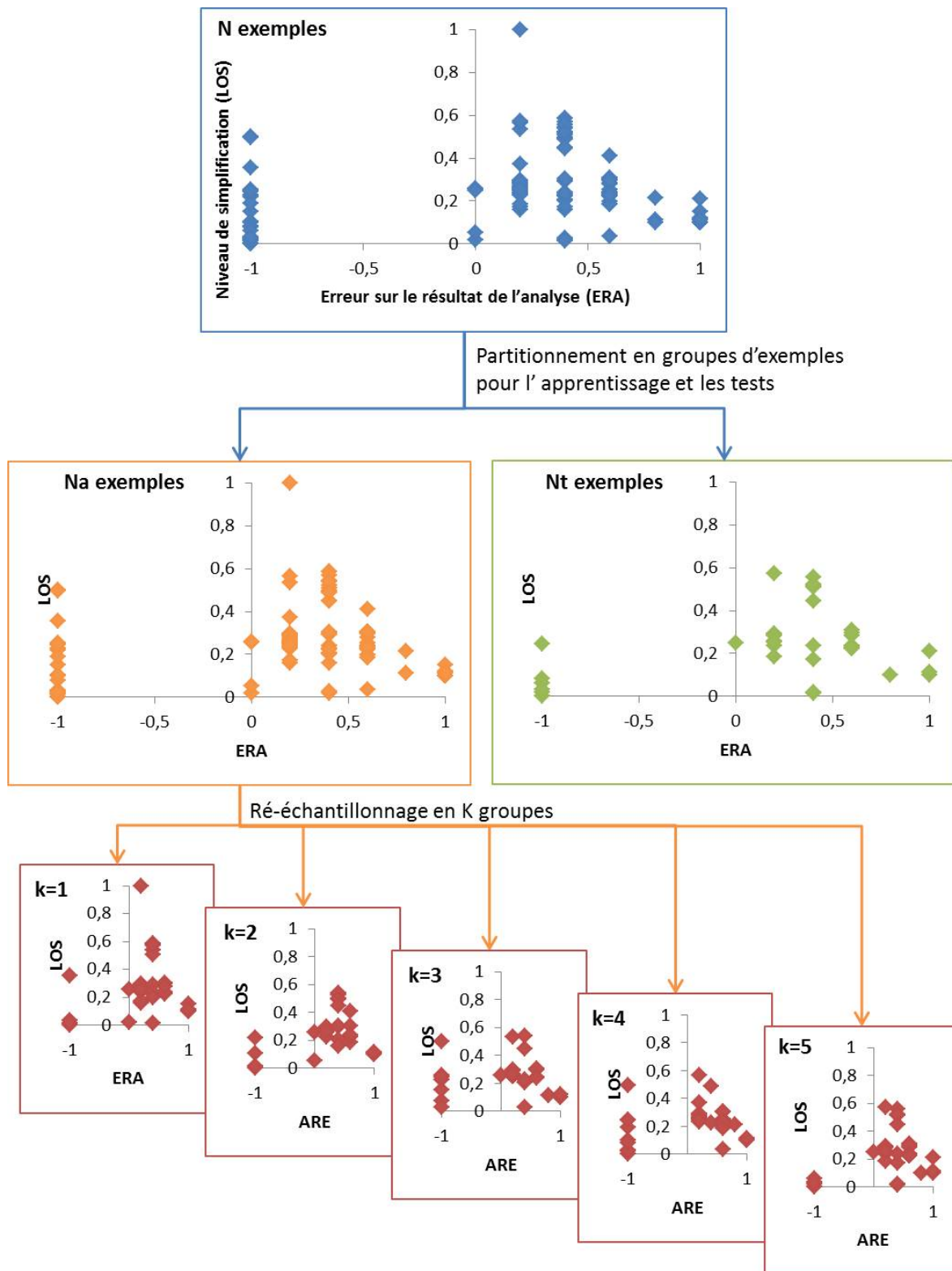


FIGURE 7.3 – Exemple de partitionnement des N exemples initiaux en N_a exemples pour l'apprentissage, N_t exemples pour les tests et ré-échantillonnage en $K=5$ groupes des cas pour la validation.

7.1.2. Implémentation et configuration des variables pour l'apprentissage

Les variables définies dans le chapitre précédent sont irrégulières, de formats multiples et peuvent contenir des valeurs aberrantes. Afin d'être intégrées dans la base d'apprentissage puis utilisées par les classifieurs, ces données doivent être transformées selon trois opérations principales (étape 2.3 de la figure 7.2) :

- la suppression de variables aberrantes ;
- la normalisation ;
- la discrétisation.

Chaque opération est détaillée dans les sous-sections suivantes. Pour certains classifieurs, les données peuvent être laissées brutes ce qui permet de réduire la durée de traitement (donc de l'apprentissage) et de réduire les imprécisions en apprenant sur des valeurs proches des valeurs réelles. Par contre, certaines architectures de classifieurs nécessitent que les données soient normalisées et/ou discrétisées.

Si le nombre de variables d'entrée n'est pas trop important, les classifieurs de type arbre de décision permettent d'identifier les variables les plus déterminantes, la sélection de variables explicatives devient alors facultative.

Il est donc nécessaire d'associer à chaque classifieur une configuration de variable adaptée.

7.1.2.1 Suppression des valeurs aberrantes

Une variable aberrante est une donnée dont la valeur est incorrecte ou très éloignées de la majorité des autres (figure 7.4 (a)). Les données aberrantes sont dues à des erreurs de saisie ou de calcul.

Si elles ne sont pas filtrées, ces variables risquent d'augmenter au final les erreurs de prédiction. En effet, puisque des opérations de traitement des données telle que la normalisation (sous-section suivante) utilisent leurs valeurs maximales et minimales, le résultat de la normalisation risque donc d'être fortement impacté. De plus, certaines méthodes de classification et certains critères utilisés pour sélectionner les variables explicatives présentés dans la section 6.3.2 sont très sensibles aux valeurs aberrantes.

Les valeurs aberrantes peuvent être identifiées grâce à des méthodes de statistiques univariées ou de tris à plat. Concrètement, une analyse visuelle des données suffit. Ces données doivent être supprimées de la base ou remplacées par des valeurs

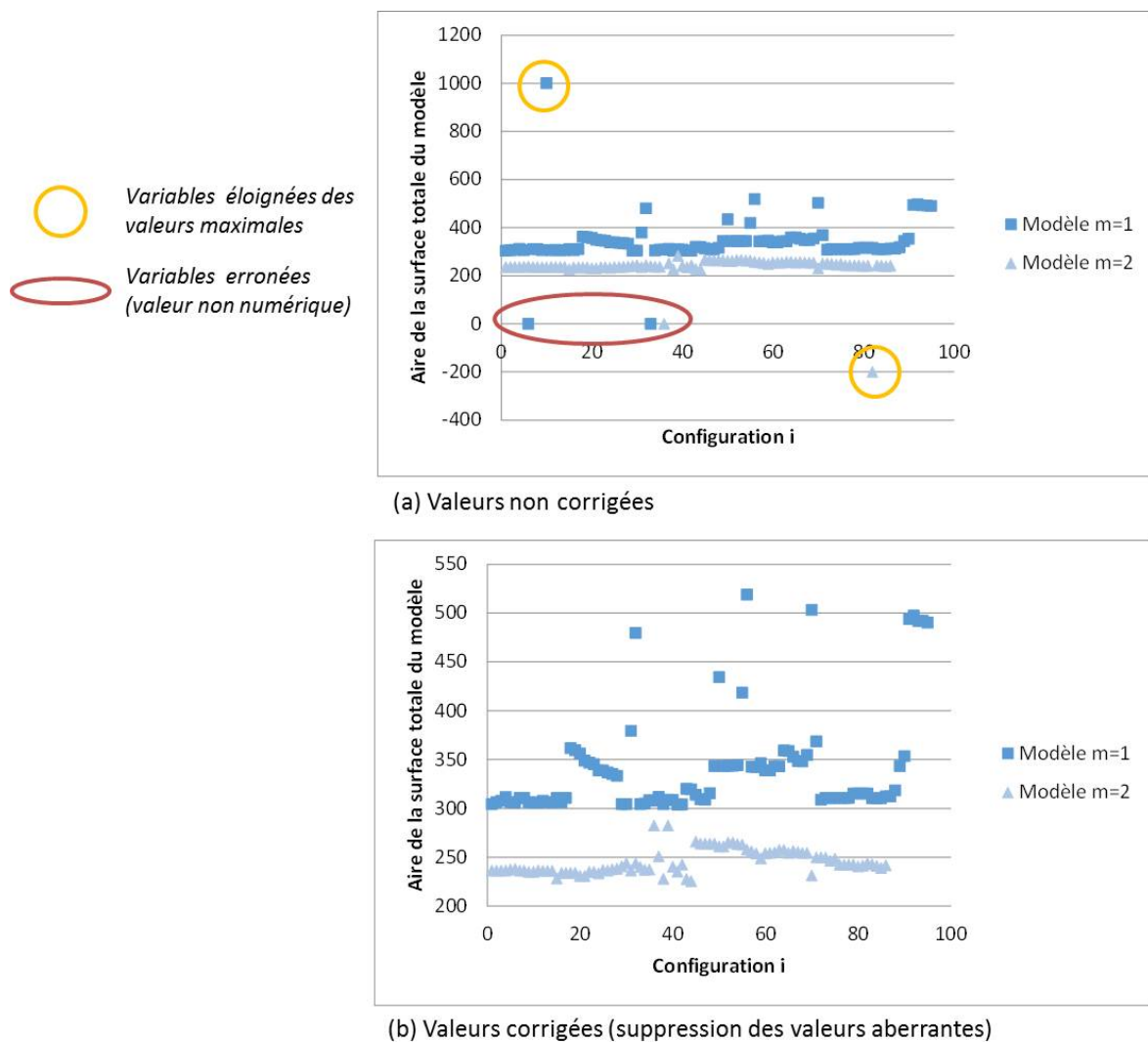


FIGURE 7.4 – Exemple de valeurs aberrantes (variable aire du modèle global $Aire_i^m$). (a) Variables non corrigées. (b) variables aberrantes supprimées ou corrigées.

7.1.2.2 Normalisation

Lorsque des variables numériques ont des unités, des ordres de grandeur ou des intervalles irréguliers, comme sur les exemples du tableau 7.1, des relations sont difficiles à trouver entre ces variables.

La normalisation consiste à ramener l'ensemble des valeurs d'une variable dans un intervalle compris entre : 0 et +1 (la valeur minimale de la variable sera ramenée à 0 et la valeur maximale à 1) ou entre -1 et +1 pour les valeurs réelles telles que les bénéfices (dans ce cas la valeur de 0 correspond à un bénéfice nul soit à une modification nulle de la géométrie).

CH	Modèle	Valeur	Aire _i ^m (mm ²)	Triangles _i ^m	Comp.volume _i ^m
	m = 1	Borne supérieure	519	673152	0,017
		Borne inférieure	304	1980	-0,014
	m = 2	Borne supérieure	282	356216	0,080
		Borne inférieure	236	12522	-0,0002

TABEAU 7.1 – Exemples de valeurs minimales et maximales des variables de deux modèles avant normalisation.

Cette opération facilite l'interprétation des données et l'obtention de relations entre les variables. Elle est indispensable lorsqu'on utilise des classifieurs de type réseaux de neurones.

La base de données contient plusieurs modèles globaux m qui peuvent être de tailles différentes. L'opération de normalisation peut être appliquée soit globalement à l'ensemble des données de la base (tableau 7.2), soit modèle par modèle (tableau 7.3). Dans ce dernier cas, la notion de taille relative entre les différentes configuration du modèle est conservée. On montrera dans la section "Construction de la base pour l'apprentissage" du chapitre "Résultat" que la normalisation sur l'ensemble des données est nécessaire pour prédire un coût et que la normalisation par modèle est préférable pour prédire l'impact de la simplification sur le résultat de l'analyse.

Modèle	Aire _i ^m (mm ²)	Triangles _i ^m	Comp.volume _i ^m
m = 1	519 → 1	673152 → 1	0,017 → 0,21
	304 → 0,24	1980 → 0	-0.014 → -1
m = 2	282 → 0,16	356216 → 0,53	0,080 → 1
	236 → 0	12522 → 0,016	-0,0002 → -0,003

TABEAU 7.2 – Exemples de normalisation globales sur l'ensemble de la base. Fonction de normalisation : $valeur_{initiale} \rightarrow valeur_{normalisée}$.

Modèle	Aire _i ^m (mm ²)	Triangles _i ^m	Comp.volume _i ^m
m = 1	519 → 1	673152 → 1	0,017 → 1
	304 → 0	1980 → 0	-0.014 → -1
m = 2	282 → 1	356216 → 1	0,080 → 1
	236 → 0	12522 → 0	-0,0002 → -1

TABEAU 7.3 – Exemples de normalisations par modèles. Fonction de normalisation : $valeur_{initiale} \rightarrow valeur_{normalisée}$.

Lorsqu'un nouveau cas se présente, le traitement de l'ensemble des données doit être relancé pour que l'ensemble des valeurs soient distribuées dans un intervalle compris entre 0 et +1 (ou -1 et +1).

Les objectifs de cette opération sont de discrétiser les variables numériques continues et de reclasser des variables discrètes réparties initialement dans un grand nombre de classes. Pour cela, la variable est partitionnée en un nombre donné de classes à partir d'intervalles proposés par l'utilisateur, ou bien à l'aide d'algorithmes qui optimisent la partition des données comme celui proposé par Fayyad et Irani [Fayyad 1993] nommé MDLPC, ou encore à l'aide de méthodes d'apprentissage non supervisées comme la méthode des K-moyennes.

Discrétiser correctement des données, consiste à choisir le nombre de classes et les bornes de classe optimales. Pour cela, on cherche à avoir un nombre minimal d'individus par classe, une variance minimale à l'intérieur de chaque classe, et une variance maximale entre les différentes classes.

Discrétisation des variables d'entrée

Les variables d'entrée textuelles ont déjà été réparties dans plusieurs classes, elles ne sont donc pas concernées par ce traitement.

La représentation des variables étant vectorielle, les variables d'entrées numériques sont discrétisées par défaut, le nombre de classes est égal au nombre d'exemples de la base (dans l'exemple sur la figure 7.5 (a) représentant la variable "aire du modèle" le nombre de classes est égal à 181).

La discrétisation permet d'accélérer la classification et d'améliorer sa qualité en limitant les risques de surapprentissage.

Le nombre de classes est déterminé à l'aide des indices de Huntsberger ($k1$) ou de Brooks-Carruthers ($k2$) qui calculent le nombre de classes N_{cl} pour un nombre N d'exemples :

$$k1 : N_{cl} = 1 + 3,3 \cdot \log_{10}(N)$$

$$k2 : N_{cl} = 5 \cdot \log_{10}(N)$$

Les bornes sont ensuite définies à l'aide d'une méthode de partition (répartition régulière, MDLPC ou méthode des k-moyennes). Si le nombre minimal d'individus par classe n'est pas suffisant (généralement 2 ou 3), la méthode est rejetée et/ou un nouveau nombre de classes est proposé.

Un coefficient F est calculé à l'aide du test de Fisher basé sur la comparaison des variances inter classes et intra classes. Si la valeur de F est acceptable, la distribution est acceptée.

$$F = \frac{Variance_{inter}}{Variance_{intra}}$$

Pour la variable "aire du modèle" de l'exemple sur la figure 7.5 les indices de Huntsberger et de Brooks-Carruthers préconisent un nombre de classes compris entre 9 et 11. La figure montre en (b), (c) et (d) les répartitions obtenues pour 10 classes respectivement avec des intervalles réguliers, avec l'algorithme MDLPC et avec la méthode des k-moyennes. La méthode des K-moyennes est retenue, c'est celle qui propose le meilleur score au test de Fisher et aucune classe avec un seul individu.

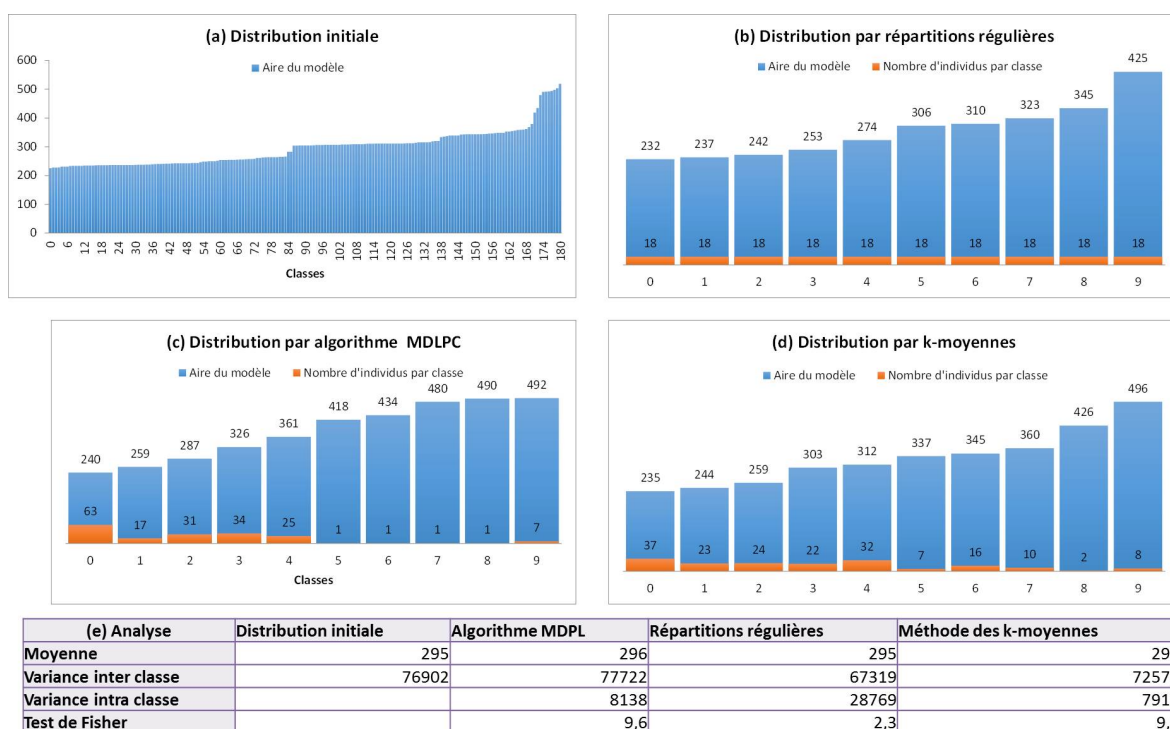


FIGURE 7.5 – Exemple de distribution. Variable aire du modèle CAO. (a) Valeurs initiales. (b) Distribution obtenue par répartitions régulières dans 10 classes. (c) Distribution obtenue par l'algorithme MDLPC [Fayyad 1993]. (d) Distribution obtenue avec la méthode des k-moyennes. (e) Calculs des variances inter / intra et du test de Fisher.

Discrétisation des variables à prédire

Les variables à prédire ont été identifiées dans la section 6.2.3. L'objectif n'est pas de prédire des valeurs précises de l'erreur sur le résultat de l'analyse ou du coût des opérations de préparation et de simulation, mais d'estimer la qualité globale du processus de préparation.

Ces variables sont donc discrétisées et réparties dans N_{cl} classes. Le nombre de classes dépendra de la précision de prédiction souhaitée, usuellement ce nombre est de 7 classes (section 4.2 "Principes de l'apprentissage artificiel").

CHAPITRE 7 APPRENTISSAGE ET PRÉDICTIONS SUR DES NOUVEAUX CAS
 Les bornes sont définies à partir de règles définies par l'utilisateur ou bien à l'aide des méthodes présentées dans la sous-section précédente.

Par exemple, concernant l'estimation de l'erreur sur le résultat de l'analyse, les valeurs sont réparties dans 7 classes dont les bornes ont été proposées par l'utilisateur. La classe "0" correspond aux cas dont l'erreur est négligeable, la dernière classe "1" correspond aux cas dont l'erreur est supérieure à un seuil donné, une classe "-1" a été ajoutée pour les cas non simulables (table 7.4 colonne 2).

Concernant les calculs de coût, la classe "1" correspond aux cas dont le coût est maximum et la classe "0" aux cas dont le coût est minimum (table 7.4 colonnes 3 et 4).

Classe	Erreurs sur le résultat de l'analyse	Coûts
-1	Simulation impossible	Simulation impossible
0	Erreur négligeable	Coût minimal
$1/(N_{cl} - 2)$	Erreur faible	Coût faible
...	...	...
$(N_{cl} - 3)/(N_{cl} - 2)$	Erreur importante	Coût important
1	Erreur inacceptable	Coût maximal

TABEAU 7.4 – Répartition des valeurs des variables à prédire.

7.1.3. Configuration des classifieurs

Les différents classifieurs à tester ont été identifiés dans la section 4.3, il s'agit de réseaux de neurones, d'arbres de décisions, de machines à vecteur de support, de classifieurs k-moyenne ou de classifieurs Naives Bayes.

7.1.3.1 Identification des paramètres des classifieurs

Configurer un classifieur consiste à déterminer les paramètres des algorithmes en minimisant une fonction coût selon la méthode présentée dans la section 4.5.2 par validation croisée *k-fold cross-validation*.

Le Tableau 7.5 récapitule, pour les classifieurs pré-sélectionnés, les principales architectures et leurs paramètres.

La méthode de validation croisée *k-fold cross-validation* consiste tout d'abord à construire un classifieur $Cl(y)$ sur un groupe de $N_a - N_k$ exemples (étape 2.4 de la figure 7.2).

Puis, l'erreur quadratique des prédictions $AQE^k(x_{ap}(y), Cl(y))$, définie dans la section 4.5.2, est calculée sur le groupe d'exemples N_k (étape 2.5 de la figure 7.2). Ainsi, l'échantillon utilisé pour l'apprentissage est bien distinct de celui utilisé pour la validation.

L'opération est répétée K fois. L'erreur globale $AQE(x_{ap}(y), Cl(y))$ est calculée sur la moyenne des erreurs des K évaluations.

7.1.3.2 Sélection des modèles d'apprentissage

Les modèles d'apprentissage $\{x_{ap}(y); Cl(y)\}$ sont évalués à partir de critères de qualité des prédictions et de durée de mise en œuvre de l'apprentissage (étape 2.6 de la figure 7.2).

Les critères de qualité des prédictions sont l'erreur quadratique moyenne AQE obtenue précédemment par validation croisée et de la surface sous la courbe ROC (*Receiver Operating Characteristic curve*) présentée dans la section 4.5.2 .

Les critères de durée de mise en œuvre de l'apprentissage $Dap(x_{ap}(y), Cl(y))$ sont déterminés à partir de la durée de préparation des données et de la durée de configuration des classifieurs.

Classifieur	Paramètres d'architecture	Paramètres de configuration
Réseau de neurones Multi-Layer Perceptron	— Type de fonction de transfert — Nombre de couches cachées de neurones — Nombre de neurones par couche	— Méthode de détermination des poids — Poids des neurones
Machine à vecteur de support	— Type d'algorithme — Type de fonction noyau kernel	— Paramètre de tolérance (biais / variance) — Paramètres spécifiques aux types d'algorithmes — Poids sur les classes
Arbres de décision type C.4.5	— Elagage	— Nombre minimal d'instance par feuille
Classifieurs Naives Bayes	— Estimateur Kernel ou distribution normalisée	
Classifieurs K-moyenne	— Nombre de classes	

TABLEAU 7.5 – Paramètres à déterminer pour la configuration des classifieurs

Les modèles d'apprentissage dont l'erreur $AQE(x_{ap}(y), Cl(y))$ est minimale et la surface sous la courbe ROC est proche de 1 sont sélectionnés (étape 2.7 de la figure 7.2).

En cas d'égalité, le classifieur retenu sera celui dont la durée $Dap(x_{ap}(y), Cl(y))$ est minimale.

7.1.3.3 Optimisation des modèles d'apprentissage

L'architecture du classifieur précédemment sélectionné $Cl(y)$ a été définie en fonction des paramètres de la colonne 2 du tableau 7.5. Les paramètres du classifieur (colonne 3) doivent être déterminés précisément.

Les paramètres des classifieurs sont optimisés à l'aide de la méthode itérative d'évaluation du gradient de la fonction coût présentée dans la section 4.5.2 "construction des modèles d'apprentissage".

Le modèle d'apprentissage est optimisé (étape 2.8 de la figure 7.2) en configurant les variables explicatives $x_{ap}(y)$ pour que l'erreur quadratique moyenne AQE soit inférieure à un seuil imposé par l'utilisateur et que la surface sous la courbe ROC soit proche de 1.

On notera $\{x_{ap}(y); Cl(y)\}_{optimal}$, le modèle d'apprentissage ainsi configuré.

Lorsque la valeur seuil de AQE $(x_{ap}(y), Cl(y))$ ne peut être atteinte pour aucun modèle, les variables d'entrées ne sont pas suffisamment déterminantes. Il sera nécessaire de reprendre la phase de sélection des variables d'entrée et d'en proposer de nouvelles.

7.1.4 Évaluation de la classification

7.1.4.1 Indicateur de confiance de la classification

Les modèles d'apprentissage $\{x_{ap}(y); Cl(y)\}$ précédemment configurés sont évalués sur les Nt exemples réservés aux tests à l'aide d'un indicateur de confiance (étape 2.9 de la figure 7.2).

L'indicateur de confiance $IC(x_{ap}(y), Cl(y))$ d'un modèle d'apprentissage correspond au nombre ou pourcentage de cas mal classés. On distinguera le pourcentage de cas mal classés acceptables et inacceptables. Une erreur de classification est considérée comme inacceptable lorsque la valeur prédite est plus optimiste que la valeur réelle. Le tableau 7.6 donne des exemples de prédictions acceptables et inacceptables pour la variable à prédire y_3 "erreur sur le résultat de l'analyse".

Valeur réelle de la variable	Valeur prédite	Type de cas mal classé
Classe -1 "simulation impossible"	Classe 0 "erreur négligeable"	Inacceptable
Classe 0 "erreur négligeable"	Classe 0.2 "erreur faible"	Acceptable
Classe 0 "erreur négligeable"	Classe 0.8 "erreur importante"	Inacceptable

TABLEAU 7.6 – Exemples de prédictions acceptables et inacceptables pour la variable y_3 (erreur sur le résultat de l'analyse).

L'indicateur de confiance des cinq principales variables à prédire doit être infé-

CHAPITRE 7. APPRENTISSAGE ET PRÉDICTIONS SUR DES NOUVEAUX CAS
 rieur à un seuil imposé par l'utilisateur (généralement 90% d'erreurs inacceptables de classification et moins de 5% d'erreurs acceptables). Si cela n'est pas le cas, le choix des variables explicatives et leur traitement devra être modifié.

7.1.4.2 Influence de l'estimation des variables intermédiaires

Parmi les variables explicatives utilisées lors de l'apprentissage certaines ne sont pas connues pour un nouveau cas (variables intermédiaires) et doivent être estimées par apprentissage. Les erreurs de classification sur les variables intermédiaires peuvent influencer le résultat de la classification des principales variables à prédire. Les valeurs réelles des variables intermédiaires sont alors remplacées dans la base d'exemples à tester par leurs valeurs obtenues par apprentissage. L'indicateur de confiance de la classification des principales variables à prédire sera ainsi plus juste.

Les variables intermédiaires interviennent à différents niveaux dans l'estimation des principales variables à prédire. On appellera variable de niveau 1 celles qui sont utilisées pour l'estimation d'une variable principale et variable intermédiaire de niveau 2 celles qui sont utilisées pour estimer une variable de niveau 1.

Plusieurs scénarios peuvent être envisagés (figure 7.6).

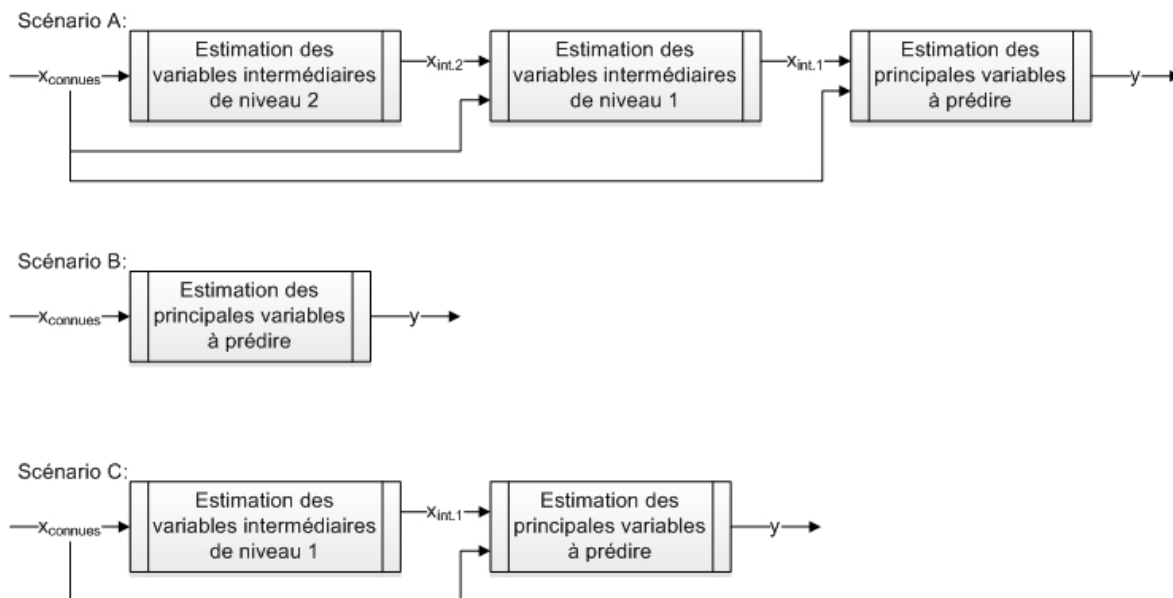


FIGURE 7.6 – Scénarios envisageables pour la prise en compte des variables intermédiaires.

- *Scénario A* : toutes les variables explicatives intermédiaires x_{int} sont estimées par apprentissage et utilisées pour estimer les variables principales y et intermédiaires de niveaux inférieurs.
- *Scénario B* : les principales variables à prédire y sont estimées uniquement à partir des variables connues $x_{connues}$.

elles sont utilisées uniquement à partir des variables connues $x_{connues}$.

Pour chacune des 5 variables à prédire, il faudra choisir le scénario qui donne le meilleur indicateur de confiance corrigé $IC_{cor}(x_{int}(y), Cl(y))$ après remplacement des valeurs réelles par leurs valeurs estimées (étape 2.10 de la figure 7.2).

7.2 Utilisation sur un nouveau cas

La notion de nouveau cas a été définie dans la section 5.1. Il peut s'agir d'un nouveau modèle CAO, d'un nouveau processus de préparation ou d'un nouvel objectif de simulation.

Concernant les nouveaux cas de type "nouveau processus de préparation", l'objectif est d'estimer les critères de qualité d'un processus de préparation proposé par l'utilisateur.

Concernant les nouveaux cas de type "nouveau modèle global" ou "nouvel objectif de simulation", l'objectif est d'identifier le meilleur processus parmi une série de processus testés.

7.2.1 Approche générale pour l'évaluation de la qualité d'un processus de préparation

La figure 7.7 synthétise les principales étapes de la prédiction de la qualité d'un processus de préparation d'un modèle pour la simulation sur un nouveau cas.

Les données propres au nouveau cas sont implémentées dans la base de données. Puis, les classifieurs définis précédemment sont utilisés pour prédire les variables intermédiaires déterminantes et finalement les principales variables de sortie au niveau de chaque sous-ensemble et au niveau de l'ensemble global.

Implémentation des données extraites du nouveau cas

L'objectif est de construire des vecteurs de variables explicatives qui caractérisent les variables à prédire.

Les données disponibles qui peuvent être extraites sont :

- la description du processus de préparation (outillage, séquençage des opérations et paramètres de réglage) proposée par l'utilisateur et représenté par le

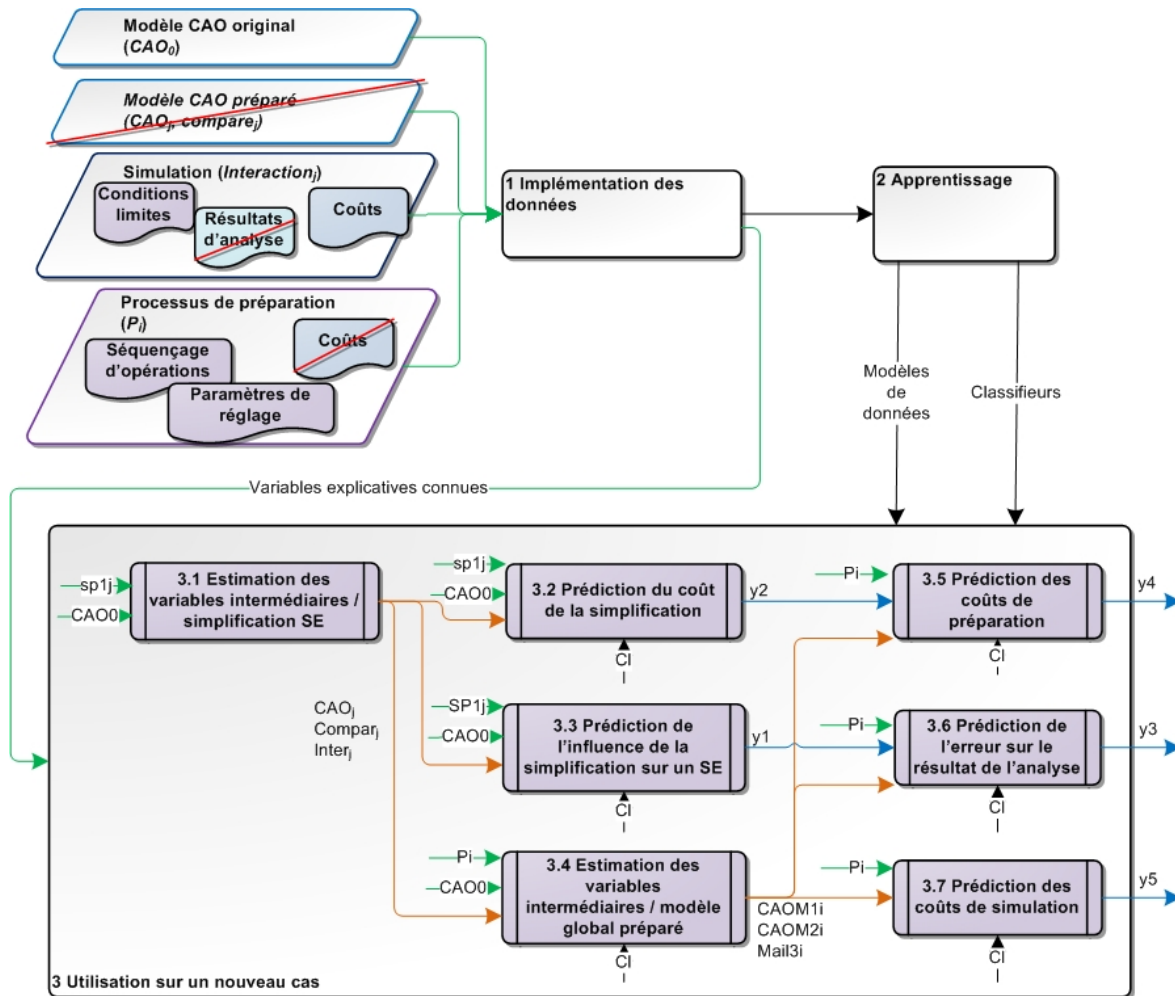


FIGURE 7.7 – Approche générale pour l'utilisation de l'apprentissage sur un nouveau cas.

vecteur P_i^m au niveau de l'ensemble global M^m et $sp1_j^n$ au niveau de la simplification des sous-ensembles SE^n ;

- la description du modèle CAO global original (vecteur de variables $CAOM1_0^m$) et la description des sous-ensembles (vecteur CAO_0^n) ;
- la description du cas de simulation (objectifs de la simulation, conditions limites, grandeurs physiques et principes mis en œuvre) représentée par le vecteur $Inter_0^n$.

Lorsqu'un nouveau cas se présente les données disponibles sont extraites et traitées (correction des valeurs aberrantes et éventuellement normalisation et discrétisation) selon la même méthode que pour les variables d'apprentissage.

Estimation des variables intermédiaires

Les variables intermédiaires relatives aux sous-ensembles n simplifiés selon une configuration j sont estimées en premier à partir des données connues. Il s'agit :

- des caractéristiques géométriques du modèle simplifié (vecteur CAO_j^n) ;

des facteurs qui caractérisent les critères de qualité sous les facteurs de comparaison des modèles simplifiés et originaux (vecteur $Compar_j^n$) et d'influence de la simplification sur le cas de simulation (vecteur $Inter_j^n$).

Au niveau de l'ensemble global, les variables intermédiaires sont estimées à partir des données connues et des variables intermédiaires relatives aux sous-ensembles précédemment estimées. Il s'agit des caractéristiques géométriques de l'ensemble m et des facteurs qui caractérisent les critères d'évaluation du processus de préparation i pour les modèles simplifiés (vecteur $CAOM1_i^m$), adaptés (vecteur $CAOM2_i^m$) et préparés (vecteur $Mail_i^m$).

Toutes les variables intermédiaires ne seront pas forcément estimées et plusieurs stratégies basées sur les trois scénarios définis dans la section précédente sont possibles. Ces stratégies seront détaillées dans les sections suivantes.

Estimation des critères de qualité d'un processus de préparation

La qualité de la simplification d'un sous-ensemble est évaluée (variables $y1$ et $y2$) pour chaque sous-ensemble n simplifié selon un sous-processus de simplification j proposé par l'utilisateur. Les variables $y1$ et $y2$ sont estimées à partir des variables connues et intermédiaires relatives aux sous-ensembles. Les étapes pour la prédiction de la qualité de la simplification d'un sous-ensemble sont décrites dans la section 7.2.2.

La qualité globale d'un processus de préparation P_i^m est évaluée à partir de l'erreur globale sur le résultat de l'analyse $y3$, des coûts de préparation $y4$ et de simulation $y5$. Ces trois variables sont estimées à partir des variables connues et intermédiaires relatives à l'ensemble global. Les étapes pour la prédiction de la qualité globale d'un processus de préparation sont décrites dans la section 7.2.3.

7.2.2 Prédiction de la qualité de la simplification d'un sous-ensemble

Il s'agit d'estimer les deux critères de qualité de la simplification d'un sous-ensemble qui sont :

- l'influence $y1$ de la simplification sur le résultat de l'analyse,
- le coût $y2$ de la simplification

La figure 7.8 propose plusieurs stratégies pour estimer la qualité de la simplification d'un sous-ensemble.

La stratégie A est basée sur le scénario 1 présenté dans la section précédente. Les deux principales variables à prédire sont estimées à partir de toutes les variables explicatives connues puis intermédiaires.

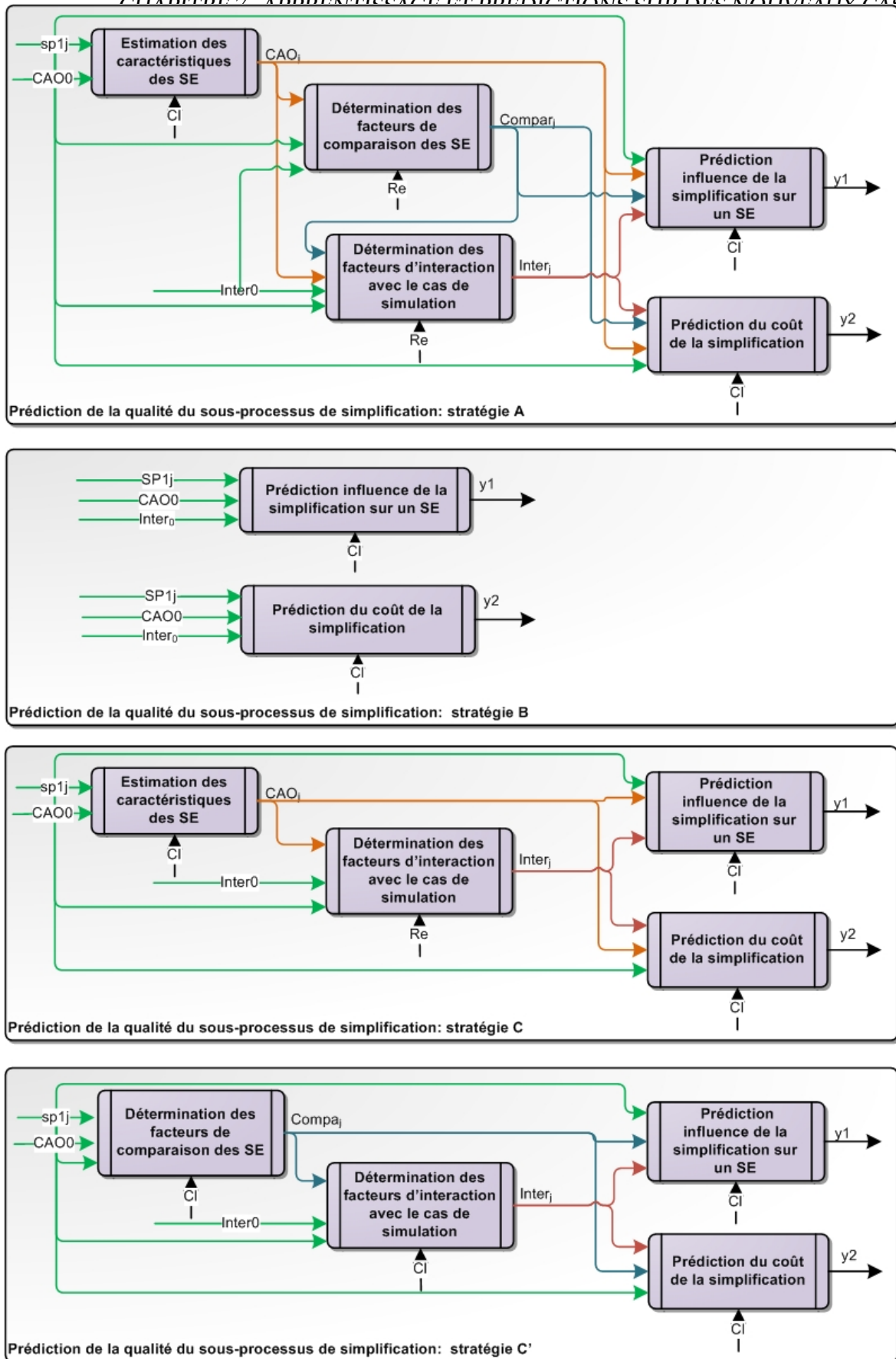


FIGURE 7.8 – Stratégies pour la prédiction de la qualité de la simplification d'un sous-ensemble sur un nouveau cas à l'aide de classifieur (Cl) ou de relations connues (Re).

Les variables connues sont la description du sous-processus de simplification $sp1_j^n$ selon la configuration j , les caractéristiques du modèle CAO original CAO_0^n et les facteurs d'interaction entre le modèle original et le cas de simulation $Inter_j^0$.

- Les variables intermédiaires sont les caractéristiques des sous-ensembles simplifiées CAO_j^n , les facteurs de comparaison des modèles simplifiés et originaux $Comp_j^n$, et les facteurs d'interaction avec le cas de simulation $Inter_j^n$.

Les caractéristiques des sous-ensembles simplifiées CAO_j^n utilisés pour la prédiction sont estimées par apprentissage à partir des variables connues $sp1_j^n$ et CAO_0^n . Connaissant les valeurs de CAO_j^n , les facteurs de comparaison des modèles $Comp_j^n$ et d'interaction avec le cas de simulation $Inter_j^n$ sont déterminés à l'aide des relations présentées dans la section 6.2.5.

La stratégie B est basée sur le scénario 2, les deux principales variables à prédire sont estimées uniquement à partir des variables connues $sp1_j^n$ et CAO_0^n .

Les stratégies C et C' sont basées sur le scénario 3, les deux principales variables à prédire sont estimées à partir de variables :

- *connues* qui sont la description du sous-processus de simplification $sp1_j^n$ selon la configuration j et les caractéristiques du modèle CAO original CAO_0^n ;
- et *intermédiaires* qui sont les facteurs d'interaction avec le cas de simulation $Inter_j^n$ et les caractéristiques des sous-ensembles simplifiées CAO_j^n ou les facteurs de comparaison des modèles simplifiés et originaux $Comp_j^n$.

Les variables intermédiaires CAO_j^n ou $Comp_j^n$ sont estimées à partir des variables connues sans utiliser de variables intermédiaires contrairement à la stratégie A

Les facteurs $Inter_j^n$ sont calculés soit à partir des variables connues et intermédiaires pour la stratégie C, soit estimées par apprentissage à partir des variables connues et des facteurs $Comp_j^n$ pour la stratégie C'.

Des expérimentations devront définir la meilleure stratégie parmi celles qui ont été proposées, c'est à dire celle dont l'erreur de prédiction $AQE(x_{ap}(y), Cl(y))$ est minimale pour la majorité des classifieurs $Cl(y1)$ et $Cl(y2)$ testés.

7.2.3 Prédiction de la qualité globale d'un processus de préparation

La qualité globale d'un processus de préparation est évaluée à partir de trois critères :

- y_3 : erreur sur le résultat de l'analyse ;
- y_4 : coût de la préparation ;
- y_5 : coût de la simulation.

La figure 7.9 propose les stratégies envisageables pour la prédiction de la qualité globale d'un processus de préparation.

La qualité de la simplification de tous les sous-ensembles qui composent l'ensemble global m a été précédemment estimée. Selon la stratégie choisie au niveau local, des variables intermédiaires peuvent avoir été déterminées.

La stratégie A consiste à estimer les trois critères de qualité d'un processus de préparation à partir :

- de variables connues qui décrivent le processus de préparation P_i^m et caractérisent les modèles CAO original $M1_0$ et simplifié $M1_i$ obtenues simplement à partir de l'ensemble des caractéristiques de chaque sous-ensemble ;
- de variables intermédiaires estimées par apprentissage qui caractérisent les modèles adapté $M2_i$ et maillé $M1_i$;
- de l'influence de la simplification globale représentée par les deux variables $Y1^m$ et $Y2^m$ calculées à partir de toutes les influences sur le résultat de l'analyse $y1^n$ et coûts de simplification $y2^n$ estimés sur chaque sous-ensemble n .

$$Y1^m = \sum_{n=1}^N y1^n$$

$$Y2^m = \frac{\sum_{n=1}^N (y2^n \times Lmoy^n)}{Lmoy^m}$$

La stratégie B estime les trois critères de qualité uniquement à partir des variables connues pour un nouveau cas.

La stratégie A nécessite d'avoir préalablement estimé par apprentissage l'ensemble des variables intermédiaires extraites des modèles adaptés $M2_i$ et maillées $M3_i$. Pour éviter ces opérations, la stratégie C propose d'estimer les variables intermédiaires relatives aux modèles adaptés et préparés uniquement à partir des variables extraites des modèles originaux $M1_0$.

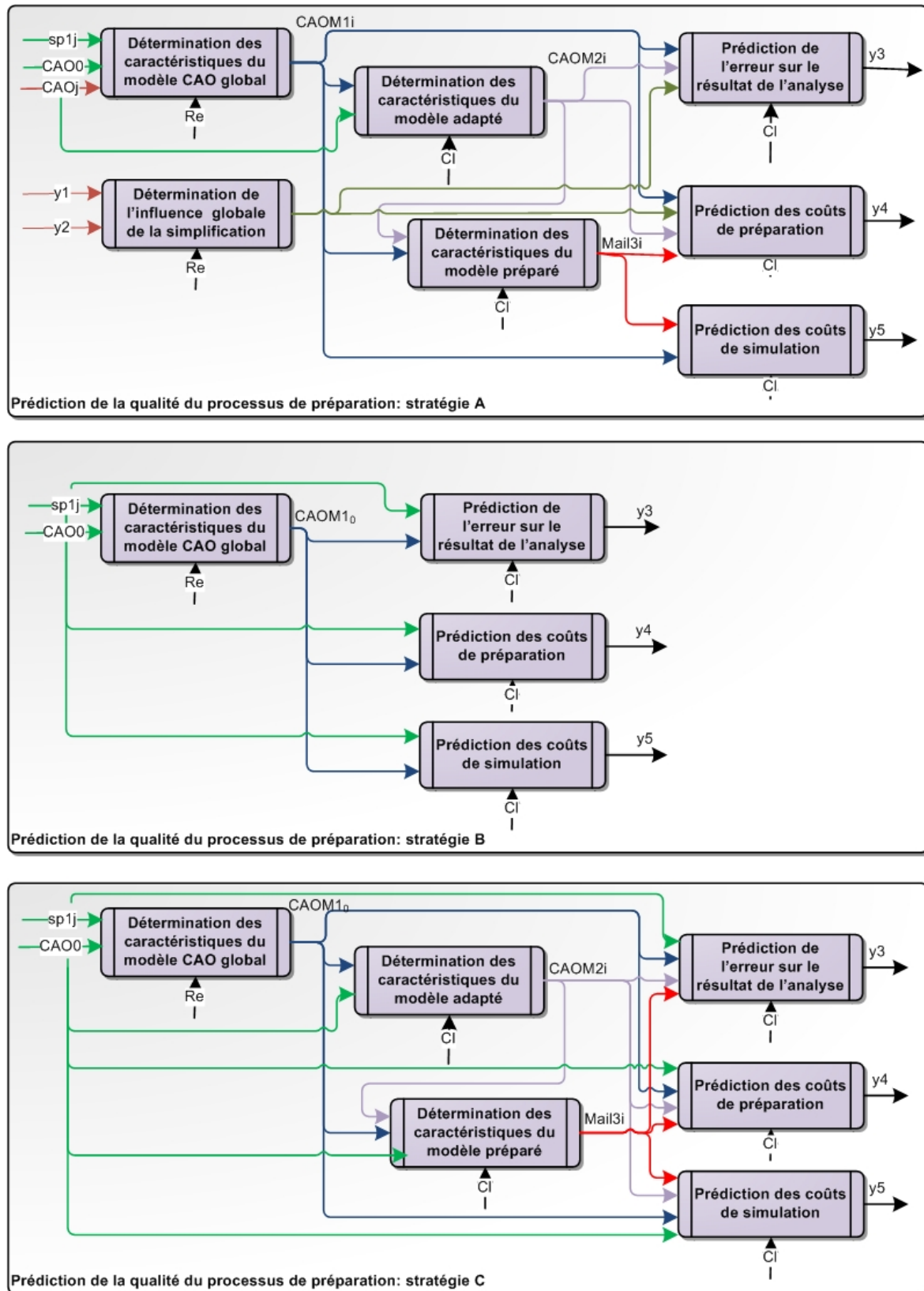


FIGURE 7.9 – Stratégies pour la prédiction de la qualité d'un processus de préparation sur un nouveau cas.

7.3 Évaluation des processus de préparation : aide à la décision

7.3.1 Indice de performance d'un processus de préparation

Les trois valeurs y_3 (erreur sur le résultat de l'analyse), y_4 (coût de la préparation) et y_5 (coût de la simulation) ont chacune un sens pour l'analyste. Il n'a pas nécessairement besoin de les combiner pour prendre une décision quant à la performance du processus de préparation évalué à partir de ces trois critères.

Combiner les estimations pondérées de l'erreur sur le résultats de l'analyse, du coût de préparation et du coût de simulation pour obtenir un indice de performance IP_i^m (équation 7.1) permet de comparer plus facilement les processus entre eux et dans de futurs travaux d'identifier un processus optimal.

$$IP_i^m = f_{IPm}(c_3 \times y_3^m, c_4 \times y_4^m, c_5 \times y_5^m) \quad (7.1)$$

Les coefficients c_3, c_4 et c_5 donnent les poids respectifs des indicateurs d'erreurs et de coûts dans le calcul de l'indice global. Ces coefficients sont proposés par l'analyse.

Un indice faible signifie que la qualité du processus de préparation est satisfaisante.

7.3.2 Identification du meilleur processus de préparation pour le simulation

Lorsqu'un nouveau modèle original ou un nouvel objectif de simulation se présente, les analystes recherchent le meilleur processus de préparation pour ce nouveau cas.

L'estimation a priori de la qualité du processus sans réaliser ni la préparation ni la simulation permet de tester un grand nombre de processus.

Pour un modèle m contenant N sous-ensembles, il existe jusqu'à $N \times 300$ processus de simplification différents. Il est possible de tester les $N \times 300$ processus de préparation et d'identifier celui dont l'indice de performance IP_i^m est minimal. Sinon, une recherche ciblée de ce meilleur processus peut être envisagée selon la méthode proposée sur la figure 7.10 et décrite dans les paragraphes suivants.

Identification du meilleur sous-processus de simplification

Cette méthode consiste dans un premier temps à identifier localement le meilleur

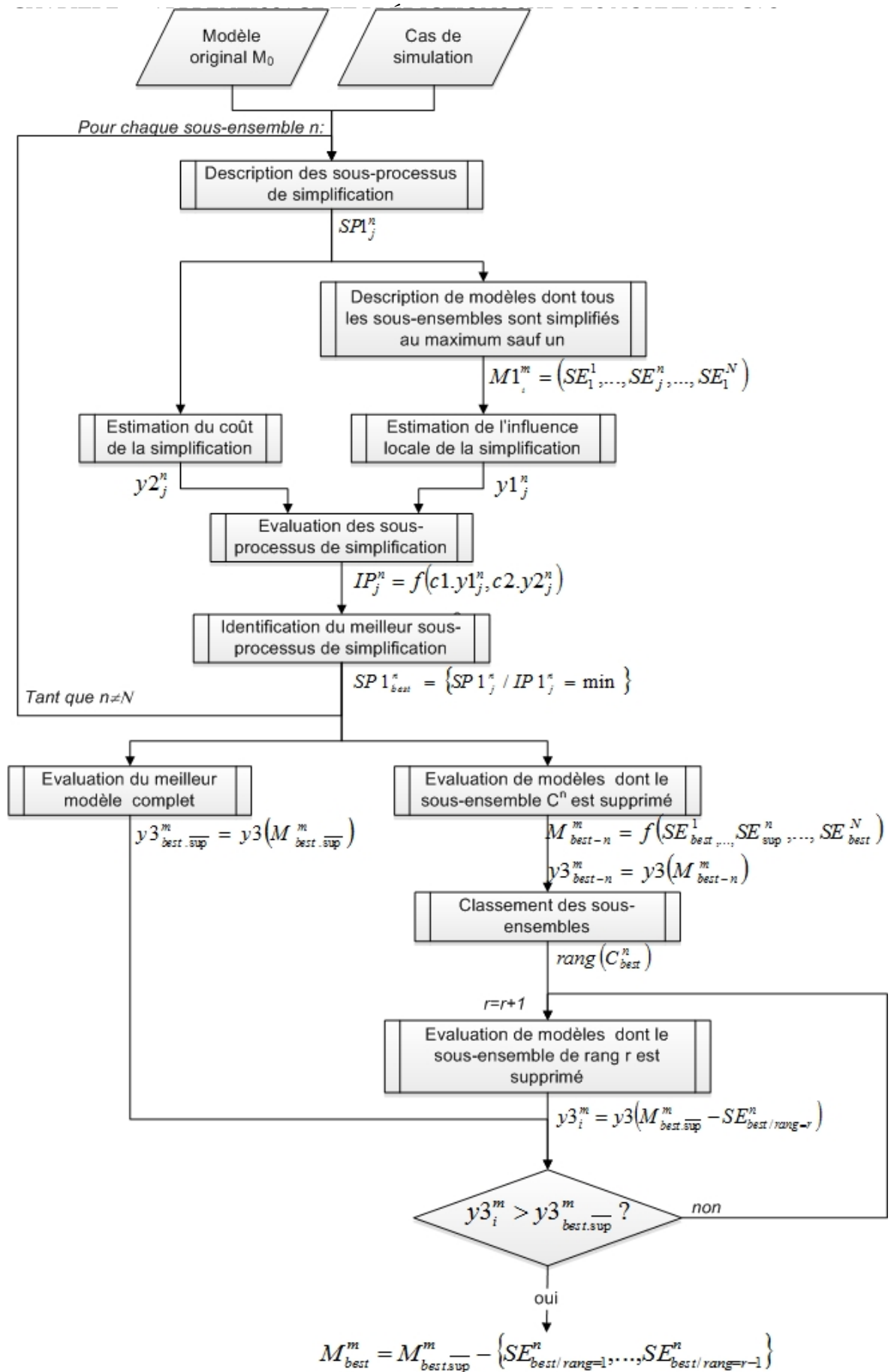


FIGURE 7.10 – Méthode de recherche du meilleur processus de préparation.

Pour chaque sous-ensemble SE_0^n , un ensemble de J sous-processus $SP1_j^n$ de simplification est proposé. Le coût de la simplification est estimé $y2_j^n$ pour chacun des sous-processus $SP1_j^n$.

L'influence de la simplification sur le résultat de l'analyse $y1_j^n$ est estimée sur des modèles dont tous les sous-ensembles sont simplifiés au maximum sauf le sous-ensemble SE_j^n qui est simplifié selon la configuration j .

Pour chaque sous-processus $SP1_j^n$, l'indice IP_j^n est déterminé à partir des estimations pondérées de l'influence $y1_j^n$ et du coût $y2_j^n$ de la simplification (équation 7.2). Le meilleur sous-processus de simplification $SP1_{best}^n$ est celui dont l'indice IP_j^n est minimal.

$$IP_j^n = f_{IPn} \left(c1 \times y1_j^n, c2 \times y2_j^n \right) \quad (7.2)$$

Identification du meilleur processus global de simplification

Le modèle ainsi obtenu n'est pas forcément le modèle optimal. Cependant cette méthode permet d'identifier rapidement un processus de préparation d'un modèle qui donnera un bon résultat d'analyse à moindre coût.

Le "meilleur" modèle global M_{best}^m est obtenu par concaténation des N sous-ensembles configurés selon leur meilleur sous-processus de simplification $SP1_{best}^n$ et par suppression des sous-ensembles dont l'influence sur le résultat de l'analyse est négligeable. Sachant que les coûts de simplification, d'adaptation et de maillage diminuent lorsqu'on supprime un composant, seul le critère d'erreur sur le résultat de l'analyse $y3_i^m$ est pris en compte pour comparer les performances de plusieurs processus.

Pour cela, l'erreur sur le résultat de l'analyse $y3_{best-n}^m$ est tout d'abord estimée pour chaque modèle M_{best-n}^m dont tous les sous-ensembles sont configurés selon leur "meilleure" configuration sauf un qui est supprimé. On estime ainsi pour chaque sous-ensemble SE_{best}^n l'influence de sa suppression sur le résultat de l'analyse.

Les sous-ensembles sont classés selon les valeurs de $y3_{best-n}^m$. Les sous-ensembles sont ensuite supprimés du modèle global un par un en fonction de leur classement.

En partant de l'hypothèse que l'erreur sur le résultat de l'analyse ne peut qu'augmenter lorsqu'on supprime un composant, le meilleur modèle global M_{best}^m est celui qui contient le moins de composants et dont l'estimation de y_3^m est identique à l'erreur sur le meilleur modèle $y_{best.sup}^m$ contenant l'ensemble des N sous-ensembles.

7.4 Conclusion sur l'apprentissage et les prédictions sur un nouveaux cas

Construire un modèle d'apprentissage consiste à configurer un couple "données pour l'apprentissage / classifieurs" en déterminant un ensemble de paramètres qui permettront d'estimer au plus juste une variable de sortie. Au niveau des données, il s'agira de partitionner les exemples en plusieurs groupes (apprentissage, validation et tests) et de définir la configuration de variables explicatives la plus adaptée au classifieur sélectionné. Au niveau des classifieurs, il s'agira d'optimiser leurs paramètres en minimisant une fonction coût.

Les principaux freins évoqués dans les chapitres sur l'état de l'art et le chapitre 5 concernaient :

- le nombre limité d'exemples de processus disponibles ;
- l'identification des variables explicatives utilisées pour estimer les variables à prédire (tout particulièrement les facteurs géométriques qui influencent le résultat de l'analyse) ;
- la prise en compte d'erreurs ou d'imprécisions de prédiction des variables intermédiaires inconnues pour un nouveau cas ;
- le choix des critères qui guident l'analyste dans le choix d'un processus de préparation.

Une méthode, basée sur le principe de la validation croisée, a été proposée (figure 7.2) afin de construire un modèle d'apprentissage à partir d'un nombre limité d'exemples.

Les modèles d'apprentissage sont évalués à l'aide d'un indicateur de confiance basé sur le nombre ou le pourcentage d'erreurs de classification. Cet indicateur permettra de vérifier le choix et la complétude des variables explicatives.

Une méthode a été proposée (figure 7.7) pour estimer la qualité d'un processus de préparation sur un nouveau cas. Cette méthode a été déclinée en plusieurs stratégies pour prendre en compte différemment les variables intermédiaires inconnues

La qualité d'un processus de préparation pour la simulation sera évaluée à l'aide d'un indice de performance. Lorsque le nouveau cas est un nouveau modèle CAO ou un nouvel objectif de simulation, une méthode a été proposée pour identifier rapidement le meilleur processus de préparation parmi un ensemble de processus évalués.

Des expérimentations devront définir l'ensemble des éléments nécessaires à la construction d'un modèle d'apprentissage, soit :

- le nombre minimal d'exemples pour l'apprentissage, la validation et les tests ;
- l'architecture de classifieur dont l'indicateur de confiance est le plus élevé ;
- les paramètres de configuration du classifieur ;
- le modèle de données le plus adapté au classifieur, soit :
 - le nombre de classes de la variable à prédire, ses bornes et le nombre minimal d'exemples par classe,
 - le niveau traitement des variables explicatives (normalisation et/ou distribution),
 - le choix des variables explicatives (seuil qui spécifie qu'une variable est déterminante ou non),
- les variables intermédiaires à utiliser, il s'agit de déterminer si :
 - toutes les variables intermédiaires sont utilisées (scénarios A des stratégies présentées sur les figures 7.8 et 7.9),
 - seules les variables connues sont utilisées (scénarios B des stratégies 7.8 et 7.9)
 - seules les variables intermédiaires déterminantes pour la prédiction des principales variables de sortie sont utilisées (scénarios C et C' des stratégies 7.8 et 7.9).

Plus largement, les expérimentations devront vérifier l'hypothèse générale qui consiste à évaluer a priori un processus de préparation à l'aide des techniques d'apprentissage artificiel en analysant les indicateurs de confiance des classifieurs. C'est tout l'objet du chapitre 8 suivant.

Chapitre 8

Expérimentations et résultats

La méthode proposée pour construire un modèle d'apprentissage et prédire la qualité d'un processus de préparation d'un modèle CAO a été appliquée à la préparation de modèles pour la simulation CFD (Computational Fluid Dynamics) couplée à un problème de transfert thermique dans une enceinte fermée. Ce chapitre a pour objectif de valider l'ensemble des hypothèses émises précédemment et de vérifier que les techniques d'apprentissage artificiel peuvent être utilisées pour évaluer a priori l'impact du processus de préparation sur les résultats et le coût de l'analyse.

La section 8.1 présente le contexte général des expérimentations, le cas de simulation, les opérations de préparation mises en œuvre et les données extraites des modèles 3D originaux et préparés.

La section 8.2 met en application les méthodes proposées précédemment pour construire une base d'exemples, sélectionner les variables explicatives déterminantes et configurer des modèles d'apprentissage (couples de données et de classifieurs configurés) capables d'estimer un indicateur de performance d'un processus de préparation.

Des tests de prédiction de l'impact de la simplification sur le résultat et le coût de l'analyse de nouveaux cas sont présentés dans la section 8.3. Ces résultats permettront de choisir la meilleure stratégie pour estimer la qualité d'un processus de préparation pour un nouveau cas. Les modèles d'apprentissage définis précédemment sont validés à partir d'indicateurs de confiance des classifieurs configurés. L'hypothèse qu'il est envisageable d'utiliser les techniques d'apprentissage artificiel et leurs limites est discutée en fin de section.

8.1 Contexte des expérimentations

L'objectif de cette section est de présenter dans un premier temps le contexte général des applications (objectifs de la préparation et modèles CAO originaux), puis de présenter les techniques utilisées et les données extraites pour chacune des phases de préparation des modèles (simplification, adaptation, maillage et simulation). La dernière sous-section concerne les outils et techniques utilisées pour l'apprentissage et les tests sur de nouveaux cas.

8.1.1 Présentation du cas de simulation

La présence de sources de chaleur dans une enceinte telle qu'un véhicule ou un bâtiment peut perturber le fonctionnement de composants embarquant des systèmes électroniques. La connaissance des champs de températures autour de ces composants permet d'optimiser leur positionnement.

Plus particulièrement, les expérimentations seront menées sur des "baies avioniques" (figure 8.1). La baie avionique est située dans la soute positionnée sous le poste de pilotage d'un aéronef. Les armoires électroniques positionnées dans cette soute sont sensibles à la chaleur émise par des harnais électriques ou des tuyaux. L'analyse CFD couplée à du transfert thermique à partir de la maquette CAO de la baie avionique a pour objectif final d'optimiser le positionnement des armoires électroniques dans la soute. La zone à étudier est délimitée par un ensemble de parois en liaison avec l'extérieur. Certains sous-ensembles sont liés aux conditions limites (CL). Il s'agit des composants cibles, des sources de chaleurs et des parois en liaison avec l'extérieur. Les autres composants hors conditions limites sont des obstacles à la circulation du fluide.

La principale grandeur physique à observer est la température autour des composants cibles. Le fluide analysé est l'air contenu dans la soute soumis à un problème de convection naturelle. L'écoulement est considéré comme étant stationnaire et turbulent. Les températures de certaines sources de chaleur sont parfois élevées, des phénomènes de rayonnements peuvent également être pris en compte. Un couplage est nécessaire entre la température des parois de l'enceinte ou des sources de chaleur et la température de l'air.

L'erreur sur le résultat de la simulation ne devrait pas dépasser un seuil par rapport à une valeur qui serait mesurée expérimentalement. Ce seuil a été fixé à 5% dans le cadre de ces travaux de thèse.

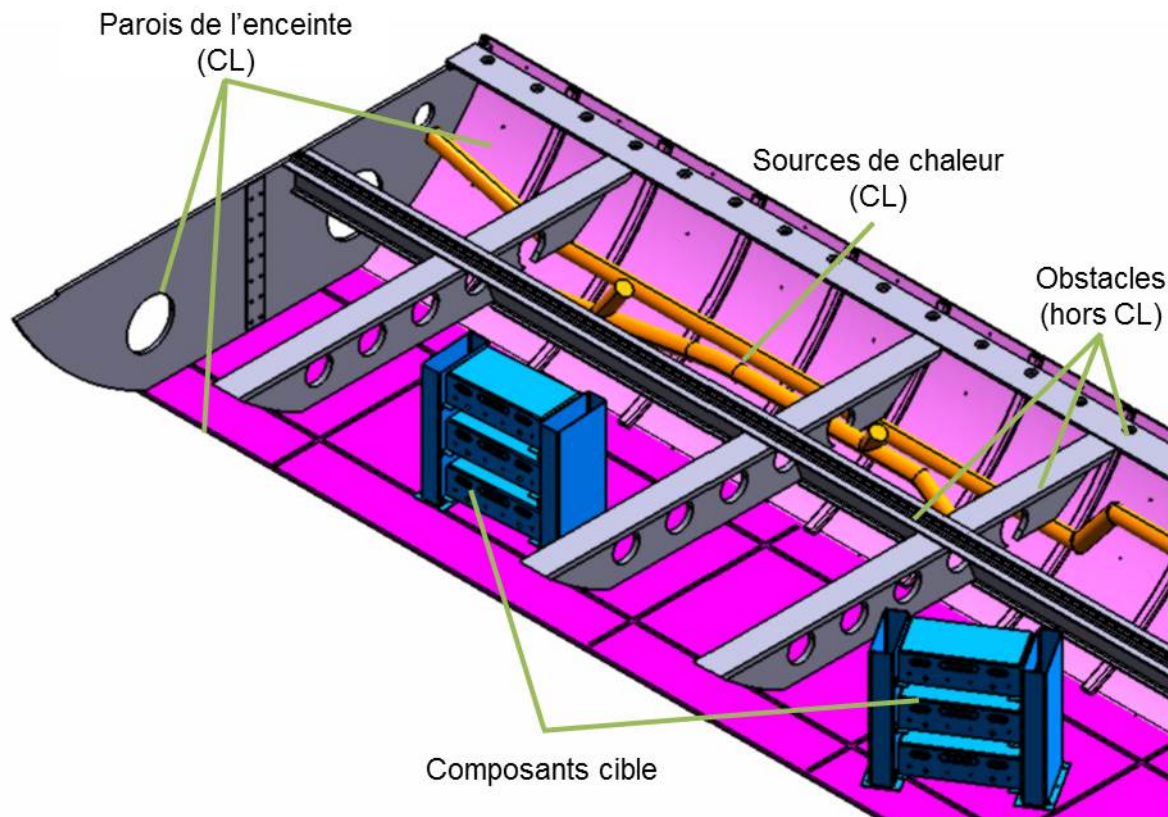


FIGURE 8.1 – Désignation des principaux composants et de quelques conditions limites (CL) d'une baie avionique pour l'analyse thermique.

8.1.2 Modèles CAO à préparer

Les expérimentations ont été réalisées à partir de quatre modèles CAO originaux (figure 8.2) comprenant plusieurs sous-ensembles (figure 8.3). La notion de détails (ou features) telle qu'elle est présentée sur cette figure correspond aux éléments définis dans la section 1.3.2 (trous, poches, bossages et congés/chanfreins). Un premier modèle "élémentaire" ne comportant que les principaux composants a été créé. La version non simplifiée de ce modèle est simulable. Les modèles 1, 2 et 3 sont inspirés de baies avioniques réelles limitées à quelques composants et quelques pièces (les éléments de fixations n'ont été conservés que pour quelques sous-ensembles) dans le but de réduire les durées d'expérimentation.

Ces ensembles ont été modélisés avec le logiciel CATIA V5R20. Certains composants ne sont disponibles que dans des formats standards (STEP) ou de visualisation (CGR).

Processus global de préparation :

Les opérations d'adaptation et de maillage ont été réalisées après les opérations de simplification selon le séquençage proposé dans l'actigramme A0 de la figure 6.4 et ceci quel que soit le modèle. Seules les opérations de simplification ont été ordonnées.

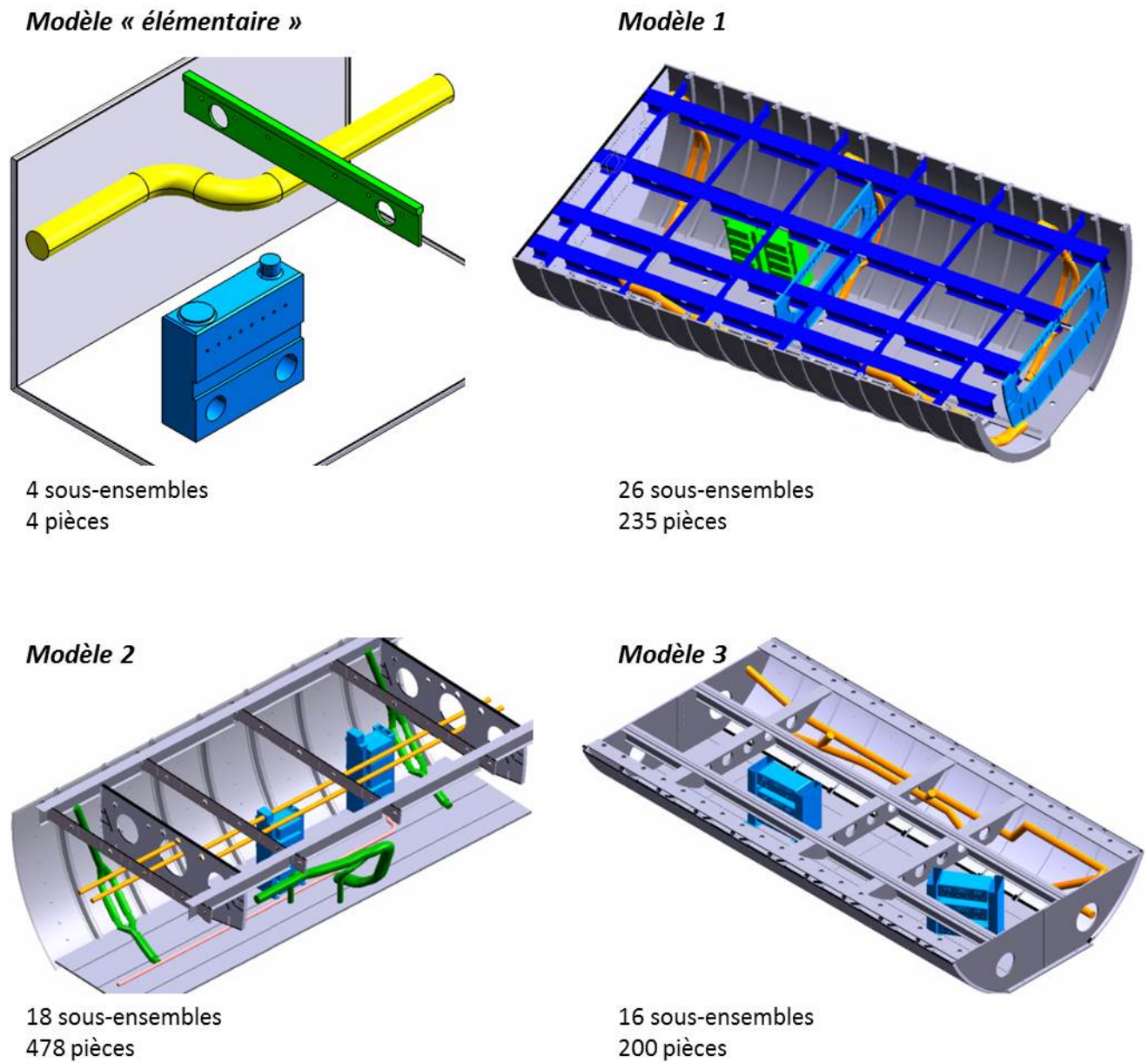
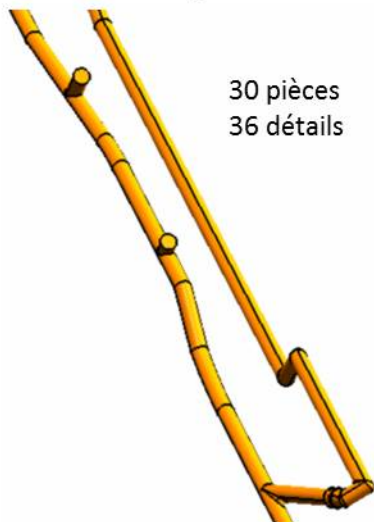


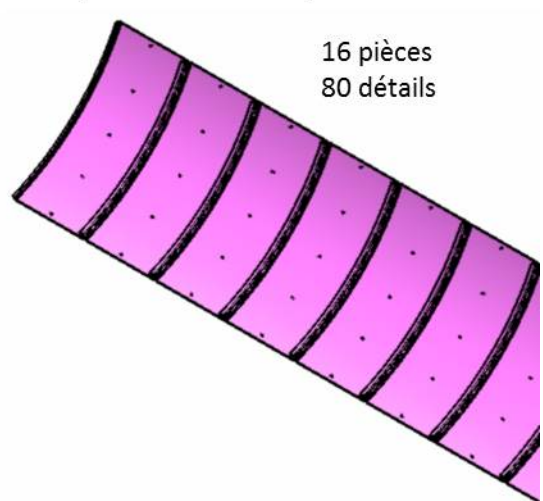
FIGURE 8.2 – Modèles CAO originaux. Le modèle 3 est le support des exemples de préparation de modèles globaux présentés dans ce chapitre.

Source de chaleur (Condition Limite)



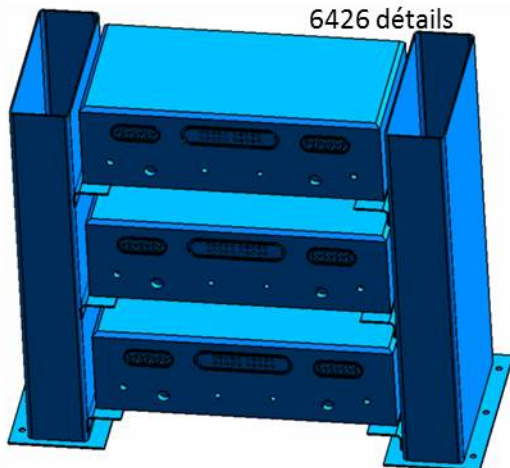
30 pièces
36 détails

Paroi (Condition Limite)



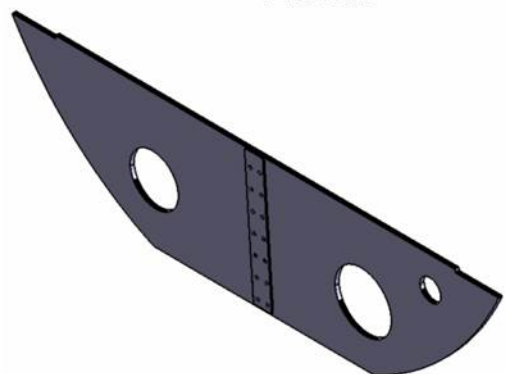
16 pièces
80 détails

Composant cible*



128 pièces
6426 détails

Obstacle



17 pièces
74 détails

FIGURE 8.3 – Exemples de sous-ensembles originaux extraits du modèle N°3. (*) Support des exemples de simplification de sous-ensemble donnés dans ce chapitre.

Les quatre modèles CAO ont été simplifiés avec différents outils :

- CATIA V5R20 pour les opérations de defeaturing des modèles enregistrés dans un format CATIA V5 natif, de suppression de sous-ensembles, de filtrage de pièces, de substitution et d'agrégation ;
- GPURE pour les opérations de décimation et de création d'enveloppes convexes ;
- NX5 pour les opérations de defeaturing des modèles enregistrés dans un format standard (STEP par exemple), de décimation et de création d'enveloppes convexes.

Sous-processus de simplification des sous-ensembles

Pour chaque sous-ensemble SE_j^n d'indice n , un nombre NJ de sous-processus de simplification j a été proposé (toutes les techniques de simplification ne sont pas adaptées à tous les types de sous-ensembles).

Étant donné qu'un nombre minimal d'exemples pour chaque sous-processus est indispensable pour l'apprentissage et que le nombre de sous-processus possibles est très important (plus de 300), un grand nombre d'exemples seraient nécessaires pour que chaque sous-processus soit suffisamment représenté. Le nombre de séquençages étudiés a donc été limité. Parmi les 40 séquençages listés en annexe A, les expérimentations se sont limitées aux 16 séquençages d'opérations les plus utilisées dans un contexte industriel. Ils sont listés dans le tableau 8.1.

Sous-processus	Seq	Sous-processus	Seq
Aucune simplification	0	Filtrage + substitution	10
Filtrage	1	Defeating + décimation	12
Defeating	2	Defeating + enveloppe	13
Décimation	3	Décimation + enveloppe	16
Création d'enveloppe	4	Agrégation + enveloppe	21
Substitution	5	Agrégation + substitution	22
Filtrage+defeating	7	Filtrage + agrégation + substitution	33
Filtrage + décimation	8	Defeating + décimation + enveloppe	35

TABLEAU 8.1 – Séquençages d'opérations utilisés lors des expérimentations. Seq = indice de séquençage.

Les sous-ensembles ont tous été simplifié avec le même outil pour une technique de simplification donnée. Un sous-processus de simplification est alors décrit par un vecteur $sp1_j^n$ réduit à 14 variables (tableau 8.2).

Opérations	Filtrage	Defeaturing	Décimation	Enveloppe	Substitution	Agrégation
Composantes de $sp1_j^n$:	$op1_j^n; op2_j^n$	$op3_j^n \dots op10_j^n$	$op11_j^n$	$op12_j^n$	$op13_j^n$	$op14_j^n$

TABLEAU 8.2 – Composantes opx_j^n de l'ensemble de vecteurs $sp1_j^n$ décrivant les opérations de simplification définies dans la section 6.2.4.1.

Des exemples de sous-processus de simplification sont donnés sur la figure 8.4 . Des exemples de représentation des sous-processus de simplification sont donnés dans le tableau 8.3.

Simplification	SE	Seq	op1	op2	op3	op4	op5	op6	op7	op8	op9	op10	op11	op12	op13	op14
Aucune	SE ₀ ¹⁵	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Agrégation + Substitution	SE ₁ ¹⁵	22	0	0	0	0	0	0	0	0	0	0	0	0	1	1
Filtrage + Substitution*	SE ₂ ¹⁵	10	0,2	1	0	0	0	0	0	0	0	0	0	0	1	0
Filtrage petites pièces	SE ₃ ¹⁵	1	0,2	1	0	0	0	0	0	0	0	0	0	0	0	0
Defeaturing petits trous	SE ₄ ¹⁵	2	0	0	0,2	0	0	0	1	0	0	0	0	0	0	0
Filtrage + Defeaturing	SE ₅ ¹⁵	7	0,2	1	0,2	0	0	0	1	0	0	0	0	0	0	0
Filtrage pièces moyennes	SE ₆ ¹⁵	1	0,5	1	0	0	0	0	0	0	0	0	0	0	0	0
Defeaturing + décimation	SE ₇ ¹⁵	7	0,2	1	0	0	0,2	0	0	0	1	0	0	0	0	0
Agrégation + Enveloppe	SE ₈ ¹⁵	21	0	0	0	0	0	0	0	0	0	0	0	1	0	1
Defeaturing + enveloppe	SE ₉ ¹⁵	13	0	0	0,2	0	0	0	1	0	0	0	0	1	0	0
Décimation + Enveloppe	SE ₁₀ ¹⁵	16	0	0	0	0	0	0	0	0	0	0	1	1	0	0
Defeaturing + Décimation + Enveloppe	SE ₁₁ ¹⁵	35	0	0	0,2	0	0	0	1	0	0	0	1	1	0	0

TABLEAU 8.3 – Description des NJ = 12 sous-processus de simplification j du sous-ensemble SE _{j} ¹⁵ (figure 8.4). Seq = référence du séquençage

(*) Exemple de description de sous-processus de simplification du sous-ensemble pour le sous-ensemble SE₂¹⁵ d'indice $n = 15$ simplifié selon la configuration $j = 2$ (simplification par filtrage et substitution) :

- $op1 = 0,2$ et $op2 = 1$: les pièces dont la longueur moyenne est inférieure à $0,2 \times$ la longueur moyenne du sous-ensemble sont supprimées quelle que soit leur distance par rapport aux conditions limites ;
- $op13 = 1$: le sous-ensemble est simplifié par substitution (parallélépipède) ;
- $op3 = 0, \dots, op12 = 0, op14 = 0$ les opérations de defeaturing, décimation, enveloppe et agrégation ne sont pas réalisées.

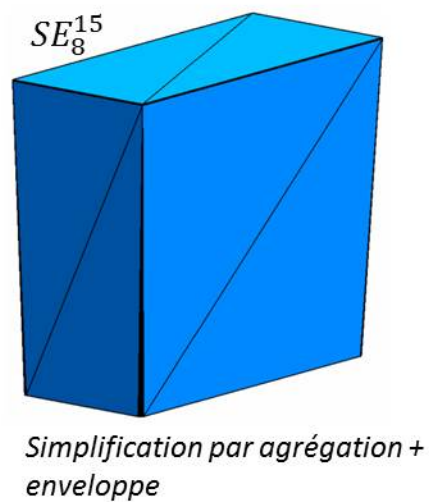
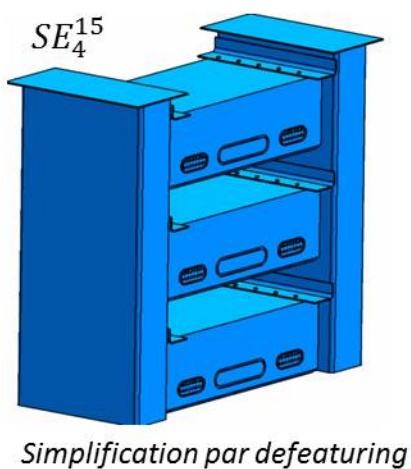
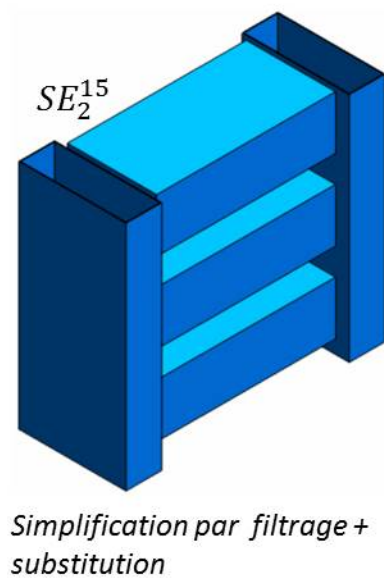
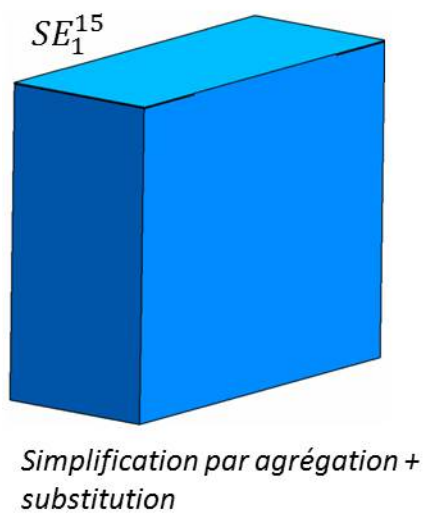
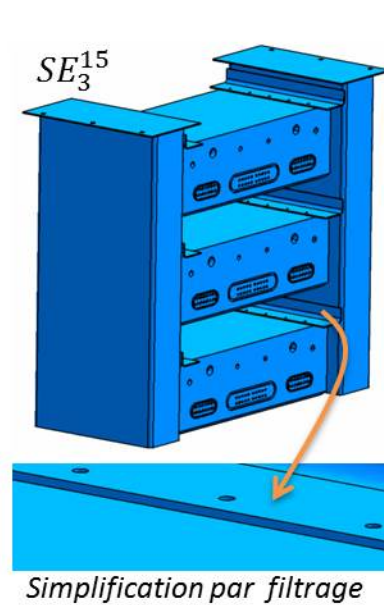
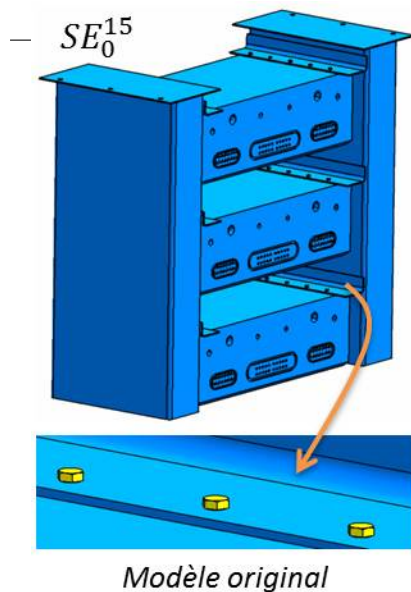


FIGURE 8.4 – Exemples de simplification du sous-ensemble SE_j^{15} pour des configurations j .

Après la simplification de chaque sous-ensemble, les modèles CAO globaux sont re-construits avec CATIA V5R20. Le processus global de simplification a été décrit par un ensemble de vecteurs $SP1_i^m$ (tableau 8.4) défini dans la section 6.2.4.2 .

Les valeurs de ces variables sont des rapports entre l'aire des éléments simplifiés et l'aire totale du sous-ensemble original.

	Filtrage	Defeaturing	Décimation	Enveloppe	Substitution	Agrégation
<i>CL</i>	<i>SP1.1</i>	<i>SP1.2</i>	<i>SP1.3</i>	<i>SP1.4</i>	<i>SP1.5</i>	<i>SP1.6</i>
<i>Cible</i>	<i>SP1.7</i>	<i>SP1.8</i>	<i>SP1.9</i>	<i>SP1.10</i>	<i>SP1.11</i>	<i>SP1.12</i>
<i>Obstacle</i>	<i>SP1.13</i>	<i>SP1.14</i>	<i>SP1.15</i>	<i>SP1.16</i>	<i>SP1.17</i>	<i>SP1.18</i>

TABLEAU 8.4 – Paramètres globaux de réglage des opérations de simplification définis dans la section 6.2.4.2. CL = Condition Limite.

La figure 8.5 et le tableau 8.5 donnent des exemples de description de processus globaux de simplification .

Var.	SP1.1	SP1.2	SP1.3	SP1.4	SP1.5	SP1.6	SP1.10	SP1.11	SP1.12	SP1.13	SP1.14	SP1.15	SP1.16	SP1.17	SP1.18
$M1_{0}^3$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$M1_{1}^3$	0,02	0	0	0	0	0,18	0	0	0	0,38	0,02	0,14	0,25	0,02	0,08
$M1_{33}^3$	0,02	0	0	0	0	0,11	0	0	0	0,02	0,02	0,08	0,02	0,02	0,08
$M1_{68}^3$	0,02	0	0	0	0	0,18	0	0	0	0,38	0,02	0,1	0,25	0,02	0,08
$M1_{71}^3$	0,02	0	0,06	0	0	0	0	0	0	0,02	0	0	0,02	0	0
$M1_{70}^3$	0,02	0	0	0	0	0	0	0	0	0,38	0,02	0,08	0,25	0,02	0,08
$M1_{55}^3$	0,02	0	0,06	0	0	0,11	0	0	0	0,02	0	0	0,02	0	0
$M1_{34}^3$	0,02	0	0,06	0	0	0,11	0	0	0	0,02	0,02	0,02	0,02	0,02	0,02

TABLEAU 8.5 – Exemples de modélisation de processus global de préparation sur le modèle $M1_i^3$ simplifié selon des configurations i (figure 8.5). Var. = composantes de l'ensemble de vecteurs $SP1_i^m$.

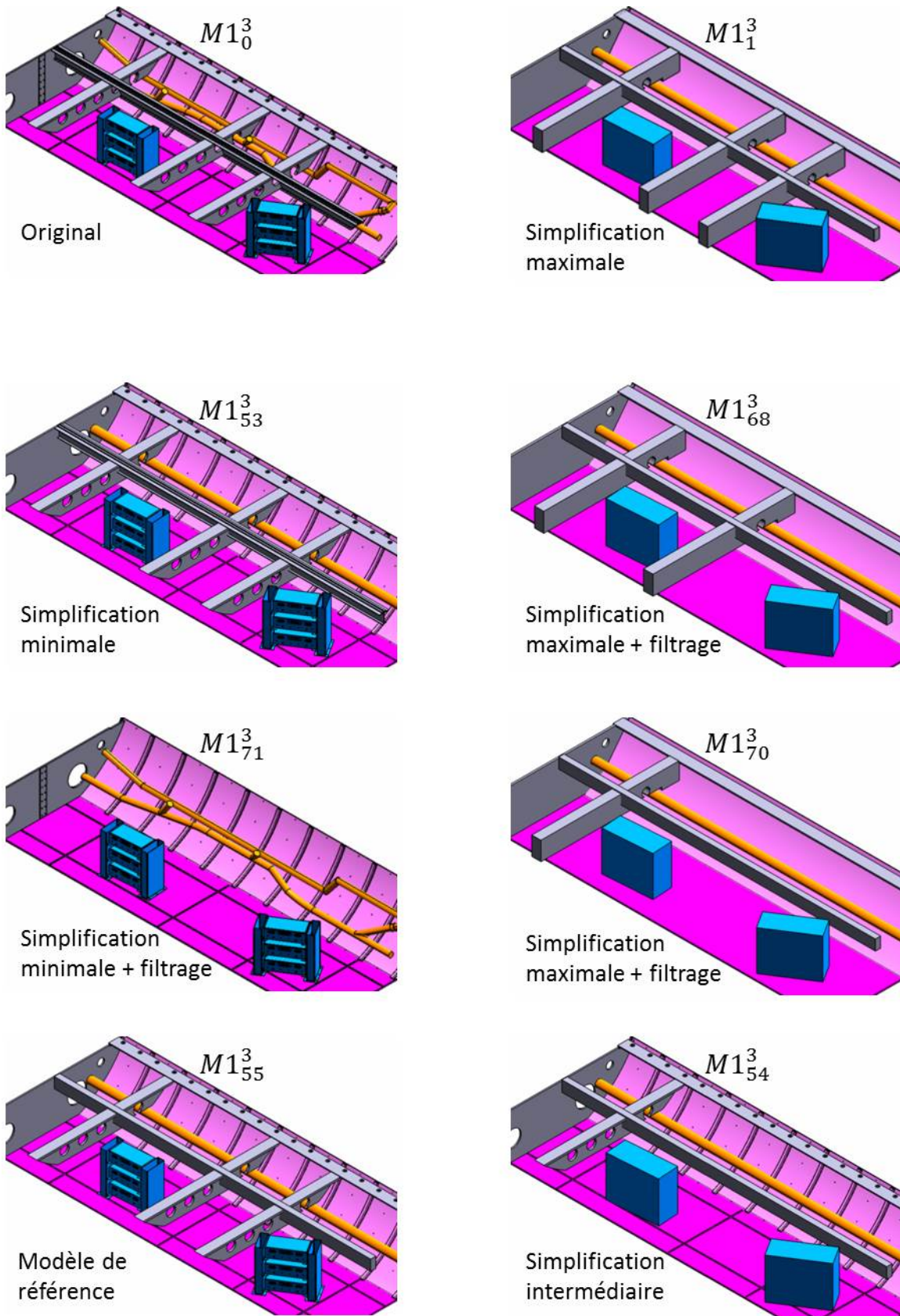


FIGURE 8.5 – Exemples de simplifications globales du modèle $M1_i^3$ simplifié selon des configurations i .

L'ensemble des variables qui décrivent les modèles CAO du tableau ont été extraites à l'aide de macros sous CATIA V5R20 (aires, volumes, nombre de pièces, nombre de détails) ou à l'aide de MeshLab¹ (nombre de triangles, nombre de patches, rayons de courbures). Les modèles CAO originaux et simplifiés sont décrits à l'aide de plusieurs vecteurs de variables $CAOM1_i^m$ présentés dans la section 6.2.5 :

- CAO_0^n : ensemble de vecteurs $CAO.x_0^n$ de variables qui décrivent la géométrie du sous-ensemble original SE_0^n (aire, volume, nombre de pièces, nombre de triangles, nombre de patches, compacité, ratio entre le volume du sous-ensemble et celui de sa boîte englobante, courbures) ;
- CAO_j^n : ensemble de vecteurs $CAO.x_j^n$ de variables qui décrivent la géométrie du sous-ensemble simplifié SE_j^n selon une configuration j ;
- $Comp_j^n$: ensemble de vecteurs $Comp.x_j^n$ de variables qui décrivent les facteurs de comparaison entre le sous-ensemble original SE_0^n et simplifié SE_j^n selon une configuration j (bénéfices sur les valeurs des variables $CAO.x_j^n$ et distances de Hausdorff) ;
- $CAOM1_i^m$: ensemble de vecteurs $CAOM1.x_i^m$ de variables qui décrivent la géométrie du modèle global $M1_i^m$ simplifié selon une configuration i .

Les tableaux 8.6 et 8.7 donnent des exemples de description et comparaison de modèles CAO au niveau des sous-ensembles. Le tableau 8.8 donne des exemples de description et comparaison de modèles CAO au niveau des ensembles globaux.

1. Istituto di Scienza e Tecnologia del l'Informazione : <http://meshlab.sourceforge.net/>

	Aire	Volume	Nb. pièces	Nb. triangles	Nb. patches	Compacité	Ratio vol. / vol. boîte	Ratio triangles / patches	Courbure k1	Courbure k2
	CAO.1 ⁿ _j (m ²)	CAO.2 ⁿ _j (m ³)	CAO.3 ⁿ _j SU	CAO.4 ⁿ _j SU	CAO.5 ⁿ _j SU	CAO.10 ⁿ _j SU	CAO.12 ⁿ _j SU	CAO.14 ⁿ _j SU	CAO.18 ⁿ _j (mm)	CAO.19 ⁿ _j (mm)
SE ¹⁵ ₀	3,78	0,059	43	29624	1344	64	0,35	22	-25	232
SE ¹⁵ ₁	1,96	0,166	1	12	6	12	0,99	2	1	27
SE ¹⁵ ₂	3,49	0,059	5	100	38	59	0,35	3	-27	22
SE ¹⁵ ₃	3,78	0,059	13	19304	924	64	0,35	21	-18	14
SE ¹⁵ ₄	3,77	0,059	43	14228	852	63	0,35	17	-14	15
SE ¹⁵ ₅	3,76	0,059	13	9428	492	63	0,35	19	-18	14
SE ¹⁵ ₆	3,51	0,059	5	15952	722	60	0,35	22	-1044	404
SE ¹⁵ ₇	3,77	0,059	13	12056	612	64	0,35	20	-3	-2
SE ¹⁵ ₈	1,87	0,155	1	76	1	12	0,92	76	-280	-4
SE ¹⁵ ₉	1,87	0,155	1	76	1	12	0,92	76	-1044	283
SE ¹⁵ ₁₀	1,86	0,155	1	84	1	12	0,92	84	-1044	1044
SE ¹⁵ ₁₁	1,86	0,155	1	84	1	12	0,92	84	-731	283

TABEAU 8.6 – Exemples de description du modèle CAO du sous-ensemble SE¹⁵_j (figure 8.4). Les valeurs sont brutes telles qu'elles sont extraites des modèles CAO avant traitements (normalisation et discrétisation). Les variables explicatives limitées aux variables déterminantes. SU = sans unité.

	Nb. pièces	Nb. triangles	Compacité	Ratio vol. / vol. boîte	Courbure k1	Courbure k2	D. Hausdorff	Distance / cl	Distance / cible	Moment distance / cl
	comp.3 ⁿ⁽¹⁾ _j SU	comp.4 ⁿ⁽¹⁾ _j SU	comp.7 ⁿ⁽¹⁾ _j SU	comp.8 ⁿ⁽¹⁾ _j SU	comp.13 ⁿ⁽¹⁾ _j SU	comp.14 ⁿ⁽¹⁾ _j SU	comp.16 ⁿ⁽²⁾ _j SU	inter.1 ⁿ⁽³⁾ _j (mm)	inter.2 ⁿ⁽³⁾ _j (mm)	inter.5 ⁿ⁽¹⁾ _j SU
SE ¹⁵ ₀	0	0	0	0	0,976	0,425	0	92	514	0
SE ¹⁵ ₁	0,977	1	0,74	-1,809	1,001	0,934	0,03	92	514	0
SE ¹⁵ ₂	0,884	0,997	0,073	0,008	0,974	0,945	0,002	92	475	0
SE ¹⁵ ₃	0,698	0,348	-0,001	0	0,983	0,965	0,001	92	475	0
SE ¹⁵ ₄	0	0,52	0,004	-0,001	0,986	0,964	0	92	475	0
SE ¹⁵ ₅	0,698	0,682	0,005	-0,001	0,983	0,964	0	92	475	0
SE ¹⁵ ₆	0,884	0,462	0,065	0,011	0	0	0,002	92	475	0
SE ¹⁵ ₇	0,698	0,593	0,003	0	0,997	1,005	0,002	92	475	0
SE ¹⁵ ₈	0,977	0,997	0,74	-1,617	0,732	1,009	0,021	92	475	0
SE ¹⁵ ₉	0,977	0,997	0,74	-1,617	0	0,299	0,021	0	299	0
SE ¹⁵ ₁₀	0,977	0,997	0,74	-1,615	0	-1,586	0,022	0	435	0
SE ¹⁵ ₁₁	0,977	0,997	0,74	-1,615	0,3	0,299	0,022	0	0	0

TABEAU 8.7 – Exemples de facteurs de comparaison $comp.x_j^n$ de sous-ensembles originaux et simplifiés SE¹⁵_j et de facteurs d'influence $inter.x_j^n$ sur le cas de simulation. Variables explicatives limitées aux variables déterminantes. (1) Bénéfices entre les valeurs du modèle simplifié et du modèle original. (2) Moyenne quadratique de la distance de Hausdorff. (3) Les valeurs sont brutes telles qu'elles sont extraites des modèles CAO avant traitements (normalisation et discrétisation). SU = sans unité

8.1.4 Adaptation des modèles CAO

Processus d'adaptation

Dans le cas de la préparation d'un modèle pour la simulation CFD couplée à une analyse thermique, l'adaptation du modèle CAO consiste à :

1. créer des volumes autour des entrées et sorties du fluide entre la soute et l'extérieur ;

	Nb. patches	Ratio vol. / sec. boîte	Aire	Volume	Nb. pièces	Nb. triangles	Ratio volume / volume boîte	Ratio vol. / sec. boîte	Ratio triangles / patches
	CAOM1.21 SU	CAOM1.28 SU	Comp1.33 SU	CompM1.34 SU	CompM1.35 SU	CompM1.36 SU	CompM1.40 SU	CompM1.41 SU	CompM1.42 SU
M1 ₀ ³	4144	0,063	0	0	0	0	0	0	0,000
M1 ₁ ³	117	0,105	0,107	-0,68	0,893	0,991	-0,68	-0,68	0,697
M1 ₅₃ ³	955	0,091	0,066	-0,44	0,773	0,599	-0,44	-0,44	-0,742
M1 ₆₈ ³	97	0,091	0,148	-0,44	0,907	0,992	-0,44	-0,44	0,653
M1 ₇₁ ³	3447	0,052	0,177	0,166	0,26	0,175	0,166	0,166	0,008
M1 ₁₀ ³	63	0,069	0,26	-0,09	0,94	0,992	-0,09	-0,09	0,506
M1 ₅₅ ³	3647	0,072	0,036	-0,14	0,213	0,109	-0,14	-0,14	-0,013
M1 ₉₄ ³	971	0,088	0,071	-0,41	0,773	0,586	-0,41	-0,41	-0,768

TABLEAU 8.8 – Exemples de facteurs descriptifs d'ensembles CAO globaux CAOM1. x_i^m (variables brutes sans unités extraites des modèles CAO) et de facteurs de comparaison Comp1. x_i^m entre le modèle original et simplifié (bénéfices entre les valeurs du modèle simplifié et celles du modèle original). Les variables explicatives sont limitées aux variables déterminantes. Les valeurs des variables sans brutes avant traitements. SU = sans unité

2. extraire les surfaces en contact avec le fluide ;
3. fermer les surfaces.

L'opération 1 est réalisée à l'aide d'un modeleur 3D (CATIA V5R20).

Les opérations 2 et 3 sont réalisées automatiquement à l'aide du logiciel ANSYS FLUENT pour tous les modèles. Néanmoins, faute de pouvoir en extraire les données depuis FLUENT, le volume de fluide a été re-modélisé à l'aide de CATIA V5R20 dans le but d'en extraire les caractéristiques géométriques.

Le processus d'adaptation est décrit par un ensemble de vecteurs SP2 _{i} ^{m} défini dans la section 6.2.4.3 qui indique quels sont les types d'éléments à extraire et à créer (1D, 2D, surfaciques ou volumiques). Tous les modèles ont été adaptés selon le même processus et les mêmes outils. Ces variables ne sont donc pas déterminantes. Le vecteur SP2 _{i} ^{m} n'est pas pris en compte pour l'apprentissage. Seules les données extraites des modèles adaptés M2 _{i} ^{m} le seront.

Un exemple de modèles adaptés est illustré sur la figure 8.6.

Données extraites des modèles adaptés

Les caractéristiques du modèle adapté (volume de fluide) sont extraites à l'aide de macros sous CATIA V5R20 et Meshlab. Les modèles adaptés sont décrits à l'aide de deux ensembles de vecteurs CAOM2 _{i} ^{m} et Comp2 _{i} ^{m} définis dans la section 6.2.5 :

- CAOM2 _{i} ^{m} : ensemble de vecteurs CAOM2. x_i^m décrivant le modèle adapté (aire,

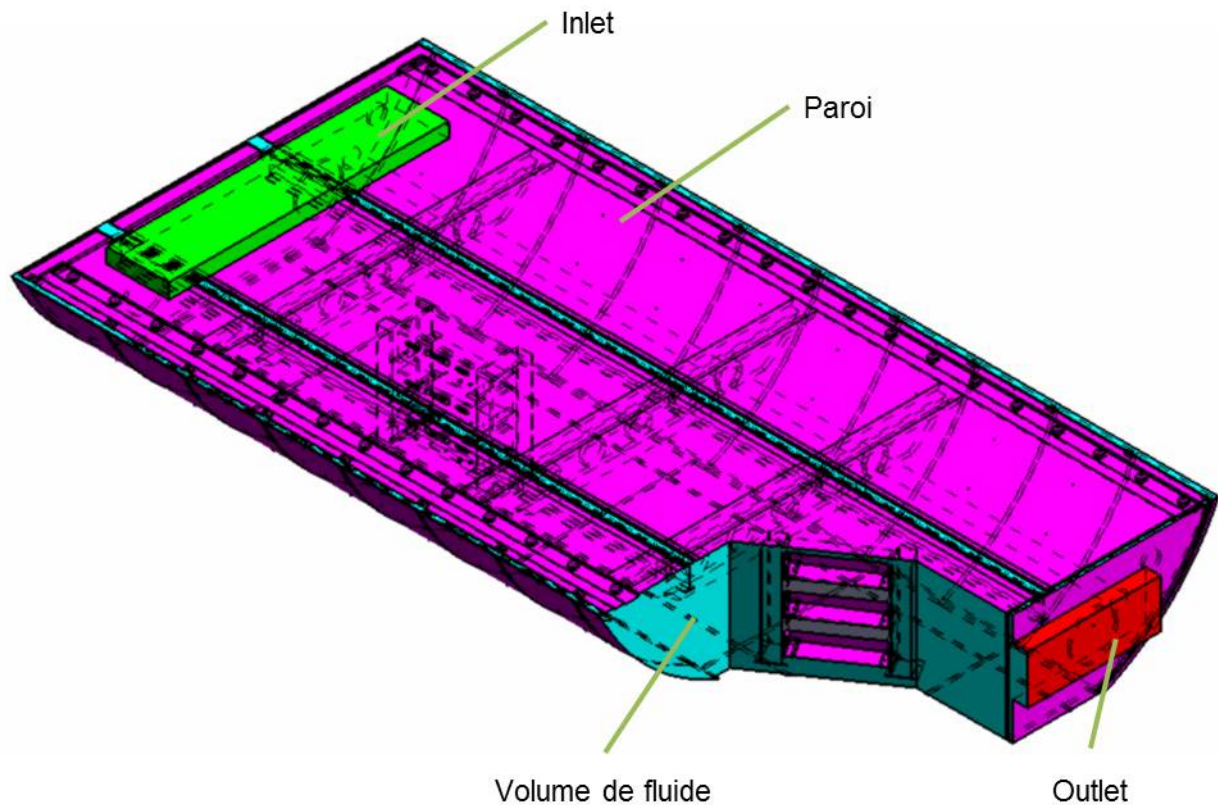


FIGURE 8.6 – Exemple de modèle adapté $M2_i^m$ (volume de fluide reconstruit par modelleur 3D pour extraction des données).

volume nombre de triangles, nombre de patches du modèle CAO représentant le volume de fluide) ;

- $CompM2_i^m$: ensemble de vecteurs $CompM2.x_i^m$ descriptifs des facteurs de comparaison entre le modèle de référence (modèle simulable le moins simplifié) et le modèle adapté après simplification selon la configuration i .

Des exemples de description de modèles adaptés sont donnés dans le tableau 8.9.

	Aire	Volume	Nb. triangles	Nb patches	Volume	Compacité	Ratio vol. / vol. boîte	D. Hausdorff
	CAOM2.1 ⁽¹⁾ (m ²)	CAOM2.2 ⁽¹⁾ (m ³)	CAOM2.3 ⁽¹⁾ SU	CAOM2.4 ⁽¹⁾ SU	CompM2.13 ⁽²⁾ SU	CompM2.17 ⁽²⁾ SU	CompM2.18 ⁽²⁾ SU	CompM2.25 ⁽³⁾ (m)
M ₀ ³	66,1	8.261	64944	4412	1	1	1	0
M ₁ ³	58,5	7.692	1940	199	0,93	0,93	0,56	0,04
M ₅₃ ³	59,8	7.909	19250	1273	0,96	0,93	0,66	0,03
M ₆₈ ³	55,9	7.874	53102	3840	0,95	0,87	0,66	0,02
M ₇₁ ³	54,6	8.377	48674	3657	1,01	0,82	1,22	0,02
M ₇₀ ³	51,2	8.132	1088	88	0,98	0,78	0,9	0,04
M ₅₅ ³	63	8.152	55198	3951	0,99	0,96	0,86	0,02
M ₅₄ ³	59,4	7.938	20126	1275	0,96	0,92	0,68	0,02

TABEAU 8.9 – Exemples de facteurs descriptifs de modèles adaptés CAOM2. x_i^m ((1)variables brutes extraites des modèles CAO adaptés M2_i³) et de facteurs de comparaison CompM2. x_i^m entre le modèle de référence et simplifié ((2) : bénéfices entre CAOM2. x_i^m et CAOM2. x_{ref}^m , (3) distance de Hausdorff). Les variables explicatives sont limitées aux variables déterminantes. Les valeurs sont brutes telles qu'elles sont extraites du modèle adapté avant traitements. SU = sans unité.

8.1.5 Maillage des modèles

Processus de maillage

Le modèle préparé dans le cadre de cette application est un modèle maillé 3D. Des exemples de modèles maillés sont donnés sur les figures 8.7 et 8.8.

Le mailleur utilisé pour tous les modèles est CFX Mesh intégré à ANSYS. Le processus de maillage est décrit à l'aide du vecteur SP3_i^m défini dans la section 6.2.4.4.

$$SP3_i^m = \{SP3.1_i^m, ..., SP3.6_i^m, SP2.3_i^m, T3_m\}$$

Les valeurs choisies pour les variables SP3.1_i^m (type d'éléments surfaciques), SP3.2_i^m (type d'éléments volumiques) et SP3.3_i^m (taille des éléments hors raffinement) sont donnés dans le tableau 8.10 .

Sous-processus de maillage	Type d'éléments surfaciques SP3.1 _i ^m	Type d'éléments volumiques SP3.2 _i ^m	Taille (mm) SP3.3 _i ^m
Modèle élémentaire	Quadrilatères	Héxaèdres	40
Modèle 1	Triangles	Tétraèdres	100
Modèle 2	Triangles	Tétraèdres	100
Modèle 3	Quadrilatères	Héxaèdres	100

TABEAU 8.10 – Description des principaux paramètres de maillage.

Lors des expérimentations, il n'a pas été recherché d'évaluer l'influence des paramètres de maillage sur le résultat de l'analyse mais seulement des paramètres de simplification.

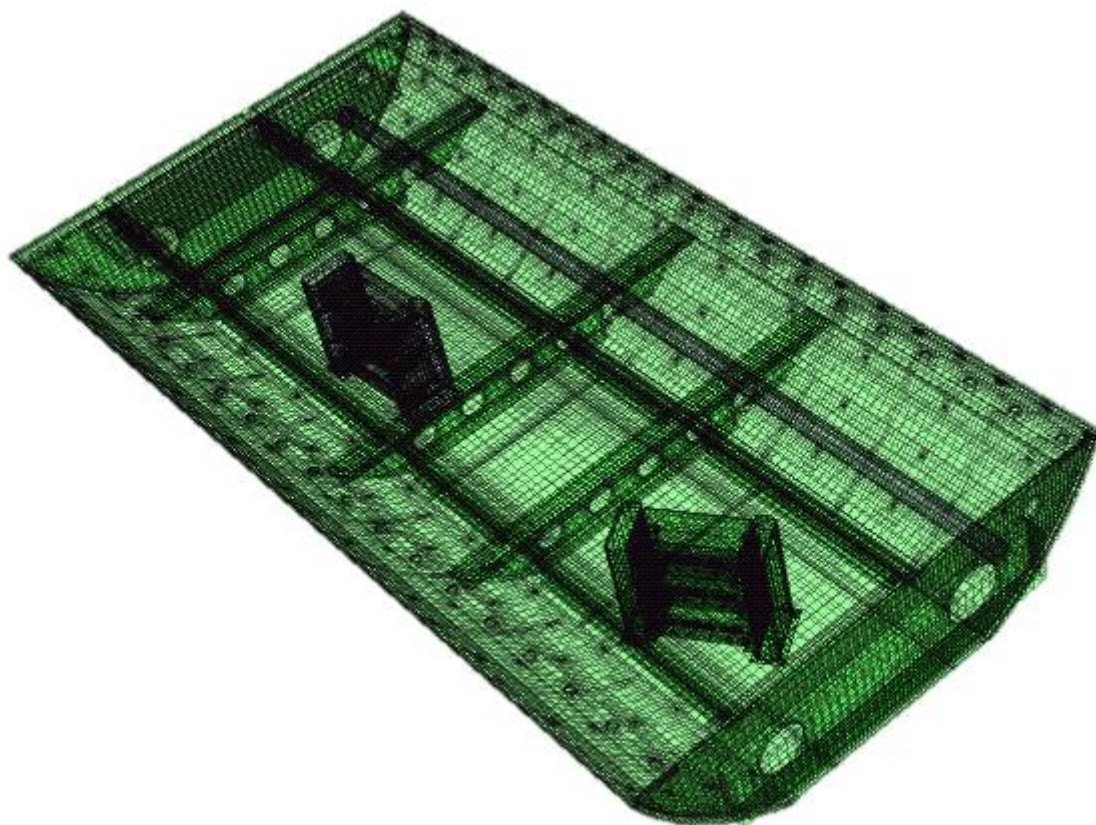


FIGURE 8.7 – Exemple de modèle maillé. Volume de fluide du modèle M_{ref}^2 (configuration de référence).

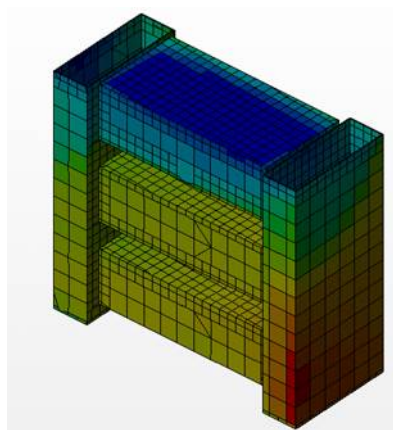


FIGURE 8.8 – Exemple de maillage des surfaces du composant cible SE_{ref}^{15} .

L'outil utilisé est le même pour tous les modèles. Le type de maillage (variables $SP3.1_i^m$ et $SP3.2_i^m$) est identique pour toutes les configurations d'un même modèle. Si possible, la taille du maillage (variable $SP3.3_i^m$) est la même pour toutes les configurations d'un même modèle. Cette taille est choisie afin que tous les critères de qualité soient respectés sur le modèle de référence (modèle simulable le plus complexe).

Les maillages sont tous raffinés (variable $SP3.4_i^m$) autour des couches limites (in-

terfaces entre le fluide et les composants). Le niveau de qualité attendu (variables $SP3.5_i^m$ et $SP3.6_i^m$) est le même pour tous les modèles et toutes les configurations. Ces trois variables ne sont donc pas déterminantes et ne sont donc pas prises en compte pour l'apprentissage. Néanmoins, la possibilité de les faire varier pourra être envisagé par la suite.

Données extraites des modèles maillés

Les données caractéristiques des modèles maillés sont extraites du mailleur, elles sont représentées par deux ensembles de vecteurs définis dans la section 6.2.5 :

- $Mail3_i^m$: ensemble de vecteurs $Mail3.x_i^m$ décrivant le modèle maillé ;
- $CompM3_i^m$: ensemble de vecteurs $CompM3.x_i^m$ décrivant les facteurs de comparaison entre le modèle de référence et le modèle maillé après simplification selon la configuration i .

Des exemples de description de modèles maillés sont donnés dans le tableau 8.11.

	Nb. éléments	Durée maillage	Nb. éléments
	$Mail3.3^{(1)}$ SU	$Mail3.4^{(1)}$ (min.)	$Comp3.7^{(2)}$ SU
M_0^3	Echec	Echec	Echec
M_1^3	95723	47	74,2
M_{53}^3	197815	235	46,7
M_{68}^3	92829	49	74,9
M_{71}^3	357732	424	3,49
M_{70}^3	91654	41	75,2
M_{55}^3	373400	446	0
M_{54}^3	199351	215	46,3

TABLEAU 8.11 – Exemples de facteurs descriptifs de maillages $Mail3.x_i^m$ ((1)variables brutes extraites des maillages $M3_i^3$) et de facteurs de comparaison $CompM3.x_i^m$ entre le modèle de référence et simplifié ((2) : bénéfices entre $CAOM3.x_i^m$ et $CAOM3.x_{ref}^m$). Les variables explicatives sont limitées aux variables déterminantes.

L'analyse CFD consiste à discrétiser et résoudre les équations aux dérivées partielles sur des petits éléments volumiques. Le modèle choisi $k - \omega$ résout les équations de conservation de l'énergie cinétique turbulente k et de la fréquence de turbulence ω . L'ensemble des données nécessaires à l'analyse CFD sont déterminées à l'avance par les analystes, il s'agit essentiellement des caractéristiques des conditions limites telles que :

- le flux de chaleur sur les parois en relation avec l'extérieur ;
- les températures et flux de chaleur sur les composants diffusant de la chaleur ;
- le profil des vitesses, la température et la pression sur l'entrée et la sortie.

Les données à extraire relatives aux cas de simulation sont :

- la durée de la simulation qui correspond au coût de l'analyse CA = y5 (les facteurs de coût relatifs à l'outillage n'ont pas été considérés) ;
- les résultats de l'analyse (températures autour des composants cibles) utilisés pour déterminer l'erreur globale sur le résultat de l'analyse ERA = y3 (équation 6.6).

La simulation a été réalisée avec les logiciels ANSYS² et STARCCM+³. Des exemples de résultats de simulation et de décisions sur le processus de préparation proposé sont donnés dans le tableau 8.12.

	Résultats simulations		Coûts		Décision
	Temp.	y3 (ERA)	y4 (CP)	y5 (CA)	processus de préparation :
M ₀ ³	Echec	Echec	0	Echec	non validé (non simulable)
M ₁ ³	313 °K	6%	518 min.	467 min.	non validé (erreur importante)
M ₅₃ ³	311 °K	1%	498 min.	930 min.	validé
M ₆₈ ³	310 °K	-2%	478 min.	446 min.	validé
M ₇₁ ³	Echec	Echec	Echec	Echec	non validé (non simulable)
M ₇₀ ³	304 °K	-16%	323 min.	401 min.	non validé (erreur importante)
M ₅₅ ³	310 °K	0%	623 min.	1830 min.	non validé (coût élevé)
M ₅₄ ³	309 °K	-1%	462 min.	940 min.	validé (meilleur processus)

TABEAU 8.12 – Exemples de résultats d'analyse et de principales valeurs à prédire : Temp. = températures maximales relevées sur la cible, ERA = y3 (erreur globale sur le résultat de l'analyse), CP = y4 (coût total de la préparation) et CA = y5 (coût de l'analyse) sur le modèle global M_i³ pour des configurations de simplification i représentées sur la figure .

Concernant la pertinence des résultats des simulations, le modèle élémentaire est simulable sans simplifications, les résultats réels de la simulation sont donc connus

2. Fluent, ANSYS, <http://www.ansys.com>

3. CD-adapco, STARCCM+, <http://www.cd-adapco.com>

et pourront être validés par des tests physiques. Pour les modèles 1, 2 et 3, on considérera que l'erreur sur les résultats des simulations sur les modèles de référence sont négligeables.

8.1.7 Apprentissage et tests sur de nouveaux cas

Parmi l'ensemble des logiciels présentés dans la section 4.2.3, le logiciel open source Weka⁴ a été choisi dans le cadre de cette étude de faisabilité pour son interface qui permet de tester aisément plusieurs algorithmes, pour ces outils de visualisation qui facilitent l'identification de variables corrélées et de valeurs aberrantes, pour ces outils de traitements des variables très variés et pour son outil de création de workflow. Il a été utilisé (figure 8.9) :

- au niveau des variables pour visualiser les données, les traiter (normalisation, discrétisation, création de classes), les classer et pour identifier les variables explicatives ;
- au niveau des classifieurs pour les configurer, les sélectionner, les tester et pour prédire une valeur sur un nouveau cas.

Les algorithmes utilisés sont des classifieurs capables de prendre en entrée à la fois des variables nominales, booléennes et numériques, et dont les variables de sortie sont discrétisées. Ces classifieurs sont proposés par l'ensemble des logiciels présentés dans la section 4.2.3. Les classifieurs testés sont listés dans le tableau 8.13.

Type de classifieur	Désignation Weka
Arbre de décision type C4.5.	J48
Machine à vecteur de support	SMO
Réseau de neurones perceptron	Multilayer Perceptron
Réseau radial basé sur la classification des k moyennes	RBFNetwork
Classifieur de type Naïves Bayésien	NaïvesBayes

TABLEAU 8.13 – Types de classifieurs utilisés lors des expérimentations.

Les résultats des expérimentations pour la configuration des classifieurs sont donnés dans la section 8.2.3.

8.1.8 Synthèse sur la construction d'exemples de processus de préparation de modèles

La figure 8.10 donne en synthèse le nombre d'exemples de processus de préparations créés à partir des quatre modèles originaux. 325 processus de préparations

4. WEKA, University of Waikato, <http://www.cs.waikato.ac.nz/ml/weka/>

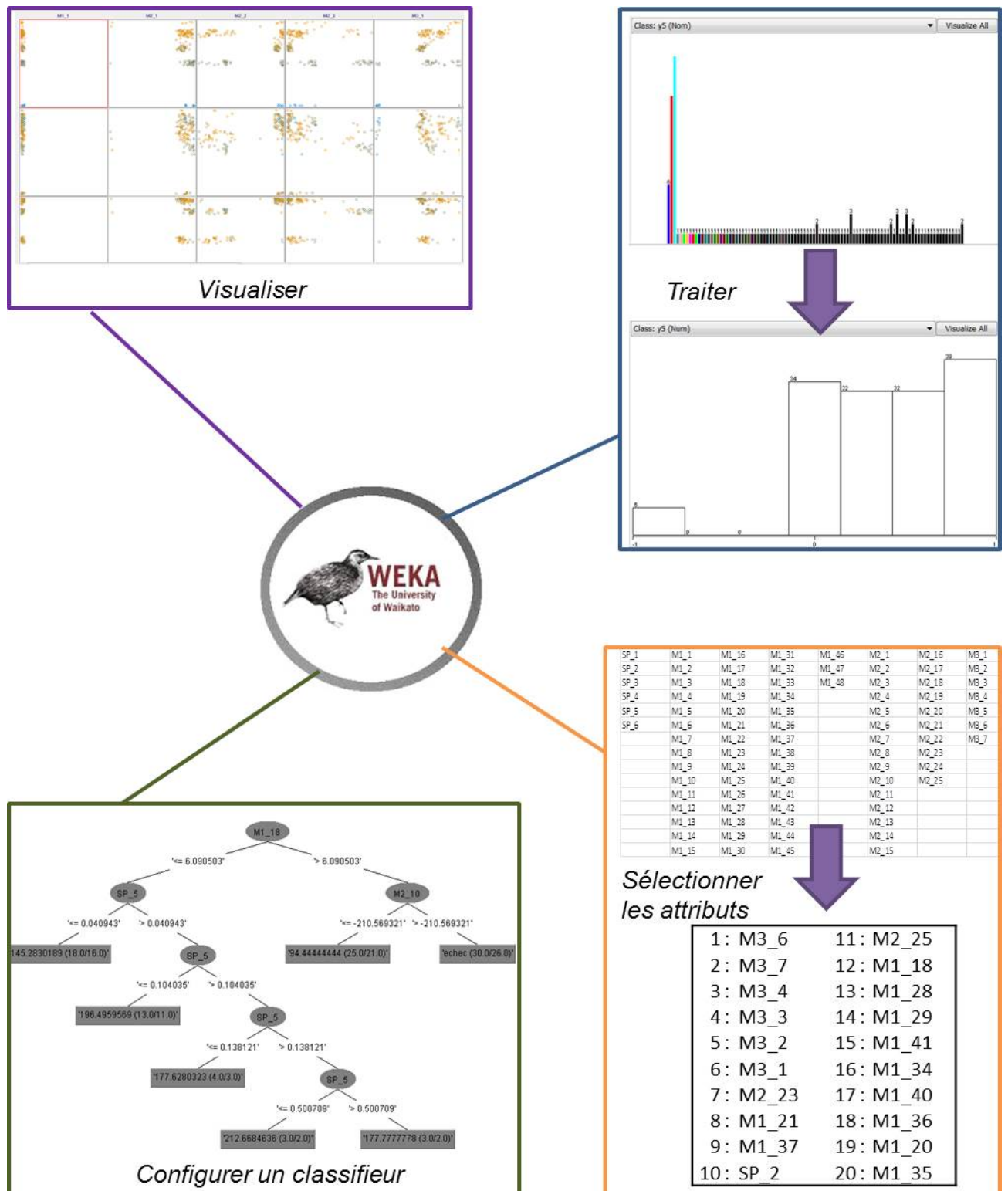


FIGURE 8.9 – Utilisations du logiciel Weka pour l'apprentissage.

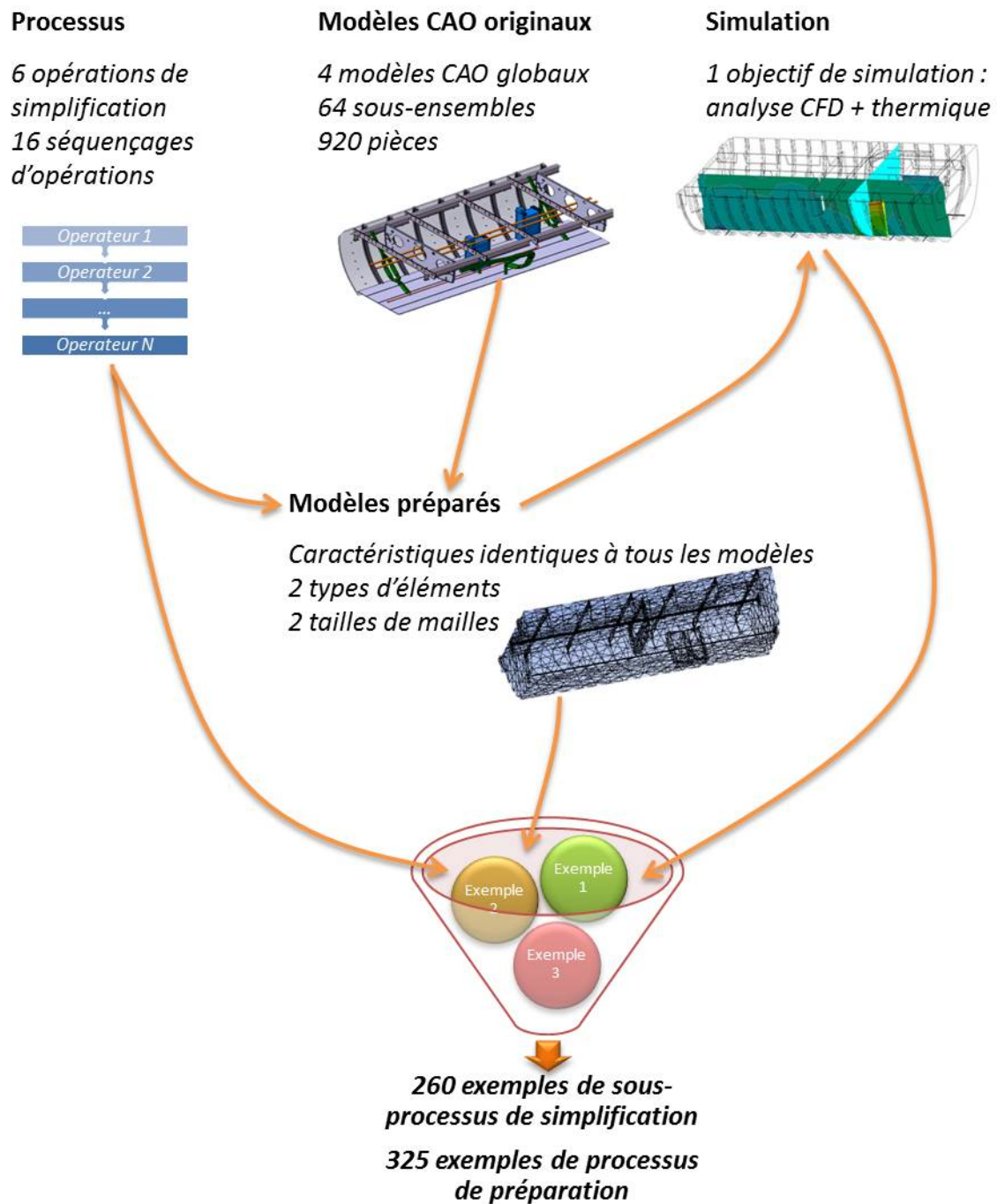


FIGURE 8.10 – Synthèse sur la construction d'exemples de processus de préparation de modèles.

Objectif de la préparation

L'objectif de la préparation des exemples construits est la simulation CFD qui nécessite, lors de la phase d'adaptation, l'extraction des surfaces bordant le volume de fluide.

260 sous-processus de simplification ont été créés à partir des 64 sous-ensembles qui constituent les quatre modèles originaux. Six opérations de simplification ont été utilisées. Le nombre d'exemples à construire étant réduit, afin d'obtenir un nombre significatif d'exemples pour chaque sous-processus de simplification, les paramètres de réglages des opérations de simplifications ont été limités à quelques valeurs. Les paramètres de réglages des opérations de simplification globales (décimation, enveloppe, substitution et agrégation) sont de type tout ou rien. Les paramètres qui caractérisent les opérations de filtrage et de defeaturing, sont limités à trois valeurs : aucun élément (pièce ou détail) n'est supprimé, seuls les éléments de petite taille (dont la longueur moyenne est inférieure à 0,2 fois la longueur moyenne du modèle) sont supprimés, et tous les éléments sont supprimés quelle que soit leur taille.

Phase d'adaptation

Un seul outil a été utilisé pour l'adaptation (extraction et fermeture des surfaces bordant le volume de fluide), tous les modèles ont été adaptés selon la même méthode.

Phase de maillage

Un seul outil a été utilisé pour le maillage, les deux seuls paramètres qui caractérisent les opérations de maillage sont le type d'éléments et leur taille. La qualité du maillage est équivalente pour tous les modèles préparés.

Il s'agit de déterminer les éléments nécessaires à la construction d'un modèle d'apprentissage en se basant sur les méthodes proposées dans les chapitres 6 et 7 et en répondant aux questions posées sur la figure 8.11.

8.2.1 Modélisation d'exemples pour l'apprentissage et les tests

Les exemples de simplification ont été construits pour chaque ensemble global m et chaque sous-ensemble n selon le protocole présenté sur la figure 6.21. 260 sous-processus de simplification ont été construits sur les 64 sous-ensembles, $Na = 195$ exemples ont été utilisés pour l'apprentissage et $Nt = 65$ exemples pour les tests. 325 processus globaux de préparation ont été construits pour les quatre modèles, $Na = 245$ exemples ont été utilisés pour l'apprentissage et $Nt = 80$ exemples pour les tests.

Il est nécessaire de s'assurer que les exemples sont répartis sur tout l'espace "niveau de simplification / erreur sur le résultat de l'analyse" et que chaque processus de simplification contient un nombre minimal d'exemples.

Il devra être montré dans cette section que le nombre et le choix des exemples proposés est suffisant pour évaluer un processus de préparation à l'aide des techniques d'apprentissage artificiel.

Vérification de la répartition des exemples

Les sous-processus de simplification construits ont été choisis afin que tout l'espace "niveau de simplification (LOS) / influence de la simplification sur le résultat de l'analyse (IS)" soit représenté, comme on peut le constater sur la figure 8.12, et qu'il y ait un nombre minimal d'exemples pour chaque opération de simplification. Concernant la répartition des exemples de processus de préparation sur les modèles globaux, la figure 8.13 montre que les exemples de processus de préparation sont régulièrement répartis dans l'espace "niveau de simplification (LOS) / erreur globale sur le résultat de l'analyse (ERA)".

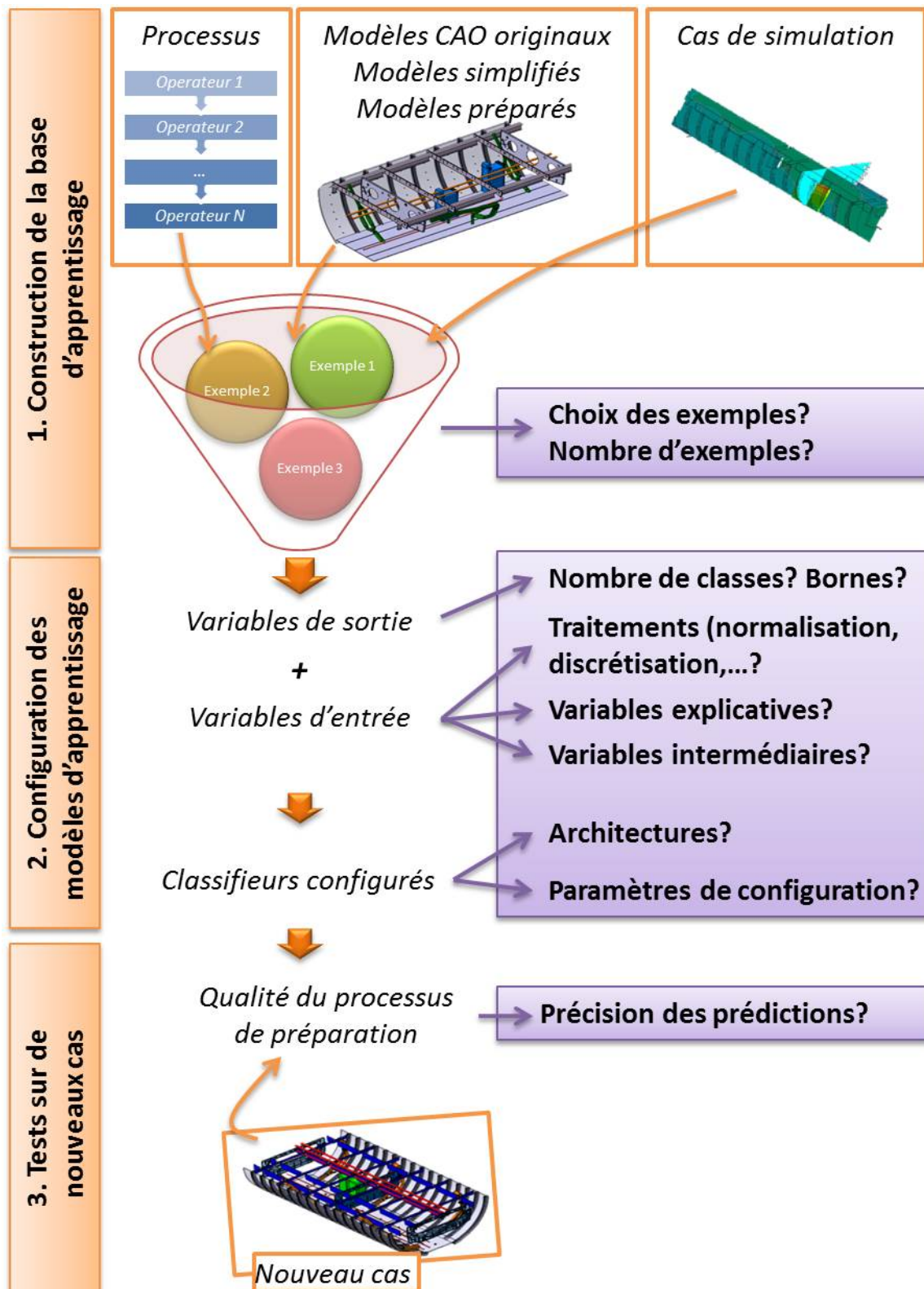
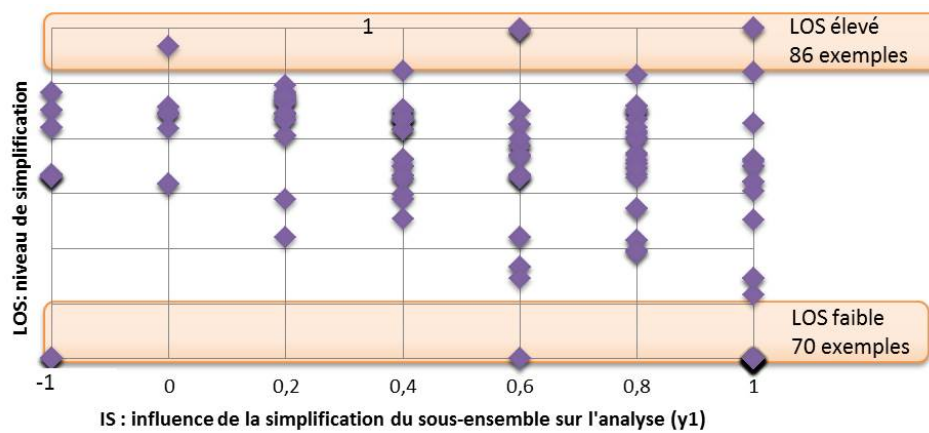
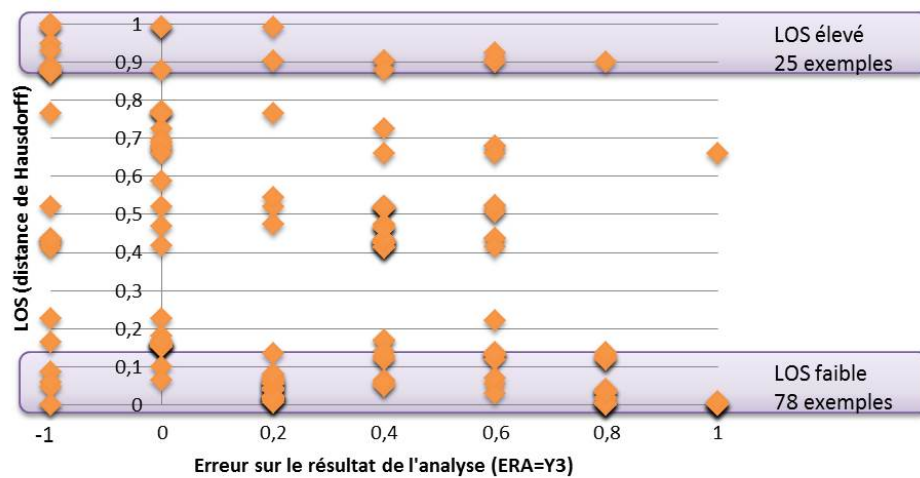


FIGURE 8.11 – Synthèse de l'approche proposée et questions pour la construction d'exemples, la configuration de modèles d'apprentissage et l'évaluation des prédictions.



Classe y1=IS	-1 (non simulable)	0 (influence importante)	0,2	0,4	0,6	0,8	1 (influence négligeable)
Nombre d'exemples	19	10	23	34	30	32	46

FIGURE 8.12 – Répartition des $N_a = 195$ exemples de **sous-processus de simplification sur un sous-ensemble** dans l'espace LOS (niveau de simplification) / IS (influence de la simplification sur l'analyse).



Classe y3=ERA	-1 (non simulable)	0 (erreur importante)	0,2	0,4	0,6	0,8	1 (erreur négligeable)
Nombre d'exemples	57	41	44	45	24	25	9

FIGURE 8.13 – Répartition des $N_a = 245$ exemples de **processus globaux de préparation** dans l'espace LOS (niveau de simplification) / ERA (erreur sur le résultat de l'analyse).

Vérification du nombre minimal d'exemples

Les paramètres de chaque opération de simplification sont décrits par un vecteur de 14 variables explicatives $op1_1 \dots op1_{14}$. Si le nombre d'exemples pour une opération de simplification n'est pas suffisant, celle-ci ne sera pas déterminante et l'erreur sur

l'erreur moyenne de classification décroît lorsque le nombre d'exemples représentant une opération de simplification augmente. Lorsque le nombre d'exemples est supérieur à 11, l'erreur moyenne de classification n'évolue plus, le nombre d'exemples n'a alors plus d'influence sur le résultat des classifications. Il est donc nécessaire d'avoir au minimum 11 exemples pour chaque opération.

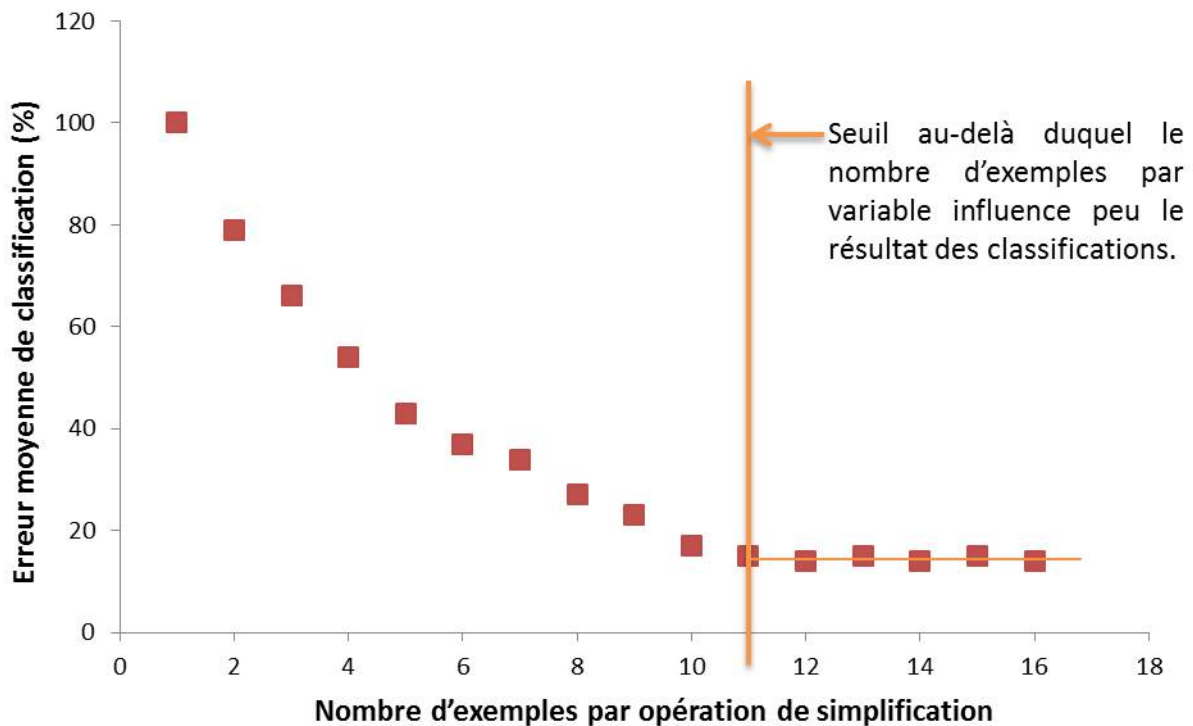


FIGURE 8.14 – Identification du nombre minimum d'exemples par variable décrivant une opération de simplification.

Les figures 8.15 et 8.16 donnent respectivement pour l'ensemble des quatre modèles le nombre d'opérations et de séquences utilisés. Les exemples sont en nombre suffisant pour chaque technique de simplification. Par contre, concernant les opérations de defeaturing et de filtrage, leurs paramètres de réglage ont été limités à quelques valeurs singulières afin de restreindre le nombre d'exemples nécessaire. Par exemple, il est actuellement possible de prendre en compte l'influence de la suppression des trous en général mais pas l'influence des trous de petite taille par rapport à ceux de taille moyenne. Au final, l'étude pourra apporter une aide aux analystes pour choisir une opération de simplification plutôt qu'une autre et pour choisir leurs paramètres généraux. Par contre, cette aide sera donc limitée dans le choix des paramètres plus précis qui pilotent ces opérations.

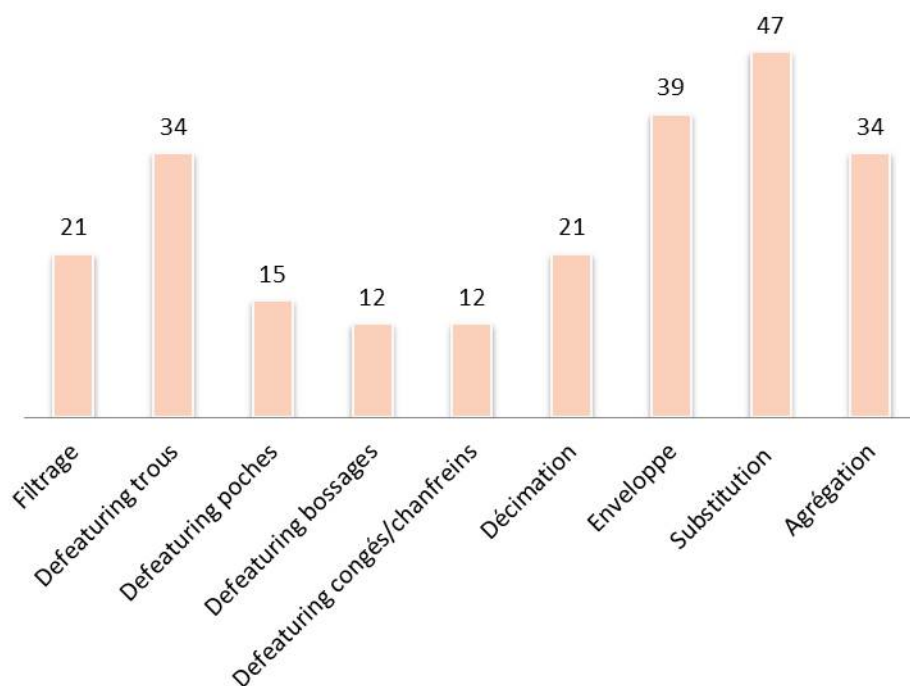


FIGURE 8.15 – Nombre d'exemples par type d'opération de simplification pour l'ensemble des modèles.

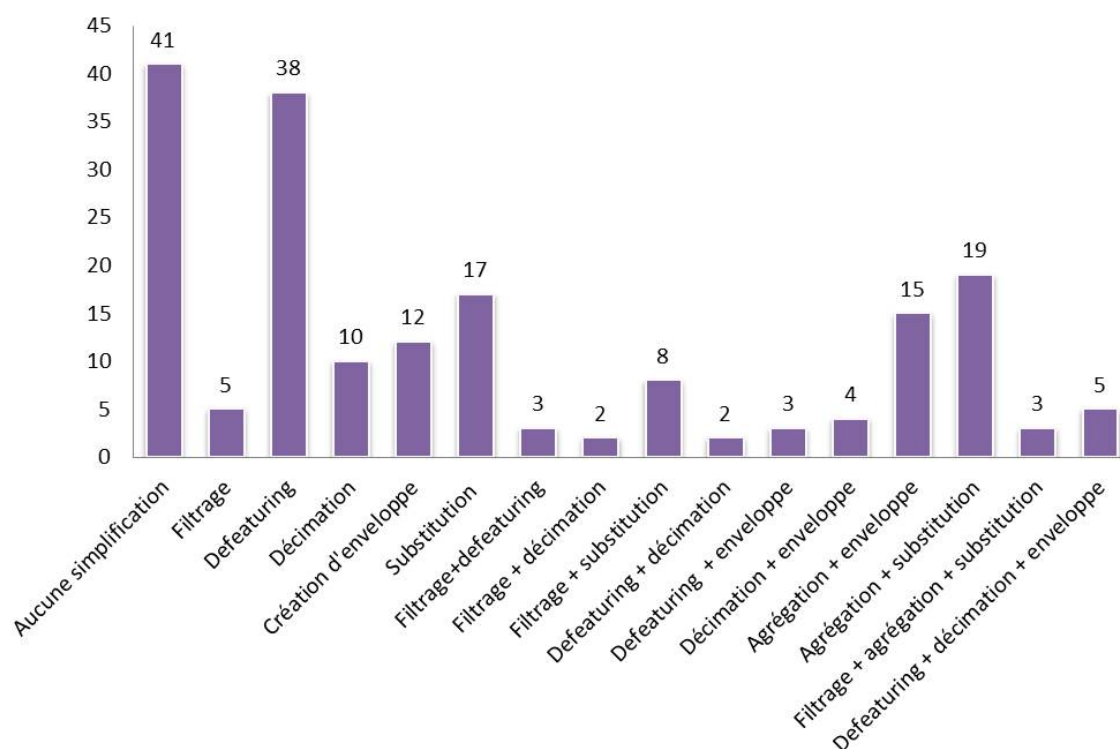


FIGURE 8.16 – Nombre d'exemples par processus de simplification (séquences d'opération utilisés) pour l'ensemble des modèles.

Nous devons répondre dans cette section à deux questions parmi celles qui ont été posées dans la première partie relative à l'état de l'art et aux verrous scientifiques :

- Comment définir un niveau de simplification global du modèle? Comment comparer les modèles intermédiaires?
- Quels sont les critères géométriques qui influencent le résultat de l'analyse numérique?

Il a été constaté dans les sections (6.2 et 6.3) sur la représentation des données et le choix des variables explicatives que les critères qui caractérisent au mieux les critères de qualité d'un processus de préparation ne sont pas connus. De nombreuses variables ont alors été proposées.

Pour chaque variable à prédire, les variables explicatives ont été sélectionnées parmi l'ensemble des variables extraites selon le protocole présenté sur la figure 6.18 qui consiste à supprimer les variables corrélées et à retenir les variables les plus déterminantes.

Les objectifs de cette section sont donc de lister les variables les plus déterminantes pour caractériser les variables à prédire et d'évaluer l'influence de la sélection des variables explicatives sur le résultat des prédictions.

8.2.2.1 Analyse des variables corrélées

Plusieurs variables avaient été proposées pour une même grandeur physique (valeur brute, valeur adimensionnelle, valeur moyenne...). Des ensembles de variables corrélées ont été identifiés, les variables les plus déterminantes pour représenter la grandeur ont été retenues après analyse à l'aide de la méthode proposée dans la section 6.3.1. Ainsi :

- la grandeur "aire" d'un sous-ensemble est caractérisée par une valeur adimensionnelle calculée à partir de l'aire totale du sous-ensemble divisé par sa longueur moyenne au carré;
- la grandeur "volume" d'un sous-ensemble est caractérisée par une valeur adimensionnelle calculée à partir du volume du sous-ensemble divisé par celui de sa boîte englobante;
- la grandeur "distance de Hausdorff" entre un modèle original et préparé d'un sous-ensemble est caractérisée par la valeur quadratique moyenne des distances.

Le tableau 8.14 montre le gain sur la précision des prédictions et sur la durée de l'apprentissage de la suppression des variables corrélées. La précision des prédictions est déterminée à partir des erreurs quadratiques moyennes (AQE equation

4.3) les prédictions. Les gains sont calculés à partir des valeurs de AQE et de durée sur le modèle brut et les modèles dont les variables corrélées ont été analysées et supprimées. L'analyse de variables corrélées permet d'améliorer de quelques % la précision des prédictions. Elle permet de diminuer significativement la durée de construction de certains classifieurs.

Avec :

$$\text{Gain}(\text{AQE}) = \text{AQE}(\text{modèle.corrigé}) - \text{AQE}(\text{modèle.brut})$$

$$\text{Gain}(\text{durée.apprentissage}) = \frac{\text{Durée}(\text{modèle.corrigé}) - \text{Durée}(\text{modèle.brut})}{\text{Durée}(\text{modèle.brut})}$$

Influence des variables corrélées	Gain sur l'AQE		Gain sur la durée d'apprentissage	
Type de classifieur :	y1 (IS)	y2 (CS)	Y1	Y2
Réseaux de neurones	0%	0%	-24%	-5,5%
Machines à support de vecteur	0%	-1,5%	-14%	-50%
K-moyennes	0%	-1,5%	-0,5%	-25%
Arbre de décision	-2,5%	-1,0%	0%	0%
Naïves Bayes	-4,5%	-6,0%	-1,0%	-1,0%

TABLEAU 8.14 – Gains sur l'erreur quadratique moyenne des prédictions (AQE) et sur la durée d'apprentissage dues à l'analyse et à la suppression des variables corrélées pour la prédiction de l'influence de la simplification d'un sous-ensemble sur l'analyse (IS) et du coût de la simplification (CS).

8.2.2.2 Sélection des variables explicatives déterminantes

Les variables explicatives déterminantes ont été sélectionnées selon le protocole proposé sur la figure 6.18. Afin de limiter le nombre de variables intermédiaires à estimer, seules les variables les plus déterminantes ont été retenues parmi les variables non connues pour nouveau cas. Après classement des variables par rapport à leur influence sur les prédictions, la sélection s'arrête lorsque l'ajout de variables n'améliore plus la précision des prédictions.

En plus de la traditionnelle méthode de validation croisée, la sélection de variables déterminantes a été réalisée dans un premier temps sur les modèles 1 et 2 puis sur les données des modèles élémentaires, 2 et 3. Il a été vérifié que les variables proposées étaient identiques dans les deux cas.

La liste des variables sélectionnées est donnée pour chaque principale variable à

Le tableau 8.15 donne, pour chaque variable principale, les erreurs de prédiction AQE pour l'ensemble des classifieurs testés après analyse et suppression des variables corrélées. Ce tableau montre l'influence de la sélection des variables sur la précisions des prédictions en fonction du scénario choisi tels qu'ils ont été proposés dans la section 7.1.4.2..

- Scénario A : les variables intermédiaires sont estimées à partir de toutes les variables explicatives.
- Scénario B : les variables intermédiaires ne sont pas utilisées, seules les variables connues sont utilisées.
- Scénario C : les variables intermédiaires sont utilisées après avoir été estimées uniquement à partir de variables connues. Les résultats du scénario C ont été obtenus en remplaçant les valeurs réelles des variables intermédiaires par les valeurs estimées.

Scénario	Variables connues	Variables intermédiaires	AQE y1	AQE y2	AQE y3	AQE y4	AQE y5
A	Toutes	Toutes	53%	45%	66%	49%	46%
A	Toutes	Déterminantes	52%	40%	65%	48%	44%
A	Déterminantes	Déterminantes	49%	41%	65%	48%	45%
B	Toutes	Aucune	68%	48%	69	48%	33%
C	Toutes	Déterminantes	66%	42%	68%	48%	32%
C	Déterminantes	Déterminantes	62%	43%	68%	48%	32%

TABLEAU 8.15 – Erreurs de prédictions moyennes par principale variable à prédire pour les 3 scénarios. "Toutes" : toutes les variables sont utilisées. "Déterminantes" : seules les variables les plus déterminantes sont sélectionnées.

Les résultats du scénario B pour les variables y4 (coût de la préparation CP) et y5 (coût de l'analyse CA) montrent que l'on peut se passer des variables intermédiaires pour la prédiction des coûts globaux. Par contre, pour les autres variables, le taux d'erreur important des modèles construits selon le scénario B montre que l'utilisation des variables intermédiaires est indispensable. Le scénario A est difficile à mettre en œuvre sur un nouveau cas dont les variables intermédiaires ne sont pas connues. Le scénario C est choisi pour l'estimation des variables y1 (influence de la simplification d'un sous-ensemble sur l'analyse IS) , y2 (coût de la simplification CS) et y3 (erreur sur le résultat de l'analyse ERA).

En conclusion :

- les variables y1(IS), y2 (CS) et y3 (ERA) sont estimées à partir des variables connues et intermédiaires limitées aux plus déterminantes, les variables intermédiaires sont estimées à partir de variables connues seulement ;

8.2.3 Configuration des modèles d'apprentissage

Le protocole expérimental pour la configuration des modèles d'apprentissage est celui de la figure 7.2. L'objectif de cette section est de configurer le meilleur modèle d'apprentissage pour chaque variable à prédire en définissant :

- les traitements à appliquer aux données ;
- le classifieur à utiliser ;
- la configuration du classifieur c'est à dire en définissant son paramétrage.

Il devra être vérifié dans cette section que la méthode proposée dans la section 7.1 et illustrée sur la figure 7.2 pour la construction de modèles d'apprentissage permet d'obtenir une précision satisfaisante des prédictions .

Afin d'identifier le meilleur modèle d'apprentissage (couple "données d'apprentissage et classifieurs") le traitement des variables se fait en parallèle du choix des classifieurs.

Pour chaque principale variable de sortie à prédire, une série de modèles d'apprentissage a été réalisée par validation croisée pour cinq types de classifieurs et pour différents traitements de variables. Ces expérimentations ont permis d'évaluer l'influence sur les prédictions des différents traitements de variables d'entrée et de choisir le classifieur le plus approprié à chaque variable de sortie.

8.2.3.1 Partitionnement des exemples

Deux séries d'essais ont été réalisées :

- Série 1 : les modèles 1 et 2 ont été utilisés pour l'apprentissage ($N_a = 160$ exemples). $N_t = 25$ exemples construits à partir du modèle 1 ont été utilisés pour les tests. Ces derniers n'ont pas été utilisés pour l'apprentissage.
- Série 2 : les modèles élémentaires, 2 et 3 ont été utilisés pour l'apprentissage ($N_a = 216$ exemples), le modèle 1 ($N_t = 75$ exemples) a été utilisé exclusivement pour les tests d'évaluation de la classification.

Les exemples pour l'apprentissage ont été ré-échantillonnés en $K = 10$ groupes dans le but d'obtenir un bon compromis entre le biais (erreur d'approximation) et la variance (erreur de généralisation).

8.2.3.2 Influence de la normalisation des variables sur la précision des prédictions

Le tableau 8.16 montre l'influence de la normalisation des variables sur la précision des prédictions et sur la durée de l'apprentissage. Ce traitement s'est révélé indispensable pour les classifieurs type réseaux de neurones et permet d'améliorer modestement le résultat des classifications. Les tests ont montré que la normalisation permet de réduire dans une proportion importante la durée de l'apprentissage lorsque celui-ci est élevé.

Influence de la normalisation		y1(IS)	y2(CS)	y3(ERA)	y4(CP)	y5(CA)
Gain moyen sur l'AQE (Avec / sans normalisation)		-0,2%	-5,2%	-2,0%	-0,4%	-2,8%
Durée moyenne d'apprentissage (s)	Sans normalisation	9	6	409	318	343
	Avec normalisation	11	1,5	6,5	8	7,5
Gain sur la durée d'apprentissage (Avec / sans normalisation)		-0,8%	0,6%	-291%	-192%	-164%

TABLEAU 8.16 – Gains moyens sur l'erreur des prédictions AQE et sur la durée de l'apprentissage pour la prédiction de y1 (influence de la simplification d'un sous-ensemble sur l'analyse IS), y2 (coût de la simplification CS), y3 (erreur sur le résultat de l'analyse ERA) y4 (coût de la préparation CP) et y5 (coût de l'analyse CA).

8.2.3.3 Sélection et configuration des classifieurs

Le tableau 8.17 donne le modèle d'apprentissage sélectionné et configuré pour chaque variable à prédire. Les résultats sont donnés pour des nombres de classes prédéterminés. Pour une variable de sortie donnée, l'intervalle des valeurs est identique pour toutes les classes.

AQE(%cl.)	y1(IS)	y2(CS)	y3(ERA)	y4(CP)	y5(CA)
Nombre de classes	7	6	7	6	7
Variables connues	Déterminantes	Déterminantes	Déterminantes	Toutes	Toutes
Variables intermédiaires	Déterminantes	Déterminantes	Déterminantes	Aucune	Aucune
Réseaux de neurones	26 (82%)	20 (86%)	41 (68%)	15 (97%)	15 (97%)
Machines à support de vecteur	88 (82%)	102 (78%)	82 (68%)	97 (96%)	93 (96%)
K-moyennes	30 (76%)	21 (81%)	60 (53%)	12 (96%)	15 (97%)
Arbre de décision C.4.5	24 (77%)	25 (78%)	60 (55%)	15 (97%)	15 (97%)
Naïves Bayes	64 (46%)	33 (69%)	58 (52%)	15 (95%)	18 (95%)

TABLEAU 8.17 – Résultats des erreurs de prédiction (AQE) et taux d'instances correctement classées (%Cl.) pour les variables à prédire (répartition régulière des valeurs).

Ainsi, les classifieurs retenus sont les réseaux de neurones pour la prédiction des variables y1, y2 et y3. Les arbres de décision de type C.4.5 seront utilisés pour prédire les variables y4 et y5.

Il s'agit ici de définir le nombre de classes dans lesquelles seront réparties les variables de sortie ainsi que les intervalles de valeurs pour chaque classe.

Les valeurs des variables y_1 et y_3 (influence de la simplification sur le sous-ensemble et erreur globale sur le résultat de l'analyse) ont été initialement réparties dans sept classes. Les variables y_2 , y_4 et y_5 qui décrivent les coûts de simplification, de préparation et de simulation n'ont pas été initialement distribuées dans un nombre fini de classes laissant ainsi la possibilité d'une estimation par régression.

Le tableau 8.18 donne le taux de classements corrects en fonction du nombre de classes et du mode de distribution des valeurs dans ces classes, à savoir :

- **distribution régulière** lorsque les intervalles de valeurs sont identiques pour toutes les classes, ce mode de distribution est utilisé en premier lieu pour sélectionner un classifieur et ses principaux paramètres ;
- **distribution par classification** lorsque les intervalles de valeurs sont définis par classification à l'aide de la méthode des k-moyennes, ce mode de distribution est réalisé et utilisé dans un second temps lors de l'optimisation des classifieurs et des variables.

La figure 8.17 donne le taux d'exemples correctement classés en fonction du nombre de classes utilisées pour chaque variable de sortie. Ces taux de classements ont été obtenus après traitement des variables et avant optimisation des classifieurs avec les classifieurs donnant les meilleurs résultats (réseaux de neurones pour les variables y_1 , y_2 et y_3 , arbres de décisions C.4.5 pour les variables y_4 et y_5). Les intervalles de valeurs sont identiques pour chaque classe. Avec une distribution régulière des valeurs, seuls les modèles à deux classes (variables booléennes) permettent d'obtenir un taux de classements corrects de 100%.

La distribution par classification est donc indispensable. Il est possible d'obtenir un taux de classifications correctes de 100% en distribuant les valeurs des variables dans 5 ou 7 classes. Ces résultats devront être vérifiés à l'aide de tests sur des nouveaux cas qui n'ont pas été utilisés pour l'apprentissage.

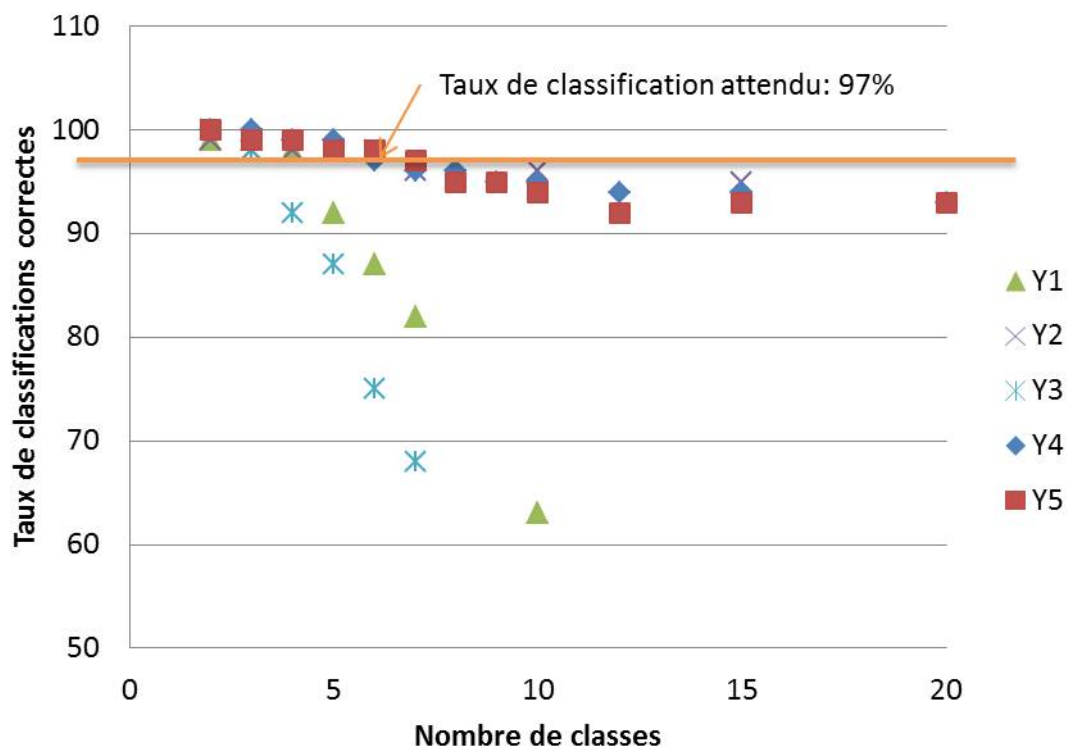


FIGURE 8.17 – Taux d'exemples correctement classés en fonction du nombre de classes pour les cinq variables de sortie.

		y1 (IS)	y2 (CS)	y3 (ERA)	y4 (CP)	y5 (CA)
Distribution régulière	Nb. classes	7	6	7	6	7
	AQE (Cl.%)	26(82%)	20(80%)	41(68%)	15(97%)	15(97%)
	Nb. classes	4	6	3	6	7
	AQE (Cl.%)	14(97%)	15(97%)	13(97%)	15(97%)	15(97%)
Distribution par classification	Nb. classes	5	7	5	7	7
	AQE (Cl.%)	0(100%)	0(100%)	0(100%)	0(100%)	0(100%)

TABEAU 8.18 – Exemples de résultats de d'erreurs des prédictions AQE et de taux de classifications correctes (Cl.%) pour les cinq principales variables à prédire avec des distributions régulières ou par classification des valeurs.

8.2.3.5 Résultats de configuration des modèles d'apprentissage

Les résultats précédents montrent que les classifieurs ont pu être correctement configurés et que le choix des variables explicatives est adapté à ces classifieurs. Le tableau 8.19 donne en synthèse les configurations choisies pour chaque variable à prédire en répondant aux questions posées sur la figure 8.11. Les résultats des classifications sont données pour des distributions régulières des variables initiales (1) et des distributions par classification (2). Les taux de cas mal classés ont été réduits en modifiant les bornes des classes des variables.

Au final, l'utilisateur devra choisir entre un modèle d'apprentissage offrant des prédictions précises (plus grand nombre de classes) avec un taux de classification cor-

reçoit ou offrant des prédictions moins précises (avec moins de classes) mais dont le taux de classification est plus fiable.

	y1 (IS)	y2 (CS)	y3 (ERA)	y4 (CP)	y5 (CA)
Nombre de classes	7 ⁽¹⁾ / 5 ⁽²⁾	6 ⁽¹⁾ / 7 ⁽²⁾	7 ⁽¹⁾ / 5 ⁽²⁾	6 ⁽¹⁾ / 7 ⁽²⁾	7 ⁽¹⁾ / 7 ⁽²⁾
Traitement	Normalisation par modèle et analyse des variables corrélées	Normalisation sur l'ensemble de la base et analyse des variables corrélées	Normalisation par modèle et analyse des variables corrélées	Normalisation sur l'ensemble de la base	Normalisation sur l'ensemble de la base
Variables explicatives (VE) sélectionnées	46 VE (figure B.1)	45 VE (figure B.1)	55 VE (figure B.2)	34 VE (figure B.3)	34 VE (figure B.3)
Variables intermédiaires (VI)	Utilisation des VI déterminantes estimées uniquement à partir de VE connues	Utilisation des VI déterminantes estimées uniquement à partir de VE connues	Utilisation des VI déterminantes estimées uniquement à partir de VE connues	VI non utilisées	VI non utilisées
Classifieur	Réseau de neurones Multilayer-Perceptron 3 couches	Réseau de neurones Multilayer-Perceptron 3 couches	Réseau de neurones Multilayer-Perceptron 3 couches	Arbre de décision C.4.5	Arbre de décision C.4.5
Durée d'apprentissage	530s	20s	160s	10s	10s
Précision des prédictions AQE(%cl.)	26(82%) ⁽¹⁾ / 0(100%) ⁽²⁾	20(86%) ⁽¹⁾ / 0(100%) ⁽²⁾	41(68%) ⁽¹⁾ / 0(100%) ⁽²⁾	15(97%) ⁽¹⁾ / 0(100%) ⁽²⁾	15(97%) ⁽¹⁾ / 0(100%) ⁽²⁾

TABLEAU 8.19 – Synthèse sur la configuration des modèles d'apprentissage pour la prédiction de y1 (influence de la simplification d'un sous-ensemble sur l'analyse IS) , y2 (coût de la simplification CS), y3 (erreur sur le résultat de l'analyse ERA) y4 (coût de la préparation CP) et y5 (coût de l'analyse CA). (1) : avec une distribution régulière des valeurs par classe. (2) : avec une distribution par classification des variables.

8.3 Validation de l'estimateur de qualité d'un processus de préparation

Les modèles d'apprentissage ont été configurés pour chaque variable principale et intermédiaire à prédire. Ces modèles ont été testés sur des nouveaux cas afin de les valider et d'identifier la meilleur stratégie pour prendre en compte les variables intermédiaires parmi les propositions illustrées sur les figures 7.8 et 7.9. Chaque nouveau cas de processus, le modèle a été préparé puis simulé afin de valider les résultats des prédictions.

8.3.1 Description des nouveau cas

Les $Nt = 75$ nouveaux cas testés sont issus du modèle N°1 qui n'a pas été utilisé pour l'apprentissage auxquels ont été ajoutés 5 nouveaux cas issus du modèle N°1. Les nouveaux cas dont certains sont représentés sur la figure 8.18 étaient représentatifs de tout l'espace "niveau de simplification et erreur sur le résultat de l'analyse".

Un « meilleur » processus de préparation (M_{60}^1) a été identifié parmi les 75 processus testés pour le modèle N°1 selon le protocole proposé sur la figure 7.10.

8.3.2 Estimation de l'impact de la simplification sur un sous-ensemble (variables y1 et y2)

Le tableau 8.20 présente, pour les variables y1 (influence de la simplification d'un sous-ensemble sur l'analyse IS) et y2 (coût de la simplification CS), les indicateurs de confiance (nombre de classifications correctes / nombres de tests) des classifieurs configurés selon les différentes stratégies proposées sur la figure 7.8 et rappelées dans la seconde colonne du tableau.

Stratégie	Utilisation des variables intermédiaires :	y1 (IS)	y2 (CS)
A	Déterminantes (toutes)	61/65	65/65
B	Aucune	44/65	64/65
C	Déterminantes (description des modèles CAO + facteurs d'influence sur l'analyse)	46/65	65/65
C'	Déterminantes (comparaison des modèles + facteurs d'influence sur l'analyse)	44/65	65/65

TABEAU 8.20 – Indicateurs de confiance des modèles d'apprentissage pour les variables y1 et y2 et les stratégies proposées.

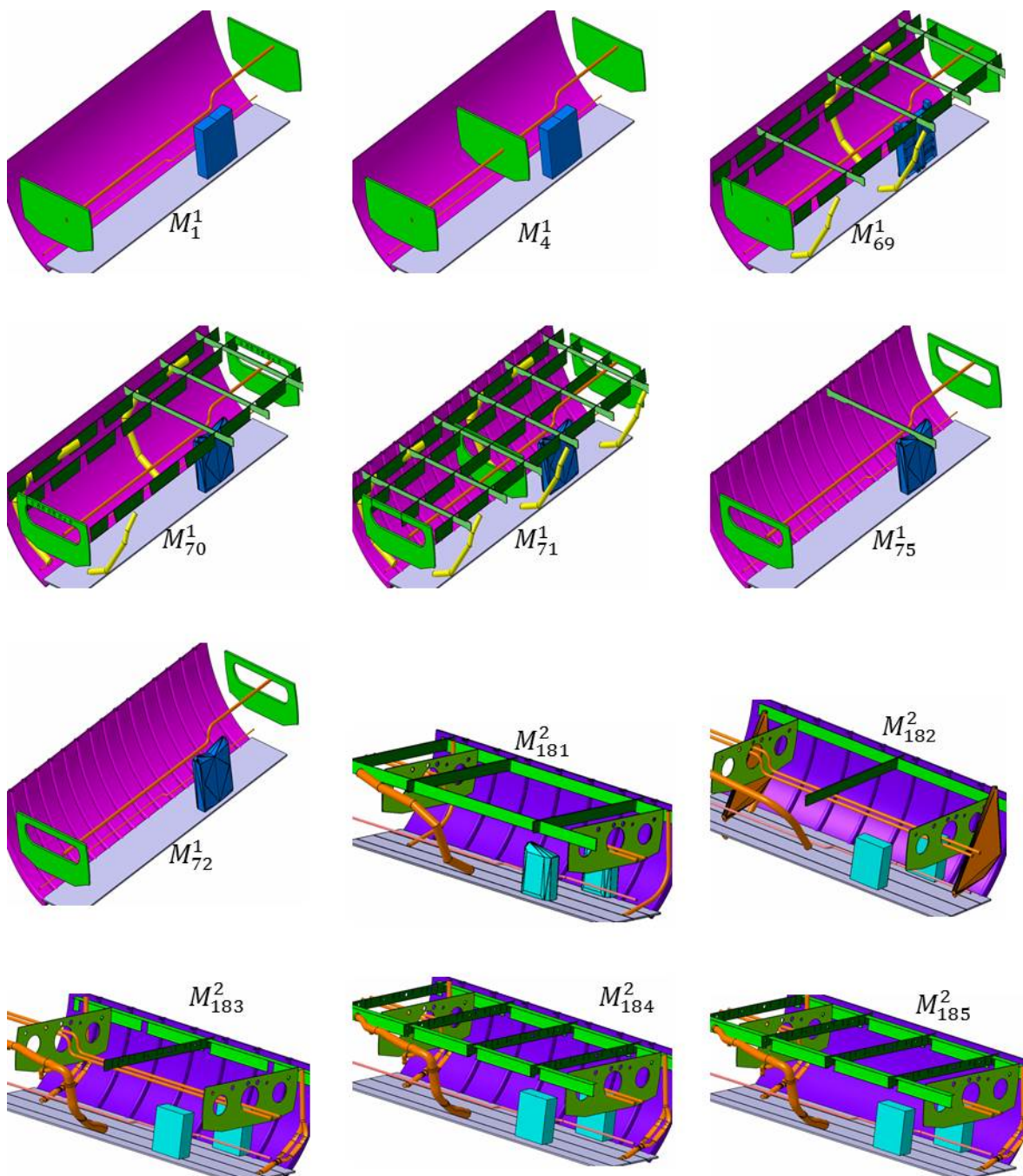


FIGURE 8.18 – Exemples de nouveau cas (non utilisés lors de l'apprentissage issus des modèles M^1 et M^2).

L'indicateur de confiance de 64/63 sur les tests relatifs à la variable y_2 montre que la stratégie B donne de bons résultats sur de nouveaux cas. On pourrait alors se passer de l'utilisation de variables intermédiaires pour estimer le coût de la simplification sur un sous-ensemble

Les stratégies C (utilisation de variables intermédiaires "CAO" et "facteurs d'influence sur l'analyse") et C' (utilisation de variables intermédiaires "comparaison des modèles" et "facteurs d'influence sur l'analyse") avaient été proposées pour limiter le nombre de variables intermédiaires. Or, il apparaît que l'on ne peut pas se passer de toutes les variables intermédiaires déterminantes. La stratégie A reste la meilleure pour estimer la variable y_1 .

Les modèles d'apprentissage configurés sont performants mais pourraient être améliorés pour la prédiction :

- de l'influence de la simplification sur un ensemble IS y_1 avec 61 nouveaux cas bien classés sur 65 (sans erreurs inacceptables) ;
- de l'erreur sur le résultat de l'analyse y_3 avec 73 nouveaux cas bien classés sur 80 (tableau 8.21) dont deux erreurs inacceptables (modèles M_4^1 et M_{72}^1 pour les quelles l'erreur prédite est plus optimiste que la valeur réelle).

Ces résultats sont dus à la prise en compte de variables intermédiaires dans le résultat final des principales variables à prédire et au manque d'exemples dans certaines classes.

8.3.3 Estimation de l'impact de la préparation sur le modèle global (variables y_3 , y_4 et y_5)

Le tableau 8.21 présente les indicateurs de confiance des classifieurs configurés pour les variables y_3 (erreur sur le résultat de l'analyse ERA) y_4 (coût de la préparation CP) et y_5 (coût de l'analyse CA). Le tableau 8.22 donne des exemples d'erreurs de prédiction pour les stratégies retenues présentées sur la figure 7.9 et rappelées dans la seconde colonne du tableau.

Comme pour le coût de la simplification y_2 , les coûts de préparation CP (y_4) et de simulation CA (y_5) peuvent être estimés uniquement à partir des variables connues (stratégie B).

Il avait été proposé des variables caractéristiques de l'influence globale de la simplification de chaque sous-ensemble sur le résultat de l'analyse et le coût global de

Stratégie	Utilisation des variables intermédiaires :	y3 (ERA)	y4 (CP)	y5 (CA)
A	Déterminantes (toutes) + influence globale de la simplification	73/80		
B	Aucune	45/80	80/80	80/80
C	Déterminantes (toutes)	73/80		

TABLEAU 8.21 – Indicateurs de confiance des modèles d'apprentissage pour les variables y3, Y4 et Y5 et les stratégies proposées.

Modèle	Type de préparation	Y3 (ERA)		y4 (CP)		y5 (CA)	
		Réel c(vb)	Prédit c	Réel c(vb)	Prédit c	Réel c(vb)	Prédit c
M ₁ ¹	Simplification maximale + filtrage tout	0 (60%)	0	2 (220)	2	1 (1751)	1
M ₄ ¹	Simplification maximale + filtrage partiel	0 (60%)	0.4 ^(!)	2 (237)	2	1 (1771)	1
M ₆₉ ¹	Simplification intermédiaire + filtrage partiel	0 (60%)	0	5 (765)	5	5 (3542)	5
M ₇₀ ¹	Simplification intermédiaire + filtrage partiel	-1 (echec)	-1	6 (1008)	6	0 (echec)	0
M ₇₁ ¹	Simplification un sous-ensemble	0,6 (0,7%)	0.4	6 (943)	6	6 (5298)	6
M ₇₂ ¹	Simplification un sous-ensemble + filtrage tout	0,6 (0,7%)	0.8 ^(!)	3 (352)	3	3 (3703)	3
M ₇₅ ¹	Simplification un sous-ensemble + filtrage partiel	0,2 (8%)	0	3 (382)	3	4 (3920)	4

TABLEAU 8.22 – Exemples de résultats de prédictions. c : classe. vb : valeur brute. (!) : erreurs inacceptables (les valeurs prédites sont plus optimistes que les valeurs réelles).

simplification. L'intégration de ces variables (stratégie A) ne permet pas d'améliorer la précision des classifications. La stratégie C sera donc retenue pour la prédiction de l'erreur sur le résultat de l'analyse.

La décision finale sur la qualité des processus de préparation est prise à partir de trois critères : l'erreur sur le résultat de l'analyse y3, le coût de la préparation y4 et le coût de la simulation y5. Tant que les erreurs de prédictions sont modérées comme celles relevées précédemment, la décision globale reste inchangée (tableau 8.23).

Modèle	Décision à partir des valeurs réelles	Décision à partir des valeurs prédites
M ₁ ¹	Erreur très importante - processus non validé	Erreur très importante - processus non validé
M ₄ ¹	Erreur importante - processus non validé	Erreur très importante - processus non validé
M ₆₉ ¹	Erreur importante - processus non validé	Erreur importante - processus non validé
M ₇₀ ¹	Erreur importante - processus non validé	Erreur importante - processus non validé
M ₇₁ ¹	Erreur très faible et coûts élevés - processus non validé	Erreur très faible et coûts élevés - processus non validé
M ₇₂ ¹	Erreur très faible et coûts moyens - processus validé	Erreur faible et coûts moyens - processus validé
M ₇₅ ¹	Erreur moyenne - processus non validé	Erreur moyenne - processus non validé

TABEAU 8.23 – Exemples de décisions obtenues à partir des résultats sur les valeurs réelles et prédites par apprentissage donnés dans le tableau 8.22.

8.4 Synthèse sur l'estimation de la qualité d'un processus de préparation

Il s'agit de synthétiser le domaine de validité des travaux et plus précisément des classifieurs du point de vue :

- de l'objectif de la préparation ;
- des opérations de préparation ;
- de l'estimation de la qualité d'un processus de préparation à l'aide de classifieurs.

8.4.1 Domaine de validité du point de vue de l'objectif de préparation

Les modèles originaux traités sont des ensembles CAO 3D contenant plusieurs sous-ensembles et des centaines de pièces. La modélisation proposée peut s'adapter à tout modèle CAO riche quel que soit le nombre sous-ensembles et de pièces qui le constituent.

Les modèles préparés sont les maillages de volumes délimités par l'ensemble des composants du modèle simplifié. Au delà de l'analyse CFD en mécanique des fluides et thermique, l'approche proposée peut s'appliquer à tout objectif de simulation qui nécessite la création d'un volume de fluide interne.

8.4.2. Domaine de validité du point de vue des opérations de préparation

L'étude qui vise à évaluer les processus de préparation se limite à l'évaluation de six opérations de simplification (filtrage, defeaturing, décimation, substitution, création d'enveloppes et agrégation). Les opérations d'adaptation et de maillage ont toutes été réalisées avec le même outil et la même méthode. Les modèles maillés ont tous les mêmes caractéristiques principales : maillage de surfaces et du volume avec des éléments de type triangles (ou tétraèdres) ou quadrilatères (ou hexaèdres), raffinement autour des couches limites et même qualité.

Au niveau des opérations de simplification, l'estimateur de qualité du processus de préparation est sensible au type d'opération utilisée quel que soit leur nombre et leur séquençage. Par contre, les opérations de simplification sont décrites par des paramètres de type "tout ou rien", l'estimateur de qualité du processus de préparation n'est pas sensible aux différents niveaux de simplification offertes par les outils.

8.4.3 Domaine de validité du point de vue de l'estimation de la qualité d'un processus de préparation à l'aide de classifieurs

Afin d'estimer a priori la qualité d'un processus de préparation, les classifieurs configurés sont capables d'estimer pour un nouveau cas :

- l'erreur sur le résultat de l'analyse $y_3 = \text{ERA}$ (écart entre le résultat de l'analyse du modèle testé et le résultat de l'analyse du modèle de référence),
- les coûts de préparation $y_4 = \text{CP}$ (durée totale des opérations de simplification, d'adaptation et de maillage),
- les coûts de simulation $y_5 = \text{CA}$ (durée du calcul).

La décision finale sur la validité d'un processus de préparation et l'identification du meilleur processus sont prises par l'analyste au vu de ces trois critères.

Les variables de sortie ont été distribuées dans plusieurs classes. Plus le nombre de classes est élevé, plus les valeurs estimées par apprentissage seront précises. Par exemple, l'indice de confiance des classifieurs peut atteindre 100% de valeurs correctement classées sur un nouveau cas avec une précision moyenne sur les valeurs estimées d'environ :

- $\pm 5\%$ pour l'erreur sur le résultat de l'analyse avec une distribution dans 4 classes ;
- $\pm 140 \text{ min}$ pour les coûts de préparation avec une distribution dans 7 classes ;
- $\pm 15 \text{ min}$ pour les coûts de simulation avec une distribution dans 7 classes.

Conclusion et perspectives

L'objectif de ces travaux de thèse était de montrer qu'il est possible d'évaluer un processus de préparation d'un modèle pour la simulation sans avoir à le réaliser et ceci en utilisant des techniques d'intelligence artificielle. L'évaluation a priori d'un processus de préparation ne dispense pas les analystes d'effectuer la simulation mais permet d'identifier dès le début le meilleur processus sans faire de sur-qualité et permet d'estimer le coût de la préparation.

Apports de la thèse

Approche pour l'évaluation a priori d'un processus de préparation

Les critères qui guident l'analyste dans le choix d'un processus de préparation sont l'erreur sur le résultat de l'analyse, le coût de la préparation et le coût de la simulation. Une nouvelle approche basée sur les techniques d'apprentissage artificiel a été développée pour évaluer **a priori** la qualité d'un processus de préparation d'un modèle CAO en utilisant ces critères comme des variables de sortie.

L'approche proposée consiste tout d'abord à capitaliser les connaissances des experts dans une base d'exemples, puis à configurer des modèles d'apprentissage capables d'estimer la qualité d'un processus de préparation pour un nouveau cas.

Au niveau de la capitalisation des connaissances et de leur exploitation, le défi était de proposer une représentation formalisée des processus de préparation, des données 3D et de l'ensemble des facteurs qui caractérisent leurs critères d'évaluation puis d'implémenter ces données dans une base pour l'apprentissage.

Au niveau de la configuration des modèles d'apprentissage, le premier défi, et certainement le plus important de cette thèse, était d'identifier les facteurs d'évaluation de l'impact de la simplification sur le résultat de l'analyse. Le second défi était de garantir la fiabilité de l'estimation de la qualité d'un processus malgré un nombre limité d'exemples et en se limitant aux données qui caractérisent les modèles origi-

Proposition d'une modélisation formalisée des processus de préparation

Les méthodes et outils pour préparer un modèle pour la simulation existent. Des techniques de simplification adaptées à la préparation de modèles riches et proposées par des logiciels de modélisation 3D ou de simplification ont été identifiées. Cependant, les critères de choix des opérations de simplification, leur paramètres et l'ordre dans lequel sont réalisées ces opérations n'étaient pas formalisés. L'utilisation de techniques d'intelligence artificielle de type classifieurs nécessite une représentation vectorielle des données. Les processus de préparation ont donc été décrits par des vecteurs sur deux niveaux : au niveau local pour représenter un sous-processus de simplification sur un sous-ensemble et au niveau global pour représenter le processus de préparation de l'ensemble du modèle. Chaque vecteur décrit les paramètres de réglages d'une opération de préparation et l'ordre dans lequel elle est réalisée par rapport aux autres opérations. Ainsi, pour un nombre d'opérations donné au départ tous les processus peuvent être représentés quel que soit le séquençage des opérations et leurs paramètres.

Identification des facteurs d'évaluation de l'impact de la simplification sur le résultat de l'analyse

Les facteurs géométriques qui influencent le résultat de la simulation numérique sont nombreux et dépendent de l'objectif de la simulation. Un grand nombre de critères possibles ont été proposés parmi les facteurs qui comparent deux modèles ou qui évaluent les écarts entre une grandeur physique et le résultat de l'analyse.

Ces facteurs sont les variables d'entrée des classifieurs. Les données extraites des modèles CAO ont été traitées pour être implémentées dans la base d'apprentissage sous forme de vecteurs. Les variables qui caractérisent au mieux l'erreur sur le résultat de l'analyse ont été identifiées grâce à une méthode proposée qui consiste à analyser des variables corrélées puis à sélectionner les facteurs les plus déterminants à l'aide d'algorithmes basés sur les techniques d'intelligence artificielle.

Les bons résultats des prédictions montrent que la sélection des variables d'entrée (donc des facteurs d'évaluation) est pertinente.

Plusieurs classifieurs ont été testés pour différentes configurations de variables afin d'identifier le meilleur modèle d'apprentissage pour chaque variable à prédire. Pour chaque variable à prédire, le choix des modèles d'apprentissages et de leur configuration ont été validés à l'aide de critères tels que l'erreur quadratique moyenne sur les prédictions et la surface sous la courbe ROC.

La préparation de modèles et la simulation sont des opérations lourdes. Le nombre d'exemples de processus de préparation est donc limité. Une méthode pour construire une base d'exemples représentative de tout l'espace "niveau de simplification / erreur sur le résultat de l'analyse" permet une répartition régulière des exemples dans toutes les classes des variables à prédire. L'approche proposée est ensuite basée sur la méthode de validation croisée pour garantir la fiabilité de l'estimation de la qualité d'un processus.

Lorsqu'un nouveau cas se présente, seules les données relatives aux modèles CAO originaux, aux processus de préparation et aux cas de simulation sont connus. L'estimation des principales variables à prédire nécessite de connaître les valeurs de variables intermédiaires relatives aux modèles simplifiés et préparés. Des stratégies ont identifiées les variables intermédiaires indispensables aux prédictions. Pour certaines variables à prédire, ces variables intermédiaires ne sont pas indispensables.

Au final, un indicateur de confiance des prédictions montre que l'approche et les méthodes proposées sont prometteuses pour évaluer a priori la qualité d'un processus de préparation sans avoir à simplifier le modèle.

Perspectives

En guidant le choix des analystes lors d'une étape de la préparation d'un modèle pour la simulation, la méthode proposée permet d'identifier dès le début un processus adapté à l'objectif de la simulation et permet d'estimer ses coûts. La méthode proposée participe ainsi en partie à la réduction du temps de développement d'un produit et de la maîtrise de sa qualité.

L'indicateur de confiance des classifications montre un taux satisfaisant de prédictions correctes pour une précision donnée sur les valeurs estimées. Ces résultats pourraient être améliorés en intégrant de nouveaux exemples. En effet, l'ajout d'exemples judicieusement choisis permettra d'augmenter le nombre de classes dans lesquelles sont distribuées les variables de sortie et donc, par conséquent,

La préparation d'un modèle, telle qu'elle a été modélisée, consiste à le simplifier, à l'adapter à l'objectif de la simulation et à le mailler. L'étude s'est limitée à la comparaison de paramètres de réglages et de séquençages des opérations de simplification.

On pourra, en ajoutant des exemples de simplification utilisant des plages de réglages différentes, estimer la qualité d'un processus non plus uniquement par rapport au choix d'opérations de simplification mais par rapport aux valeurs de leurs paramètres. L'analyste connaîtra ainsi pour un processus évalué, les paramètres précis à appliquer pour chaque opération.

L'approche proposée pourrait ensuite être étendue aux opérations d'adaptation et de maillage. Tout d'abord, de nouveaux processus pourraient être proposés avec des séquençages différents (par exemple en réalisant des opérations de simplification après le maillage). Concernant les opérations de maillage, un plus grand nombre de paramètres pourraient être pris en compte ainsi que différents niveaux de qualité.

Appliquée à l'évaluation des processus de préparation pour l'analyse CFD et thermique, l'approche pourrait être généralisée à l'ensemble du processus de développement d'un produit. On pourrait par exemple, appliquer cette méthode lors des processus d'industrialisation ou de fabrication pour l'aide au choix d'outils et de paramètres de réglages.

Le processus de préparation identifié comme étant le "meilleur", n'est pas le processus optimal mais le meilleur parmi un ensemble de processus testés. De futurs travaux pourraient intégrer une boucle d'optimisation afin de prédire non plus la qualité du processus mais ses variables descriptives optimales pour un objectif de préparation donné en minimisant les coûts de préparation, les coûts de simulation et les erreurs sur le résultat de l'analyse. L'utilisation d'un outil tel que Weka permet de configurer et de tester un classifieur rapidement (10 minutes au maximum avec un réseau de neurones lorsque les variables sont normalisées et quelques secondes avec un arbre de décision C.4.5). Il serait alors possible d'envisager de réaliser un grand nombre d'estimations dans le cadre d'optimisations. L'utilisation combinée de techniques d'intelligence artificielle telles que des classifieurs de type réseaux de neurones et d'algorithmes génétiques pourraient être une piste d'identification d'un processus optimal.

Bibliographie

- [1] ISO 10303-242 :2014. Systèmes d'automatisation industrielle et intégration – Représentation et échange de données de produits – Partie 242 : Protocole d'application : Gestion des modèles 3d d'ingénierie, 2014.
- [2] Mark Ainsworth and Alan Craig. A posteriori error estimators in the finite element method. *Numerische Mathematik*, 60(1) :429–463, 1991.
- [3] C. Andújar, P. Brunet, and D. Ayala. Topology-reducing surface simplification using a discrete solid representation. *ACM Transactions on Graphics*, 21(2) :88–105, 2002.
- [4] Cecil G. Armstrong. Modelling requirements for finite-element analysis. *Computer-Aided Design*, 26(7) :573 – 578, 1994.
- [5] I. Babuška and T. Strouboulis. *The Finite Element Method and Its Reliability*. Numerical mathematics and scientific computation. Clarendon Press, 2001.
- [6] C.B. Barber, D.P. Dobkin, and H. Huhdanpaa. The Quickhull Algorithm for Convex Hulls. *ACM Transactions on Mathematical Software*, 22(4) :469–483, 1996.
- [7] A. Barr and E.A. Feigenbaum. *The Handbook of Artificial Intelligence*. Number vol. 1 in The Handbook of Artificial Intelligence. Heuristech Press, 1981.
- [8] Jérémy-Maxime Berthet. Nouvelles approches pour la maîtrise des phases de génération de maillages éléments finis à partir d'assemblages issus de maquettes numériques. Projet de fin d'études ENSAM Aix en Provence, 2011.
- [9] Jaap Bloem, Pascal Sire, Yves Coze, Nicolas Kowski, Torsten Kulka, and Philippe Sottocasa. *Virtual Concept - Real Profit*. Sogeti, 2009.
- [10] L. Breiman, J. Friedman, C.J. Stone, and R.A. Olshen. *Classification and Regression Trees*. The Wadsworth and Brooks-Cole statistics-probability series. Taylor & Francis, 1984.

BIBLIOGRAPHIE

- [11] A. Buecar, M. Nagode, and M. Fajdiga. An improved neural computing method for describing the scatter of S-N curves. *International Journal of Fatigue*, 29(12) :2125–2137, 2007.
- [12] W.L. Chan, M.W. Fu, and J. Lu. An integrated FEM and ANN methodology for metal-formed product design. *Engineering Applications of Artificial Intelligence*, 21(8) :1170–1181, 2008.
- [13] S. Chen, S. A. Billings, and W. Luo. Orthogonal least squares methods and their application to non-linear system identification. *International Journal of Control*, 50 :1873–1896, 1989.
- [14] Y. Chen, G.A. Kopp, and D. Surry. Prediction of pressure coefficients on roofs of low buildings using artificial neural networks. *Journal of Wind Engineering and Industrial Aerodynamics*, 91(3) :423–441, 2003.
- [15] C. S. Chong, A. Senthil Kumar, and K. H. Lee. Automatic solid decomposition and reduction for non-manifold geometric model generation. *Computer-Aided Design*, 36(13) :1357 – 1369, 2004.
- [16] Cornuéjols, Christel Vrain, and Zucker Jean-Daniel. *Panorama de l'intelligence artificielle*, volume 2 Algorithmes pour l'intelligence artificielle. Cépaudès, Toulouse, 2014.
- [17] J. Dompierre, P. Labbé, F. Guibault, and R. Camarero. Critères de qualité pour les maillages simpliciaux. In Hermès, editor, *Maillage et adaptation*, pages 311–348. 2001.
- [18] G. Dreyfus. *Réseaux de neurones : méthodologie et applications*. Algorithmes (Paris). Eyrolles, 2004.
- [19] Rosalinda Ferrandes. *Shape Operators and Mechanical Criteria in the Preparation of Components for Engineering Analysis*. Theses, Institut National Polytechnique de Grenoble - INPG, May 2008.
- [20] Vincent Feuillard. Scikit-learn for predictive maintenance at Airbus. Paris, France, 2015.
- [21] Lionel Fine. *Processus et méthodes d'adaptation et d'idéalisation de modèles dédiés à l'analyse de structures mécaniques*. PhD thesis, Institut National Polytechnique de Grenoble - INPG, 2001.
- [22] Gilles Foucault, Jean-Christophe Cuillière, Vincent François, Jean-Claude Léon, and Roland Maranzana. Adaptation of CAD model topology for finite element analysis. *Computer-Aided Design*, 40(2) :176 – 196, 2008.

-
- [23] Shuming Gao, Wei Zhao, Hongwei Lin, Fanqin Yang, and Xiang Chen. Feature suppression based CAD mesh model simplification. *Computer-Aided Design*, 42(12) :1178–1188, December 2010.
- [24] Herbert Martins Gomes, Armando Miguel Awruch, and Paulo André Menezes Lopes. Reliability based optimization of laminated composite structures using genetic algorithms and Artificial Neural Networks. *Structural Safety*, 33(3) :186 – 195, 2011.
- [25] S.H. Gopalakrishnan and K. Suresh. A formal theory for estimating defeaturing-induced engineering analysis errors. *CAD Computer Aided Design*, 39(1) :60–68, 2007.
- [26] J.P. Haton. *Le raisonnement en intelligence artificielle : modèles, techniques et architectures pour les systèmes à bases de connaissances*. Collection iia. Inter-Editions, 1991.
- [27] G. Huang, H. Wang, and G. Li. A reanalysis method for local modification and the application in large-scale problems. *Structural and Multidisciplinary Optimization*, 49(6) :915–930, 2014.
- [28] Keisuke Inoue, Takayuki Itoh, Atsushi Yamada, Tomotake Furuhashi, and Kenji Shimada. Face clustering of a large-scale {CAD} model for surface mesh generation. *Computer-Aided Design*, 33(3) :251 – 261, 2001.
- [29] Natraj Iyer, Subramaniam Jayanti, Kuiyang Lou, Yagnanarayanan Kalyanaraman, and Karthik Ramani. Three-dimensional shape searching : state-of-the-art review and future trends. *Computer-Aided Design*, 37(5) :509 – 530, 2005. Geometric Modeling and Processing 2004.
- [30] A. Jahangirian and A. Shahrokhi. Aerodynamic shape optimization using efficient evolutionary algorithms and unstructured {CFD} solver. *Computers & Fluids*, 46(1) :270 – 276, 2011. 10th {ICFD} Conference Series on Numerical Methods for Fluid Dynamics (ICFD 2010).
- [31] S. Jayanti, Y. Kalyanaraman, and K. Ramani. Shape-based clustering for 3d CAD objects : A comparative study of effectiveness. *CAD Computer Aided Design*, 41(12) :999–1007, 2009.
- [32] Athar Kharal and Ayman Saleem. Neural networks based airfoil generation for a given using Bezier–PARSEC parameterization. *Aerospace Science and Technology*, 23(1) :330 – 344, 2012. 35th ERF : Progress in Rotorcraft Research.
- [33] Sun-Jeong Kim, Chang-Hun Kim, and David Levin. Surface simplification using a discrete curvature norm. *Computers & Graphics*, 26(5) :657 – 663, 2002.

- [34] K.W. Lee, T.H. Chong, and G.-J. Park. Development of a methodology for a simplified finite element model and optimum design. *Computers and Structures*, 81(14) :1449–1460, 2003.
- [35] Ming Li, Shuming Gao, and Ralph R. Martin. Engineering analysis error estimation when removing finite-sized features in nonlinear elliptic problems. *Solid and Physical Modeling 2012*, 45(2) :361–372, February 2013.
- [36] P.A.M. Lopes, H.M. Gomes, and A.M. Awruch. Reliability analysis of laminated composite structures using finite elements and neural networks. *Composite Structures*, 92(7) :1603–1613, 2010.
- [37] Ruding Lou. *Modification of semantically enriched FE mesh models*. Theses, Arts et Métiers ParisTech, June 2011.
- [38] Farrukh Mazhar, Abdul Munem Khan, Imran Ali Chaudhry, and Mansoor Ahsan. On using neural networks in {UAV} structural design for {CFD} data fitting and classification. *Aerospace Science and Technology*, 30(1) :210 – 225, 2013.
- [39] Thomas M. Mitchell. *Machine Learning*. McGraw-Hill, Inc., New York, NY, USA, 1 edition, 1997.
- [40] J. Tinsley Oden and Serge Prudhomme. Estimation of Modeling Error in Computational Mechanics. *Journal of Computational Physics*, 182(2) :496 – 515, 2002.
- [41] National Institute of Standards and Technology. Standard for Integration Definition For Function Modeling (IDEF0), 1993.
- [42] Mehran Ali Azizi Oroumieh, S. Mohammad Bagher Malaek, Mahmud Ashrafizaadeh, and S. Mahmoud Taheri. Aircraft design cycle time reduction using artificial intelligence. *Aerospace Science and Technology*, 26(1) :244 – 258, 2013.
- [43] Sylvain PIECHOWIAK. Intelligence artificielle et diagnostic. *Techniques de l'ingénieur Automatique avancée*, base documentaire : TIB393DUO.(ref. article : s7217), 2015.
- [44] J.R. Quinlan. Induction of Decision Trees. *Machine Learning*, 1(1) :81–106, 1986.
- [45] Christopher K. Riesbeck and R.C. Schank. *Inside Case-based Reasoning*. Artificial Intelligence Series. L. Erlbaum, 1989.
- [46] Olivier Ruf. Critères de pilotage du processus de génération des modèles de simulation E.F., application au cas FitFlex pour la simulation de pièces aéronautique. Projet de fin d'études ENSAM Aix en Provence, 2012.

-
- [47] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. *Learning representations by back-propagating errors*. *Nature*, 323(6088) :533–536, October 1986.
- [48] A. Sheffer. Model simplification for meshing using face clustering. *CAD Computer Aided Design*, 33(13) :925–934, 2001.
- [49] Bao-Quan Shi, Jin Liang, and Qing Liu. Adaptive simplification of point cloud using -means clustering. *Computer-Aided Design*, 43(8) :910–922, August 2011.
- [50] I. Steinwart and A. Christmann. *Support Vector Machines*. Information Science and Statistics. Springer New York, 2008.
- [51] R. Sun, S. Gao, and W. Zhao. An approach to B-rep model simplification based on region suppression. *Computers and Graphics (Pergamon)*, 34(5) :556–564, 2010.
- [52] V.B. Sunil and S.S. Pande. Automatic recognition of machining features using artificial neural networks. *International Journal of Advanced Manufacturing Technology*, 41(9-10) :932–947, 2009.
- [53] J. Tang, S. Gao, and M. Li. Evaluating defeaturing-induced impact on model analysis. *Mathematical and Computer Modelling*, 57(3-4) :413–424, 2013.
- [54] Etienne Tassin and Jean-Philippe Pernot. Intelligence artificielle pour le traitement éléments finis de maquettes numériques. PJE Arts et Métiers ParisTech centre d’Aix en Provence PJEx-Ai-12-2-003, 2014.
- [55] Atul Thakur, Ashis Gopal Banerjee, and Satyandra K. Gupta. A survey of CAD model simplification techniques for physics-based simulation applications. *Computer-Aided Design*, 41(2) :65 – 80, 2009.
- [56] S. Tichkiewitch, M. Tollenaere, and P. Ray. *Advances in Integrated Design and Manufacturing in Mechanical Engineering II*. Springer Netherlands, 2010.
- [57] S. Tufféry. *Data mining et statistique décisionnelle : l’intelligence des données*. Editions Technip, 2010.
- [58] Y. Uematsu and R. Tsuruishi. Wind load evaluation system for the design of roof cladding of spherical domes. *Journal of Wind Engineering and Industrial Aerodynamics*, 96(10-11) :2054–2066, 2008.
- [59] P. Véron and J. C. Léon. Static polyhedron simplification using error measurements. *Computer-Aided Design*, 29(4) :287 – 298, 1997. Reverse Engineering of Geometric Models.

BIBLIOGRAPHIE

- [80] A. Yuan, J. Hongfan, and W. Yu. A neural network approach to surface blending based on digitized points. *Journal of Materials Processing Technology*, 120(1-3) :76–79, 2002.

Annexe A

Séquençages d'opérations de simplification

ANNEXE A. SÉQUENÇAGES D'OPÉRATIONS DE SIMPLIFICATION

Opération 1	Opération 2	Opération 3	Opération 4
0	Aucune		
1	Filtrage		
2	Defeaturing		
3	Decimation		
4	Enveloppes		
5	Substitution		
6	Agrégation		
7	Filtrage	Defeaturing	
8	Filtrage	Decimation	
9	Filtrage	Enveloppes	
10	Filtrage	Substitution	
11	Filtrage	Agrégation	
12	Defeaturing	Decimation	
13	Defeaturing	Enveloppes	
14	Defeaturing	Agrégation	
15	Decimation	Agrégation	
16	Decimation	Enveloppes	
17	Enveloppes	Agrégation	
18	Substitution	Agrégation	
19	Agrégation	Defeaturing	
20	Agrégation	Decimation	
21	Agrégation	Enveloppes	
22	Agrégation	Substitution	
23	Filtrage	Defeaturing	Decimation
24	Filtrage	Defeaturing	Agrégation
25	Filtrage	Defeaturing	Enveloppes
26	Filtrage	Decimation	Agrégation
27	Filtrage	Decimation	Enveloppes
28	Filtrage	Enveloppes	Agrégation
29	Filtrage	Substitution	Agrégation
30	Filtrage	Agrégation	Defeaturing
31	Filtrage	Agrégation	Decimation
32	Filtrage	Agrégation	Enveloppes
33	Filtrage	Agrégation	Substitution
34	Defeaturing	Decimation	Agrégation
35	Defeaturing	Decimation	Enveloppes
36	Defeaturing	Agrégation	Decimation
37	Agrégation	Defeaturing	Decimation
38	Filtrage	Defeaturing	Decimation Agrégation
39	Filtrage	Defeaturing	Agrégation Decimation
40	Filtrage	Agrégation	Defeaturing Decimation

TABEAU A.1 – Séquençages réalistes d'opérations de simplification d'un sous-ensemble.

Annexe B

Variables explicatives sélectionnées

Sélection des variables explicatives	Estimation de Y1 = influence de la simplification d'un SE	Estimation de Y2 = coût de la simplification d'un SE
Description du processus de simplification (vecteurs $Op.x_j^n$)	Description des simplifications : toutes variables(*) $\{Op1_j^n, \dots, Op14_j^n\}$	Description des simplifications : toutes variables(*) $\{Op1_j^n, \dots, Op14_j^n\}$
Description du sous-ensemble (vecteurs $CAO.x_j^n$)	Caractéristiques du modèle original(*) $\{CAO.1_0^n, \dots, CAO.16_0^n\}$ + Caractéristiques du modèle simplifié(**) : 1. courbures, 2. nombre de triangles, 3. nombre de patches, 4. volume (adimensionnel / taille du SE), 5. aire (adimensionnel / taille du SE).	Caractéristiques du modèle original(*) $\{CAO.1_0^n, \dots, CAO.16_0^n\}$ + Caractéristiques du modèle simplifié(**) 1. courbures, 2. nombre de patches, 3. ratio nombre de triangles / patches, 4. nombre de triangles, 5. aire (adimensionnel / taille du SE), 6. compacité (type 1).
Facteurs de comparaison des modèles CAO (vecteurs $Compar.x_j^n$)	Bénéfices(**) sur: 1. nombre de patches, 2. ratio nombre de triangles / patches, 3. volume, 4. compacité (type 2), 5. nombre de triangles, 6. distance de Hausdorff (moyenne quadratique), 7. courbures, 8. nombre de pièces.	Bénéfices(**) sur: 1. ratio nombre de triangles / patches, 2. nombre de patches, 3. distance de Hausdorff (moyenne), 4. nombre de triangles, 5. compacité (type 1), 6. aire, 7. nombre de pièces, 8. volume, 9. courbures.
Facteurs d'influence avec le cas de simulation (vecteurs $Inter.x_j^n$)	Distances(*) entre le SE et : 1. les conditions limite (CL), 2. la cible. Bénéfice du moment distance / CL(**).	Aucun

FIGURE B.1 – Variables explicatives déterminantes pour la prédiction des variables Y1 (influence de la simplification d'un sous-ensemble) et de Y2 (coût de la simplification).(*) : variables connues pour un nouveau cas. (**) : variables intermédiaires non connues pour un nouveau cas.

ANNEXE B. VARIABLES EXPLICATIVES SÉLECTIONNÉES

Sélection des variables explicatives	Estimation de Y3 = erreur sur le résultat de l'analyse
Description du processus de préparation (vecteurs $SP1.x_i^m$)	Description des simplifications : toutes variables ^(*) $\{SP1.1_i^m, \dots, SP1.18_i^m\}$ Description du maillage : taille et type d'éléments
Description et comparaison des modèles globaux CAO (vecteurs $M1.x_i^m$)	Caractéristiques du modèle original ^(*) $\{M1.1_i^m, \dots, M1.16_i^m\}$ + Caractéristiques du modèle simplifié ^(**) : 1. compacité (type 2), 2. ratio nombre de triangles / patches, 3. nombre de pièces. + Bénéfices ^(**) sur : 1. ratio nombre de triangles / patches, 2. compacité (type 2), 3. volume, 4. aire, 5. nombre de pièces, 6. nombre de triangles. + Bénéfice du moment distance / CL ^(**) . Bénéfice du moment distance / cible ^(**) .
Description et comparaison des modèles adaptés (vecteurs $M2.x_i^m$)	Caractéristiques du modèle simplifié ^(**) : 1. Nombre de patches, 2. volume (adimensionnel / taille boîte), 3. courbures, 4. compacité (type 1), 5. nombre de triangles. + Bénéfices ^(**) sur : 1. distance de Hausdorff (moyenne quadratique), 2. compacité (type 1), 3. volume, 4. aire 5. nombre de triangles, 6. courbures, nombre de patches.
Description et comparaison des modèles maillés (vecteurs $M3.x_i^m$)	Nombre d'éléments ^(**) Bénéfice sur le nombre d'éléments ^(**)

FIGURE B.2 – Variables explicatives déterminantes pour la prédiction de la variable Y3 (erreur sur le résultat de l'analyse).(*) : variables connues pour un nouveau cas. (**) : variables intermédiaires non connues pour un nouveau cas.

Sélection des variables explicatives	Estimation de Y4 = coût global de la préparation	Estimation de Y5 = coût dfe la simplification
Description du processus de préparation (vecteurs $SP1.x_i^m$)	Description des simplifications : toutes variables ⁽¹⁾ $\{SP1.1_i^m, \dots, SP1.18_i^m\}$ Description du maillage : taille et type d'éléments	
Description et comparaison des modèles globaux CAO (vecteurs $M1.x_i^m$)	Caractéristiques du modèle original ^(*) $\{M1.1_i^m, \dots, M1.16_i^m\}$ + Caractéristiques du modèle simplifié ^(**) : 1. nombre de patches, 2. compacité (type 2), 3. volume (adimensionnel / taille boîte) 4. aire (brut) + Bénéfices ^(**) sur : 1. ratio nombre de triangles / patches, 2. compacité (type 2), 3. volume, nombre de patches.	Caractéristiques du modèle original ^(*) $\{M1.1_i^m, \dots, M1.16_i^m\}$ + Caractéristiques du modèle simplifié ^(**) : nombre de patches, 1. volume (adimensionnel / taille boîte) 2. compacité (type 2), 3. aire (brut) + Bénéfices ^(**) sur : 1. compacité (type 2), 2. ratio nombre de triangles / patches, 3. volume, 4. nombre de pièces, + Bénéfice du moment distance / cible ^(**) .
Description et comparaison des modèles adaptés (vecteurs $M2.x_i^m$)	Caractéristiques du modèle simplifié ^(**) : 1. volume (adimensionnel / taille boîte) 2. aire, + Bénéfices ^(**) sur : 1. volume (adimensionnel / taille boîte) 2. courbures, 3. compacité (type 2)	Caractéristiques du modèle simplifié ^(**) : 1. aire, 2. compacité (type 1), 3. volume (adimensionnel / taille boîte) 4. nombre de patches, 5. nombre de triangles, 6. courbures, + Bénéfices ^(**) sur : 1. Volume, 2. compacité (type 1), 3. aire, 4. courbures 5. nombre de patches, 6. distance de Hausdorff (moyenne), 7. nombre de triangles.
Description et comparaison des modèles maillés (vecteurs $M3.x_i^m$)	Nombre d'éléments ^(**) Bénéfice sur le nombre d'éléments ^(**)	Nombre d'éléments ^(**) Bénéfice sur le nombre d'éléments ^(**)

FIGURE B.3 – Variables explicatives déterminantes pour la prédiction des variables Y4 (coût global de la préparation) et Y5 (coût de la simplification).(*) : variables connues pour un nouveau cas. (**) : variables intermédiaires non connues pour un nouveau cas.

Liste des abréviations et notations

Abréviation de termes pour la modélisation et la simulation

B-rep	Boundary Representation
CAO	Conception Assistée par Ordinateur
CAD	Computer Aided Design
CAE	Computer Aided Engineering
CFD	Computational Fluid Dynamics
CL	Condition Limite
DMU	Digital Mock-Up
EF	Eléments Finis
IGES	Initial Graphics Exchange Specification
STEP	Standard for the Exchange of Product Model Data
STL	STéréo-Lithographie
VRML	Virtual Reality Markup Language

Abréviations et notations pour les techniques d'intelligence artificielles

IA	Intelligence Artificielle
AA	Apprentissage Artificiel
CBR	Case Based Reasoning (raisonnement à base de cas)
RN	Réseau de Neurones
SVM	Support Vector Machine (machine à vecteur de support)
N	nombre d'instances (exemples)
y	vecteur de variables de sortie
x	vecteur de variables d'entrée
h	hypothèse
$J(w)$	fonction coût
ROC	Receiver Operating Characteristic Curve (courbe ROC)
AQE	Average Quadratic Error (erreur quadratique moyenne)

ANNEE 2019-2020
**Abréviations et notations pour la description des modèles
et des processus de préparation**

Abréviations et notations générales

M_i^m	Modèle global m
i	Indice de configuration du modèle global
SE_j^n	Sous-ensemble n simplifié selon la configuration
j	Indice de configuration du sous-ensemble
Nse	Nombre total de sous-ensembles
CS	Coût de la Simplification
CP	Coût de la Préparation
CA	Coût de l'Analyse
ERA	Erreur sur le Résultat de l'Analyse
IS	Influence de la Simplification sur un sous-ensemble
LOS	Level OF simplification (niveau de simplification)

Notations pour la modélisation d'un processus de préparation

P_i^m	processus global de préparation
$SP1_i^m$	processus global de simplification
$SP1.x^m$	opération " x " de simplification globale
$sp1_j^n$	sous-processus de simplification du sous-ensemble n
$op1_x^n$	opération de simplification " x "
$t1_x^n$	outillage de la simplification pour l'opération " x "
$SP2_i^m$	processus d'adaptation
$SP2.x^m$	caractéristique du processus d'adaptation
$T2^m$	outillage d'adaptation
$SP3_i^m$	processus de maillage
$SP3.x^m$	caractéristique du processus de maillage
$T3^m$	outillage de maillage

ANNEXE B : VARIABLES EXPLICATIVES SÉLECTIONNÉES

Notations pour la description des modèles

$CAOM1_i^m$	caractéristiques géométrique du modèle global
CAO_j^n	caractéristiques géométrique d'un sous-ensemble
$Comp_j^n$	facteurs de comparaison entre le modèle original et simplifié
$Inter_j^n$	facteurs d'influence sur les conditions limites ou les objectifs de la simulation
$CompM1_i^m$	facteurs de comparaison entre le modèle CAO original et simplifié
$CAOM2_i^m$	caractéristiques géométrique du modèle adapté
$CompM2_i^m$	facteurs de comparaison entre deux modèles préparés
$Mail_i^m$	caractéristiques du modèle maillé
$CompM3_i^m$	facteurs de comparaison entre deux modèles maillés

.

Liste des figures

1.1	Cycle de vie d'un produit et exploitations multiples de la maquette numérique	8
1.2	Exemples de représentations intermédiaires du produit	9
1.3	Exemple de processus de préparation de modèles CAO pour l'analyse CFD	12
1.4	Activités de préparation des représentations intermédiaires (RI) et de simulation dans la phase de conception. (a) Méthode traditionnelle. (b) Méthode avec simplification du modèle maillé. (c) Méthode avec estimations a priori du modèle préparé.	13
1.5	Composition du modèle CAO. (a) Ensemble global constitué de conditions limites, de cibles et d'obstacles. (b) Exemple de sous-ensemble. (c) Exemples de pièces	14
1.6	Exemples de détails	14
1.7	Types de formats utilisés du modèle original au modèle intermédiaire pour la simulation numérique	15
1.8	Exemples de modèles préparés. Scénario 1 : extraction de surfaces bordant le volume de fluide puis préparation du modèle pour l'analyse CFD. Scénario 2 : extraction du contour du produit puis préparation du modèle pour le calcul de structure.	17
1.9	Multitude de processus possibles pour la préparation d'un modèle de simulation pour un même objectif de simulation. [8]	19
1.10	Problématique générale : quel est le meilleur processus de préparation ?	23
2.1	Defeaturing. Méthodologie et illustration de l'approche proposée par GAO et al. [23]	30
2.2	(a) Méthodologie de defeaturing. (b) Illustration de l'approche proposée par Sun et al. [51]	30
2.3	Exemples de defeaturing. (a) modèle original d'une pièce. (b) modèle simplifié par suppression des petits trous. (c) modèle simplifié par suppression des bossages. (d) modèle simplifié par suppression des petits détails. (e) modèle simplifié par suppression de tous les détails. .	31

2.4	Exemples de simplification par filtrage de pièces. (a) Modèle original d'un sous-ensemble. (b) Modèle d'un sous-ensemble filtré. Illustration d'une perte de connexion entre pièces.	33
2.5	Exemples de simplification par substitution. (a1, a2, a3) Modèles originaux. (b1) Modèle substitué par un parallélépipède. (b2) Modèle substitué par des cylindres. (b3) Modèle substitué par un volume autour de la surface médiane.	34
2.6	Décimation de faces. [21].	36
2.7	Exemples de décimation. (a) modèle original. (b) Modèle décimé par décimation avec conservation des arêtes extérieures. (c) Modèle simplifié avec précision réglée au minimum.	37
2.8	Exemples de simplifications par enveloppes convexes. (a) Modèles originaux. (b) Modèles simplifiés par enveloppe convexe.	38
2.9	Exemples d'agréations. (a) Modèle original. (b) Modèle agrégé après simplification par enveloppe convexe. (c) Modèle agrégé avant simplification par enveloppe convexe.	39
2.10	Distance de Hausdorff. Distances entre un modèle X et un modèle Y. .	41
2.11	Calcul de moments tridimensionnels sur des modèles continus et discrétisés proposés par Iyer et al. [29]	42
3.1	Méthode "Medial Axis Transform" et suppression de détails, par Armstrong et al. [4]	49
3.2	(a) A 2-D heat block assembly. (b) A defeatured 2-D block assembly. Par Gopalakrishnan et al. [25]	51
3.3	Rapport des rayons inscrit et circonscrit. Par Dompierre [17].	52
3.4	Transformation d'un élément régulier en élément réel. Par Dompierre [17].	53
3.5	Définition de l'angle solide θ_i . Par Dompierre [17].	53
3.6	Critères de qualité d'un maillage éléments finis. Par Gilles Foucault [22] .	54
4.1	Principe d'un système à base de cas (CBR)	58
4.2	Apprentissage supervisé. Prédiction d'une variable de sortie pour une nouvelle instance à l'aide d'une hypothèse obtenue par apprentissage. .	60
4.3	Algorithmes de classifications. (a) Arbre de décision. (b) Réseau de neurones. (c) Machine à support de vecteurs.	61
4.4	Apprentissage non supervisé. Regroupement d'instances proches dans un cluster.	61
4.5	Schéma d'un neurone à n entrées x et une sortie y.	66

4.6	Exemple d'un problème de classification à deux classes $y = 1$ ou $y = -1$ avec hyperplan optimal de séparation d'équation $h(x) = wx + w_0$. . .	67
4.7	Exemple de séparation non linéaire. (a) séparation non linéaire. (b) exemple de linéarisation par fonction noyau polynomiale.	68
4.8	exemples de courbes ROC pour l'analyse de "vrais-positifs" / "faux-positifs".	73
5.1	Exemples de cas existants de préparation de modèles pour la simulation.	79
5.2	Structuration de l'approche. (a) Approche type « systèmes à base de cas ». (b) Approche par identification d'un processus optimal.	80
5.3	Approche générale proposée (niveau 1 de la figure 5.2.	81
5.4	Etape 1 de l'approche proposée. Construction de la base de données pour l'apprentissage.	82
5.5	Etape 2 de l'approche proposée. Sélection et configuration de classifieurs par apprentissage.	83
5.6	Etape 3 de l'approche proposée. Utilisation sur un nouveau cas. Prédiction des critères d'évaluation d'un processus de préparation pour la simulation. VE = variables explicatives connues. VI = variables intermédiaires. CS = coût de la simplification. CP = coût de la préparation. CA = coût de l'analyse. IS = influence de la simplification d'un sous-ensemble sur le résultat. ERA = erreur sur le résultat de l'analyse.	84
5.7	Représentation matricielle des données et éléments à identifier pour prédire une variable de sortie par apprentissage.	85
6.1	Définition d'un processus de préparation de modèles pour la simulation	89
6.2	Actigramme A-0. Processus de préparation d'un modèle pour la simulation numérique.	90
6.3	Exemple de modèle original constitué de quatre sous-ensembles non simplifiés.	91
6.4	Actigramme A0. Processus détaillé de préparation d'un modèle pour la simulation numérique.	92
6.5	Actigramme A1. Simplification du modèle pour la simulation numérique.	93
6.6	Actigramme A1.m.n. Simplification d'un sous-ensemble SE^n par l'opération m	94
6.7	Actigramme A2. Adaptation du modèle simplifié au maillage.	95
6.8	Actigramme A3. Maillage.	95

6.9	Niveaux de représentation. (a) Niveau global pour la description du modèle complet et pour l'évaluation du processus de préparation. (b) Niveau local pour la description du sous-ensemble et pour l'évaluation du sous-processus de simplification.	99
6.10	Etape 1 construction de la base de données pour l'apprentissage. 1.1 Représentation formalisée des connaissances.	100
6.11	Contenu de la matrice de données locale au niveau des sous-ensembles.	102
6.12	Contenu de la matrice de données globale au niveau de l'ensemble complet.	103
6.13	Exemples de sous-processus j du sous-ensemble N°2.	109
6.14	Exemples de simplifications du modèle basique.	112
6.15	Contenu de la matrice de données locale au niveau des sous-ensembles.	128
6.16	Contenu de la matrice de données globale au niveau de l'ensemble complet.	129
6.17	Etape 1 construction de la base de données pour l'apprentissage (figure 5.4).1.2 Sélection des variables explicatives et variables à prédire.	130
6.18	Méthode pour l'implémentation et la sélection de variables explicatives.	132
6.19	Exemples de corrélation. (a) Variables non corrélées. (b) Variables faiblement corrélées. (c) Variables fortement corrélées. $\rho(a), \rho(b), \rho(c)$: coefficients de corrélation.	134
6.20	Espace "Erreur sur le résultat de l'analyse / Niveau de simplification du modèle"	136
6.21	Méthode pour la construction d'exemples de simplification de sous-ensembles	140
6.22	Cinq éléments à déterminer pour la configuration d'un modèle d'apprentissage : avancement après la phase de construction de la base de données.	144
7.1	Approche générale pour la construction de modèles d'apprentissage. .	147
7.2	Méthode pour la construction de modèles d'apprentissage.	149
7.3	Exemple de partitionnement des N exemples initiaux en N_a exemples pour l'apprentissage, N_t exemples pour les tests et ré-échantillonnage en $K=5$ groupes des cas pour la validation.	150
7.4	Exemple de valeurs aberrantes (variable aire du modèle global $Aire_i^m$). (a) Variables non corrigées. (b) variables aberrantes supprimées ou corrigées.	152

7.3	Exemple de distribution. Variable aire du modèle CAO. (a) Valeurs initiales. (b) Distribution obtenue par répartitions régulières dans 10 classes. (c) Distribution obtenue par l'algorithme MDLPC [Fayyad 1993]. (c) Distribution obtenue avec la méthode des k-moyennes. (e) Calculs des variances inter / intra et du test de Fisher.	155
7.6	Scénarios envisageables pour la prise en compte des variables inter-médiaires.	160
7.7	Approche générale pour l'utilisation de l'apprentissage sur un nouveau cas.	162
7.8	Stratégies pour la prédiction de la qualité de la simplification d'un sous-ensemble sur un nouveau cas à l'aide de classifieur (Cl) ou de relations connues (Re).	164
7.9	Stratégies pour la prédiction de la qualité d'un processus de préparation sur un nouveau cas.	167
7.10	Méthode de recherche du meilleur processus de préparation.	169
8.1	Désignation des principaux composants et de quelques conditions limites (CL) d'une baie avionique pour l'analyse thermique.	175
8.2	Modèles CAO originaux. Le modèle 3 est le support des exemples de préparation de modèles globaux présentés dans ce chapitre.	176
8.3	Exemples de sous-ensembles originaux extraits du modèle N°3. (*) Support des exemples de simplification de sous-ensemble donnés dans ce chapitre.	177
8.4	Exemples de simplification du sous-ensemble SE_j^{15} pour des configurations j	180
8.5	Exemples de simplifications globales du modèle $M1_i^3$ simplifié selon des configurations i	182
8.6	Exemple de modèle adapté $M2_i^m$ (volume de fluide reconstruit par modeleur 3D pour extraction des données).	186
8.7	Exemple de modèle maillé. Volume de fluide du modèle M_{ref}^2 (configuration de référence).	188
8.8	Exemple de maillage des surfaces du composant cible SE_{ref}^{15}	188
8.9	Utilisations du logiciel Weka pour l'apprentissage.	192
8.10	Synthèse sur la construction d'exemples de processus de préparation de modèles.	193
8.11	Synthèse de l'approche proposée et questions pour la construction d'exemples, la configuration de modèles d'apprentissage et l'évaluation des prédictions.	196

8.12 Répartition des $N_a = 195$ exemples de sous-processus de simplification sur un sous-ensemble dans l'espace LOS (niveau de simplification) / IS (influence de la simplification sur l'analyse).	197
8.13 Répartition des $N_a = 245$ exemples de processus globaux de préparation dans l'espace LOS (niveau de simplification) / ERA (erreur sur le résultat de l'analyse).	197
8.14 Identification du nombre minimum d'exemples par variable décrivant une opération de simplification.	198
8.15 Nombre d'exemples par type d'opération de simplification pour l'ensemble des modèles.	199
8.16 Nombre d'exemples par processus de simplification (séquençages d'opération utilisés) pour l'ensemble des modèles.	199
8.17 Taux d'exemples correctement classés en fonction du nombre de classes pour les cinq variables de sortie.	206
8.18 Exemples de nouveau cas (non utilisés lors de l'apprentissage issus des modèles M^1 et M^2)	209
 B.1 Variables explicatives déterminantes pour la prédiction des variables Y_1 (influence de la simplification d'un sous-ensemble) et de Y_2 (coût de la simplification).(*) : variables connues pour un nouveau cas. (**) : variables intermédiaires non connues pour un nouveau cas.	V
B.2 Variables explicatives déterminantes pour la prédiction de la variable Y_3 (erreur sur le résultat de l'analyse).(*) : variables connues pour un nouveau cas. (**) : variables intermédiaires non connues pour un nouveau cas.	VI
B.3 Variables explicatives déterminantes pour la prédiction des variables Y_4 (coût global de la préparation) et Y_5 (coût de la simplification).(*) : variables connues pour un nouveau cas. (**) : variables intermédiaires non connues pour un nouveau cas.	VII

Liste des tableaux

2.1	Outils et techniques de defeaturing	31
2.2	Outils et techniques de filtrage	32
2.3	Outils et techniques de substitution	34
2.4	Simplification par voxelisation	35
2.5	Simplification par décimation	36
2.6	Outils et techniques d'enveloppe convexe	38
2.7	Outils et techniques d'agrégation	39
3.1	Principales techniques d'évaluation des écarts entre les grandeurs physiques réelles et le résultat de l'analyse dus à des modifications sur la géométrie.	47
6.1	Matrice d'antériorité des opérations de simplification. Avant : l'opération (ligne) doit être réalisée avant l'opération (colonne). Après : l'opération (ligne) doit être réalisée après l'opération (colonne). Av/Ap : l'opération (ligne) peut être réalisée indifféremment avant ou après l'opération (colonne). X : les deux opérations ne sont pas utilisées sur un même composant.	96
6.2	Description des sous-processus de simplification d'un sous-ensemble. Composantes du vecteur $sp1_j^n$ (opérations).	108
6.3	Description des J = 8 sous-processus de simplification du sous-ensemble (SE) N°2 (figure 6.13). Seq = référence du séquençage (annexe A)(1) Tous les trous dont la longueur moyenne est inférieure à 0,2 fois la longueur moyenne du modèle global sont supprimés (quelle que soit leur distance par rapport aux conditions limites).	108
6.4	Composantes décrivant l'opération de defeaturing	110
6.5	Paramètres globaux de réglage des opérations de simplification	113
6.6	Exemples de modélisation de processus global de préparation sur le modèle M^b (figure 6.14). Var. = niveaux de simplifications (définis dans le tableau 6.5) pour chaque opération (Op.) et chaque type de composant (Cp.).	113

6.7	Composantes du vecteur CAO_j^n . Facteurs descriptifs de la géométrie d'un sous-ensemble.	118
6.8	Description du modèle CAO du sous-ensemble N°2 (figure 6.13). Exemples de variables explicatives limitées aux variables déterminantes.	119
6.9	Composantes du vecteur $comp_j^n$. Facteurs de comparaison entre le sous-ensemble simplifié et original.	119
6.10	Exemples de comparaison $comp.x_j^n$ de sous-ensembles originaux et simplifiés (figure 6.13) et de facteurs d'influence $inter.x_j^n$ sur le cas de simulation. Variables explicatives limitées aux variables déterminantes.	120
6.11	Composantes du vecteur $Interactions_j^n$. Facteurs d'interactions entre le sous-ensemble et le cas de simulation.	121
6.12	Composantes des vecteurs $CAOM1_i^m$ et $CAOM2_i^m$. Facteurs descriptifs de la géométrie du modèle global et adapté. $\emptyset$: variable non utilisée (la notion de pièces et de détails ne s'applique pas au modèle adapté).	123
6.13	Exemples de description CAOM1 et de comparaison $CompM1$ d'ensembles CAO représentés sur la figure 6.14. Variables explicatives limitées aux variables déterminantes.	123
6.14	Exemples de description de modèles adaptés CAOM2, de description de maillages $MailM3$, de comparaison de modèles adaptés $CompM2$, de comparaison de maillages $CompM3$. Variables explicatives limitées aux variables les plus déterminantes	124
6.15	Coefficients de corrélation entre les distances de Hausdorff (maximale, moyenne ou quadratique) et les variables à prédire.	133
7.1	Exemples de valeurs minimales et maximales des variables de deux modèles avant normalisation.	153
7.2	Exemples de normalisation globales sur l'ensemble de la base. Fonction de normalisation : $valeur_{initiale} \rightarrow valeur_{normalisée}$	153
7.3	Exemples de normalisations par modèles. Fonction de normalisation : $valeur_{initiale} \rightarrow valeur_{normalisée}$	153
7.4	Répartition des valeurs des variables à prédire.	156
7.5	Paramètres à déterminer pour la configuration des classifieurs	158
7.6	Exemples de prédictions acceptables et inacceptables pour la variable y_3 (erreur sur le résultat de l'analyse).	159
8.1	Séquençages d'opérations utilisés lors des expérimentations. Seq = indice de séquençage.	178
8.2	Composantes opx_j^n de l'ensemble de vecteurs $sp1_j^n$ décrivant les opérations de simplification définies dans la section 6.2.4.1.	179

LISTE DES TABLEAUX	
8.3 Description des NJ = 12 sous-processus de simplification j du sous-ensemble SE_j^{15} (figure 8.4). Seq = référence du séquençage	179
8.4 Paramètres globaux de réglage des opérations de simplification définis dans la section 6.2.4.2. CL = Condition Limite.	181
8.5 Exemples de modélisation de processus global de préparation sur le modèle $M1_i^3$ simplifié selon des configurations i (figure 8.5). Var. = composantes de l'ensemble de vecteurs $SP1_i^m$	181
8.6 Exemples de description du modèle CAO du sous-ensemble SE_j^{15} (figure 8.4). Les valeurs sont brutes telles qu'elles sont extraites des modèles CAO avant traitements (normalisation et discrétisation). Les variables explicatives limitées aux variables déterminantes. SU = sans unité.	184
8.7 Exemples de facteurs de comparaison $comp.x_j^n$ de sous-ensembles originaux et simplifiés SE_j^{15} et de facteurs d'influence $inter.x_j^n$ sur le cas de simulation. Variables explicatives limitées aux variables déterminantes. (1) Bénéfices entre les valeurs du modèle simplifié et du modèle original. (2) Moyenne quadratique de la distance de Hausdorff. (3) Les valeurs sont brutes telles qu'elles sont extraites des modèles CAO avant traitements (normalisation et discrétisation). SU = sans unité	184
8.8 Exemples de facteurs descriptifs d'ensembles CAO globaux $CAOM1.x_i^m$ (variables brutes sans unités extraites des modèles CAO) et de facteurs de comparaison $Comp1.x_i^m$ entre le modèle original et simplifié (bénéfices entre les valeurs du modèle simplifié et celles du modèle original). Les variables explicatives sont limitées aux variables déterminantes. Les valeurs des variables sans brutes avant traitements. SU = sans unité	185
8.9 Exemples de facteurs descriptifs de modèles adaptés $CAOM2.x_i^m$ ((1) variables brutes extraites des modèles CAO adaptés $M2_i^3$) et de facteurs de comparaison $CompM2.x_i^m$ entre le modèle de référence et simplifié ((2) : bénéfices entre $CAOM2.x_i^m$ et $CAOM2.x_{ref}^m$, (3) distance de Hausdorff). Les variables explicatives sont limitées aux variables déterminantes. Les valeurs sont brutes telles qu'elles sont extraites du modèle adapté avant traitements. SU = sans unité.	187
8.10 Description des principaux paramètres de maillage.	187

8.12 Exemples de résultats d'analyse et de principales valeurs à prédire :
Temp. = températures maximales relevées sur la cible, ERA = y_3 (erreur globale sur le résultat de l'analyse), CP = y_4 (coût total de la préparation) et CA = y_5 (coût de l'analyse) sur le modèle global M_i^3 pour des configurations de simplification i représentées sur la figure 190

8.13 Types de classifieurs utilisés lors des expérimentations. 191

8.14 Gains sur l'erreur quadratique moyenne des prédictions (AQE) et sur la durée d'apprentissage dues à l'analyse et à la suppression des variables corrélées pour la prédiction de l'influence de la simplification d'un sous-ensemble sur l'analyse (IS) et du coût de la simplification (CS). 201

8.15 Erreurs de prédictions moyennes par principale variable à prédire pour les 3 scénarios. "Toutes" : toutes les variables sont utilisées. "Déterminantes" : seules les variables les plus déterminantes sont sélectionnées. 202

8.16 Gains moyens sur l'erreur des prédictions AQE et sur la durée de l'apprentissage pour la prédiction de y_1 (influence de la simplification d'un sous-ensemble sur l'analyse IS) , y_2 (coût de la simplification CS), y_3 (erreur sur le résultat de l'analyse ERA) y_4 (coût de la préparation CP) et y_5 (coût de l'analyse CA). 204

8.17 Résultats des erreurs de prédiction (AQE) et taux d'instances correctement classées (%Cl.) pour les variables à prédire (répartition régulière des valeurs). 204

8.18 Exemples de résultats de d'erreurs des prédictions AQE et de taux de classifications correctes (Cl.%) pour les cinq principales variables à prédire avec des distributions régulières ou par classification des valeurs. 206

8.19 Synthèse sur la configuration des modèles d'apprentissage pour la prédiction de y_1 (influence de la simplification d'un sous-ensemble sur l'analyse IS) , y_2 (coût de la simplification CS), y_3 (erreur sur le résultat de l'analyse ERA) y_4 (coût de la préparation CP) et y_5 (coût de l'analyse CA). (1) : avec une distribution régulière des valeurs par classe. (2) : avec une distribution par classification des variables. 207

8.20 Indicateurs de confiance des modèles d'apprentissage pour les variables y_1 et y_2 et les stratégies proposées. 208

8.21 Indicateurs de confiance des modèles d'apprentissage pour les variables y3, Y4 et Y5 et les stratégies proposées.	211
8.22 Exemples de résultats de prédictions. c : classe. vb : valeur brute. (!) : erreurs inacceptables (les valeurs prédites sont plus optimistes que les valeurs réelles).	211
8.23 Exemples de décisions obtenues à partir des résultats sur les valeurs réelles et prédites par apprentissage donnés dans le tableau 8.22. . . .	212
A.1 Séquençages réalistes d'opérations de simplification d'un sous-ensemble. III	

Traitement de maquettes numériques pour la préparation de modèles de simulation en conception de produits à l'aide de techniques d'intelligence artificielle

Résumé : Maitriser le triptyque coût-qualité-délai lors des phases du Processus de Développement d'un Produit (PDP) est un enjeu majeur pour l'industrie. Le développement de nouvelles méthodes pour adapter une représentation du produit à une activité du PDP est une piste d'amélioration des plus prometteuses. Dans le domaine du transfert de modèles de Conception Assistée par Ordinateur (CAO) vers des activités de simulations numériques, des méthodes et des outils de préparation existent. Cependant, ces processus de préparation sont des tâches complexes qui reposent souvent sur les connaissances des experts. Pouvoir estimer a priori l'impact de la préparation d'une maquette numérique sur le résultat de la simulation permettrait d'identifier dès le début le meilleur processus et assurerait une meilleure maîtrise des processus et des coûts de préparation. Cette thèse a pour objectif de relever ce défi en utilisant des techniques d'intelligence artificielle. L'idée principale est d'utiliser des exemples de préparation de maquettes numériques comme entrées d'algorithmes d'apprentissage pour évaluer la performance d'un processus. Une méthode a été développée pour construire une base d'exemples représentatifs, identifier les variables d'entrée déterminantes et configurer des modèles d'apprentissage. Le premier défi de l'approche proposée est d'extraire les données des modèles 3D puis d'identifier les variables explicatives les plus déterminantes. Un autre défi est de configurer des modèles d'apprentissage capables d'évaluer avec précision la qualité d'un processus malgré un nombre limité d'exemples et de données disponibles. Au final, l'estimateur de la performance d'un processus proposé permet d'aider les analystes dans le choix des opérations de préparation de modèles CAO.

Mots clés : Évaluation a priori de processus, simplification de modèles, transfert CAD-CAE, apprentissage artificiel, formalisation des connaissances.

A priori evaluation of simulation models preparation processes using artificial intelligence techniques

Abstract : Controlling the triptych costs, quality and time during the different phases of the Product Development Process (PDP) is an everlasting challenge for the industry. The development of new methods to adapt to the models all along the PDP is one of the most challenging improvement area. This is particularly true for the adaptation of CAD (Computer-Aided Design) models to CAE (Computer-Aided Engineering) applications. Today, even if methods exist, such a preparation phase still requires a deep knowledge. Thus, being able to estimate a priori the impact of DMU preparation process on the simulation results would help identifying the best process right from the beginning, and this will ensure a better control of processes and preparation costs. This thesis addresses such a difficult problem by using Artificial Intelligence (AI) techniques to learn and accurately predict behaviors from carefully selected examples. The main idea is to estimate by learning methods the quality of a preparation process without having to perform it. To reach this objective, a method to build a representative database of examples has been developed, the right input and output variables have been identified, then the learning model and its associated control parameters have been tuned. The first challenge of the proposed approach is to extract and select most relevant input variables from the 3D models. Another challenge is to configure learning models able to assess with good accuracy the quality of a process, despite a limited number of examples of preparation processes and data available (the only data known to a new case are the data that characterize the original CAD models and simulation case). In the end, the estimator of the process' performance will help analysts in the selection of CAD model preparation operations.

keywords : A priori process evaluation, large CAD models simplification, link CAD-CAE, artificial intelligence techniques, knowledge formalization.