



Conception automatisée de scénarios d'apprentissage réalistes et variés, pour l'acquisition et la consolidation d'expertise en anesthésie, assistées par le numérique

Hugo Boisaubert

► To cite this version:

Hugo Boisaubert. Conception automatisée de scénarios d'apprentissage réalistes et variés, pour l'acquisition et la consolidation d'expertise en anesthésie, assistées par le numérique. Informatique. Nantes Université, 2022. Français. NNT : 2022NANU4058 . tel-04047040

HAL Id: tel-04047040

<https://theses.hal.science/tel-04047040>

Submitted on 27 Mar 2023

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

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

THÈSE DE DOCTORAT DE

NANTES UNIVERSITÉ

ÉCOLE DOCTORALE N° 601
*Mathématiques et Sciences et Technologies
de l'Information et de la Communication*
Spécialité : *Informatique*

Par

Hugo BOISAUBERT

Conception automatisée de scénarios d'apprentissage réalistes et variés, pour l'acquisition et la consolidation d'expertise en anesthésie, assistées par le numérique.

Thèse présentée et soutenue à Nantes, le 14/12/2022

Unité de recherche : Laboratoire des Sciences du Numérique de Nantes (LS2N) - UMR 6004

Rapporteurs avant soutenance :

Maria RIFQI Professeur des universités à Panthéon-Assas Université
Cyrille BERTELLE Professeur des universités à l'Université Le Havre Normandie

Composition du Jury :

Président :	Cyrille BERTELLE	Professeur des universités à l'Université Le Havre Normandie
Examineurs :	Maria RIFQI	Professeur des universités à Panthéon-Assas Université
	Mounira HARZALLAH	Maître de conférences (HDR) à Nantes Université
	Raphael CINOTTI	Professeur des Universités - Praticien Hospitalier au CHU de Nantes
	Frédéric BLANCHARD	Maître de conférences à l'Université de Reims Champagne-Ardenne
Dir. de thèse :	Christine SINOQUET	Maître de conférences (HDR) à Nantes Université

Invitée :

Corinne LEJUS-BOURDEAU Professeur des Universités - Praticien Hospitalier au CHU de Nantes

REMERCIEMENTS

Mes premiers et plus profonds remerciements s'adressent au Dr Christine SINOQUET, pour avoir parié sur le profil atypique que je suis et avoir cru en moi pour cette thèse.

Je tiens à remercier le Pr Corinne LEJUS-BOURDEAU qui nous à ouvert les portes du monde de l'anesthésie et de la simulation en santé, et dont la patience et la pédagogie m'ont permis de développer ma compréhension de ces domaines.

Ce travail n'aurait pu être mené à bien sans le financement conjoint de la Région Pays de la Loire et de Nantes Université au travers du RFI OIC.

Ces travaux s'appuient largement sur la contribution de personnes qui sont désignées par "experts" dans cette thèse. Je remercie ainsi les chirurgiens, étudiantes et étudiants infirmiers anesthésistes avec qui nous avons collaboré.

Merci au Pr Pierre-Antoine GOURRAUD et à la Clinique des données du CHU de Nantes pour avoir rendu possible la consultation des données anesthésiques.

Merci aux collègues doctorants du laboratoire pour former une précieuse communauté de partage et d'écoute.

Merci aussi à tous les collègues du LS2N et du département informatique qui ont contribué par leurs diverses fonctions à faciliter la réalisation de cette thèse, tout particulièrement : Christine BRUNET et Chaléry VILAR.

Mes remerciements vont aussi à tous mes proches pour leur présence. Je pense en premier à ma très grande famille, que je n'ai vu que trop peu souvent ces dernières années. Merci à tous ceux dont j'ai le plaisir de partager le temps : Adrien, Sophie, Cyprien, Oscar, Néfis, Enora, Anna et à la famille PODEVIN. Merci à tous pour votre présence autour de moi.

Merci à toutes les associations où je suis passé, passe et passerais. Merci aux membres, anciens comme actuels, de SciNaPSE et Lab Sciren. Vous êtes ma joie et ma fierté. Vous me rappelez chaque jour que je ne regrette pas mes choix et mes engagements.

Merci aux membres de Planètes Sciences, pour vos encouragements et votre patience, et à tous les bénévoles de la Coupe de France de robotique pour réussir à organiser cette folie chaque année depuis 30 ans. Vous avez été un espace de respiration salubre pendant cette thèse.

Remerciements particuliers à Frédéric COURANT et Jamy GOURMAUD qui m'ont transmis, comme à de nombreux autres, la curiosité pour la nature des choses, la passion des sciences et l'envie de découverte.

Mes ultimes remerciements vont à Jacques OFFENBACH dont *La grande-Duchesse de Gerolstein*, a accompagné les derniers moments de la rédaction de cette thèse et tempéré les ultimes frustrations dues à l'exploitation difficile des données anesthésiques.

"Quand on n'a pas ce que l'on aime, il faut aimer ce que l'on a."
(*La Grande-Duchesse*, Acte 4, Scène 3)

TABLE DES MATIÈRES

Introduction	7
Contexte	7
Objectifs des travaux de thèse	8
Enjeux	9
Plan du manuscrit	11
1 L'anesthésie et la formation par la simulation en santé	13
1.1 L'anesthésie et ses grands principes	14
1.1.1 Le contexte d'une anesthésie	14
1.1.1.1 Préparation d'une anesthésie programmée	15
1.1.1.2 Facteurs de risques anesthésiques et indicateurs	16
1.1.2 Principe d'une anesthésie générale	18
1.1.2.1 Éléments de compréhension sur le système nerveux humain	18
1.1.2.2 Les anesthésies générales	19
1.1.3 La structure et le déroulement type d'une anesthésie générale . . .	20
1.1.3.1 L'entrée du patient	21
1.1.3.2 L'induction anesthésique	22
1.1.3.3 La procédure de l'intervention	23
1.1.3.4 La sortie du patient	23
1.1.4 Les principaux gestes médicaux et médicaments utilisés	24
1.2 La formation par la simulation en santé	28
1.2.1 Le développement de la simulation en santé en France	28
1.2.1.1 Historique et précurseurs	29
1.2.1.2 État des lieux de la simulation en santé en 2012	30
1.2.1.3 Guide de l'HAS et développements récents	31
1.2.2 La simulation au SiMU de Nantes	31
1.2.2.1 La simulation haute fidélité interprofessionnelle	32
1.2.3 Le cadre et les pratiques pédagogiques des sessions de simulation au SiMU de Nantes	33

1.2.3.1	Contexte pédagogique	33
1.2.3.2	Le déroulement des sessions de simulation	34
1.2.3.3	Les scénarios	36
2	Profils anesthésiques et données de santé	39
2.1	Les données des profils anesthésiques	40
2.1.1	Un outil numérique central : Pégase	40
2.1.2	Les dossiers médicaux des patients	41
2.1.3	Profils anesthésiques	41
2.1.3.1	Caractéristiques d’une intervention	42
2.1.3.2	Événements anesthésiques	42
2.1.3.3	Séries temporelles	43
2.2	La transformation et la sélection des données	45
2.2.1	Les transformations structurelles	46
2.2.2	Sélection des données anesthésiques	46
2.2.3	Les extractions de cohortes de patients	47
2.3	Le cadre légal des données de Santé et leur utilisabilité	49
2.3.1	Type de données des profils anesthésiques	49
2.3.2	Contraintes appliquées aux données de santé	50
3	Générateur de profils anesthésiques réalistes	55
3.1	Formalisation informatique d’une anesthésie	56
3.1.1	Formalisation médicale classique	56
3.1.2	Formalisation informatique par une grammaire formelle	57
3.2	Création d’une typologie d’événements basée sur l’expertise des soignants	60
3.2.1	Récolte des éléments d’expertise	60
3.2.2	Classification des événements	62
3.3	Génération de profils anesthésiques réalistes	67
3.3.1	Approche en deux étapes	67
3.3.2	Génération d’une séquence d’événements	70
3.3.3	Génération des séries temporelles	71
3.3.4	Validation des profils anesthésiques réalistes	72
3.3.4.1	Analyse de similarité	73
3.3.4.2	Analyse statistique	78

4	Similarité entre séries temporelles multivariées	81
4.1	Motivations pour une analyse comparative	82
4.2	Repères sur les mesures de dissimilarité entre séries temporelles	84
4.2.1	Approches Lock-step	84
4.2.2	Approches élastiques	85
4.2.3	Approches par transformées	88
4.2.4	Approches basées sur des <i>features</i>	89
4.3	Comparaison de neuf mesures de dissimilarité	90
4.3.1	Protocole expérimental	92
4.3.2	Résultats et principales conclusions	95
5	Simulation d'un patient virtuel basée sur des données	99
5.1	État de l'art	100
5.1.1	Simulation jointe de série temporelles et de séquences d'événements d'un patient virtuel	100
5.1.2	Raisonnement à base de cas	101
5.1.3	Score de similarité multivarié	102
5.1.4	Prédiction	104
5.2	Simulation par une approche par raisonnement à base de cas	107
5.2.1	Approche par raisonnement à base de cas	107
5.2.2	Définition d'un cas	107
5.2.3	Algorithmes	108
5.3	Recherche de patients similaires	111
5.3.1	Patient virtuel et cohorte de patients réels	111
5.3.2	Contexte et motif multivarié contextualisé	112
5.3.3	Pré-sélection des patients réels	114
5.3.4	Score de similarité multivarié	115
5.3.4.1	Similarité entre les séquences d'événements	116
5.3.4.2	Similarité entre les séries temporelles multivariées	117
5.4	Simulation du patient virtuel	119
5.4.1	Stratégies de prédiction	119
5.4.2	Évaluation de la qualité de la simulation	120
5.4.2.1	Protocole	120
5.4.2.2	Données expérimentales	121

5.4.2.3	Implémentation et paramétrage	121
5.4.2.4	Résultats	122
6	Représentations synthétiques	127
6.1	Représentation d'un ensemble de séquences d'événements par un graphe orienté	128
6.2	Construction de cohortes de profils anesthésiques	131
6.2.1	Partitionnement d'une cohorte de profils anesthésiques	133
6.2.1.1	Repères sur le clustering par cartes auto-organisées	133
6.2.1.2	Partitionnement d'une cohorte de profils anesthésiques par cartes auto-organisées	136
6.2.2	Critères estimatifs de la complexité d'un graphe	140
6.2.2.1	Critères sur les séquences	140
6.2.2.2	Critères sur les graphes	143
6.2.2.3	Expérimentation	146
6.2.3	Association de clusters de profils anesthésiques pour la constitution de sous-cohortes idéales	149
6.2.4	Application aux profils anesthésiques réalistes	153
6.3	Construction d'une représentation synthétique	155
6.3.1	Approche	155
6.3.2	Construction du graphe et association partielle des séquences similaires	159
6.3.2.1	Grammaire squelette	159
6.3.2.2	Alignement sur la grammaire	160
6.3.2.3	Construction du graphe	162
6.3.3	Traitement des séries temporelles, création de séries consensus et enrichissement des nœuds	165
6.4	Opportunités pour l'anonymisation de données	169
6.4.1	Repères sur l'anonymisation de données	169
6.4.2	Perspective pour des représentations synthétiques anonymisées de profils anesthésiques	171
7	Génération et simulation de scénarios d'apprentissage contextualisés	175
7.1	Les scénarios d'apprentissage contextualisés	176
7.1.1	Conditionnement du contexte médical	177

7.1.2	Altération du déroulement de la simulation dans un but pédagogique	177
7.1.2.1	Types d'altérations	178
7.1.2.2	Contrainte de crédibilité	178
7.2	Génération et simulation de scénarios à partir d'une représentation synthétique	179
7.2.1	Limites des modalités de simulation	179
7.2.2	Simulation d'un patient virtuel à partir d'une représentation synthétique	180
7.2.2.1	Approche à base de cas appliquée aux représentations synthétiques	180
7.2.2.2	Qualité de la simulation et performances	181
7.2.3	Altération d'une représentation synthétique	184
7.2.3.1	Modification des effets d'un événement	184
7.2.3.2	Inclusion d'événements et de déclencheurs	185
7.2.3.3	Modification des probabilités d'instanciation	186
7.3	Simulateur réactif	187
7.3.1	Contraintes	187
7.3.2	Construction du simulateur	189
7.3.2.1	Aspects fonctionnels	189
7.3.2.2	Interface	190
Conclusion		193
	Éléments clefs	194
	Perspectives	196
Bibliographie		199
Annexes		200
A Profils anesthésiques		200
A.1	Exemple de dossier médical	200
A.2	Exemple d'une séquence d'événements	203
A.3	Paramètres physiologiques enregistrés	204
A.4	Exemple d'une série temporelle multivariée	205
A.5	Caractéristiques d'une intervention chirurgicale et d'une anesthésie	207

B	Grammaire anesthésique et typologie des événements	208
B.1	Grammaire anesthésique - Niveau 1	208
B.2	Grammaire anesthésique - Niveau 2	208
B.3	Grammaire - Hernie inguinale	209
B.4	Grammaire - Prothèse totale de genou	210
B.5	Typologie des événements anesthésiques	211
C	Étude comparative de neuf dissimilarités entre séries temporelles univariées	213
C.1	Description des jeux de données	213
C.1.1	Jeux de profils anesthésiques simulés	213
C.1.2	Jeux de la <i>UCR archive</i>	214
C.2	Résultats complémentaires	215
C.2.1	Jeux de profils anesthésiques	215
C.2.2	Jeu <i>inlineskate</i> de la <i>UCR archive</i>	217
D	Simulation du patient virtuel	218
D.1	Qualité de la simulation	218
E	Partitionnement d'une cohorte de profils anesthésiques par cartes auto-organisées	221
E.1	Jeux expérimentaux	221
E.2	Estimation du nombre de cycles d'apprentissage	225
F	Critères estimatifs de la complexité d'un graphe	227
F.1	Corrélations entre critères sur les séquences et critères sur les graphes . . .	227
F.2	Critères calculés sur les clusters constitués après regroupement ascendant hiérarchique pour les jeux expérimentaux	234
F.3	Critères calculés sur les clusters constitués après regroupement ascendant hiérarchique pour un ensemble de profils anesthésiques	241

INTRODUCTION

Contexte

En 2012, la Haute Autorité de Santé (HAS)¹ a publié un nouveau guide de bonnes pratiques sur l’usage de la simulation en santé [1]. Dans le cadre de ce document, elle indique un nouveau principe qui doit se généraliser dans les centres de formation en santé : « *Jamais la première fois sur le patient* ».

Ce rapport a encouragé le développement de la formation par la simulation qui était émergente dans les centres de formation en santé français. C’est dans ce mouvement qu’a été créé en 2011 le *Laboratoire Expérimental de Simulation de Médecine Intensive de Nantes Université* (LE SiMU), dont les installations sont utilisées dans le cadre de différentes formations médicales, notamment pour la formation des soignants en anesthésie (internes, infirmiers).

En 2010, presque 12 millions d’interventions chirurgicales en France impliquaient une anesthésie générale [2], dont plus de 52 000 au CHU de Nantes [3].

L’anesthésie est une pratique médicale qui permet la suppression de la douleur pour permettre la réalisation d’un acte chirurgical, obstétrical ou médical. Depuis l’antiquité, différentes drogues ont été utilisées pour permettre l’analgésie, c’est-à-dire la suppression des sensations douloureuses. Le développement de cette pratique médicale commence à se formaliser à partir de la fin du XIX^e siècle avec l’utilisation de nouveaux produits comme l’éther ou le protoxyde d’azote². L’anesthésie moderne a depuis été rendue possible avec le développement de nouvelles drogues et procédures qui permettent des soins toujours plus sûrs.

Comme de nombreuses pratiques médicales, l’anesthésie n’est pas sans risque. Pour des raisons médico-légales, toute procédure anesthésique réalisée est documentée, *via* l’utilisation de *feuilles d’anesthésie*, qui gardent une trace des différentes actions anesthésiques,

1. L’HAS est une autorité publique française indépendante dont les missions sont l’évaluation des soins, la recommandation de bonnes pratiques et la certification des établissements de santé.

2. Plus connu sous le terme de « gaz hilarant »

sous la forme d'une séquence d'actions, et l'évolution des paramètres physiologiques du patient lors de la procédure, sous la forme de séries temporelles multivariées. Les centres de soins ont numérisé progressivement ces feuilles d'anesthésie au début des années 2000. Pour le Centre Hospitalier Universitaire (CHU) de Nantes cette numérisation est effective depuis 2004, *via* un outil dédié, ce qui lui a permis d'accumuler des données spécifiques à ces procédures.

Objectifs des travaux de thèse

L'anesthésie est aujourd'hui une pratique médicale courante dont la formation est assistée par la simulation. Au SiMU de Nantes, cette formation en anesthésie par la simulation est principalement basée sur l'expertise des anesthésistes pour la création de scénarios de formation ainsi que pour le pilotage des installations de simulation. Toutefois, cette modalité de simulation induit plusieurs problèmes :

- Les scénarios de simulation étant conditionnés par l'expérience des anesthésistes, ils sont peu diversifiés, et présentent une faible offre de variation dans leurs contenus.
- Le pilotage des installations de simulations est réalisé manuellement par les formateurs. La réponse du patient simulé à une action médicale ou chirurgicale gagnerait probablement à s'appuyer sur les variations observées sur des patients réels soumis à cette même action.

L'exploitation des données réelles, enregistrées dans le cadre de l'archivage médico-légal des procédures anesthésiques, pourrait permettre de répondre à ces problèmes et venir compléter utilement l'expertise des anesthésistes lors de la formation par la simulation.

Les travaux dans le cadre de cette thèse ont principalement pour but de proposer des méthodes de création de scénarios afin de consolider l'expertise en anesthésie grâce un simulateur, en s'appuyant sur les données réelles.

Dans cette nouvelle modalité de simulation, la création de scénarios implique une formalisation informatique d'une anesthésie, ce qui n'a jamais été réalisé pour le moment, et représente ainsi une étape préalable. De plus, à partir de cette formalisation, il faut pouvoir construire un scénario à partir d'un ensemble de données anesthésiques. Enfin, la création de scénarios implique la construction d'un simulateur réactif pour pouvoir les jouer et permettre l'interaction avec l'apprenant.

L'utilisation de données réelles dans ce projet implique d'identifier les différents éléments constitutifs d'une procédure anesthésique au sein de ces données. Cela implique aussi de reproduire les différentes séries temporelles, qui représentent les variations des paramètres physiologiques du patient, de manière réaliste, tout en tenant compte des différentes dépendances entre la procédure chirurgicale, la procédure anesthésique et l'état de santé du patient, ainsi que ces antécédents, qui nous sont donnés par son dossier médical.

La consolidation de l'expertise anesthésique par la simulation ne peut se faire qu'en permettant l'évolution d'un patient virtuel dans le cadre d'un contexte médical et pédagogique maîtrisé et spécifique.

Les scénarios de simulation doivent ainsi prendre en compte cette contrainte médicale et pédagogique et être en capacité de traiter les éventuelles contradictions que cela peut impliquer. Pour conserver l'intérêt pédagogique de la formation par la simulation, un scénario peut ainsi altérer le réalisme médical d'une situation. Un scénario ne prévoit jamais le décès d'un patient virtuel, bien que sa situation soit médicalement critique, mais préférera plutôt l'intervention d'un formateur ou d'un autre soignant.

Il est de plus nécessaire de pouvoir mesurer les éléments de progression d'un apprenant dans le cadre de la simulation. Il faut pour cela lui permettre une liberté de choix lors de ses décisions qui soit similaire à celle d'une procédure se déroulant en situation réelle.

Enjeux

Ces différents objectifs se confrontent à différents enjeux scientifiques, techniques et pratiques.

Les données anesthésiques enregistrées par les centres de soins sont des données issues généralement d'outils en exploitation conçus pour le soin et ne sont donc pas parfaitement adaptées à une exploitation scientifique. Cela induit un enjeu de transformation de format pour passer de données de production à des données scientifiques. A cet enjeu s'ajoute celui de la qualité des données, celles-ci ayant été récoltées avec une autre finalité que celle de nos travaux. L'exploitation de ces données de santé se confronte aussi à un cadre légal contraignant qui rend leur exploitation complexe. Cette complexité d'exploitation provient principalement des modalités d'accès aux données et des moyens techniques pour

leur exploitation, mais aussi de la difficulté d’anonymiser des données temporelles présentant des dépendances entre elles. Ces contraintes ont conduit, dans le cadre de ces travaux, à la création d’un générateur de données réalistes, en ayant ainsi pour enjeu de réaliser une modélisation de l’expertise des soignants et d’assurer le réalisme des données générées.

La création de scénarios à partir de ces données réelles, composées de séquences d’actions et de séries temporelles multivariées, nous impose de mobiliser des techniques spécifiques de minage afin d’identifier des motifs au sein des séquences d’actions mais aussi au sein des séries temporelles. Les données anesthésiques sont multivariées, et ces variables sont parfois de types hétérogènes, ce qui nous conduit à proposer de nouvelles méthodes pour ce type de données. Le souhait d’utiliser les données réelles comme base de la simulation a conduit à la conception et au développement d’une nouvelle méthode, inspirée du raisonnement à base de cas, mais prenant en compte, pour la première fois dans la littérature, des données temporelles multivariées, de divers types, et interdépendantes. Notre contribution est axée sur un objectif de formation en anesthésie par la simulation assistée par le numérique, mais son intérêt dépasse le cadre de la simulation en milieu médical.

Sur des aspects plus techniques, l’utilisation de données réelles en grande dimension pour la simulation pose des enjeux de structure de données et de gestion de la taille de ces données qui rendent pertinentes la construction de représentations synthétiques pour permettre le bon fonctionnement des algorithmes et méthodes envisagés.

L’utilisation de scénarios basés sur des données réelles dans le cadre d’un simulateur pose la question du réalisme de la simulation aux yeux des apprenants et de la réactivité du simulateur. En effet, celui-ci doit être réactif aux actions des apprenants, et pour cela assurer la recherche de motifs temporels complexes dans des données réelles de grande taille, pour ensuite simuler le patient virtuel de manière réaliste en prenant en compte les contraintes d’un cadre pédagogique précis, comme la maîtrise d’une technique spécifique, par exemple l’intubation, ou les spécificités d’un type de patient, par exemple le patient diabétique.

Les travaux menés dans le cadre de cette thèse ont permis d’explorer et d’apporter des solutions à ces différents enjeux.

Plan du manuscrit

Ce manuscrit s'articule autour de différents chapitres, décrits ci-après, qui présentent des réalisations en réponse aux objectifs, et des propositions de solutions aux enjeux présentés plus haut.

Le chapitre 1 a pour but de décrire le contexte médical dans lequel ces travaux se placent. Il introduit les grands principes d'une procédure anesthésique et sa formalisation médicale. Il présente de plus la formation par la simulation et précise les modalités de simulation utilisées dans la formation des soignants en anesthésie au SiMU de Nantes.

Le second chapitre s'intéresse plus particulièrement aux données qui sont récoltées lors d'une procédure anesthésique et fournit la définition des profils anesthésiques. Ce chapitre présente de plus les importants traitements réalisés sur les données issues du soin pour les convertir en données exploitables. Il propose aussi un focus plus particulier sur le cadre légal contraignant des données de santé et sur la complexité de leur exploitation.

Le chapitre 3 présente la formalisation informatique conçue dans le cadre de ces travaux ainsi qu'une typologie des événements anesthésiques et de leurs impacts sur un patient. L'utilisation jointe de ces deux éléments nous a permis de construire un générateur de profils anesthésiques réalistes. Ce générateur nous affranchit des lourdes contraintes légales et techniques de l'exploitation des données de santé.

Nos travaux impliquant des séries temporelles, il s'est avéré nécessaire de déterminer une distance entre séries temporelles qui soit adaptée à notre contexte et à nos approches. Le chapitre 4 présente et compare de multiples distances afin de déterminer la plus adaptée.

Le chapitre 5 décrit les méthodes qui permettent, grâce à un raisonnement à base de cas, la simulation d'un patient virtuel directement à partir de données réelles, ce patient virtuel évoluant selon les décisions d'un apprenant dans un contexte de simulation.

Le chapitre 6 présente la méthode de construction d'une représentation synthétique qui associe les profils de patients réels localement similaires et mémorise localement, sous

la forme de consensus, les séries temporelles des profils anesthésiques d’une intervention. Ce chapitre se conclut sur une discussion sur l’opportunité apportée par ce type de représentation pour l’anonymisation de profils anesthésiques.

Le chapitre 7 décrit la construction de scénarios d’apprentissage contextualisés pédagogiquement et médicalement à partir d’une représentation synthétique. Il présente comment sont associés les différents éléments présentés dans ces travaux pour permettre la simulation d’un scénario d’apprentissage *via* un simulateur réactif.

L'ANESTHÉSIE ET LA FORMATION PAR LA SIMULATION EN SANTÉ

Introduction

L'anesthésie est une pratique médicale courante depuis la seconde moitié du XX^e siècle et dont l'usage est en progression. Si 8 millions d'anesthésies étaient réalisées en France en 1996 [4], 11.3 millions ont été réalisées en 2010 [2]. C'est de plus une pratique de plus en plus sûre : entre 1980 et 1999 le nombre de décès liés à une procédure anesthésique a été divisé par 10 en France (de 1 décès pour 13 200 anesthésies en 1980 à 1 décès pour 145 000 anesthésies en 1999) [5].

Cette progression est elle aussi continue au CHU de Nantes où les procédures anesthésiques sont en augmentation de 6.6% par an environ pour atteindre 67 942 interventions en 2016 [3].

Cette procédure médicale s'appuie sur des gestes médicaux et des médicaments spécifiques qui sont utilisés dans le cadre d'une procédure médicalement formalisée que nous présenterons dans la première partie de ce chapitre.

Dans une seconde partie, nous nous intéresserons à la formation des soignants en anesthésie par la simulation, avec un regard sur le contexte national et un focus particulier sur les modalités et pratiques pédagogiques au SiMU de Nantes.

1.1 L'anesthésie et ses grands principes

La société française d'anesthésie-réanimation (SFAR) donne pour *anesthésie* la définition [6] suivante :

L'anesthésie permet la réalisation d'un acte chirurgical, obstétrical ou médical (endoscopie, radiologie...), en supprimant la douleur provoquée pendant et en l'atténuant après l'intervention dans des conditions optimales de sécurité.

Dans le cadre d'un acte, l'anesthésie intervient pour supprimer la douleur provoquée par l'acte et pour faciliter cet acte. Elle peut ainsi impliquer une immobilisation totale ou partielle du patient, ainsi qu'une perte de conscience. Ces trois aspects : l'analgésie¹, l'immobilisation et la perte de conscience, et le fait qu'elles soient totales ou partielles, permettent une pluralité de modalités d'anesthésies pour s'adapter aux spécificités d'un patient ou de l'acte à réaliser.

La prise en charge d'un patient inconscient ou traumatisé rentre dans le champ de la réanimation. Si les techniques et médicaments utilisés sont proches, ces situations ne seront pas traitées dans le cadre des présents travaux de thèse.

1.1.1 Le contexte d'une anesthésie

L'anesthésie et sa pratique sont légalement encadrées par le *Décret n°94-1050 du 5 décembre 1994 relatif aux conditions techniques de fonctionnement des établissements de santé en ce qui concerne la pratique de l'anesthésie* [7]. Ce texte indique très précisément le déroulement du processus anesthésique et précise l'environnement technique et organisationnel nécessaire à son bon déroulement. Ce décret impose notamment la réalisation d'une consultation et d'une visite pré-anesthésique et demande aux établissements de santé de disposer de l'ensemble des moyens nécessaires pour réaliser une anesthésie, sa surveillance continue après l'intervention et d'être en capacité de faire face aux complications éventuelles.

1. La perte de la sensation de douleur.

Une anesthésie peut être nécessaire pour différentes raisons. L'atténuation de la douleur ressentie par le patient peut ainsi permettre :

- la bonne réalisation d'un soin ou sa facilitation,
- une procédure chirurgicale,
- la réalisation d'un examen médical invasif.

Les contextes d'utilisation de l'anesthésie sont ainsi très diversifiés. En 2010, c'est l'endoscopie digestive² qui représentait l'utilisation d'une anesthésie la plus fréquente (16.2%) suivie par les accouchements et césariennes (soins obstétricaux) (8.3%) [2].

La majorité des anesthésies sont programmées et suivent ainsi la procédure du décret de 1994.

1.1.1.1 Préparation d'une anesthésie programmée

Pour une intervention programmée, le décret de 1994 impose une consultation pré-anesthésique, et une visite pré-anesthésique avant toute procédure. Ces deux temps ont pour but de prendre en compte les spécificités du patient, d'adapter les gestes médicaux et médicaments utilisés, ainsi que les dosages, aux spécificités de la procédure et aux antécédents médicaux du patient.

La consultation pré-anesthésique doit avoir lieu plusieurs jours avant l'intervention. Réalisée par un médecin anesthésiste, elle a pour but de proposer la meilleure stratégie pour l'intervention prévue. Cette stratégie comprend la technique anesthésique, la gestion de la douleur post-opératoire, mais aussi éventuellement les aspects liés à une transfusion si l'intervention peut en nécessiter une. C'est un temps privilégié pour la bonne information du patient sur la procédure à venir. La consultation pré-anesthésique est surtout un temps dédié pour identifier et évaluer les facteurs de risques anesthésiques qui sont ensuite consignés dans le dossier médical du patient.

La visite pré-anesthésique qui a lieu quelques heures avant l'intervention a pour but de confirmer ou d'adapter, la stratégie formulée lors de la consultation.

Lors d'une situation d'urgence, il n'est pas possible de réaliser la consultation et la visite pré-anesthésique imposées par le décret de 1994.

Des procédures spécifiques sont ainsi mises en place pour minimiser les risques d'une anesthésie en situation d'urgence. Ces procédures utilisent des gestes médicaux et des médicaments spécifiques.

2. Un examen médical consistant à l'exploration du système digestif avec un tube souple muni d'une petite caméra.

1.1.1.2 Facteurs de risques anesthésiques et indicateurs

Les risques anesthésiques sont évalués grâce à plusieurs scores et indicateurs qui permettent une bonne préparation de l'intervention :

- l'évaluation d'une potentielle intubation difficile,
- la typologie du patient et ses éventuels antécédents ou comorbidités,
- les éventuelles allergies,
- le score ASA.

Intubation difficile Si l'intervention nécessite une intubation, l'évaluation d'une potentielle intubation difficile peut se faire *via* plusieurs indicateurs :

- le signe du prier : la capacité à joindre ses deux mains en signe de prière, parfaitement ou non, est un indicateur de la raideur articulaire,
- l'ouverture de la bouche : une distance inférieure à 35 mm entre les deux mâchoires est indicatrice d'une potentielle intubation difficile,
- le score de Mallampati [8], qui classe les patients selon l'encombrement de leurs voies aériennes. Les classes 1 et 2 présagent d'une intubation *a priori* facile, les classes 3 et 4 d'une intubation difficile (voir table 1.1 et figure 1.1).

Classes	Définition
I	Toute la luette et les amygdales sont visibles.
II	La luette est partiellement visible.
III	Le palais mou est visible.
IV	Seul le palais dur est visible.

Table 1.1 – Description des différentes classes du score de Mallampati, d'après *Mallampati et al.* [8]

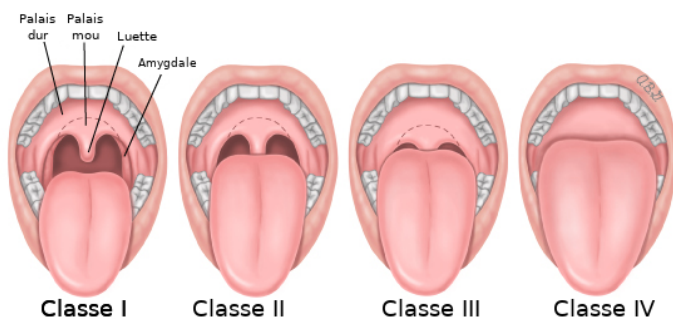


Figure 1.1 – Illustration des classes du score de Mallampati, d'après *Samsoon et al.* [9]

Typologie de patient et comorbidité Certains types de patients, ou pathologies, impliquent une surmortalité post-opératoire ou une procédure plus complexe. Les situations les plus courantes générant des complications sont :

- un âge avancé,
- une insuffisance cardiaque ou rénale,
- une hypertension artérielle,
- le diabète,
- l'obésité,
- le tabagisme.

L'appartenance d'un patient à ces typologies ou la présence d'une de ces pathologies est identifiée lors de la consultation pré-anesthésique.

Risques allergiques La majorité des accidents allergiques lors d'une anesthésie implique les curares³. Ils sont responsables de 62.6% des chocs anaphylactiques⁴, suivis par les allergies au latex (13.8%). [10] La recherche d'antécédents allergiques chez le patient, ou ses proches, ainsi que l'identification de groupes à risque, font partie de la consultation pré-anesthésique. Si nécessaire, le médecin anesthésiste peut prescrire un bilan allergologique.

Score ASA Lors de la consultation pré-anesthésique, l'interrogation du patient, l'analyse de ses antécédents et la récolte des différents indicateurs et facteurs de risques permet d'estimer son score ASA. Le score ASA⁵ correspond à une classification internationale des patients [11] et donne une estimation globale du risque anesthésique. Les différents scores et définitions sont présentés dans la table 1.2.

Score	Définition
1	Patient en bonne santé
2	Affection peu grave sans limitation fonctionnelle
3	Affection sévère avec limitation fonctionnelle
4	Affection sévère constituant une menace vitale permanente
5	Patient moribond avec une espérance de vie inférieure à 24 heures
6	Patient en mort cérébrale faisant l'objet d'un prélèvement pour le don d'organes

Table 1.2 – Table descriptive du score ASA

3. Les curares sont utilisés pour relâcher les cordes vocales et faciliter l'intubation. Voir 1.1.4

4. Un choc anaphylactique est la réaction physiologique exacerbée du corps en contact avec un agent allergène.

5. D'après *American Society of Anesthesiologists*.

1.1.2 Principe d'une anesthésie générale

Si l'anesthésie implique systématiquement l'atténuation de la douleur, la perte de conscience et l'immobilisation du patient ne sont pas systématiques, ce qui permet de différencier deux grands types d'anesthésies :

- les anesthésies locorégionales, où le patient reste conscient,
- les anesthésies générales, où le patient est totalement endormi.

Dans le cadre de ces travaux de thèse, nous nous intéressons plus spécifiquement aux anesthésies générales qui sont les procédures qui font l'objet d'une formation par la simulation.

Il existe des approches combinées d'anesthésies générales et locorégionales, mais elles sont spécifiques à certaines chirurgies et, comme les anesthésies locorégionales, elles ne seront pas discutées ici.

Quel que soit le type d'anesthésie, il doit suivre les mêmes règles de sécurité et de surveillance et la même procédure globale indiquée dans le décret de 1994.

1.1.2.1 Éléments de compréhension sur le système nerveux humain

L'analgésie permet de bloquer les signaux de la douleur au sein du système nerveux. Les paragraphes suivants présentent de manière synthétique les éléments principaux du système nerveux humain.

Le système nerveux Le système nerveux est responsable de la communication entre les différentes parties du corps, sur lequel il exerce un contrôle conscient et inconscient. Il conduit l'information sensorielle et permet la coordination musculaire.

Il est composé de deux parties spécifiques : le *système nerveux central* et le *système nerveux périphérique*.

Le système nerveux central Ce système a pour principal rôle la conduction et le traitement des informations. Il se compose principalement du cerveau, de la moelle épinière et du tronc cérébral⁶ qui fait la jonction entre les deux précédents. L'ensemble du système nerveux central est protégé par une enveloppe osseuse formée par le crâne et la colonne vertébrale, et par *les méninges*⁷.

6. Le tronc cérébral est situé sous le cerveau et en avant du cervelet, dans la fosse crânienne postérieure.

7. Les méninges sont situées entre les membranes qui enveloppent la moelle épinière et la cavité osseuse constituée par le canal vertébral, et baigne dans un liquide, le *liquide cébrospinal*.

Le système nerveux périphérique Il est composé des différents nerfs (appelé aussi *tronc nerveux*) qui partent de la moelle épinière et du tronc cérébral jusqu'aux organes. Il transporte l'information sensitive et motrice. L'entrecroisement de fibres nerveuses forme un *plexus nerveux*, une zone d'échanges intenses d'informations.

Neurones Le système nerveux est, d'un point de vue cellulaire, principalement composé de cellules particulières, les neurones, qui sont très diversifiés. Ces cellules spécialisées peuvent transmettre un signal électrochimique aux neurones adjacents. Ce signal consiste en la polarisation/dépolarisation de la membrane du neurone. Ce signal est ensuite transmis grâce à un neurotransmetteur chimique.

Polarisation cellulaire La membrane des neurones est électrochimiquement polarisée grâce à une différence de concentration (un gradient) en ions entre l'intérieur et l'extérieur de la cellule. Ce sont principalement les ions sodium (Na^+) qui participent à cette polarisation. Les ions sont transportés *via* des canaux ioniques à travers la membrane cellulaire.

1.1.2.2 Les anesthésies générales

L'anesthésie générale consiste en l'utilisation d'antalgiques puissants, les *morphiniques*, sur un patient inconscient. Elle est préconisée quand le patient ne peut pas supporter l'intervention consciemment. La perte de conscience d'un patient implique l'usage de médicaments spécifiques, les *hypnotiques*⁸, et des gestes médicaux spécialisés afin de l'initier et de la maintenir.

Selon la profondeur de la sédation⁹, le patient peut perdre son réflexe respiratoire. Il est donc nécessaire de l'intuber et de le relier à un respirateur mécanique pour lui assurer une assistance respiratoire. L'intubation peut nécessiter l'utilisation de médicaments comme les *curares*.

Pendant toute la durée de l'intervention, l'anesthésie est entretenue par l'injection continue, ou régulière, d'hypnotiques, de morphiniques et de curares si nécessaire.

8. Substance capable d'induire et/ou de maintenir le sommeil.

9. Perte de la vigilance et de la conscience par l'usage de médicament. La sédation peut être plus ou moins profonde, de la somnolence au coma.

Une anesthésie générale implique de nombreux risques qui sont généralement liés à la chirurgie. La perte de conscience, quand elle implique une perte du réflexe respiratoire, peut engendrer des complications respiratoires. La plus courante est l'intubation difficile où le patient peut être exposé à un temps d'apnée long.

L'utilisation de curare est la cause de la majorité de risques allergiques, mais l'utilisation de certains hypnotiques peut engendrer, dans des cas très rares, des réactions allergiques graves et généralement imprévisibles.

1.1.3 La structure et le déroulement type d'une anesthésie générale

Une anesthésie générale se déroule en quatre grandes étapes :

- **l'entrée du patient** : le patient arrive au bloc et est préparé pour l'intervention.
- **l'induction anesthésique** : l'analgésie est réalisée et la perte de conscience est induite. Le patient est placé sous assistance respiratoire.
- **la procédure chirurgicale ou médicale** : les actes prévus lors de l'intervention sont réalisés. La perte de conscience et l'analgésie sont entretenues.
- **la sortie du patient** : l'entretien de l'anesthésie est arrêté et le patient est préparé pour le réveil.

Chacune de ces étapes correspond à un temps spécifique de l'anesthésie. Elles ont pour caractéristique commune qu'au début et à la fin de chaque étape, le patient est dans un état stable. En d'autres termes, si l'équipe médicale ne déclenche pas de nouvelles actions sur le patient, ses paramètres physiologiques resteront stables.

Une représentation de ces grandes étapes est donnée en figure 1.2.

Grandes étapes	Étapes	Sous étapes	
Entrée	Mise en place	Entrée du patient Mise en place du patient	
	Monitoring	Monitoring de la fréquence cardiaque Monitoring de la pression artérielle Monitoring de la saturation en O2	
		Préparation	Monitoring Bi-Spectral (BIS) Contrôle TOF Bair hugger Installation de la voie veineuse périphérique Propylaxie Remplissage vasculaire
			Induction
Procédure	Mise en place pour la chirurgie Auscultation pulmonaire Contrôle des points de support Mise à disposition Incision Pose du trocart ombilical Insufflation du pneumopéritoine Auscultation pulmonaire Manœuvre de recrutement alvéolaire Dissection de l'espace préperitonéal Dissection du cordon spermatique Dissection de la zone herniaire Mise en place de la prothèse Déflation du pneumopéritoine Fermeture Pansement		
	Sortie	Contrôle de la décubitarisation Décubitarisation Mise en place du patient Sortie	

Figure 1.2 – Structure classique d’une anesthésie générale. Illustration dans le cas de la cure d’une hernie inguinale, sous coelioscopie, avec pose de prothèse.

1.1.3.1 L’entrée du patient

L’entrée est la première grande étape d’une anesthésie générale. Elle est composée de 3 sous-étapes principales qui consistent en la mise en place du patient dans la salle, l’installation des différents capteurs de monitoring et la préparation de l’induction par la *préoxygénation*.

Mise en place Lors de son entrée dans la salle d’opération, le patient est éveillé et mis en place sur la table d’opération.

Monitoring L’enregistrement des paramètres physiologiques du patient est une nécessité pour assurer le bon déroulement de l’intervention. Différents capteurs sont mis en place afin de récolter ces paramètres physiologiques (fréquence cardiaque, pression artérielle, composition et débit des gaz). Des tests et examens sont aussi réalisés pour servir de référence pour la suite de l’intervention (généralement un test TOF¹⁰).

Préparation Sur la table d’opération, les patients se refroidissent progressivement. Pour contrecarrer ce refroidissement naturel, une couverture chauffante est mise en place sur le patient.

Une voie veineuse périphérique est aussi installée sur le patient afin de réaliser les injections intraveineuses.

10. *Train of four*, voir sous-section 1.2.4.

Selon les types d'intervention, des fluides peuvent aussi être injectés au patient (solutions physiologiques ou ioniques) ou un antibiotique en prophylaxie¹¹.

1.1.3.2 L'induction anesthésique

Une fois le patient en place, il est possible d'initier le *sommeil anesthésique*. Une fois le patient inconscient, si l'intervention est longue et la sédation profonde, il est nécessaire de le mettre sous assistance respiratoire et donc pour cela de l'intuber.

L'intubation est réalisée en apnée. La durée d'apnée tolérable par le patient dépend directement de sa saturation sanguine en oxygène¹². Si une situation d'intubation difficile imprévue se présente, la durée d'apnée peut s'allonger et dépasser les réserves du patient. Ainsi, avant toute induction d'une anesthésie, les réserves du patient sont augmentées lors d'une étape de préoxygénation.

Préoxygénation La préoxygénation est réalisée avec de l'oxygène pur, grâce à un masque facial. La préoxygénation est considérée terminée quand la saturation sanguine en oxygène du patient et la concentration en oxygène des gaz sortants sont maximales (> 90%) et après au moins 3 minutes d'inhalation d'oxygène pur. Pendant la préoxygénation, les premiers morphiniques sont injectés au patient.

Médication Une fois les réserves du patient optimisées par la préoxygénation, le patient est prêt pour l'endormissement.

Pour induire la perte de conscience, les hypnotiques sont injectés majoritairement *via* la voie veineuse périphérique, plus rarement par inhalation sous forme gazeuse.

Intubation Le patient est ensuite *curarisé*, par l'injection d'un curare, afin de relâcher les cordes vocales et ainsi faciliter l'intubation qui permet l'assistance respiratoire.

Contrôle et fin d'induction Une fois le patient endormi et intubé, les différents éléments en place sont contrôlés et la bonne intubation est vérifiée grâce à une auscultation pulmonaire.

11. Approche qui consiste à donner un médicament à titre préventif.

12. La quantité d'oxygène dissoute dans le sang.

La concentration en oxygène inspiré est descendue à 30% et l'entretien de l'anesthésie est initié par l'injection continue (perfusion) ou répétée (bolus¹³) d'hypnotiques, de morphiniques et éventuellement de curares si l'intervention le nécessite.

À la fin de l'induction, le patient est inconscient et dans un état stable qui permet sa «mise à disposition» au chirurgien.

1.1.3.3 La procédure de l'intervention

Pour des raisons logistiques et organisationnelles, il peut parfois s'écouler un certain temps entre la fin de l'induction et le début de la procédure.

Préparation Un contrôle supplémentaire des éléments en place et une auscultation pulmonaire sont ainsi généralement réalisés avant le début de la procédure chirurgicale ou médicale.

Procédure Une anesthésie bien réalisée maintient un patient stable tout au long de la procédure chirurgicale ou médicale. Pendant cette étape, l'entretien anesthésique anticipe les actes chirurgicaux les plus douloureux, ou les gestes médicaux ayant un fort impact sur le patient, afin de le maintenir le plus stable possible.

1.1.3.4 La sortie du patient

Une fois la procédure réalisée, la dernière étape d'une intervention consiste à préparer le patient pour la sortie du bloc opératoire.

Décurarisation Il est recommandé que le patient sorte de la salle d'opération totalement décurarisé [12]. Un contrôle de la décurarisation est donc réalisé, en comparant les résultats avec le test contrôle réalisé en début d'intervention. Si à la fin de la procédure, le patient n'est pas naturellement décurarisé et que la curarisation est encore profonde, il est nécessaire d'attendre ou d'utiliser un antidote. Une fois la décurarisation réalisée, le patient est transféré en salle de réveil avec un maintien du monitoring.

13. Un bolus est une dose administrée au patient de manière complète en une seule fois. En médecine moderne, c'est généralement réalisé par une injection intraveineuse.

1.1.4 Les principaux gestes médicaux et médicaments utilisés

Pendant les quatre grandes étapes d'une anesthésie générale, des médicaments et des gestes médicaux spécifiques sont utilisés. Cette partie s'intéresse spécifiquement aux gestes médicaux et aux médicaments anesthésiques, indépendamment de l'intervention nécessitant une anesthésie.

Mise en place du monitoring Le monitoring du patient se concentre sur deux principaux aspects : le monitoring hémodynamique et le monitoring des gaz.

Le monitoring de la fréquence cardiaque et de la saturation sanguine en oxygène est réalisé grâce à un capteur spécifique (généralement placé au bout d'un doigt), relié à un moniteur, le *scope*. Le monitoring de la pression artérielle est réalisé par une mesure par brassard à intervalles réguliers gérés automatiquement par le *scope*.

Le monitoring des gaz entrants et sortants est réalisé directement par le respirateur, grâce à la présence de différents capteurs dans le circuit de distribution des gaz au patient, à proximité de la sonde d'intubation.

Pour certaines chirurgies ou interventions spécifiques, d'autres paramètres peuvent être monitorés, comme la glycémie.

Indice Bi-spectral Dans certains cas, une sonde spécifique peut être installée afin de mesurer l'indice bi-spectral (ou BIS). Le BIS est un indicateur de la profondeur de la sédation. Il est basé sur une modélisation statistique s'appuyant sur l'électroencéphalogramme (EEG) du patient. La sonde et les logiciels liés sont sous licence propriétaire et ne sont donc utilisés que ponctuellement.

TOF Le TOF (*Train Of Four*) est un examen de mesure de la décurarisation qui consiste en la réalisation de 4 stimulations nerveuses répétitives, puis en la mesure de la force de la réponse pour chaque stimulation. En cours d'intervention, l'objectif est de 0 à 2 réponses. La décurarisation est considérée complète quand le rapport des réponses n°4 (*Train 4* ou T_4) et n°1 (*Train 1* ou T_1), ou rapport T_4/T_1 est supérieur ou égal à 90% [13].

Voie veineuse périphérique Il s'agit d'un dispositif médical qui permet l'introduction d'un tube (le *cathéter*) dans une veine périphérique. Il permet l'injection de produits anesthésiques par voie intraveineuse.

Prophylaxie antibiotique La prophylaxie antibiotique est une mesure de prévention qui permet de prévenir une éventuelle infection liée à l'intervention, notamment pour les interventions à risques importants d'infection post-opératoire (principalement les chirurgies). Le type d'antibiotique et la posologie sont spécifiques aux différents types de chirurgies.

Remplissage vasculaire Le remplissage vasculaire est un apport de liquide par voie veineuse. Il permet de corriger ou prévenir une hypotension.

Préoxygénation Une bonne pré-oxygénation dure au moins 3 minutes avec de l'oxygène pur et un débit en gaz frais¹⁴ de 5L/min [13]. L'objectif, pour régénérer les réserves du patient, est d'avoir à la fois une saturation sanguine en oxygène de plus de 90% mais aussi une concentration en oxygène dans les gaz expirés de 90%.

Hypnotiques Ils assurent l'*endormissement*, puis l'entretien de la perte de conscience. Ils sont généralement introduits par voie intraveineuse, et plus rarement par inhalation.

Le produit le plus souvent utilisé est le *propofol* pour sa facilité d'élimination, bien que son injection soit douloureuse. Il est généralement utilisé en dose (ou bolus) pour l'induction puis par perfusion continue pour l'entretien anesthésique.

Les produits volatiles, comme le *sévoflurane*, ou l'*halothane*, sont vaporisés par l'appareillage respiratoire à l'aide de systèmes dédiés et calibrés.

Les systèmes AIVOC (Anesthésie Intra-veineuse à Objectif de Concentration) sont des systèmes automatisés qui permettent une induction et un entretien par référence à une concentration plasmatique cible représentative du niveau d'anesthésie souhaité.

Morphiniques Les morphiniques sont des antalgiques puissants de la même famille que la *morphine* tel que le *sufentanil*, l'*alfentanil* ou le *fentanyl*. Ils sont introduits par voie intraveineuse au patient.

La *morphine* n'est pas utilisée pendant les interventions en raison de son délai d'action trop long avant les premiers effets (10 minutes).

Curares Les curares sont utilisés pour relâcher les cordes vocales et faciliter l'intubation, et ainsi diminuer les complications traumatiques éventuelles. Les curares sont injectés par

14. Un *gaz frais* est un gaz qui est issu d'une bonbonne et qui n'a pas déjà été utilisé, par opposition à un *gaz recyclé*.

voie intraveineuse. Leur choix dépend de la durée de l'intervention et du délai d'action souhaité.

Les curares «dépolarisants» comme la *succinylcholine* provoquent une dépolarisation de la jonction neuromusculaire pour tous les muscles avant d'entraîner une paralysie. Avec un délai d'action très rapide (inférieur à 1 minute), c'est le seul curare utilisable dans le cas où le patient n'est pas à jeun (et donc en cas d'urgence), malgré un fort risque allergique. L'effet est d'environ 10 minutes.

Les curares «non dépolarisants», comme l'*atracurium* ou le *rocuronium* ont un délai d'action de 3 minutes. Pendant ce délai, le patient ne peut déjà plus respirer par lui-même ; il a donc besoin d'une assistance respiratoire apportée grâce au masque facial. Ces curares ont une durée d'action variable mais plus longue que celle de la *succinylcholine*.

Décurarisation La mesure du niveau de curarisation, et donc de décurarisation par référence à un test contrôle, est réalisée grâce à plusieurs moyens incluant notamment le TOF.

Si la curarisation est trop profonde, un antidote comme l'*atropine* ou la *prostigmine* peut être utilisé.

Intubation L'intubation permet le maintien de la ventilation¹⁵ du patient. Une sonde est placée dans le larynx, après les cordes vocales, pour faire circuler l'air et l'étanchéité est assurée par un petit ballonnet gonflé.

Une intubation trop profonde peut engendrer une intubation sélective¹⁶. Une autre complication potentielle est une intubation oesophagienne, où l'oesophage est intubé au détriment des poumons.

Quand l'intubation n'est pas possible, en cas d'*intubation difficile* par exemple, il est possible d'utiliser un masque laryngé. C'est un tube doté d'un petit coussinet qui se place sur les cordes vocales.

Auscultation pulmonaire L'auscultation pulmonaire permet de vérifier la bonne intubation par détection des bruits respiratoires normaux.

15. La ventilation est la capacité d'un patient à respirer. Elle est naturelle si le patient respire par lui-même, ou mécanique si elle nécessite l'assistance d'un appareillage.

16. En étant placée trop loin, la sonde d'intubation se trouve alors dans le pharynx et un seul poumon est intubé.

Manoeuvre de recrutement alvéolaire Sous l'effet du respirateur et de la position du patient, les alvéoles pulmonaires peuvent s'écraser légèrement en formant des *atélectasies*, ce qui réduit la capacité respiratoire du patient et les échanges gazeux. Une manoeuvre de recrutement alvéolaire permet de re-mobiliser (recruter) ces alvéoles par une augmentation de la pression de l'air dans les poumons, ce qui «re-gonfle» les alvéoles pulmonaires écrasées.

1.2 La formation par la simulation en santé

La pratique de l'anesthésie implique inévitablement un patient qui fait l'objet d'une intervention. La formation des soignants en anesthésie a, par le passé, impliqué la présence d'apprenants lors de procédures réelles, et la pratique de novices sur des patients, une approche discutable d'un point de vue éthique et de qualité des soins.

Depuis 2012, et le principe *«jamais la première fois sur le patient»* encouragé par la Haute Autorité de Santé, l'usage de la simulation s'est systématisé dans la formation des soignants.

1.2.1 Le développement de la simulation en santé en France

Le SiMU de Nantes indique à propos de la simulation [14] :

La simulation est un outil pédagogique moderne dédié à l'apprentissage et au maintien des compétences dans le domaine des sciences de la santé. Il permet aux apprenants de s'entraîner à la prise en charge de situations réalistes dans un environnement sécuritaire, sans risque pour le patient.

Les approches de simulation sont multiples et spécifiques aux savoir-faire simulés. Pour la simulation en santé, on peut identifier plusieurs éléments communs à toutes ces approches [15] :

- l'utilisation d'un objet support, organique ou non, vivant ou non, qui est l'objet de l'action des apprenants (acteur, mannequin, etc.),
- l'immersion plus ou moins importante des apprenants, que ce soit par le contexte de l'action réalisée (scénario, dossier médical) ou l'environnement (réaliste, avec plongement en réalité virtuelle),
- l'échelle, qui peut être globale au niveau de la prise en charge complète d'un patient ou spécifique par la réalisation d'un geste médical précis.

On peut ainsi classer les différentes modalités de simulation selon plusieurs catégories, selon l'objet support de la simulation [16].

Simulation animale L'expérimentation sur des animaux [17, 18] permet la formation aux gestes chirurgicaux simples comme complexes.

Simulation Humaine Sous la forme de jeux de rôles [19], il est possible de travailler avec les apprenants sur les compétences relationnelles en les confrontant à des patients types, *les patients standardisés* [20]. Ce type de simulation implique un acteur, incarnant un patient dont la situation médicale et le type de personnalité est fixe (le *patient standardisé*), afin de confronter les apprenants à des scénarios identiques notamment vis-à-vis des réactions du patient. Cette modalité déroule un scénario permettant la simulation de consultations.

L'utilisation de cadavres peut permettre la formation aux techniques de la chirurgie, de l'anesthésie-réanimation et de la médecine d'urgence [21].

Simulation synthétique Cette modalité de simulation mobilise des appareillage spécifiques comme les *simulateurs patients* [22], qui représentent des patients sous la forme de mannequins en haute ou basse fidélité. La formation à des gestes particuliers par la répétition, par exemple la pose d'une perfusion, ou l'intubation, est possible grâce à des *simulateurs procéduraux* en haute ou basse fidélité qui permettent l'entraînement à ces procédures médicales.

Si la simulation est historiquement employée depuis très longtemps, son développement et son usage large en santé sont très récents, notamment en France.

1.2.1.1 Historique et précurseurs

L'usage de la simulation dans l'histoire est à rapprocher de celui du jeu et de l'entraînement qui sont pratiqués dans les activités guerrières depuis l'antiquité (manoeuvre, joutes, etc) [23].

L'usage de la simulation en santé est historiquement lié à la formation. Des mannequins sont utilisés pour la formation des sages-femmes de campagne aux manoeuvres obstétricales, à l'initiative de Madame Du Coudray, au XVII^e siècle [24].

Si l'utilisation d'acteurs pour simuler des patients dit «standardisés», commence dans les années 70 aux États-Unis [25] pour la formation des soignants, le développement de la simulation en santé à l'époque moderne, puis contemporaine, est directement lié aux améliorations technologiques des mannequins. Au long du XX^e siècle, différents mannequins sont développés par des soignants et ingénieurs pour faciliter l'apprentissage des gestes techniques. On peut ainsi noter la création du *Sim One*, premier mannequin contrôlé par ordinateur [26]. Le développement de logiciels spécialisés participe aussi au développement

de la simulation, notamment le logiciel *GasMan* (1984) [27] qui permet la simulation des interactions entre des produits anesthésiques et un patient, et la simulation du comportement de ces produits anesthésiques.

Les différents dispositifs de simulation développés ensuite apportent systématiquement un gain en réalisme ou en capacité d'interaction, pour aboutir aux mannequins haute fidélité actuels [28].

La publication du rapport "*To err is human*" en 2001 [29] aux États-Unis, sur l'implication des facteurs humains dans les erreurs médicales, met en avant la simulation comme un moyen de réduire la fréquence de ce type d'erreurs, et positionne les États-Unis en tête de l'utilisation de la simulation pour la formation des soignants.

1.2.1.2 État des lieux de la simulation en santé en 2012

En 2010, le Pr Jean-Claude GRANRY et le Dr Marie-Christine MOLL sont missionnés par la Haute Autorité de Santé (HAS) pour réaliser un état des lieux des initiatives et expériences en cours en France et à l'international sur la simulation en santé. Leur travail aboutit à la publication d'un rapport en janvier 2012. [1]

Ce rapport souligne l'utilisation large de la simulation en Amérique du Nord, où elle est un outil pédagogique à part entière, utilisé de manière routinière dans l'enseignement initial des professions médicales et paramédicales.

Les auteurs indiquent de plus, que si la simulation en santé est bien implantée en Europe, avec un développement plus récent qu'en Amérique du Nord, les moyens sont plus faibles et les inégalités vis-à-vis de ces moyens sont importantes.

En 2012, en France, la simulation est encore une activité émergente, avec peu de centres de simulation, mais une répartition assez homogène sur le territoire national. Les centres sont de petite taille avec peu de diversité dans les mannequins haute fidélité utilisés. Les simulateurs chirurgicaux sont rares et la simulation ayant recours à des environnements virtuels est quasiment inexistante. La pratique de la simulation en santé en France manque alors de bonnes pratiques formalisées et souffre d'une grande dispersion méthodologique.

Le rapport indique dans ses conclusions 10 propositions dont notamment :

- l'intégration de la simulation dans *«tous les programmes d'enseignement des professionnels de santé à toutes les étapes de leur cursus (initial et continu)»*,
- un objectif éthique prioritaire : *«jamais la première fois sur le patient »*,
- la mise en place d'une politique nationale de formation par la simulation, et donc de son développement.

1.2.1.3 Guide de l'HAS et développements récents

En décembre 2012, dans la continuité de l'état des lieux, la Haute Autorité de Santé publie un guide de bonnes pratiques en matière de simulation en santé [30].

Dans ce guide, l'HAS indique ses ambitions de développer la simulation en santé, notamment comme outil de développement professionnel continu (DPC) des soignants, en permettant sa bonne structuration par des bonnes pratiques.

Elle y structure ses recommandations et bonnes pratiques sur 5 aspects :

- les programmes de simulation, leurs conception et leurs mise en place,
- l'organisation des centres de simulation,
- les infrastructures et équipements,
- les axes de recherche et modalités d'organisation de la recherche sur la simulation,
- l'évaluation des programmes, formateurs et apprenants.

La diffusion de ce guide de bonnes pratiques a encouragé le développement des centres de simulation en santé en France et a standardisé les pratiques.

En 2012, l'état des lieux de l'HAS indiquait l'existence de 34 centres de simulation implantés ou en projets. En 2022, la Société Francophone de Simulation en Santé (SoFra-SimS) dénombre 95 centres de simulation en Santé en France [31], soit un développement de 270% en 10 ans.

Dans la continuité des bonnes pratiques publiées en 2012, l'HAS a publié en mai 2015 un « *Guide pour l'évaluation des infrastructures de simulation en santé* » [32], afin d'encourager les centres de simulation à s'engager dans une démarche d'amélioration continue de la qualité.

1.2.2 La simulation au SiMU de Nantes

Nantes Université dispose, depuis 2011, d'un centre de simulation en santé : le laboratoire expérimental de simulation de médecine intensive de Nantes Université, ou LE SiMU.

LE SiMU intervient dans la formation initiale des étudiants en médecine (1^{er}, 2^e et 3^e cycles), odontologie (2^e cycle) et maïeutique¹⁷ (2^e cycle), sur des thématiques et scénarios spécifiques à ces formations.

17. Science médicale exercée par les sages-femmes et qui traite de la grossesse, de l'accouchement ainsi que de la santé sexuelle et reproductive des femmes.

Dans le cadre du développement professionnel continu (DPC), LE SiMU propose des sessions de formation régulières sur différentes thématiques précises (anesthésie-réanimation, médecine d’urgence, etc.)

LE SiMU met principalement en œuvre des sessions de simulation haute fidélité, généralement dans un contexte interprofessionnel.

1.2.2.1 La simulation haute fidélité interprofessionnelle

Il est difficile de mesurer l’impact sur les apprentissages de la fidélité d’une simulation à la réalité. La littérature consacrée à la pédagogie oppose la *fidélité d’ingénierie*, ou authenticité, correspondant aux similitudes physiques entre la simulation et la réalité, et la *fidélité psychologique* qui correspond aux éléments clefs d’une situation réelle. Ces deux notions sont empruntées à la simulation aéronautique [33]. La littérature spécialisée actuelle suggère que la fidélité psychologique serait plus efficace pour l’apprentissage [34].

La littérature suggère qu’une trop grande différence entre l’environnement d’apprentissage et l’environnement d’application nuit au transfert des compétences [35, 36] et qu’une forte similarité perçue de ces environnements influence favorablement les potentialités de transfert [37].

Une représentation la plus réaliste possible de l’environnement et des faits pendant une simulation est donc souhaitable, notamment lors de l’apprentissage de tâches complexes [38].

Environnement haute fidélité Les environnements de haute fidélité s’attachent à reconstituer l’ensemble des éléments d’un environnement de travail. Ainsi le SiMU reconstruit les environnements habituels des soignants selon les scénarios et disciplines (bloc opératoire, salle d’accouchement, etc.) et dispose pour cela de 3 salles de simulation (adulte, salle de naissance, pédiatrie).

Mannequin haute fidélité Les mannequins peuvent être classés en plusieurs catégories selon la réponse qu’ils proposent aux actions des apprenants, d’aucune réponse (basse fidélité) aux réponses complexes (haute fidélité). Les mannequins haute fidélité proposent un pilotage informatisé large et une réponse physiologique intégrée aux actions des apprenants [22]. Les larges potentialités des mannequins haute fidélité en font des outils très polyvalents pour les centres de simulation en santé.

LE SiMU dispose de 7 simulateurs de patients haute fidélité (adultes, parturiente, enfant, nourrissons et nouveau-né) ainsi qu'un appareillage de simulation de bloc opératoire complet.

Contexte interprofessionnel La recommandation «*Jamais la première fois sur le patient*» s'applique aussi aux compétences relationnelles et à la collaboration interprofessionnelle. La simulation haute fidélité dans un contexte interprofessionnel permet une action collective dans la réalisation du soin. La coopération interprofessionnelle permet une meilleure performance face aux soins complexes [39] et ainsi une stabilisation des pratiques collectives.

Utilisation La simulation haute fidélité interprofessionnelle a montré son impact positif dans l'amélioration de l'enseignement [40] et s'intègre dans les pratiques de gestion des risques et des facteurs humains. Ses effets positifs pour entretenir les compétences réactionnelles des soignants sont démontrés depuis les années 60, notamment en anesthésie [41].

1.2.3 Le cadre et les pratiques pédagogiques des sessions de simulation au SiMU de Nantes

La simulation est un moyen qui sert des objectifs spécifiques selon les publics et contextes dans lesquels elle est utilisée.

En formation initiale, comme une formation continue, les scénarios et le contexte pédagogique sont adaptés aux objectifs de formation.

1.2.3.1 Contexte pédagogique

Le contexte pédagogique est l'ensemble des traits décrivant et conditionnant la situation d'enseignement. Le contexte pédagogique d'une séance de simulation comprend ainsi les approches, les parti-pris des formateurs ainsi que l'écriture des scénarios et les moyens mis en place pour atteindre les objectifs de la séance.

Si une séance a pour objectif l'acquisition d'une compétence pointue, comme la réalisation d'une intubation, alors le contexte pédagogique se composera d'un scénario qui présente une situation qui nécessite une intubation et du matériel de simulation adapté.

En revanche, pour une séance qui cherche à faire travailler les apprenants en collaboration avec d'autres soignants, une situation mettant en avant des problématiques médicales nécessitant cette collaboration sera proposée aux apprenants.

Le contexte pédagogique, pour permettre d'atteindre les objectifs de la séance, peut venir altérer le réalisme médical de la simulation, par exemple en limitant les effets sur le patient virtuel d'un geste mal maîtrisé, ou encore en contraignant les choix de l'apprenant pour le pousser à utiliser une technique spécifique.

Posture et éthique Les séances de simulation sont parfois des temps intenses et stressants. 20% des participants à une formation en anesthésie par la simulation ont indiqué avoir déjà été stressés ou intimidés¹⁸ [42].

Sur les aspects éthiques, un sujet largement débattu est le décès du patient virtuel. En dehors de scénarios spécifiques sur la gestion des décès, dans un scénario avec un risque vital, si la performance des apprenants n'est pas à la hauteur des objectifs du scénario, le patient doit-il mourir ou la simulation doit-elle être interrompue par l'intervention d'un autre soignant qui prend le patient en charge ? Plus largement, les questions éthiques en simulation portent sur la charge émotionnelle et le stress que peut provoquer la simulation, ce qui nuirait à la motivation des apprenants [43, 44]. Les recommandations encouragent l'application du principe *primum non nocere*¹⁹ en l'adaptant en *erudere sine nocere*²⁰.

Le cadre pédagogique et les formateurs ont un rôle crucial pour assurer le bon déroulement des séances, la mise en confiance des apprenants et la conformité de la formation aux principes éthiques du centre de simulation. Cet accompagnement pédagogique s'illustre dans l'organisation des sessions de simulation et la place laissée à l'échange en amont et en aval de la pratique simulée.

1.2.3.2 Le déroulement des sessions de simulation

Une session de simulation est généralement thématique et cherche à faire travailler les apprenants sur un sujet ou une situation précise, par exemple la gestion de situations critiques en anesthésie adulte.

18. La principale cause donnée par les apprenants est la crainte du jugement des formateurs ou de leurs pairs.

19. *En premier, ne pas nuire*, premier principe de prudence appris aux soignants lors de leur formation. Si la locution est d'apparition tardive (XIX^e siècle) elle s'insère dans la tradition hippocratique : *faire du bien, ou au moins ne pas faire de mal*.

20. *Eduquer sans nuire*

Une session de simulation rassemble un groupe d'une dizaine d'apprenants, généralement sur une journée. Une session permet aux apprenants d'enchaîner plusieurs séances de simulation avec des scénarios différents mobilisant quelques apprenants, les autres étant placés dans une situation d'observateurs.

Après l'accueil des participants, la session est introduite par un briefing. Cette étape préparatoire est indispensable pour construire un climat d'apprentissage favorable [42].

Une séance de simulation est une entité pédagogique qui s'articule autour d'un scénario et d'un groupe d'apprenants qui y participent. Généralement, une séance comprend :

- une préparation, un échange court dont l'objectif est de mettre en contexte les participants sur le scénario et les tâches à réaliser. Le patient et l'intervention sont présentés pour contextualiser le scénario.
- une pratique simulée, qui est déroulée selon les paramètres prévus par le scénario en impliquant quelques-uns des apprenants.
- un débriefing, afin de faire un retour collectif sur la simulation.

Les participants qui ne sont pas impliqués dans le scénario observent le déroulement de la simulation dans la salle de briefing grâce à une ou plusieurs caméras.

Le débriefing est un temps qui cherche à faire réfléchir collectivement les participants sur la prestation en simulation et à comparer celle-ci aux standards et recommandations. C'est un temps dédié à l'analyse des éventuels écarts aux bonnes pratiques, et au feedback direct aux apprenants.

D'un point de vue pédagogique, le débriefing permet à la fois la *contextualisation* des connaissances, par une application à un cas concret présenté dans le scénario, mais aussi une *décontextualisation* en permettant une généralisation des connaissances lors des échanges pour pouvoir les transférer à d'autres contextes.

La littérature spécialisée a démontré le rôle du débriefing et du feedback dans la construction des apprentissages [45, 46]. C'est la réflexion collective, en lien avec la pratique simulée qui rend ce temps d'échange efficace dans son rôle formateur [47, 48].

Les séances de simulations sont de structure fixe ; le scénario est l'élément de variation qui permet une progression des acquis des apprenants par enchaînement des scénarios de complexité croissante.

1.2.3.3 Les scénarios

Les scénarios sont spécifiques aux apprenants. Selon leur niveau de compétence et leur discipline, les tâches à réaliser et compétences à mobiliser sont différentes. Un scénario de simulation cherche principalement à faire progresser les apprenants sur une ou deux notions ou situations spécifiques. Il implique donc rarement la simulation d'une intervention complète.

Il existe plusieurs structures de scénarios [49, 50] mais les éléments suivants sont généralement présents :

- le contexte de prise en charge et l'environnement, ce qui fournit les paramètres de départ du simulateur et détermine éventuellement la thématisation de l'espace.
- la description du patient avec les éléments clef de son dossier médical. La découverte d'un élément clef en cours de simulation (par un examen ou la transmission d'une information) peut être un élément de rebondissement.
- la progression du scénario, qui décrit la prise en charge du patient avec éventuellement plusieurs branches, en fonction des actions des apprenants. Cette progression comprend les éventuels éléments perturbateurs du scénario.

Lors de la simulation, les formateurs pilotent le simulateur haute fidélité, en conformité avec les paramètres décrits dans le scénario, pour faire évoluer les réactions et paramètres physiologiques du patient virtuel. Ce pilotage doit faire preuve d'une certaine souplesse afin de s'adapter aux actions des apprenants qui peuvent être non prévues par le scénario.

Conclusion

L'anesthésie est une pratique médicale désormais courante, et dont la procédure est légalement réglementée. Cet encadrement légal impose notamment un enregistrement et une surveillance des patients. L'anesthésie utilise différents gestes médicaux et médicaments spécifiques pour sa réalisation, chacun ayant un impact sur le patient. Les différents types d'anesthésie sont médicalement formalisés par grandes étapes. Une anesthésie générale est ainsi structurée en quatre grandes étapes successives : l'entrée, l'induction, la procédure et la sortie.

Depuis 2012, la pratique de la simulation dans la formation en anesthésie s'est largement développée tant pour la formation initiale que pour la formation professionnelle continue des soignants. Au SiMU de Nantes, la principale modalité de simulation est la simulation haute fidélité interprofessionnelle. Cette modalité de simulation s'appuie sur des scénarios de simulation, basés sur l'expertise des soignants, et sur un mannequin haute fidélité contrôlé par les formateurs.

On peut toutefois conjecturer que la pratique de la simulation en anesthésie gagnerait à s'appuyer sur les données de santé qui sont récoltées pendant les surveillances de patients. En effet, ces données de santé peuvent permettre la construction automatique de scénarios d'apprentissage qui augmenteraient considérablement la bibliothèque existante de scénarios basés sur l'expertise. De plus ces données peuvent permettre de gérer automatiquement l'évolution des paramètres physiologiques d'un patient virtuel et donc d'apporter un supplément de réalisme et de réactivité par rapport au contrôle manuel des mannequins de simulation.

PROFILS ANESTHÉSQUES ET DONNÉES DE SANTÉ

Introduction

L'anesthésie est une pratique médicale légalement encadrée dans son déroulement avant, pendant, et après une intervention. Ce cadre légal impose de plus une surveillance continue des patients pendant tout ce processus.

Un logiciel dédié, Pégase, a été mis en service, à partir de 2004, au CHU de Nantes pour suivre toute la procédure anesthésique et enregistrer les événements ainsi que les variations des paramètres physiologiques du patient lors d'une intervention. En 2016, Pégase avait enregistré plus d'1.1 million d'interventions. Dans ce chapitre, nous nous intéresserons aux données de santé enregistrées lors de ces interventions et contenues dans les profils anesthésiques, à leur utilisation et à leur cadre légal.

Dans la première partie de ce chapitre, nous nous intéresserons spécifiquement à qualifier les données présentes dans les profils anesthésiques et enregistrées lors de la surveillance d'une intervention.

La seconde partie s'intéressera plus spécifiquement aux traitements réalisés sur les données issues du soin pour les transformer en données exploitables par la recherche scientifique, et plus spécifiquement pour les présents travaux de thèse.

La dernière partie s'intéressera plus particulièrement au cadre légal des données de santé et à leurs utilisabilités dans le cadre de travaux de recherche.

2.1 Les données des profils anesthésiques

Le décret de 1994, qui encadre la pratique de l'anesthésie, impose une surveillance continue des patients. Cette surveillance est réalisée par l'utilisation d'une feuille d'anesthésie qui consigne les principaux événements anesthésiques survenus pendant l'intervention ainsi que l'évolution des paramètres physiologiques du patient. Dans le cas d'une anesthésie générale, ce monitoring commence dès l'entrée du patient dans la salle d'opération, et se poursuit jusqu'à son réveil.

La généralisation de l'outil informatique au début du XXI^e siècle a encouragé le CHU de Nantes à devenir précurseur et à expérimenter la numérisation de la surveillance anesthésique et des consultations pré-anesthésiques dès 2000, pour ensuite systématiser cette pratique en 2004.

2.1.1 Un outil numérique central : Pégase

Un outil numérique spécifique a été construit en interne au CHU de Nantes pour la récolte, le traitement, la consultation et l'archivage des données anesthésiques : le logiciel Pégase.

Pégase est un outil au cœur de la procédure anesthésique. Il permet de centraliser l'ensemble des données du process de suivi anesthésique, de la prescription de l'intervention à sa réalisation.

Pégase permet la consultation des dossiers médicaux des patients et la récolte des informations lors de la consultation pré-anesthésique préalable à toute intervention programmée. Cet outil sert ainsi d'interface pour les médecins anesthésistes et permet de consigner les différents éléments impactant le déroulement de l'anesthésie.

C'est aussi à travers cet outil qu'est monitorée l'intervention prévue et que les événements anesthésiques, tout comme les paramètres physiologiques du patient, sont enregistrés. Ces éléments sont saisis par les soignants ou enregistrés directement à partir des appareils de surveillance du bloc opératoire.

En cas d'incident pendant l'intervention, ou en cas d'effets secondaires de l'anesthésie, le patient peut être amené à être reçu en consultation après l'intervention. Lors de cette consultation, les éventuels effets indésirables de la procédure sont consignés et si nécessaire, le médecin anesthésiste peut prescrire des examens supplémentaires.

2.1.2 Les dossiers médicaux des patients

Le dossier médical d'un patient est une notion globale qui correspond à l'ensemble des informations sur la santé et les soins d'un patient. Il n'y a pas de dossier médical unique, mais un par praticien ou centre de santé. Les dossiers médicaux consultables via Pégase intègrent uniquement les éléments relatifs à l'anesthésie et aux interventions pratiquées sur le patient. Ils sont synchronisés régulièrement avec le système d'information de l'hôpital.

On y retrouve les éléments d'état civil et d'adresse, les éventuels antécédents médicaux et les différents éléments de l'examen clinique s'ils ont déjà été renseignés lors d'une précédente consultation de pré-anesthésie. Les composants d'état civil et les antécédents médicaux d'un patient sont présentés en annexe A.1 et A.2.

À ces éléments s'ajoutent ceux renseignés, ou mis à jour, pendant la consultation pré-anesthésique. Le médecin anesthésiste renseigne pendant l'examen clinique 23 variables (par exemple, la présence de pathologie ou d'appareillage dentaire) et 30 facteurs de risque. Sont aussi renseignés les valeurs normales de la pression artérielle, de la fréquence cardiaque du patient ainsi que les éléments relatifs à une transfusion (groupe sanguin, précédente transfusion, etc.). Les éléments récoltés pendant la consultation sont présentés en annexe A.3.

2.1.3 Profils anesthésiques

Tout au long de l'intervention sont saisis par le médecin anesthésiste, ou par les IADE¹, les différents événements et leurs horodatages. Sont aussi enregistrées, de manière automatique, plusieurs variables concernant les paramètres physiologiques du patient ou de l'intervention. À ce titre, Pégase peut enregistrer 40 variables différentes selon l'intervention, sous la forme de séries temporelles. Après l'intervention, la séquence des événements anesthésiques et les séries temporelles de surveillance sont ajoutées au dossier médical numérique du patient.

Les données relatives à une intervention, les informations sur le patient issues de son dossier médical et la surveillance pendant l'intervention, forment un *profil anesthésique*. Un profil anesthésique est ainsi spécifique à un patient et à une intervention. La suite de cette section présente les différents aspects des données qui composent ces profils anesthésiques.

1. Infirmiers Anesthésistes Diplômés d'État. Les IADE sont des infirmiers ayant suivi une formation supplémentaire de grade master.

2.1.3.1 Caractéristiques d’une intervention

Une intervention est caractérisée avant tout par la procédure chirurgicale ou médicale réalisée. Celle-ci est précisée principalement par son intitulé et un code unique (tous deux normalisés par la *Classification Commune des Actes Médicaux* ou CCAM) ainsi que par la date et les heures de début et fin de l’intervention. L’ensemble des variables caractérisant une intervention dans un profil anesthésique est présenté en annexe A.5.

Cette intervention implique une anesthésie qui est elle aussi décrite par des informations spécifiques concernant son déroulement et ses caractéristiques :

- les horodatages et durées des grandes étapes anesthésiques (entrée en salle, induction, etc.),
- les horodatages et durées de certains gestes ayant un impact hémodynamique²,
- le parcours de soin avec le service d’origine et le service de destination,
- les liquides entrant et sortant du patient et leurs volumes, comme les perfusions, pertes et transfusions,
- les informations sur l’intubation, le matériel utilisé et la technique mise en oeuvre.

Un profil anesthésique comprend ainsi 48 variables qui décrivent spécifiquement l’intervention et l’anesthésie réalisée et leurs contextes.

2.1.3.2 Événements anesthésiques

En dehors des actions de soin, le rôle des soignants en anesthésie pendant la procédure est de consigner les différents événements anesthésiques. Ces différents événements sont enregistrés en temps continu et ils sont donc associés à un horodateur. Ils sont de plusieurs types dont les deux principaux sont les gestes médicaux et l’administration des médicaments. Un exemple de séquence d’événements est donné en table 2.1. Une séquence complète est fournie en annexe A.2.

Temps (s)	ID	Événement	Temps (s)	ID	Événement
0	AGE-insta01	Mise en place du patient	2010	AGE-auspu01	Auscultation pulmonaire
30	AGE-mofca01	Monitoring FC	2040	AGE-supoa01	Surveillance points d’appui
60	AGE-mopam01	Monitoring PAM	2190	ADG-sevoc01	SEVOFLURANE continu
90	AGE-mospo01	Monitoring SPO2	2220	ADG-sufec01	SUFENTANIL continu
120	AGE-mobis01	Monitoring BIS	2280	AGE-insci01	Incision
150	AGE-povvp01	Pose VVP (cathlon)	2760	AGE-troca01	Trocard sous-ombilical
390	AGE-baihu01	Bair hugger	2820	AGE-inpne01	Inflation pneumopéritoine
450	ADG-cefab01	CEFAZOLINE bolus	3270	ADG-atrac01	ATRACURIUM continu

Table 2.1 – Exemple de séquence d’événements pour l’étape de l’entrée de la cure d’une hernie inguinale sous coelioscopie pour un homme de 30 ans sans antécédents médicaux.

2. Par exemple les clampages ou les garrots

Les événements de type *geste médical* rendent compte de la réalisation d'un geste médical spécifique de l'anesthésie sur le patient. Les soignants utilisent pour saisir les événements des listes exhaustives d'actions proposées par Pégase, et plus rarement si nécessaire des champs de texte libre.

L'injection de médicaments au patient est elle aussi consignée. La saisie est réalisée à partir de listes standardisées en y précisant le médicament et le dosage administré ainsi que la méthode d'administration.

Ces différents événements ont un impact direct sur le patient et donc sur ses paramètres physiologiques. Ceux-ci sont enregistrés sous la forme de séries temporelles.

2.1.3.3 Séries temporelles

Les différents paramètres physiologiques du patient et de l'intervention qui sont enregistrés, systématiquement ou non, par Pégase sous la forme de séries temporelles sont présentés ci-après par catégorie. Des séries temporelles sont présentées en annexe A.4. Selon le paramètre, la fréquence d'échantillonnage peut être différente en fonction de la méthode d'enregistrement. La table présentée en annexe A.3 présente une revue de l'ensemble des paramètres, les unités dans lesquelles ils sont exprimés, la fréquence d'enregistrement et le caractère systématique ou non de leur enregistrement.

Paramètres hémodynamiques Les paramètres hémodynamiques sont ceux relatifs à la circulation du sang. Ils sont des indicateurs directs du fonctionnement du système circulatoire. Ils comprennent :

- la fréquence cardiaque (FC), en battements par minute,
- les 2 composantes de la pression artérielle : la pression artérielle systolique³ (PAS) et la pression artérielle diastolique⁴ (PAD), ce qui permet d'estimer la pression artérielle moyenne (PAM)⁵,
- de manière occasionnelle, la pression artérielle pulmonaire (PAP) et ses déclinaisons (systolique, diastolique et moyenne) ou la pression veineuse centrale (PVC).

3. Valeur de la pression dans l'artère au moment où le cœur se contracte.

4. Valeur de la pression dans l'artère au moment où le cœur est au repos.

5. Les appareils de mesure automatique procèdent différemment, la PAS et la PAM sont mesurées et la PAD est estimée par un algorithme.

Paramètres respiratoires Les paramètres respiratoires sont relatifs à la ventilation du patient et reflètent le fonctionnement de l'appareil respiratoire du patient, en interaction avec le respirateur, ainsi que la capacité du patient à capter l'oxygène. On y retrouve les paramètres suivants :

- la saturation en oxygène du sang, mesurée sur le pouls (SPO₂),
- les fractions de gaz dans l'air inspiré (Fi) et expiré (Fe), notamment pour l'oxygène (O₂) et le dioxyde de carbone (CO₂), en pourcentages,
- la fréquence respiratoire (FR), en cycles minute, et le volume respiratoire moyen (VM), en millilitres,
- le débit de gaz frais (O₂),
- des paramètres occasionnels comme la saturation veineuse en oxygène (SVO₂) ou la pression respiratoire (PM).

Paramètres anesthésiques Les paramètres anesthésiques sont directement liés au déroulement de l'anesthésie. On y retrouve principalement des gaz hypnotiques halogénés et des indicateurs. On retrouve notamment :

- la température (T), en degrés Celsius,
- les fractions dans l'air inspiré et expiré en gaz hypnotiques (FiHal/FeHal) ou en protoxyde d'azote (N₂O), quand de tels gaz sont utilisés,
- l'indice bi-spectral (BIS) et son niveau de qualité,
- le niveau de curarisation du patient (en %) et les variables liées, comme le nombre de réponses à un test TOF.

Paramètres spécifiques Pour certaines interventions spécifiques, des paramètres particuliers peuvent être monitorés, par exemple la composition du sang (hématocrite⁶, hémoglobine, glycémie) ou encore certains paramètres du rythme cardiaque (segment ST⁷).

6. Mesure du volume de sang occupée par les cellules (globules rouges, globules blancs et plaquettes).

7. Partie spécifique d'un battement cardiaque sur un électrocardiogramme. Le segment ST donne une information sur la santé du muscle cardiaque

2.2 La transformation et la sélection des données

Pégase est un outil conçu et destiné à l'enregistrement médico-légal des anesthésies et au suivi anesthésique des patients. Il s'insère dans les activités quotidiennes de l'hôpital et sa finalité est donc avant tout le soin.

L'exploitation des données de santé était une pratique inexistante en dehors de la recherche au moment de la mise en service de cet outil en 2004. Une telle exploitation n'a donc pas été envisagée et prise en compte lors de la conception de Pégase.

La conception de Pégase, centrée sur le soin, a impacté la construction de ses différents composants et les solutions techniques qui ont été choisies pour sa réalisation. Les données sont ainsi structurées d'une manière spécifique pour le fonctionnement de l'outil, et donc pas sous la forme de profils anesthésiques directement utilisables. De plus, les choix de construction et des solutions techniques ont été réalisés avec les technologies disponibles lors de la mise en place de l'outil. Celles-ci n'ont pas évolué de manière synchrone avec les nouvelles technologies des systèmes d'information.

Pégase ne permet donc pas directement l'accès et le traitement de ses données à des fins scientifiques. Il a donc été nécessaire de réaliser une extraction des données à partir de la base de données relationnelle de Pégase. L'infrastructure fonctionnelle de Pégase étant externalisée chez un prestataire de service, l'extraction qui a été réalisée a été livrée sous la forme de plusieurs lots comme suit :

- lot 1, pour les données de santé de l'année 2016,
- lot 2, pour les données de santé des autres années,
- lot 3, pour les données relatives au fonctionnement de Pégase.

Ces lots sont composés de tables extraites d'une base de données, sans mention sur le type de variables, la structure relationnelle et sans documentation.

Une rapide consultation des données brutes a de plus permis de constater de nombreuses hétérogénéités dans les formatages, des redondances entre les lots extraits et des artefacts de dysfonctionnement de l'outil (ou de l'extraction), sous la forme d'éléments de profils anesthésiques manquants.

Pour palier ces défauts, des traitements très importants sur les données se sont avérés nécessaires : d'abord, des transformations structurelles pour obtenir une base compatible avec un usage scientifique, puis une sélection des données pertinentes pour extraire des cohortes de patients et leurs profils anesthésiques.

2.2.1 Les transformations structurelles

Les données brutes sont composées de 170 fichiers csv de très grande taille (la table *Patients* contient 1 942 368 enregistrements), qui représentent les 121 différentes tables de la base de données relationnelle de Pégase.

L’infrastructure numérique mise à disposition par le CHU de Nantes pour la consultation et le traitement de données s’est limitée à une machine dédiée à la bureautique et donc avec une puissance de calcul très restreinte. La fusion des différents lots a été réalisée à l’aide de scripts Python 3.10 et a été particulièrement fastidieuse en s’étirant sur plusieurs semaines. À partir des données rassemblées, une base de données a été reconstruite *via* un moteur de base de données léger (SQLite), afin d’être compatible avec les contraintes techniques de l’environnement numérique aux capacités limitées mis à disposition.

Cette base de données est un outil technique pour le bon fonctionnement de Pégase et non une base de données de santé. Les données relatives à l’anesthésie sont ainsi noyées dans des données techniques, des thésaurus et des données relatives à la réanimation ou à d’autres types de soins qui ne font pas l’objet des présents travaux.

Un travail d’identification et de sélection des données anesthésiques au sein de la base a donc dû être réalisé à la main parmi les 121 tables. Nous avons ainsi identifié 16 tables directement liées à l’anesthésie et pertinentes pour nos travaux.

2.2.2 Sélection des données anesthésiques

L’absence d’une documentation nous a contraints à deviner le rôle et l’usage de chaque table d’après les données qu’elle contiennent. Cette étude de la base nous a mobilisé plus d’une semaine.

Les différentes tables identifiées comme relatives à l’anesthésie sont décrites dans la table 2.2. Pour chacune de ces tables, nous avons associé les données identifiées à l’un des composants d’un profil anesthésique : le dossier médical, ou les événements anesthésiques ou encore les séries temporelles.

Ce travail d’identification réalisé, il nous a alors été possible de reconstituer les profils anesthésiques. Toutefois, l’absence de puissance de calcul, les performances limitées du moteur de base de données et l’importante quantité d’informations à traiter nous ont contraints à abandonner toute approche horizontale, qui aurait permis un traitement complet de la base. Nous avons donc eu une approche verticale en nous focalisant sur une cohorte précise de patients ayant subi la même intervention.

Nom	Enregistrements	Description	Composant
ALR	196 164	Détails des anesthésies loco-régionales	Dossier médical
ANESTHESIE	650 369	Détails des procédures anesthésiques	Dossier médical
COMPLICATIONS	2 702	Éventuelles complications et événements indésirables lors d'une intervention	Dossier médical et événements
CONSULTATION	619 723	Informations récoltées pendant la consultation pré-anesthésique	Dossier médical
CONSULTATION_PA	1292	Informations récoltées pendant l'éventuelle consultation post-anesthésique	Dossier médical
EVENEMENTS	10 748 717	Événements anesthésiques pendant une intervention	Événements
INTERVENTION	1 114 643	Caractéristiques d'une intervention	Dossier médical
INTUBATION_DIF	26 121	Caractéristiques liées à une intubation difficile identifiée lors de la consultation	Dossier médical
INTUBATION_TECH	35 741	Historique des intubations réalisées selon la technique et leurs succès	Dossier médical
MATERIEL	2 542 159	Matériel utilisé	Événements
MEDICAMENTS	5 831 422	Médicaments utilisés et dosages	Événements
PERFUSIONS	1 341 942	Perfusions utilisées	Événements
PERTES	461 188	Éventuelles pertes pendant une intervention	Événements
PATIENTS	1 942 368	État civil des patients	Dossier médical
SURVEILLANCE	47 680 997	Surveillance des paramètres physiologiques des patients pendant une intervention	Séries temporelles
TRAITEMENTS	927 903	Traitements des patients, passés ou en cours, au moment de l'intervention	Dossier médical

Table 2.2 – Tables anesthésiques présentes dans la base de données de Pégase. Sont précisés le nombre d'enregistrements, une description des données contenues et le rattachement à l'un des éléments d'un profil anesthésique

2.2.3 Les extractions de cohortes de patients

Nous avons cherché à identifier dans les tables présentées en 2.2 toutes les données relatives à une intervention pour réaliser une nouvelle base de données spécifique. Ainsi, cette sous-base relative à une intervention contient un nombre moins important de données et permet d'envisager les traitements de déduplication et d'extraction des profils anesthésiques en un temps humainement acceptable vis-à-vis de la puissance de calcul mise à disposition. Une première déduplication a ainsi été réalisée sur cette sous-base afin d'identifier et d'éliminer les patients et interventions en double.

Les différents composants des profils anesthésiques sont répartis dans différentes tables de la base de données, comme le présente la table 2.2. Pour extraire les profils anesthésiques, il est nécessaire d'associer les différentes données dispersées relatives à ces trois composants. Cette extraction est réalisée pour différents types de patients spécifiquement selon leur classe d'âge, sexe et antécédents médicaux. L'annexe A présente un exemple d'un profil anesthésique tel qu'il est utilisé dans le cadre de nos travaux. Lors de cette extraction, une seconde déduplication a été réalisée pour retirer les redondances au sein des profils anesthésiques. Une vérification de la complétude du profil est aussi réalisée et les profils incomplets sont exclus de la cohorte finale de patients.

Si un travail de sélection, de nettoyage et de formatage est inhérent à tout projet de *data science*, des contraintes techniques aussi importantes, comme la qualité très hétérogène des données brutes, n'étaient pas prévisibles et ont impacté lourdement le déroulement des travaux de cette thèse. Ainsi, les traitements réalisés sur ces données se sont focalisés sur leur utilisabilité dans le contexte de ces travaux et n'ont pas été réalisés de manière horizontale sur l'ensemble de données, mais seulement sur deux interventions.

Nos travaux se sont donc concentrés principalement sur deux interventions de durée moyenne et sur un profil de patient qui n'est pas générateur de complications : la cure d'une hernie inguinale⁸ sous coelioscopie⁹, avec pose de prothèse¹⁰, et le remplacement de l'articulation du genou par prothèse tricompartimentaire¹¹, pour un homme d'environ 30 ans sans antécédents médicaux.

En dehors des contraintes propres à la structuration et à la qualité des données, celles-ci ne sont pas des données anodines. Ces données étant issues du soin, et donc relatives à des patients, elles entrent dans le cadre des données personnelles et sont soumises à un cadre réglementaire spécifique.

8. Une hernie inguinale est une grosseur présente au niveau de l'aîne qui apparaît quand une portion de l'intestin ressort entre les muscles abdominaux, généralement en raison d'un effort inhabituel.

9. La coelioscopie est une technique chirurgicale qui consiste à gonfler l'abdomen d'un patient avec du CO₂ puis à utiliser des appareils vidéos, des endoscopes, et des outils chirurgicaux spécifiques pour réaliser l'intervention sans réaliser d'incision importante.

10. La prothèse permet de consolider l'intervention et d'éviter une récurrence.

11. L'articulation du genou est composée de 3 éléments : la tête du fémur, la rotule et la tête du tibia. Une prothèse tricompartimentaire, ou prothèse totale, vient remplacer ces trois éléments.

2.3 Le cadre légal des données de Santé et leur utilisabilité

La mise en place progressive à la fin du XX^e siècle des traitements informatiques dans les différents aspects de la société a motivé le législateur français à donner un cadre légal à la protection des données à caractère personnel via la *loi relative à l'informatique, aux fichiers et aux libertés* du 6 janvier 1978 [51]. Cette loi constitue par la même occasion une autorité indépendante : la Commission nationale de l'informatique et des libertés (CNIL).

Ce cadre réglementaire a évolué en cohérence avec le cadrage européen, d'abord en 2004 avec la *directive sur la protection des données personnelles* [52], puis en 2018 avec l'application du Règlement Général sur la Protection des Données (RGPD) [53]. Le RGPD est aujourd'hui le cadre de référence de la protection des données à caractère personnel dans les traitements informatiques mis en œuvre sur le territoire français. La CNIL est l'autorité en charge de son application en France.

Il n'est pas possible d'être « propriétaires » de nos données personnelles. Ce principe a été juridiquement exclu et inscrit dans la loi informatique et liberté de 1978 [51]. Ainsi, les profils anesthésiques présents dans Pégase sont collectés et gérés par le CHU de Nantes qui en est le gardien pour les patients qu'ils concernent.

Les données qui composent les profils anesthésiques ne sont pas anodines et correspondent à une typologie particulière, celle des données à caractère personnel. Ces données sont ainsi encadrées par des réglementations et pratiques spécifiques. De plus, comme elles sont relatives à la santé d'un patient, ces données sont des données à caractère personnel *sensibles*. Conséquemment, leur récolte, stockage et utilisation sont régies par des spécificités réglementaires et des techniques spécifiques. Les autorités régulatrices, la CNIL en France et le *European Data Protection Board* (EDPB) en Europe, éditent à ces fins des référentiels de bonnes pratiques pour la gestion et l'exploitation de ces données particulières.

2.3.1 Type de données des profils anesthésiques

Comme l'indique le RGPD dans son article premier, il «*protège les libertés et droits fondamentaux des personnes physiques, et en particulier leur droit à la protection des données à caractère personnel*». À ce titre, ce règlement donne plusieurs définitions concernant les types de données et les traitements qu'il est possible de leurs appliquer.

RGPD, Chapitre 1, Article 4, 1. Données à caractère personnel

Toute information se rapportant à une personne physique identifiée ou identifiable [...] directement ou indirectement, notamment par référence à un identifiant, tel qu'un nom, un numéro d'identification, des données de localisation, un identifiant en ligne, ou à un ou plusieurs éléments spécifiques propres à son identité physique, physiologique, génétique, psychique, économique, culturelle ou sociale.

Ainsi la notion de « données personnelles » est à comprendre de façon très large. Les profils anesthésiques qui sont l'objet de ces travaux se rapportant directement à un patient, une personne physique, ils entrent bien évidemment dans le cadre des « données personnelles ». Pourtant, ce ne sont pas des données personnelles comme les autres car elle sont relatives à la santé des patients. Pour ce type de données, le RGPD donne une définition spécifique :

RGPD, Chapitre 1, Article 4, 15. Données concernant la santé

Les données à caractère personnel relatives à la santé physique ou mentale d'une personne physique, y compris la prestation de services de soins de santé, qui révèlent des informations sur l'état de santé de cette personne.

Là encore, la notion de donnée de santé se considère de manière large. Sont comprises dans les données de santé les informations relatives aux soins mais aussi toutes informations qui permettraient de déduire l'état de santé d'un patient par recoupement avec d'autres informations. Une donnée qui ne permettrait pas de déduire une information sur l'état de santé d'un patient, avec recoupement ou non, n'est pas une donnée de santé.

2.3.2 Contraintes appliquées aux données de santé

Les données des profils anesthésiques sont particulièrement sensibles, et sont donc encadrées par un régime juridique particulier et elles sont notamment soumises au secret médical.

Il est donc impossible pour une personne autre que les soignants de consulter le dossier médical d'un patient sans avoir son consentement. Toutefois, il est possible d'utiliser les données des patients sans leur consentement à des fins de recherche scientifique dans un cadre formalisé et encadré par la CNIL [54]. Notre accès aux données est ainsi placé sous l'égide du cadre réglementaire et sécuritaire donné par le CHU.

Les données auxquelles nous avons accès ont été pseudonymisées, en conformité avec les recommandations du RGPD pour limiter les risques liés au traitement de données personnelles. La notion de pseudonymisation se différencie conceptuellement et techniquement de l'anonymisation.

Le RGPD donne une définition précise de la pseudonymisation :

RGPD, Chapitre 1, Article 4, 5. pseudonymisation

La «pseudonymisation» est le traitement de données à caractère personnel de telle façon que celles-ci ne puissent plus être attribuées à une personne concernée précise sans avoir recours à des informations supplémentaires, pour autant que ces informations supplémentaires soient conservées séparément et soumises à des mesures techniques et organisationnelles afin de garantir que les données à caractère personnel ne sont pas attribuées à une personne physique identifiée ou identifiable ;

En pratique, la pseudonymisation consiste à remplacer les données directement identifiantes, comme les noms et dates de naissance, par des données indirectement identifiantes comme un numéro unique ou une troncature de la donnée d'origine. Ainsi, la pseudonymisation réduit le risque d'identification directe d'un patient, mais elle ne fait pas disparaître le caractère identifiant des données car un recroisement avec d'autres informations reste possible.

A contrario, l'anonymisation est une technique qui consiste à supprimer de manière *irréversible* tout caractère identifiant d'un ensemble de données, la charge de la preuve de cette irréversibilité étant au responsable des données anonymisées. L'anonymisation de données complexes comme les profils anesthésiques est particulièrement difficile. Des éléments supplémentaires sur l'anonymisation de profils anesthésiques sont discutés dans le chapitre 6.

Ainsi, comme les données pseudonymisées des profils anesthésiques restent potentiellement identifiantes l'accès ne peut se faire que dans un espace physique dédié, au CHU, sur une infrastructure sécurisée à accès réglementé. Malheureusement, l'exploitation par l'intelligence artificielle des données de santé réelles est encore récente et donc l'infrastructure de consultation n'est pas adaptée d'un point de vue technique et logiciel.

De plus, le caractère transversal de nos travaux, avec la collaboration entre un laboratoire d'informatique et le CHU de Nantes est novateur et ne s'insérait pas dans les procédures de fonctionnement interne du CHU. Cette situation a eu pour conséquences de longues tractations et difficultés organisationnelles pour nous donner un accès à l'espace de consultation.

L'ensemble des différentes contraintes présentées dans cette section, tant réglementaires que techniques ou pratiques, a motivé un travail de modélisation de l'anesthésie pour produire un générateur de profils anesthésiques réalistes présenté dans le Chapitre 3.

Conclusion

Les interventions impliquant une anesthésie utilisent, pour leur préparation et bon déroulement, des données spécifiques, présentes dans les dossiers médicaux des patients qui font l'objet de l'intervention. Une anesthésie est aussi une procédure génératrice de données sur les caractéristiques de l'intervention, et par le monitoring de l'ensemble de la procédure.

Le dossier médical d'un patient, le déroulement de l'intervention et les séries temporelles qui rendent compte de l'évolution des paramètres physiologiques du patient, forment un *profil anesthésique* qui est archivé dans une base utilisée pour le soin et l'archivage médico-légal.

Ces données sont des données personnelles sensibles, car liées à la santé des patients. Elles sont soumises à un cadre réglementaire spécifique qui contraint leurs exploitations.

L'exploitation des données de santé est une pratique récente qui n'a pas été prévue dans le développement de Pégase, l'outil de récolte de ces données. Cet outil étant ancien, un important travail est nécessaire avant toute tentative d'exploitation. De plus, ces exploitations sont des pratiques émergentes dans les CHU, ainsi la collaboration interdisciplinaire entre différents établissements est balbutiante, ce qui amène son lot supplémentaire de contraintes techniques et pratiques.

Les difficultés rencontrées pour l'exploitation des profils anesthésiques réels rendent alors pertinente la création d'un générateur de données de synthèse.

GÉNÉRATEUR DE PROFILS ANESTHÉSQUES RÉALISTES

Introduction

L'exploitation de données de santé réelles est très difficile en raison des contraintes qui découlent de leur cadre réglementaire particulier. De plus, cette pratique relativement récente est balbutiante dans les centres de santé qui n'ont pas forcément l'infrastructure adaptée.

En s'appuyant sur l'expertise des anesthésistes, il est envisageable de produire un générateur de profils anesthésiques. Pour un type de patient et une intervention spécifique, construire un générateur de profils anesthésiques demande de produire une séquence d'événements et la série temporelle multivariée correspondante.

Si la comparaison de ces profils simulés par le générateur avec des profils réels ne montre pas de différences statistiques significatives, alors nous pourrions affirmer que ces profils sont réalistes.

La première partie de ce chapitre présente comment il est possible de formaliser informatiquement le déroulement des anesthésies et plus particulièrement la séquence d'événements anesthésiques au moyen d'une grammaire.

La seconde partie présente la construction d'une typologie d'événements, sa classification et la modélisation des différents impacts de ces événements sur le patient.

La dernière partie montre comment, à partir des deux composants précédents, la grammaire d'une anesthésie et la typologie des événements anesthésiques, il est possible de construire un générateur de profils anesthésiques qui sont statistiquement similaires aux profils réels.

3.1 Formalisation informatique d'une anesthésie

La formalisation médicale de l'anesthésie en quatre grandes étapes permet de donner un cadre à l'ensemble des anesthésies générales indépendamment de la procédure médicale, ou chirurgicale, réalisée pendant l'intervention.

3.1.1 Formalisation médicale classique

Les quatre grandes étapes d'une anesthésie correspondent à différents moments d'activité de l'intervention. C'est le niveau d'abstraction le plus élevé de la formalisation médicale d'une anesthésie. L'entrée et la sortie sont des temps de préparation à la continuité du parcours de soins. L'induction est un temps important, car il assure la transition entre deux états de conscience du patient et un passage sous le contrôle direct des soignants de certaines fonctions vitales. Enfin, la procédure est le cœur même de l'intervention ; c'est pendant cette période que la chirurgie, ou le soin, sont réalisés par un praticien spécialisé.

Ces quatre grandes étapes correspondent ainsi à des contextes médicaux différents, où les gestes et les médicaments utilisés sont spécifiques.

Pour chacune de ces grandes étapes, des étapes précises sont réalisées par les soignants dans un ordre déterminé et fixe. Ces différentes étapes sont présentées dans la figure 3.1.

Ces étapes correspondent à la préparation d'un soin ou à l'exécution d'un soin. Ainsi la *préoxygénation* correspond à un soin qui permet la recharge des réserves en oxygène du patient.

Sous étapes	Mise en place	Monitoring	Préparation
Entrée du patient	Mise en place du patient	Monitoring de la fréquence cardiaque	Monitoring de la pression artérielle
		Monitoring de la saturation en O ₂	Monitoring BIS-Spectral (BIS)
		Contrôle TOF	Bair hugger
		Installation de la voie veineuse périphérique	Propylaxie
			Remplissage vasculaire
			Pose du masque facial
			Préoxygénation
			Morphiniques
			Analgésie
			Hypnotiques
			Curare
			Ventilation mécanique
			fermeture des yeux
			Intubation
			Ventilation contrôlée
			Entretien de l'anesthésie
			Oxygène 30%
			Auscultation pulmonaire
			Vérification de la pression du ballonnet
			Manoeuvre de recrutement alvéolaire
			Monitoring de la température
			Mise en place pour la chirurgie
			Auscultation pulmonaire
			Contrôle des points de support
			Mise à disposition
			Incision
			Pose du trocart ombilical
			Inflation du pneumopéritoine
			Auscultation pulmonaire
			Manoeuvre de recrutement alvéolaire
			Dissection de l'espace préperitonéal
			Dissection du cordon spermatique
			Dissection de la zone herniaire
			Mise en place de la prothèse
			Déflation du pneumopéritoine
			Fermeture
			Pansement
			Contrôle de la décurarisation
			Décurarisation
			Mise en place du patient
			Sortie
Grandes étapes	Entrée	Induction	Procédure
			Sortie

Figure 3.1 – Étapes anesthésiques d'une intervention : la cure d'une hernie inguinale sous coelioscopie avec pose de prothèse.

Chacune de ces étapes, relative à un soin, est composée de sous-étapes qui correspondent à différents temps dédiés à des actions médicales. Elles sont réalisées par l'équipe médicale et varient selon l'intervention et le profil du patient. Ce sont ces actions, réalisées par l'équipe médicale, qui sont enregistrées comme événements pendant la surveillance de l'intervention.

La structure d'une anesthésie générale se caractérise ainsi par une séquence de grandes étapes, et, à l'intérieur de chaque grande étape, une séquence d'étapes. Ces deux types de séquences sont de contenu et d'ordonnancement fixes : ces deux niveaux de séquence sont communs à toutes les chirurgies pratiquées sous anesthésie générale. Une plus grande variabilité est présente dans les *sous-étapes* et dans les actions médicales qui composent ces sous-étapes. La table 3.1 récapitule les 4 niveaux de structuration d'une anesthésie.

Niveau	Composition	Description
I	Fixe	Grandes étapes anesthésiques
II	Fixe	Étapes de l'intervention
III	Variable	Sous-étapes contenant les actions
IV	Ponctuelle	Actions multiples d'une sous-étape

Table 3.1 – Table des différents niveaux de structuration d'une anesthésie

3.1.2 Formalisation informatique par une grammaire formelle

Il est nécessaire de proposer une formalisation informatique qui puisse prendre en compte et satisfaire les différents éléments suivants :

- la conservation de la structure d'une anesthésie telle que présentée dans la formalisation médicale,
- la prise en compte de toutes les variations potentielles dans une intervention,
- un cadre formel homogène pour toutes les interventions,
- une lecture aisée par un non-spécialiste de l'informatique afin de permettre une validation par les experts.

Une grammaire formelle est un formalisme qui permet tous ces éléments et donc la formalisation informatique d'une anesthésie. Les 4 niveaux de structuration, présentés dans la formalisation médicale, se retrouvent alors dans la grammaire sous la forme de 4 niveaux d'imbrication de non-terminaux.

La grammaire d'une anesthésie est composée d'un alphabet d'événements anesthésiques (symboles terminaux), d'un ensemble de non-terminaux correspondant aux diffé-

rents niveaux de structuration, ainsi qu’aux étapes de ces niveaux, et d’un ensemble de règles de production dont nous donnons un exemple pour chacun des niveaux de la hiérarchie.

Au premier niveau de la hiérarchie, l’axiome de la grammaire se réécrit ainsi :

`<anesthésie> ::= <entrée> <induction> <procédure> <sortie>`

Chacune de ces grandes étapes contient des étapes qui sont elles aussi fixes et systématiques ; c’est le second niveau de structuration. Ainsi, pour le non-terminal du premier niveau `<entrée>`, on peut formuler la règle suivante :

`<entrée> ::= <mise_en_place> <monitorage> <prémedication>`

Chacune de ces étapes est elle-même composée de sous-étapes, qui constituent le troisième niveau de structuration. La composition de ce niveau est spécifique à chaque type d’intervention. Le non-terminal de niveau deux `<prémedication>` peut ainsi se réécrire comme suit :

`<prémedication> ::= <bair_hugger> <vvp> <prophylaxie> <rv>
|<vvp> <prophylaxie> <rv>
|<vvp> <prophylaxie>`

où `<vvp>` correspond à la pose de la voie veineuse périphérique et `<rv>` au remplissage vasculaire. Les différentes règles alternatives correspondent aux variations possibles dans une procédure chirurgicale.

Les sous-étapes du niveau trois correspondent généralement directement à une action médicale et donc à un événement anesthésique. Les événements anesthésiques correspondant à ces sous-étapes sont hautement contextualisés par spécification du matériel, des médicaments et doses utilisés. Plusieurs événements similaires sont possibles pour une même sous-étape. Ainsi, au non terminal de niveau trois `<prophylaxie>` correspond une action médicale qui administre au patient un antibiotique. Cette action peut être représentée par plusieurs terminaux :

`<prophylaxie> ::= Médicament_A | Médicament_B`

où `<Médicament_A>` et `<Médicament_B>` sont des événements anesthésiques correspondant à l'administration chez le patient de médicaments différents, mais à l'effet similaire.

Toutefois, dans certaines interventions, ce troisième niveau de structuration ne suffit pas pour tenir compte de la variabilité de la procédure. Il est alors ponctuellement précisé au moyen d'un quatrième niveau de structuration. Ainsi, le non-terminal `<rv>` peut être réécrit *via* les règles suivantes, pour prendre en compte le cas d'une transfusion :

$$\begin{aligned} \text{<rp>} &::= \text{<solution_ionique>} \\ &| \text{<solution_ionique> } \text{<sang>} \end{aligned}$$

En combinant les différents ingrédients et les principes présentés ci-dessus, il est possible de proposer une grammaire qui peut correspondre à toutes les anesthésies générales.

Les niveaux I et II de la grammaire sont présentés en annexe B.1 et B.2. Les niveaux de structuration III et IV sont spécifiques à une intervention ou à type de patient. Ces niveaux sont donnés pour un patient «normal»¹ et pour deux interventions, la *cure d'une hernie inguinale sous cœlioscopie avec pose de prothèse* et le *remplacement de l'articulation du genou par une prothèse totale*, en annexe B.3 et B.4 respectivement.

En dérivant les règles de production d'une de ces grammaires spécifiques à une intervention, il est possible de produire un mot qui est une séquence d'événements anesthésiques. En utilisant les événements spécifiques au type d'un patient qui sont compatibles avec ses antécédents, on peut produire une séquence d'événements spécifiques à ce type de patient et à cette intervention. Il faut toutefois disposer des différents événements correspondants à une action.

De plus, pour pouvoir produire un profil anesthésique complet, il reste nécessaire de produire la série temporelle multivariée qui représente l'évolution des paramètres physiologiques du patient qui subit la séquence d'événements.

1. Homme de 30 ans en bonne santé et sans aucun antécédent médical.

3.2 Création d’une typologie d’événements basée sur l’expertise des soignants

La grammaire proposée utilise comme terminaux les différents événements anesthésiques. Pour pouvoir dériver la grammaire et produire une séquence d’événements anesthésiques, il faut au préalable disposer d’une typologie de ces événements.

Des techniques de *data mining* peuvent être mobilisées pour déterminer cette typologie. Cependant, l’accès restreint aux données anesthésiques, l’absence de puissance de calcul, et les perturbations des centres de santé par le COVID nous ont imposé d’autres alternatives.

La réalisation d’une telle typologie peut aussi se faire directement à partir de l’expertise médicale. La construction de cette typologie a ainsi été réalisée en se basant sur l’expertise des soignants du CHU de Nantes.

3.2.1 Récolte des éléments d’expertise

Le déroulement d’une procédure anesthésique lors d’une intervention est certes légalement encadré par le décret de 1994, mais ce décret impose seulement des principes organisationnels. La pratique de l’anesthésie est régulée à partir des recommandations émises par la SFAR² sur la base d’études scientifiques ou cliniques dans les principes de la médecine fondée sur les faits (*Evidence-Based Medicine*³ [55]).

Néanmoins, la variabilité des usages et des pratiques est importante au sein des centres de santé, et même au sein des différents services qui pratiquent l’anesthésie.

L’extraction de l’expertise des soignants, et sa formalisation d’un point de vue informatique a été possible grâce à la collaboration avec des étudiants IADE⁴ et des médecins anesthésistes du service d’anesthésie et réanimation chirurgicale de l’hôpital femme-enfant-adolescent du CHU de Nantes.

IADE Dans le cadre de leur formation, les étudiants IADE doivent réaliser un stage au sein d’une équipe de recherche lors de leur seconde année de formation.

Ces étudiants ont été mobilisés en 2019 et 2022 pour participer à la création de la

2. Société française d’anesthésie et de réanimation.

3. Combinaison de l’expertise des soignants, des meilleures données cliniques et scientifiques disponibles dans la prise de décisions médicales

4. Infirmiers Anesthésistes Diplômés d’État

typologie des événements pour deux chirurgies : la *cure d'une hernie inguinale sous cœlioscopie avec pose de prothèse* (2019) et le *remplacement de l'articulation du genou par une prothèse totale* (2022).

Médecin anesthésiste La principale médecin anesthésiste collaboratrice de ces travaux a été le Pr Corinne LEJUS-BOURDEAU PU-PH⁵ HDR, anesthésiste au sein du service d'anesthésie et de réanimation chirurgicale de l'hôpital femme-enfant-adolescent du CHU de Nantes et responsable du SiMU. La collaboration avec le Pr LEJUS-BOURDEAU concernant la construction de la typologie des événements anesthésiques s'est principalement concentrée sur son expertise concernant les mécanismes d'actions de ces événements sur le patient, leurs impacts sur les paramètres physiologiques, et leurs interactions.

Récolte d'information et sources primaires La récolte des informations s'est essentiellement basée sur l'expertise des soignants sous la forme d'entretiens autour de sujets spécifiques ou de listes de questions.

Ce travail s'est aussi appuyé sur de la documentation issue des procédures internes du CHU et de la littérature technique, clinique ou scientifique de l'anesthésie. Notamment :

- les modes opératoires des procédures chirurgicales, documents internes au CHU de Nantes,
- les bonnes pratiques et recommandations éditées par la SFAR,
- les protocoles de poche MAPAR⁶ [13] édités par le département d'Anesthésie-Réanimation de l'hôpital de Bicêtre,
- le référentiel d'anesthésie-réanimation et de médecine péri-opératoire [56], édité par le CNEAR⁷ pour les ECN⁸,
- *Anesthesia*, ouvrage de référence sur l'anesthésie réalisé sous la direction du Pr Ronald Miller [57].

5. Professeur des Universités - Praticien Hospitalier

6. Mise Au Point en Anesthésie Réanimation

7. Collège National des Enseignants d'Anesthésie et de Réanimation

8. Épreuves Classantes Nationales, le concours de l'internat

3.2.2 Classification des événements

À partir de l'ensemble des informations récoltées auprès des soignants, la construction de la typologie a été réalisée au moyen d'une classification des différents types d'événements anesthésiques. Une modélisation des impacts de ces événements sur le patient a également été produite.

Type d'événements Parmi les événements anesthésiques enregistrés lors d'une intervention, on peut identifier plusieurs catégories selon l'impact sur le patient et les caractéristiques propres de ces événements. Une synthèse de ces catégories est présentée dans la table 3.2.

La surveillance d'une intervention cherche à donner une vision précise de son déroulement. En plus des actions de l'équipe médicale, des informations sur le déroulement de l'intervention sont aussi enregistrées. On y retrouve des réactions du patient qui ne font pas directement suite à une action médicale, ainsi que des jalons propres à l'intervention et au protocole opératoire qui est réalisé.

Cette différenciation forme les trois principales catégories d'événements anesthésiques :

- les *actions*, qui correspondent aux actions réalisées par l'équipe médicale,
- les *réactions*, qui comprennent les réactions du patient qui ne sont pas directement imputables à une action,
- les *descripteurs*, qui informent sur le déroulement de la procédure, mais n'ont pas d'impact sur le patient.

Parmi les *actions* réalisées par l'équipe médicale, on peut ajouter un niveau de différenciation en dissociant d'un côté les *gestes médicaux*, et de l'autre les *médicaments*. En effet, ces deux types d'événements, s'ils sont tous deux des interventions de l'équipe médicale, sont très différents en ce qui concerne leurs impacts sur le patient et les informations qui les caractérisent. Cela nous permet de formuler deux sous-catégories correspondantes : les *actions «geste médical»* et les *actions «médicament»*.

Codification Le traitement informatique de ces événements est facilité par l'utilisation de bases de données. L'affectation d'un identifiant unique à chaque événement est un élément facilitant les traitements informatisés.

Type	Sous-type	Code	Description
Action	Médicament	AGE	Médicament administré par l'équipe médicale
Action	Geste médical	ADG	Geste réalisé par l'équipe médicale
Réaction	-	REA	Réaction non provoquée du patient
Descripteur	-	ES	Informe du déroulement de l'intervention

Table 3.2 – Types et sous-types de la classification des événements anesthésiques réalisée d'après l'expertise des soignants

Les événements sont ainsi codés selon un identifiant composé de trois éléments :

[type] – [id] – [occurrence]

Le composant [Type] correspond au type d'événement, codé sur 3 caractères majuscules :

- ADG, pour les événements de type *action* «*médicament*» (Action-DruG),
- AGE, pour les événements de type *action* «*geste médical*» (Action-GEsture),
- DES, pour les événements de type *descripteur*,
- REA, pour les événements de type *réaction*.

L'élément [id] est spécifique à un événement, ainsi l'événement «*Pose de la voie veineuse périphérique*» a pour identifiant unique **povvp**.

Toutefois, selon les interventions et les typologies de patients, un même événement peut exercer des impacts sur deux patients suffisamment différents pour qu'on considère qu'il est pertinent de dissocier ces impacts. Le dernier composant de l'identifiant, [occurrence], permet cette différenciation en rendant possible le codage de plusieurs événements similaires.

Ainsi, l'identifiant **AGE-povvp-01** se réfère de manière unique à l'événement «*Pose de la voie veineuse périphérique*» sur un patient «normal»⁹, là où l'identifiant **AGE-povvp-02** se réfère spécifiquement à l'événement «*Pose de la voie veineuse périphérique*» sur un patient à *capital veineux pauvre*¹⁰.

Modélisation des impacts sur le patient Sur la base de connaissances expertes complétées par la consultation de la littérature, nous avons pu établir des règles sur la nature de l'impact des événements *action* et *réaction* sur un paramètre physiologique donné (croissance, décroissance, absence d'impact), ainsi que le délai, la durée et l'intensité de cet impact.

9. Homme de 30 ans en bonne santé et sans aucun antécédent médical

10. Le capital veineux correspond à la qualité du réseau veineux. Un capital veineux pauvre complique la pose de la voie veineuse périphérique ce qui rend le geste plus pénible pour le patient

Pour chaque paramètre physiologique du patient (voir 2.1.3.3) l'impact de l'événement est formulé selon 3 modalités de variations :

- un effet stationnaire, quand l'événement n'a pas d'effet sur cette variable,
- un effet à plateau, avec un effet positif ou négatif sur le paramètre, qui correspond à une phase d'apparition progressive de l'effet suivie d'une phase de plateau qui dure un certain temps, puis éventuellement une disparition de l'effet,
- un effet linéaire, croissant ou décroissant.

Les deux dernières modalités sont définies grâce à plusieurs paramètres. Le choix d'une modalité, et son paramétrage ont été faits en se basant sur l'expertise des soignants et sur la littérature. Les différents types de variations et leurs différents paramètres sont présentés dans la table 3.3 et sur la figure 3.2. L'ensemble des 61 événements modélisés est présenté dans l'annexe B.5

Modalité	Paramètres	Type	Unité	Description
Effet à plateau	Délai	Réel	s ou mn	Délai d'apparition de l'effet maximal sur le paramètre physiologique
	Coefficient	Réel	-	Coefficient de l'effet sur le paramètre physiologique
	Durée	Réel	s ou mn	Durée du plateau de maintien de l'effet maximal
	Estompage	Booléen	-	Si vrai, l'effet s'estompe, si faux, alors le plateau a une durée infinie
Effet linéaire	DélaiL	Réel	s ou mn	Délai d'apparition de l'effet maximal sur le paramètre physiologique
	CoefficientL	Réel	-	Coefficient linéaire de l'effet sur le paramètre
	Unité	Texte	-	Unité prise en compte dans le calcul linéaire (minute ou seconde)

Table 3.3 – Paramètres décrivant les différentes modalités d'impact des actions sur les paramètres physiologiques.

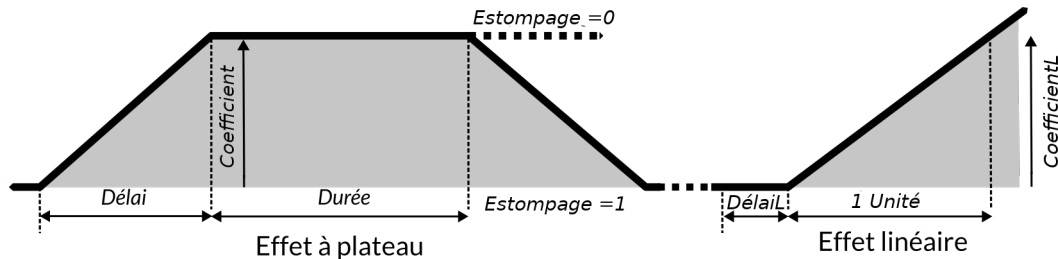


Figure 3.2 – Schéma des modalités de variations linéaire et à plateau d'un paramètre physiologique soumis à l'effet d'un événement

Spécificité des médicaments L'impact d'un médicament sur le patient est directement lié à sa dose et à sa méthode d'administration. Le contexte d'utilisation du médicament influence la dose et le mode d'administration, qui peuvent varier au sein de l'intervention. Un médicament peut être utilisé dans 3 contextes différents :

- pendant l'induction, ce qui implique des dosages et méthodes d'administration spécifiques,
- en entretien continu, où le médicament est administré de manière continue au patient (de la même manière qu'une perfusion),
- en entretien par bolus, où le médicament est administré au patient par doses successives.

Un médicament utilisé lors de l'induction et le même médicament utilisé lors de l'entretien, va potentiellement avoir des effets différents sur les paramètres physiologiques du patient.

Selon les médicaments et les recommandations d'usage, la dose est calculée selon 3 types de poids :

- le poids réel (P),
- le poids idéal (P_i) [58, 59],
- le poids ajusté (P_a) [60].

Les formules du poids idéal (P_i) et du poids ajusté (P_a) sont données ci-après :

$$P_i = P + 0,91x(T - 152,4)$$
$$P_a = P_i + C_a(P - P_i)$$

où $x = 45.5$ pour une femme, et $x = 50$ pour un homme, et C_a est un coefficient d'ajustement qui dépend des médicaments (généralement 0.3 ou 0.4).

Les événements de type action «médicament» ont ainsi besoin de paramètres supplémentaires en plus des paramètres des différentes modalités d'effet présentées plus haut :

- le type de poids utilisé pour le calcul du dosage (réel, idéal ou ajusté) et l'éventuel coefficient d'ajustement,
- la posologie¹¹ pour les 3 contextes d'utilisation (induction, entretien continu, entretien par bolus),
- le temps entre les réinjections dans le cas de l'entretien par bolus,
- les éventuelles prévalences¹² de risque allergique selon le sexe¹³.

11. Dosage et fréquence d'administration d'un médicament.

12. Nombre de cas d'une pathologie dans une population, autrement dit : la fréquence de cette pathologie.

13. Le risque allergique n'est généralement pas le même pour les hommes et les femmes.

Descripteurs Pour décrire le déroulement de la procédure, les événements descripteurs sont renseignés dans les feuilles d’anesthésie. Ainsi, l’entrée et la sortie du patient du bloc opératoire sont renseignées bien qu’ils n’aient aucun impact sur le patient. De même, après la période de l’induction, un événement est renseigné, la «*mise à disposition du patient*», pour signifier que le patient est prêt pour la procédure opératoire. Quand cette procédure commence, un événement spécifique est renseigné pour l’indiquer.

Ce travail d’analyse, particulièrement chronophage, s’est étalé sur plusieurs mois. Il a permis de couvrir l’ensemble des principaux événements anesthésiques non spécifiques à une chirurgie. Des événements spécifiques ont été modélisés pour deux chirurgies : *La cure d’une hernie inguinale sous cœlioscopie avec pose de prothèse* et *le remplacement de l’articulation du genou par une prothèse totale*. Une table des différents événements et de leurs caractéristiques est donnée en annexe B.5.

La grammaire d’une anesthésie et la typologie des événements sont les deux éléments qui permettent la génération de profils anesthésiques.

3.3 Génération de profils anesthésiques réalistes

Les deux briques élémentaires que sont la grammaire d’une intervention et la typologie des événements anesthésiques nous permettent de construire un générateur de profils anesthésiques. Ce générateur a été dénommé DBLBS¹⁴.

La construction de la typologie des événements a été particulièrement chronophage, tout particulièrement la modélisation des effets de chaque événement sur les paramètres physiologiques du patient.

La construction du générateur DBLBS a donc été réalisée en deux temps qui ont produit deux versions successives. Une première version s’est appuyée sur la grammaire construite pour la *cure d’une hernie inguinale sous cœlioscopie avec pose de prothèse* et la typologie des événements, pour simuler les quatre principales variables surveillées pendant une anesthésie : la fréquence cardiaque et les pressions artérielles systolique, diastolique et moyenne.

Cette version a ensuite été étendue aux 13 autres variables dont la surveillance est systématique pendant une intervention, une fois leur modélisation réalisée. Avec cette seconde version nous avons pu produire plusieurs jeux de données pour la *cure d’une hernie inguinale sous cœlioscopie avec pose de prothèse* et pour le *remplacement de l’articulation du genou par une prothèse totale*, dont les grammaires sont présentées en annexe B.3 et B.4 respectivement.

Les jeux générés sont composés de 1000 patients simulés, soit 1000 séquences d’événements et 1000 séries temporelles multivariées d’une longueur d’environ 200 pas de temps de 30 secondes.

3.3.1 Approche en deux étapes

DBLBS fonctionne selon une approche en deux étapes. La première étape consiste à produire une séquence horodatée d’événements à partir des caractéristiques d’un patient saisies par l’utilisateur, et de la grammaire d’une intervention telle que présentée en 3.1.2.

Pour produire les séries temporelles du profil anesthésique, la seconde étape s’appuie sur la séquence horodatée d’événements produite et la typologie des événements présentée en 3.2. Dans cet objectif, on simule un patient virtuel qui subit l’effet d’un événement de la séquence. Les algorithmes 1 et 2 présentent spécifiquement le fonctionnement des deux étapes de cette approche.

14. Du nom des différents collaborateurs ayant permis sa création : DAMA, BOISAUBERT, LEJUS-BOURDEAU, SINOQUET

DBLBS permet la production d'un profil anesthésique complet comportant la séquence d'événements et les séries temporelles des quatre variables physiologiques principales : la fréquence cardiaque et les pressions artérielles diastolique, systolique et moyenne. Le générateur simule aussi les séries temporelles de 13 autres variables physiologiques qui sont enregistrées systématiquement lors d'une intervention. Ces 17 variables sont présentées dans l'annexe A.3.

Algorithme 1 : Construction d'une séquence d'événements horodatée

Entrées :

- *Grammaire* : Grammaire d'une intervention
- *Événements* : Ensemble des événements anesthésiques
- *Patient* : Dossier médical d'un patient

Variables :

- *Instance_grammaire* : Grammaire instanciée sous la forme d'un graphe
- *Événements_compatibles* : Ensemble des événements compatibles avec les antécédents et le profil du patient
- *Noeud* : nœud courant en cours d'instanciation
- *Événement_sélectionné* : Événement correspondant à une action choisie parmi les symboles terminaux d'un nœud

Sorties :

- *Sequence* : Séquence d'événements horodatée

FONCTION Génération_sequence(*Grammaire*, *Événements*, *Patient*)

```

1  Instance_grammaire ← Instanciation_grammaire_selon_antécédents_patient(Patient, Grammaire);
2  Événements_compatibles ← Selection_événements_compatibles_patient(Patient, Événements) ;
3  Noeud ← Premier_noeud(Instance_grammaire);
4  tant que Existe(Noeud.fils) faire
5      Noeud ← Choisir_noeud_suivant(Instance_grammaire, Noeud);
6      Événement_sélectionné ← Choisir_événement(Noeud, Événements_compatibles);
7      Ajoute(Sequence.événements, Événement_sélectionné);
8      Ajoute(Sequence.horodatage, Durée(Événement_sélectionné));
9      Adapte(Instance_grammaire, Événement_sélectionné);
10 fin
10 retourner Sequence

```

Algorithme 2 : Construction d'une série temporelle multivariée**Entrées :**

- *Sequence* : Séquence d'événements anesthésiques horodatée d'une intervention
- *Typologie* : Typologie des événements anesthésiques
- *Patient* : Dossier médical d'un patient
- *Pas* : Pas de temps séparant deux points de mesure successifs de la série temporelle

Variables :

- *Horodatage* $\leftarrow 0$: Horodatage de la série temporelle en cours de construction
- *Événements* : Événements présents dans un intervalle entre deux pas de temps
- *Événement* : Un des événements présent dans un intervalle entre deux pas de temps
- *Effets_instanciés* : Effets des événements impactant les paramètres physiologiques simulés entre deux pas de temps
- *Effet* : Effet d'un des événements impactant les paramètres physiologiques simulés entre deux pas de temps
- *Valeurs* : Valeur des paramètres physiologiques simulés entre deux pas de temps

Sorties :

- *Séries_temporelles* : Une série temporelle multivariée correspondant à la séquence d'événements

FONCTION Génération_série_temporelle(*Sequence, Typologie, Patient, Pas*)

```

1 Ajoute(Séries_temporelles, Horodatage, Tirage_valeurs_initiales());
2 tant que Horodatage  $\leq$  Sequence.horodatage_final faire
3   | Événements  $\leftarrow$  Extraction_événements_dans_intervalle(Sequence, Horodatage, Pas);
4   | pour chaque Événement dans Événements faire
5     | si Événement dans Effets_instanciés alors
6       | si Événement.type = MEDICAMENT alors Modifie_effet(Effets_instanciés,
7         |   | Événement) ;
8         | sinon si Événement.type = DOULEUR alors
9           |   | Substitue_effet_douleur(Effets_instanciés, Événement) ;
10          | sinon
11            |   | Effets_instanciés  $\leftarrow$  Instancie_effets_événements(Événements);
12          | fin
13        | fin
14      | fin
15    | pour chaque Effet dans Effets_instanciés faire
16      |   | Valeurs  $\leftarrow$  Impacte_valeurs(Valeurs, Effet)
17      | fin
18    | Ajoute(Séries_temporelles, Horodatage, Valeurs);
19    | Horodatage  $\leftarrow$  Horodatage + Pas;
20    | Suppression_effets_terminés(Effets_instanciés, Horodatage)
21  fin
22 retourner Séries_temporelles

```

3.3.2 Génération d’une séquence d’événements

La grammaire d’une intervention contient toutes les règles de production qui permettent la représentation du déroulé d’une anesthésie. Le niveau le plus bas de cette grammaire contient des terminaux représentant les différents événements anesthésiques.

Nous commentons dans cette section l’algorithme 1. La première étape de la génération d’une séquence d’événements consiste en l’instanciation de la grammaire de l’intervention, en tenant compte du dossier médical du patient (ligne 1). La grammaire est instanciée sous la forme d’un graphe orienté correspondant au niveau le plus bas de la grammaire. La génération de la séquence d’événements est réalisée en se déplaçant de nœud en nœud dans ce graphe. Chaque nœud correspond à une action anesthésique et contient une liste de tous les événements propres à cette action. Les événements sont associés à une probabilité d’instanciation déterminée avec les anesthésistes. Ainsi, par exemple, pour une même action impliquant l’utilisation d’un type de médicament peut correspondre deux événements distincts pour deux médicaments différents. Le choix d’un événement ou d’un autre est réalisé selon la probabilité d’instanciation de chacun des événements.

Cependant, les éléments qui composent le dossier médical exercent un impact sur la grammaire instanciée. Les différents antécédents du patient qui pourraient avoir un impact sur le déroulement de l’intervention sont pris en compte par modification des choix des nœuds suivants, et donc des actions suivantes, et par modification des probabilités d’instanciation d’un nœud ou des événements. Ainsi, dans un contexte de bonne pratique, si le patient est allergique à un médicament, alors la probabilité d’instanciation d’un événement impliquant ce médicament est mise à zéro.

De plus, tous les événements présents dans les nœuds qui ne sont pas compatibles avec le type du patient sont retirés (ligne 2).

À partir de la grammaire instanciée, et altérée par le dossier médical du patient, il est possible de générer une séquence d’événements en dérivant ses règles de production. Pour dériver les règles de production, le générateur se place au début de l’instance de la grammaire (ligne 3), puis se déplace successivement de nœud en nœud.

Chaque nœud possède un nombre n_f ($1 \leq n < \infty$) de nœuds fils ; une probabilité est associée à chaque nœud et donc à chaque fils. Cette probabilité est déterminée selon les prévalences des actions dans la séquence d’événements représentés par les nœuds et a été déterminée à partir de l’expertise des anesthésistes. Pour déterminer le nœud suivant, on teste par un tirage aléatoire la probabilité de l’action représentée par le nœud et on se

déplace sur le nœud ayant reçu un test favorable (ligne 5).

On teste ensuite la probabilité d’instanciation des événements présents dans ce nœud et on sélectionne celui au test favorable (ligne 6). Si un événement est associé à un événement déclencheur, et que l’événement déclencheur n’est pas présent dans la séquence d’événements alors sa probabilité est au préalable mise à zéro lors de la sélection de l’événement à instancier.

L’événement choisi est ensuite ajouté à la séquence d’événements (ligne 7). À chaque événement correspond une durée typique qui est ajoutée à la séquence d’événements pour former l’horodatage du prochain événement (ligne 8).

Si l’événement se caractérise par son unicité, alors il ne peut plus être instancié une nouvelle fois. Après l’instanciation d’un événement, le générateur vérifie son unicité et adapte, éventuellement, les probabilités des événements similaires dans la grammaire pour les mettre à zéro (ligne 9). Le générateur se déplace ainsi de nœud en nœud dans le graphe jusqu’à produire une séquence complète d’événements.

3.3.3 Génération des séries temporelles

À partir de la séquence d’événements horodatée relative à une intervention, il est possible de produire des séries temporelles pour chacune des variables physiologiques du patient simulé. Pour cela, on applique sur les variables physiologiques du patient simulé les différents impacts des événements.

Nous commentons dans cette section l’algorithme 2. À chaque simulation, la valeur initiale des paramètres du patient virtuel sont tirées aléatoirement dans des intervalles de valeurs fournies par les experts (ligne 1). Ensuite, on itère sur chaque pas de temps, dont la durée est définie grâce à un paramètre (ligne 2). Si l’horodatage d’un événement est présent entre deux pas de temps (ligne 3) alors son effet est instancié (ligne 8) pour pouvoir impacter les paramètres.

Selon les effets des événements sur le patient virtuel, ceux-ci peuvent se chevaucher. Pour pouvoir simuler l’effet combiné de ces événements sur l’évolution des paramètres physiologique du patient virtuel, nous avons défini plusieurs règles avec l’aide des anesthésistes :

- les effets opposés se compensent en impactant le paramètre (ligne 10). Ainsi l’augmentation de la fréquence cardiaque due à une douleur est compensée par la diminution de la fréquence cardiaque due à un médicament.
- la durée de l’effet des médicaments est proportionnelle à la dose reçue. Si le patient

- reçoit deux doses successives, alors la durée de l’effet est doublé (ligne 6),
- les effets de douleurs mineures successives ne se cumulent pas et sont remplacés par l’effet d’une douleur plus forte (ligne 7).

À partir de ces différentes règles, les valeurs des paramètres physiologiques sont impactées selon les effets des événements (lignes 10) et sont ajoutées à la série temporelle (ligne 11).

Quand la durée de l’effet d’un événement est expirée, alors l’effet est supprimé des effets instanciés (ligne 13).

Le processus précédemment décrit est ainsi itéré sur les pas de temps successifs, en même temps qu’est prise en compte la séquence d’événements horodatées. L’impact des effets successivement appliqués au patient virtuel détermine l’évolution des paramètres physiologiques de ce dernier.

3.3.4 Validation des profils anesthésiques réalistes

DBLBS s’appuie sur la grammaire d’une intervention et sur un profil de patient pour générer les profils anesthésiques simulés pour un type de patient et une intervention. Pour déterminer le caractère *réaliste* des profils anesthésiques simulés, il faut donc les comparer vis-à-vis de leurs homologues réels.

Une première validation a été réalisée par les experts en anesthésie qui ont comparé une dizaine de profils anesthésiques produits par notre générateur avec des profils anesthésiques réels. L’évaluation qualitative qu’ils ont produite nous confirme le caractère réaliste et cohérent des profils anesthésiques produits.

La validation par les experts, bien que très fiable, ne permet pas une validation en grande dimension. Nous avons donc effectué une analyse de similarité et une analyse statistique pour comparer les profils anesthésiques réels et simulés.

Pour cette validation, nous nous sommes basés sur un patient «normal» et nous avons choisi comme intervention la cure d’une hernie inguinale sous coelioscopie avec pose de prothèse. Entre 2004 et 2019, 255 patients ont été opérés d’une hernie inguinale sous coelioscopie avec pose de prothèse. Sur ces 255 patients, 50 correspondent à la description du patient suivant : homme de 30 à 50 ans en bonne santé sans antécédents impactant l’intervention. Ces patients sont appelés ci-après cohorte de *référence*.

Comme nous l’indiquons plus haut, avec le générateur DBLBS nous avons produit

1000 profils anesthésiques correspondant à des patients virtuels de type similaire pour la même intervention. Nous avons sélectionné au hasard 50 patients simulés pour constituer une cohorte de taille identique. Ces patients sont désignés ci-après cohorte *simulée*.

3.3.4.1 Analyse de similarité

Pour un patient et une intervention donnés, les éléments de différenciation entre un profil anesthésique réel et simulé résident dans la séquence d'événements et la série temporelle multivariée. Ainsi, si nos profils anesthésiques simulés ne sont pas réalistes, alors les séquences d'événements et les séries temporelles multivariées issues des patients simulés seront sensiblement différentes de celles des patients de références .

En comparant, grâce à des mesures de similarité, la séquence d'événements et la série temporelle des différents profils anesthésiques, nous pouvons déterminer ce niveau de dissimilarité entre profils anesthésiques réel et simulé.

Pour déterminer la similarité Sim_{seq} entre deux séquences d'événements, on peut utiliser un score d'alignement des deux séquences horodatées selon l'approche de SMITH-WATERMANN [61] (voir chapitre 6) qui récompense les *matches* entre événements, ainsi plus ce score sera élevé plus les deux séquences comparées seront similaires. Un score de 0 signifie que les traces ne s'alignent pas. Les séquences comparées sont composées des événements anesthésiques, représentés par leur identifiant unique (voir 3.2.2), et de points pour les intervalles de temps entre ces événements. Ainsi la séquence d'événements S_E présentée ci-dessous, avec son horodatage S_H , permet de construire la séquence comparative S_C .

$$\begin{aligned} S_E &= \{A; B; C; D\} \\ S_H &= \{3; 4; 2\} \\ S_C &= \{A; ...; B;; C; ..; D\} \end{aligned}$$

On calcule ainsi, pour l'ensemble des séquences comparatives issues des profils anesthésiques des cohortes simulée et de référence, une distribution de scores. Une distribution des scores normalisés entre 0 et 1 a été calculée comme suit :

$$Norm_{seq} = \frac{Sim_{seq}}{max_score}$$

où max_score est le score de la plus longue séquence alignée avec elle-meme.

La distribution des scores normalisés est présentée en figure 3.3.

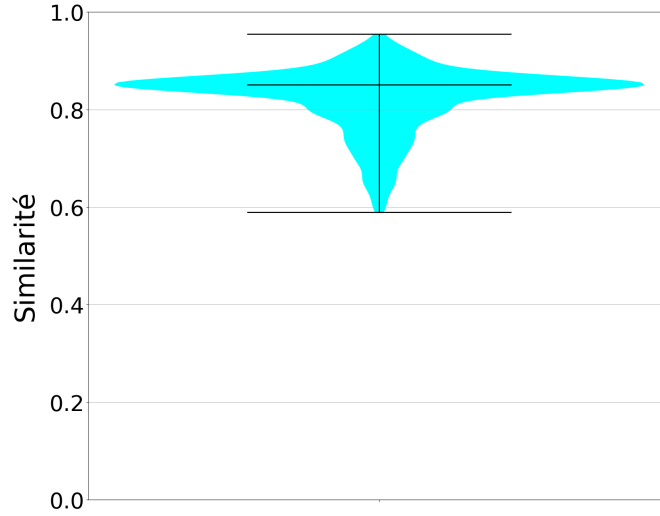


Figure 3.3 – Distribution des scores d’alignement normalisés entre les séquences d’événements simulées et réelles

Pour les séries temporelles multivariées, une mesure de dissimilarité Dis_{st} comme la DTW¹⁵ permet de comparer les séries temporelles univariées (voir Chapitre 4). Si cette mesure est calculée sur chacune des dimensions de la série temporelle multivariée, on peut ensuite réaliser un moyenne des dissimilarités univariées pour estimer la dissimilarité entre deux séries temporelles multivariées.

Toutefois, la série temporelle multivariée d’un profil anesthésique est directement déterminée par la séquence d’événements. Il nous faut pouvoir évaluer la dissimilarité entre les séries temporelles multivariées réelle et simulée, sans biaiser cette évaluation par la séquence d’événements. Pour cela, nous avons extrait les séquences d’événements des patients de référence, et à partir de ces séquences, nous avons simulé des séries temporelles multivariées en utilisant notre générateur. Ainsi, en comparant la série temporelle réelle et la série temporelle simulée, pour la même séquence d’événements, la même intervention et le même patient, nous pouvons déterminer si notre générateur produit des séries temporelles multivariées simulées similaires aux séries temporelles multivariées réelles.

Nous avons donc calculé, pour chacune des dimensions des séries temporelles multivariées simulée et de référence, une distribution de dissimilarités en utilisant la DTW. La distribution des valeurs de dissimilarité est présentée en figure 3.4 (a). Une distribution de similarités normalisées a été calculée comme suit :

$$Sim_{st} = \frac{Dis_{st}}{\max(Dis_{st})}$$

15. *Dynamic Time Warping*, voir Chapitre 4.

La figure 3.4 (b) présente cette distribution. La distribution des valeurs des médianes des dissimilarités est présentée en figure 3.5 (a), et la distribution des valeurs des médianes des similarités normalisées en figure 3.5 (b). Les tables 3.4 et 3.5 présentent les différentes caractéristiques de ces distributions.

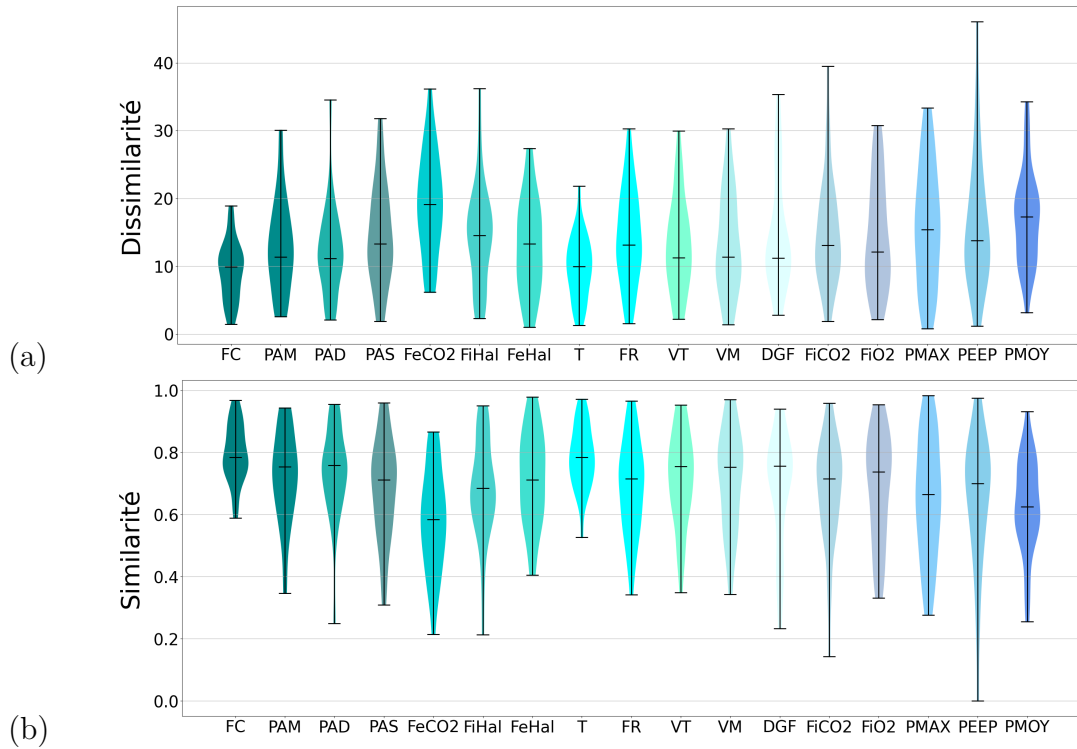


Figure 3.4 – Distribution des dissimilarités (a) et similarités normalisées (b) entre les 50 séries temporelles univariées simulées et les 50 séries temporelles réelles pour les 16 variables simulées. FC : Fréquence cardiaque, PAM : Pression artérielle moyenne, PAD : Pression artérielle diastolique, PAS : Pression artérielle systolique, FeCO₂ : Fraction expirée en CO₂, FiHal : Fraction inspirée en gaz halogéné, FeHal : Fraction expirée en gaz halogéné, T : Température, FR : Fréquence respiratoire, VT : Volume de ventilation, VM : Volume moyen, DGF : Débit de gaz frais, FiCO₂ : Fraction inspirée en CO₂, FiO₂ : Fraction inspirée en O₂, PMAX : Pression maximale, PEEP : Pression expiratoire positive, PMOY : Pression moyenne

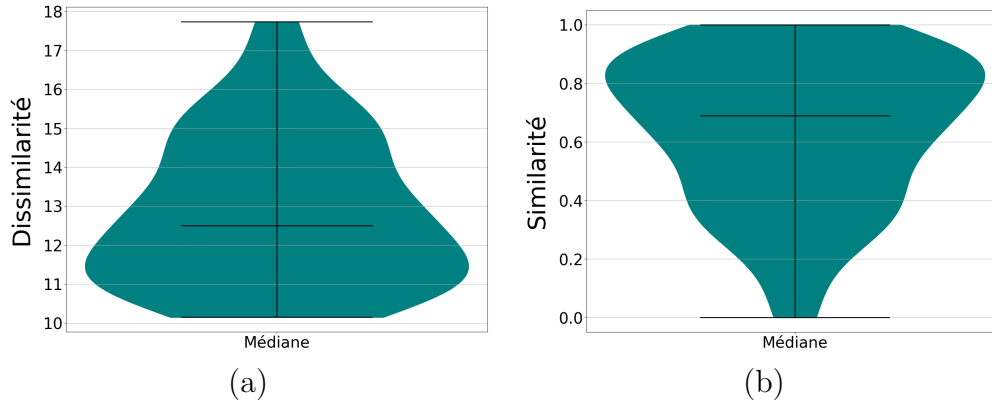


Figure 3.5 – Distribution des médianes des dissimilarités (a) et des médianes des similarités normalisées (b) entre les 50 séries temporelles multivariées simulées et réelles

Caractéristiques	Séquences d'événements		Séries temporelles	
	Sim_{seq}	$Norm_{seq}$	$Med(Dist)$	$Med(Sim_{st})$
Moyenne	145.04	0.82	12.87	0.64
Médiane	149.74	0.85	15.50	0.68
Écart type	11.52	0.065	1.91	0.25

Table 3.4 – Caractéristiques des distributions de scores d'alignement et scores normalisés entre séquences d'événements de référence et simulées, et des distributions de dissimilarités moyennes et similarités normalisées moyennes entre séries temporelles multivariées de référence et simulées

Comme on peut le constater sur la figure 3.3, les scores d'alignements normalisés entre les séquences d'événements simulées et réelles se concentrent autour d'une similarité médiane de 85% et qu'elles sont de plus assez peu dispersées (Écart-type de 0.065).

La figure 3.5 nous montrent que les différentes séries temporelles simulées sont similaires aux séries temporelles de référence avec lesquelles elles partagent leur séquence d'événements (Moyenne = 0.64, Médiane = 0.68). Une analyse plus fine des dissimilarités et similarités normalisées pour les différentes variables de la série temporelles univariée est présentée à la figure 3.4. On remarque ainsi que notre générateur produit des séries temporelles univariées simulées similaires (entre 70% et 80% de similarité) aux séries de référence pour une majorité des paramètres physiologique. Toutefois, cette similarité est légèrement moins bonne sur les variables respiratoires. La similarité médiane pour la *fraction inspirée en CO₂* est ainsi de 58%. La fraction inspirée en CO₂ est un paramètre physiologique qui rend directement compte du fonctionnement métabolique du patient et lui est très spécifique, c'est un paramètre pour lequel il est très difficile de déterminer une modélisation générique.

Caractéristiques	Séries temporelles univariées							
	FC		PAM		PAD		PAS	
	$Dis_{st}()$	$Norm_{st}()$	$Dis_{st}()$	$Norm_{st}()$	$Dis_{st}()$	$Norm_{st}()$	$Dis_{st}()$	$Norm_{st}()$
Moyenne	9.33	0.79	12.74	0.72	11.73	0.74	14.43	0.68
Médiane	9.93	0.78	11.36	0.75	11.13	0.75	13.29	0.71
Écart type	4.16	0.09	6.57	0.14	5.90	0.12	7.12	0.15
Caractéristiques	FeCO2		FiHal		FeHal		T	
	$Dis_{st}()$	$Norm_{st}()$	$Dis_{st}()$	$Norm_{st}()$	$Dis_{st}()$	$Norm_{st}()$	$Dis_{st}()$	$Norm_{st}()$
Moyenne	19.11	0.58	14.31	0.68	12.92	0.71	9.74	0.78
Médiane	19.12	0.58	14.53	0.68	13.29	0.71	9.95	0.78
Écart type	7.25	0.15	7.00	0.15	6.72	0.14	4.40	0.09
Caractéristiques	FR		VT		VM		DGF	
	$Dis_{st}()$	$Norm_{st}()$	$Dis_{st}()$	$Norm_{st}()$	$Dis_{st}()$	$Norm_{st}()$	$Dis_{st}()$	$Norm_{st}()$
Moyenne	14.19	0.69	12.90	0.71	13.09	0.71	14.37	0.68
Médiane	13.12	0.71	11.25	0.75	11.38	0.75	11.23	0.75
Écart type	6.65	0.14	6.59	0.14	7.15	0.15	8.07	0.17
Caractéristiques	FiCO2		FiO2		PMAX		PEEP	
	$Dis_{st}()$	$Norm_{st}()$	$Dis_{st}()$	$Norm_{st}()$	$Dis_{st}()$	$Norm_{st}()$	$Dis_{st}()$	$Norm_{st}()$
Moyenne	14.61	0.68	14.37	0.68	15.98	0.65	15.72	0.65
Médiane	13.11	0.71	12.09	0.73	15.41	0.66	13.79	0.70
Écart type	7.32	0.15	7.89	0.17	8.66	0.18	9.00	0.19
Caractéristiques	PMOY							
	$Dis_{st}()$	$Norm_{st}()$						
Moyenne	16.98	0.63						
Médiane	17.28	0.62						
Écart type	6.64	0.14						

Table 3.5 – Caractéristiques des distributions de dissimilarités et similarités normalisées entre séries temporelles univariées de référence et simulées. Les différentes dimensions sont désignées par leur abréviation comme décrit en 2.1.3.3

Toutefois, ces résultats nous permettent seulement d’affirmer avec confiance que nos profils anesthésiques simulés sont *similaires* aux profils anesthésiques réels. Pour aller plus loin dans notre validation, nous avons réalisé une analyse statistique.

3.3.4.2 Analyse statistique

Le but de cette analyse statistique est de déterminer si les profils anesthésiques simulés peuvent être qualifiés de réalistes ou non.

Nous avons appliqué deux méthodes d’analyse statistique aux jeux de profils anesthésiques simulés et réels mentionnés plus haut. Nous avons calculé la fonction d’autocorrélation pour chacune des deux cohortes, et pour chaque variable. L’autocorrélation est la corrélation entre une variable et sa version décalée dans le temps. Nous avons observé que les fonctions d’autocorrélation des deux cohortes sont proches, comme le montre la figure 3.7 pour les 4 principaux paramètres physiologiques. De même, nous avons calculé la fonction de répartition pour chacune des deux cohortes, et pour chaque variable. Les fonctions de répartition empirique se sont révélées très similaires, comme le montre la figure 3.6 pour les 4 principaux paramètres physiologiques.

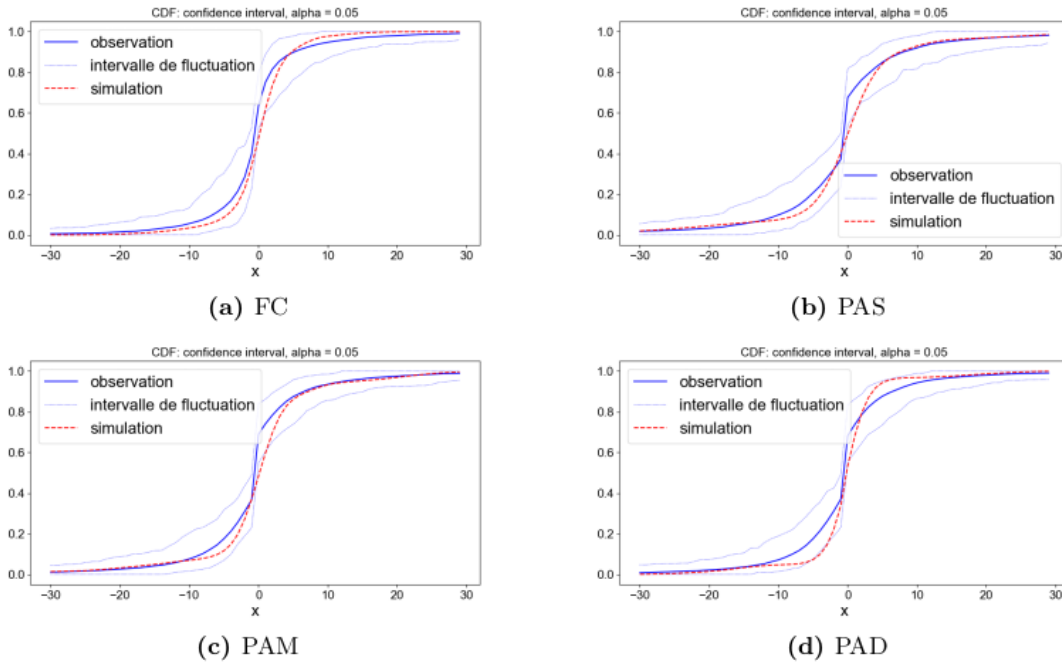


Figure 3.6 – Fonction de répartition pour les 4 variables hémodynamiques des séries temporelles des cohortes de référence et simulée

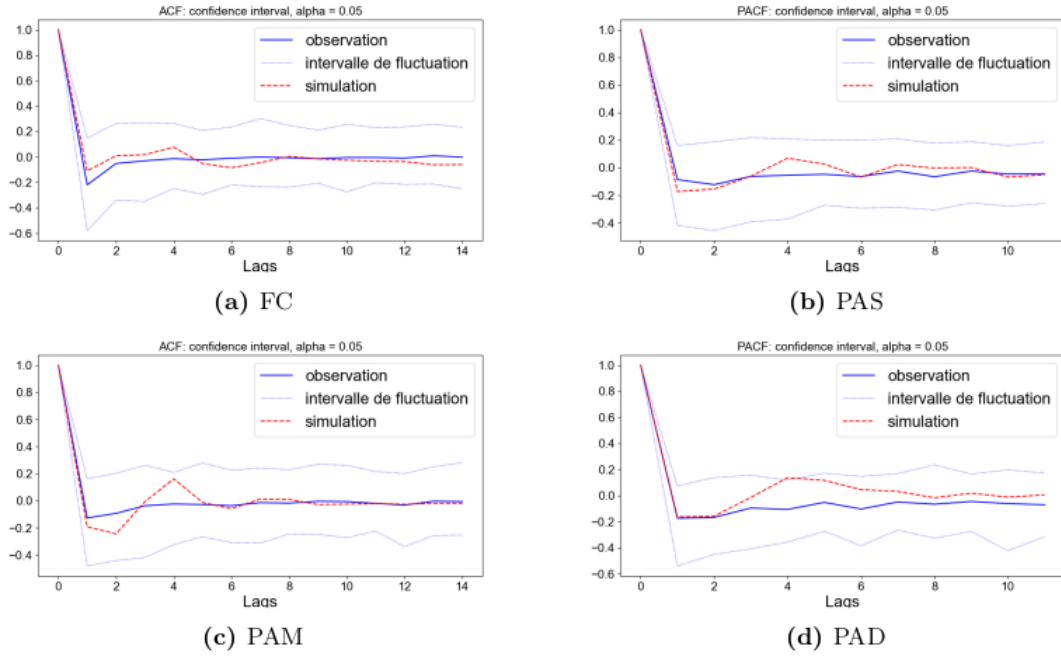


Figure 3.7 – Autocorrélation pour les 4 variables hémodynamiques des séries temporelles des cohortes de référence et simulée

Conclusion

En nous appuyant sur la formalisation médicale d’une anesthésie, nous avons proposé une formalisation informatique basée sur une grammaire formelle. Cette grammaire comporte plusieurs niveaux d’abstraction imbriqués ce qui permet de donner un cadre commun à toutes les anesthésies.

Faute de moyens techniques adaptés pour apprendre l’effet des différents événements anesthésiques sur un patient, nous avons utilisé l’expertise des professionnels de santé pour produire une typologie des événements anesthésiques. Cette typologie classe les événements en quatre grands types. Nous modélisons par ailleurs l’effet de ces événements sur toutes les variables d’un patient.

En nous appuyant sur ces deux éléments, nous avons construit le générateur de profils anesthésiques DBLBS, qui produit une séquence d’événements et simule la série temporelle multivariée d’un patient virtuel soumis à ces événements. Les séries temporelles produites ont été comparées avec des séries temporelles de patients réels soumis aux mêmes séquences d’événements. La validation par des experts, une analyse de similarité et une analyse statistique, ont conclu à la similarité des séries temporelles. Les séries temporelles simulées peuvent donc être qualifiées de réalistes et cohérentes. DBLBS nous a alors permis de produire plusieurs jeux de données réalistes de 1000 profils anesthésiques pour deux interventions.

SIMILARITÉ ENTRE SÉRIES TEMPORELLES MULTIVARIÉES

Introduction

La composante la plus importante d'un profil anesthésique est sa série temporelle multivariée en cela qu'elle représente directement l'effet des actions d'une équipe médicale sur l'état vital du patient. Ainsi, dans un contexte de simulation, la qualité de la série temporelle multivariée simulée est centrale dans le réalisme de la situation.

L'ambition de réalisme portée par ces travaux nous impose de pouvoir confronter l'évolution simulée des paramètres physiologiques d'un patient virtuel avec celle des paramètres des patients réels. De plus, les différentes approches de simulation que nous proposons s'appuient sur des mesures de dissimilarité entre séries temporelles. Enfin, la similarité entre séries temporelles est un élément d'évaluation des différentes approches de simulation que nous proposons. Il est donc nécessaire de déterminer une mesure de similarité entre séries temporelles multivariées adaptée à nos travaux.

La première partie de ce chapitre indique le contexte nous motivant à réaliser une expérimentation pour déterminer une mesure de similarité entre séries temporelles multivariées appropriée.

La seconde partie donne des repères issus de la littérature sur les différentes approches et méthodes que nous comparons pour identifier une mesure de similarité ou de dissimilarité appropriée.

La dernière partie de ce chapitre présente l'étude comparative des mesures de dissimilarité que nous avons réalisée, ainsi que ses résultats.

4.1 Motivations pour une analyse comparative

Les mesures de dissimilarité entre séries temporelles multivariées s'appuient majoritairement sur les mesures univariées. Dans la littérature, plusieurs études comparatives ciblent des applications telles que la classification, le clustering ou la détection de changements dans les séries temporelles. Ainsi, les études rapportées dans [62] et [63] utilisent un classificateur pour ensuite comparer les performances des différentes mesures *via* la précision de la classification. Cependant, ces études ne conviennent à nos travaux que très partiellement.

Premièrement, notre objectif n'est pas de regrouper ou de classer les séries temporelles, mais de quantifier la différence entre les séries temporelles.

Deuxièmement, les mesures de dissimilarité considérées par les études comparatives de la littérature ne sont pas toujours pertinentes pour le type de séries temporelles que nous traitons. Nos données sont spécifiques de l'anesthésie. Ainsi, elles sont multivariées et correspondent à des variables aux comportements très différents entre elles. Elles sont appariées à une séquence d'événements, et elles ne montrent aucune périodicité.

Troisièmement, la création de scénarios d'apprentissage s'appuie sur la simulation, réactive aux actions d'un apprenant, et d'un patient virtuel. Les approches que nous proposons pour atteindre cet objectif dans les chapitres suivants impliquent de reconnaître des séries temporelles similaires sous contrainte de temps réel. Ainsi, dans ce cadre d'utilisation, la mesure de similarité doit permettre de comparer des séries temporelles de taille courtes (estimées entre 10 et 30 points de mesure par les anesthésistes). Par conséquent, nous n'étions pas certains que les conclusions des études portant sur des séries temporelles de «longueurs habituelles» (environ 150 points) fonctionneraient dans notre cas.

Enfin, nous travaillons dans un contexte de *big data*, avec des variations potentiellement infimes entre les séries temporelles. Nous nous intéressons donc aux mesures de dissimilarité sensibles au contraste. En effet, notre objectif est d'identifier rapidement les différences entre patients, du point de vue de leurs séries temporelles. Notre besoin est ainsi centré sur la notion de *contraste*. Nous définissons comme *contraste* le comportement par lequel une augmentation du degré de perturbation conduit à une augmentation de la dissimilarité entre une série et sa version perturbée.

L'étude comparative des mesures de dissimilarité entre séries temporelles que nous avons réalisée cherche à étudier comment une augmentation du degré de perturbation d'une série temporelle conduit à une augmentation de la dissimilarité entre la série temporelle initiale et la version perturbée.

Même si certaines études comparatives appliquent diverses perturbations à des séries temporelles initiales, leur but est d'évaluer l'insensibilité des mesures de dissimilarité à des perturbations de natures et d'amplitudes diverses. Ainsi, [64] effectue une comparaison sur des séries temporelles satellitaires spécifiquement collectées et perturbées pour décrire la dynamique des écosystèmes (type de végétation, pratiques d'utilisation des terres). Un autre exemple est celui de [65] qui n'a pris en compte que les signaux d'électrocardiogramme dans ses ensembles de données de référence avant de les perturber.

Aucune autre comparaison empirique ne s'était focalisée sur les mêmes objectifs que la nôtre. Nous avons ainsi comparé diverses mesures basées sur des approches différentes.

4.2 Repères sur les mesures de dissimilarité entre séries temporelles

La littérature sur les séries temporelles est large et transversale à de multiples disciplines. (économétrie, traitement du signal, etc.). Pour les besoins spécifiques de ces différentes disciplines, de multiples mesures de similarité entre séries temporelles ont été mises au point [66]. Ces mesures se classent selon plusieurs catégories. On y retrouve notamment les approches :

- à pas fixe (*Lock-step*), où une distance est calculée point à point,
- élastiques, qui prennent en compte une compression ou un étirement du temps,
- par transformées, qui utilise les outils mathématiques spécifiques que sont les transformées.

Parmi ces différentes catégories, nous avons sélectionné plusieurs approches pour notre étude comparative.

4.2.1 Approches Lock-step

Les approches à pas fixe ou (*Lock-step*) calculent des similarités en comparant, pour chaque série temporelle, chacune des valeurs aux différentes positions. Les similarités les plus triviales de ces approches sont celles basées sur des distances [67].

Similarité simple Une similarité simple calculée à partir de la distance entre deux valeurs x et y , respectivement issues des séries temporelles X et Y , peut se définir comme suit :

$$distance(x, y) = 1 - \frac{|x| - |y|}{|x| + |y|}$$

On peut ensuite déterminer différentes similarités pour les séries X et Y de n valeurs, tel que la moyenne (S_m) ou encore la similarité entre pics (S_p) [68] :

$$S_m(X, Y) = \frac{1}{n} \sum_{i=0}^n distance(x_i, y_i)$$

$$S_p(X, Y) = \frac{1}{n} \sum_{i=0}^n \left[1 - \frac{|x_i| - |y_i|}{2 \max(x_i, y_i)} \right]$$

Distances Il est aussi possible d'utiliser des distances comme la distance euclidienne (D_e), la distance de Manhattan (D_m), qui sont toutes des variations de la distance de Minkowski (D_{min}) :

$$D_e(X, Y) = \sqrt{\sum_{i=0}^n (x_i - x_y)^2}$$

$$D_m(X, Y) = \sum_{i=0}^n |x_i - x_y|$$

$$D_{min}(X, Y) = \sqrt[p]{\sum_{i=0}^n (x_i - x_y)^p}$$

Les différentes approches *Lock-step* ont comme principales faiblesses de devoir s'appliquer à des séries de même longueur, et d'être particulièrement sensibles au bruit ou aux valeurs aberrantes. Toutefois, ces faiblesses ne concernent pas nos travaux car les séries temporelles anesthésiques sont absentes de bruit en raison de leur résolution temporelle faible (pas de 30 à 150 secondes entre chaque valeur).

4.2.2 Approches élastiques

Les approches élastiques s'affranchissent de la comparaison point à point des approches *Lock-step* et prennent en compte d'éventuelles déformations de la série temporelle en fonction du temps. Le temps n'est plus à pas fixe mais élastique.

Dynamic Time Warping Il est possible de chercher à associer deux séries temporelles selon leur forme et à estimer ainsi leur similarité, c'est le cas de la *Dynamic Time Warping* (DTW) [69, 70]. Cette approche construit un alignement des séries temporelles en cherchant à associer, à un ou plusieurs points d'une série, un ou plusieurs points de l'autre série. La figure 4.1 présente le principe de cette approche en comparaison avec la distance euclidienne.

Pour une série $X = x_1, \dots, x_n$ et une série $Y = y_1, \dots, y_m$, le calcul de la dtw utilise une matrice d'alignement B , de taille $n \times m$. $B(i, j)$ correspond à l'alignement partiel optimum entre la sous-série x_1, \dots, x_i et la sous série y_1, \dots, y_j . Le score d'alignement optimal entre les deux séries X et Y est mémorisé dans la case $B(n, m)$.

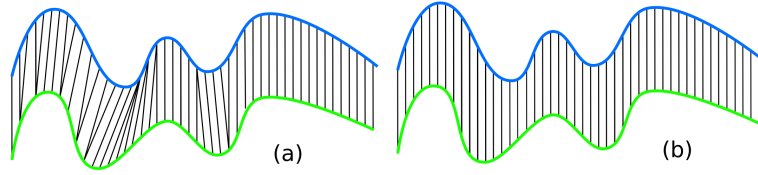


Figure 4.1 – Principe d'affectation des points pour la *Dynamic Time Warping* (a) en comparaison à la Distance euclidienne (b)

Des techniques de programmation dynamique permettent de calculer la matrice d'alignement B [71], telle que chaque valeur $B(i, j)$ est calculée comme suit :

$$B(i, j) = D(X_i, Y_j) + \min(B(i-1, j-1), B(i, j-1), B(i-1, j)),$$

où $D(i, j)$ correspond à la distance euclidienne entre le $i^{\text{ème}}$ point de X et le $j^{\text{ème}}$ point de Y .

Dans B les mouvements autorisés sont des mouvements verticaux, horizontaux et diagonaux de la case $B(0, 0)$ à $B(n, m)$.

Le chemin W est déterminé en reliant la case $B(n, m)$ à $B(0, 0)$ et en identifiant le mouvement qui avait conduit à l'obtention des valeurs minimum $B(i, j)$. La figure 4.2 présente la notion de chemin d'alignement minimal.

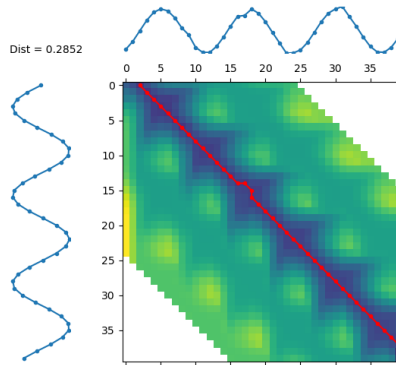


Figure 4.2 – Matrice d'alignement et chemin minimal pour un alignement de deux séries temporelles sinusoïdales de 40 points. Le chemin d'alignement qui relie la case en haut à gauche de la matrice à la case en bas à droite, est construit par déplacement sans retour en arrière sur les cases de valeurs minimales.

Il arrive parfois que cette méthode associe un seul point à un nombre important de points de la série à comparer. Pour contrer cela, il est possible de restreindre le chemin à

une bande diagonale de la matrice M . Cette bande, dite de SAKOE-CHIBA [72], contraint le chemin W dans une bande autour de la diagonale en choisissant un seuil S pour $i - j$, tel que :

$$i - j < S$$

Ainsi, $B(i, j)$ est désormais calculée comme suit :

$$B(i, j) = \begin{cases} D(X_i, Y_j) + \min(B(i-1, j-1), B(i, j-1), B(i-1, j)) & \text{si } i - j < S \\ \infty & \text{sinon.} \end{cases}$$

où S est un seuil fixé

D'autres restrictions existent telle que le parallélogramme d'ITAKURA [73], qui contraint le chemin W dans un parallélogramme autour de la diagonale. Ces contraintes sont présentées sur la figure 4.3.

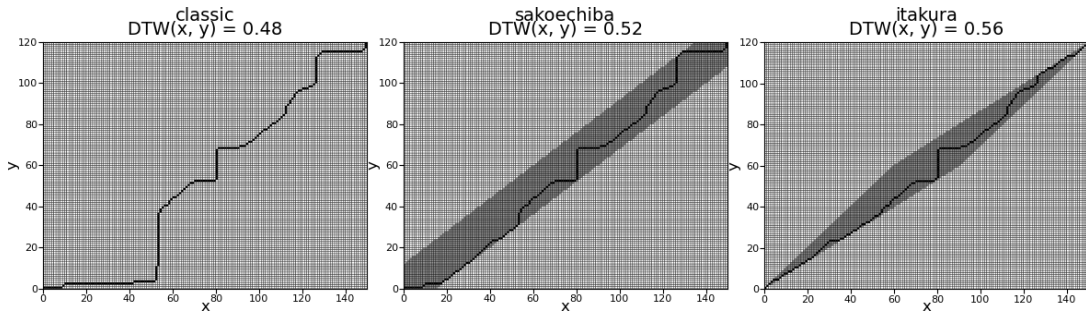


Figure 4.3 – Matrice d’alignement et chemin minimal pour deux séries temporelles de 120 et 150 points selon une DTW classique, avec une bande de SAKOE-CHIBA et un parallélogramme d’ITAKURA

D’autres variantes de la DTW ont été formulées pour réduire sa tendance à associer un seul point d’une série à de nombreux points d’une autre série, tel que la *Derivative dynamic time warping* (DDTW). Des approches combinées DTW et DDTW ont aussi été formulées pour augmenter les performances d’association [74].

Plus longue sous-séquence commune Une autre approche consiste à chercher la plus longue sous-séquence commune (*Longest Common SubSequence* ou LCSS) [75]. Cette approche est une variation de la distance d’édition et est plus résiliente au bruit¹ que la DTW. Cette approche cherche à faire correspondre deux séries en permettant à certains

1. Le bruit dans une série temporelle correspond à des changements brusques de valeurs

éléments d'être sans correspondance, alors que pour une DTW, tous les éléments doivent être appariés, même les valeurs aberrantes.

Pour une série $X = x_i, \dots, x_n$ et une série $Y = y_j, \dots, y_m$, la LCSS utilise une matrice de similarité M de taille $m \times n$, où $M(i, j)$ correspond à la distance entre le i -ème point de X et j -ème point de Y .

On peut à partir de M constituer une matrice L de taille $m \times n$ tel que chaque valeur $L(i, j)$ soit calculée comme suit :

$$L(i, j) = \begin{cases} 0 & \text{Si } i = 0 \\ 0 & \text{Si } j = 0 \\ 1 + L(i - 1, j - 1) & \text{Si } M(i, j) \leq S \\ \max(L(i - 1, j), L(i, j - 1)) & \text{Sinon} \end{cases}$$

où S est un seuil fixé

$L(n, m)$ contient la longueur de la plus longue sous-séquence commune ℓ entre X et Y . On peut déterminer une distance (DL) avec la formule suivante [75] :

$$DL(X, Y) = \frac{n+m+2\ell}{n+m}$$

4.2.3 Approches par transformées

On distingue les transformations *numériques* et les transformations *symboliques*. La similarité est déterminée sur la base d'une distance entre les représentations produites par ces transformations.

Transformée de Fourier discrète L'approche par transformée de FOURRIER discrète, *Discrete Fourier Transform* (DFT), est une transformation numérique, l'une des premières proposées dans la littérature [76].

Cette méthode considère un signal comme une somme de sinusoidales (les séries de FOURRIER). Chaque sinusoidale est caractérisée par un nombre complexe, le coefficient de FOURRIER [77, 78].

La force de cette approche est la possibilité de reconstruire le signal initial en utilisant seulement les sinusoidales présentant les coefficients les plus importants. Les autres sinusoidales peuvent être écartées sans perte significative d'information.

À partir des séries de FOURRIER d'une série temporelle, il est possible de calculer des distances métriques.

Transformée en ondelettes discrètes De manière similaire à la DFT, la Transformée en ondelettes discrète, *Discrete Wavelet transform* (DWT), est une transformation numérique qui cherche à décomposer un signal en une somme de fonctions prototypes (*wavelets*, ou ondelettes). Toutefois, la DWT présente l'avantage de pouvoir identifier les contributions locales de certaines ondelettes, là où la DFT implique forcément une contribution globale des séries de FOURRIER [78].

L'ondelette la plus communément utilisée est *Haas*. Comme pour la DFT, il est possible de réduire la série temporelle aux coefficients les plus importants et ensuite de calculer des distances métriques [79].

Transformées symboliques On peut mentionner l'approche par *Piecewise Aggregate Approximation* (PAA) [80] qui découpe une série temporelle en des segments de taille fixe qui sont représentés par la moyenne de la sous-série temporelle qu'ils recouvrent.

Cette approche a donné naissance à la *Symbolic Aggregate approXimation* (SAX) [81] qui reprend le principe de la PAA mais en groupant les points selon leurs valeurs. La représentation de la série est obtenue en associant à chaque valeur un symbole, généralement une lettre, en fonction d'une fourchette où se situe la valeur.

4.2.4 Approches basées sur des *features*

Les dissimilarités basées sur les *features*² tirent également parti des mesures de dissimilarité existantes [82]. Dans cette lignée, certaines approches s'appuient sur une mesure de dissimilarité déjà existante pour obtenir une nouvelle représentation des séries temporelles. Par exemple, les vecteurs de *features* sont produits sous forme de vecteurs de distances à d'autres séries temporelles, et la distance euclidienne peut ensuite être utilisée pour comparer deux de ces représentations.

2. Faute de pouvoir traduire la notion de *feature* en français, sans provoquer de confusion pour le lecteur, nous avons fait le choix de garder le terme anglais

4.3 Comparaison de neuf mesures de dissimilarité

Nous avons concentré notre analyse comparative sur une liste de 9 mesures de dissimilarité. Six critères ont été utilisés pour présélectionner ces mesures :

- l'adéquation à notre type de données et à notre application,
- la diversité des approches de dissimilarité entre séries temporelles,
- la disponibilité des implémentations en Python,
- le nombre de paramètres (idéalement aucun),
- la qualité de la documentation (fournissant notamment un paramétrage par défaut),
- la rapidité de calcul, requise par la contrainte de réactivité temps réel des scénarios.

La table 4.1 présente les mesures sélectionnées et explique nos choix. La table 4.2 met en évidence les principales caractéristiques des neuf mesures sélectionnées.

Méthode	Pertinence	Implémentations	Paramètres	Documentation	Vitesse
Mesures basées sur les modifications					
6 variantes de DTW	Mesure élastique	Pyts, dtaidistance	Si des paramètres sont requis, des valeurs par défaut sont fournies	Oui	Implémentation de la méthode originale, une version rapide (DTWD-tai) et quatre heuristiques
Mesures basées sur les transformations					
Transformée de FOURIER discrète combinée à une distance euclidienne	Reste pertinente pour des signaux non périodiques	numpy.fft.fft	Aucun	Oui	$n \log(n)$
Mesures basées sur les <i>features</i>					
MPDist	Correspondances entre sous-séries	MatrixProfile	Taille de la fenêtre de lecture w , avec $w = 0,05S$, où S est la somme des tailles de deux séries temporelles	Oui	Supposée rapide, mais impossible à savoir avant expérimentation
<i>tsfresh</i> combinée à une distance euclidienne sur les <i>features</i>	Large ensemble de <i>features</i>	Package Python tsfresh	Aucun	Oui	Inconnue avant expérimentation

Table 4.1 – Mesures de dissimilarité sélectionnées et argumentaire de sélection

Abréviation	Mesure	Librairie utilisée	
DTWc	DTW classique	Pyts	Alignement global de deux séries temporelles [72]
DTWdtai	DTW classique	dtadistance	Calcul exact de DTW implémenté en Cython
DTWf	DTW rapide	pyts	Alignement global des deux séries temporelles en alignant répétitivement les sous-séries des deux premières [83]
DTWsc	DTW Sakoe-Chiba	pyts	Alignement global des deux séries temporelles où le chemin d'alignement est limité par la bande de Sakoe-Chiba. Elle est définie comme une région de largeur constante le long de la diagonale de la matrice d'alignement. [72]
DTWi	DTW Itakura	pyts	Alignement global des deux séries temporelles où le chemin d'alignement est limité par le parallélogramme d'Itakura. Il est défini par deux paramètres contrôlant sa largeur et son étendue. [73]
DTWm	DTW multiscale	pyts	Alignement global utilisant un sous-échantillonnage des deux séries temporelles, puis une projection à l'échelle d'origine du chemin optimal ainsi obtenu. [84]
MPDist	MPDist	MatrixProfile	Mesure de dissimilarité basée sur les <i>features</i> utilisant les similitudes entre les sous-séries des deux séries comparées [85]
tsfresh	Mesure basée sur tsfresh	tsfresh	Mesure de dissimilarité basée sur les <i>features</i> où la distance euclidienne est appliquée aux représentations numériques des deux séries temporelles, obtenues grâce aux fonctions d'extraction du package tsfresh [86]
DFT	Transformée de FOURIER discrète	numpy.fft.fft	Mesure de dissimilarité basée sur les transformées où la distance euclidienne est appliquée aux vecteurs d'amplitude des composantes obtenues par transformée de FOURIER

Table 4.2 – Comparaison des neuf mesures de dissimilarités sélectionnées.

Nous avons ainsi sélectionné 6 variantes dans la catégorie des mesures élastiques, à savoir 2 implémentations exactes et 4 heuristiques de la *Dynamic Time Warping* (DTW). Bien que la DTW soit une approche ancienne et que son calcul nécessite un temps d'exécution quadratique, cette mesure n'a pas été supplantée par des propositions plus récentes pour classer les séries temporelles. Même si nous recherchons un autre type de performance (contraste) que la précision de la classification dans notre application, il reste pertinent de considérer ce type de mesure.

Nous avons également sélectionné *MPDist*, une mesure de dissimilarité basée sur des *features* récemment proposée par [85], dont le calcul est réputé rapide. Deux séries temporelles sont similaires selon *MPDist* si elles partagent de nombreuses sous-séries similaires au sens de la distance euclidienne, quelle que soit la position des sous-séries correspondantes dans leurs séries d'origine.

Nous avons étudié une deuxième mesure basée sur les *features*. En combinant 63 caractéristiques sur les séries temporelles, le package tsfresh [86] permet de calculer un total de 794 combinaisons de *features*. Ces *features* couvrent un large spectre, y compris la variance, l'autocorrélation, et les caractéristiques du spectre de la transformée de FOURIER

absolue. La distance euclidienne a été utilisée pour comparer les représentations des séries temporelles obtenues par extraction de ces *features* qui prennent la forme de vecteurs de taille 794.

Enfin, nous avons sélectionné une mesure basée sur la transformée de FOURIER discrète. Étant donné les deux séries temporelles univariées x et y , une mesure de dissimilarité est définie comme :

$$DFT(x, y) = \sqrt{\sum_{k=0}^m (A_x(k) - A_y(k))^2}, \quad (4.1)$$

où $A_z(k)$ représente la $k^{\text{ème}}$ composante d'amplitude dans la transformée de FOURIER de la série temporelle z .

4.3.1 Protocole expérimental

Nous souhaitons examiner comment les neuf mesures se comportent lorsque les séries temporelles sont perturbées avec un degré croissant de perturbation, pour six types de perturbations différentes.

Notre expérimentation consiste en 4 étapes qui sont résumées dans la figure 4.4 : (i) à partir de jeux de données sont produits des jeux expérimentaux de séries temporelles courtes qui serviront de références, (ii) les séries temporelles de référence sont perturbées grâce à 6 types de perturbations, et selon différents degrés, (iii) les séries perturbées et leurs références sont ensuite comparées, pour chacune des 9 dissimilarités sélectionnées.

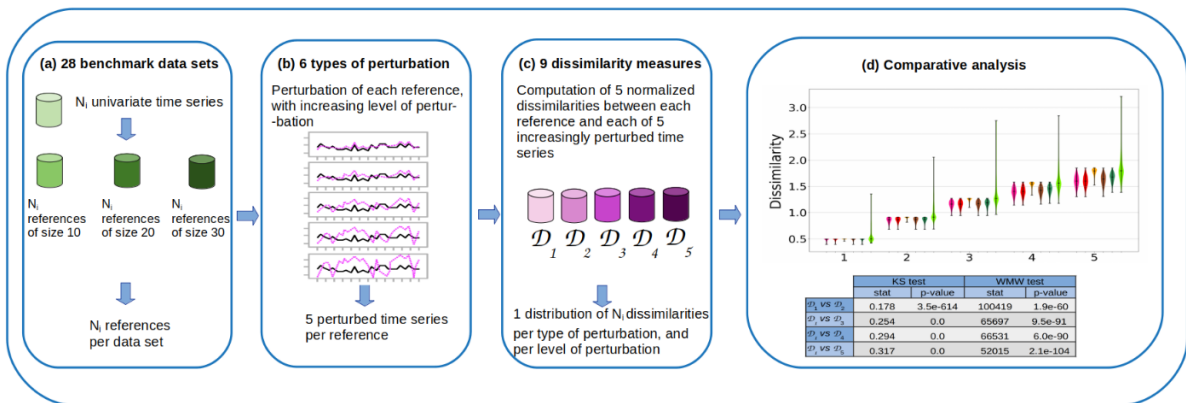


Figure 4.4 – Protocole expérimental appliqué dans l'analyse comparative des 9 mesures de dissimilarités sélectionnées.

Jeux de données Nous avons sélectionné 26 jeux de données de référence issues de la *UCR Time Series Classification Archive* mise à disposition pour l'étude des séries temporelles [87]. Ces jeux ont été sélectionnés, en écartant les jeux de données manifestement différents (par inspection visuelle) des séries temporelles observées lors du suivi des patients au bloc opératoire. Par exemple, nous avons exclu les jeux de données avec des séries temporelles cycliques ou saisonnières, ou encore ceux montrant des croissances ou décroissances fréquentes.

Nous avons également généré 2 ensembles de données de 1000 profils anesthésiques par simulation, un en utilisant le générateur DBLBS présenté en 3.3, l'autre produit par un modèle *Partially Hidden Markov Chain Linear AutoRegressive* (PHMC-LAR) [88]. Les 28 jeux de données sont décrits dans l'annexe C.1.

Ensuite, 3 sous-séries de taille 10, 20 et 30 points de mesure ont été tirées au hasard pour chaque série temporelle de ces 28 jeux de données comme le montre la figure 4.4 (a). Nous avons ainsi obtenu $3 \times 28 = 84$ ensembles de sous-séries, pour chaque jeu de données initial. Ces ensembles sont dénommés ci-après ensembles de référence.

Perturbations Dans un second temps, nous avons appliqué successivement 6 perturbations sur chaque sous-série de référence, en augmentant à chaque fois le degré de perturbation. Nous avons considéré 5 degrés. Ainsi, pour un type de perturbation donné, nous obtenons 5 séries temporelles perturbées pour chaque référence (Figure 4.4(b)). La figure 4.5 montre l'impact de ces perturbations sur une série temporelle.

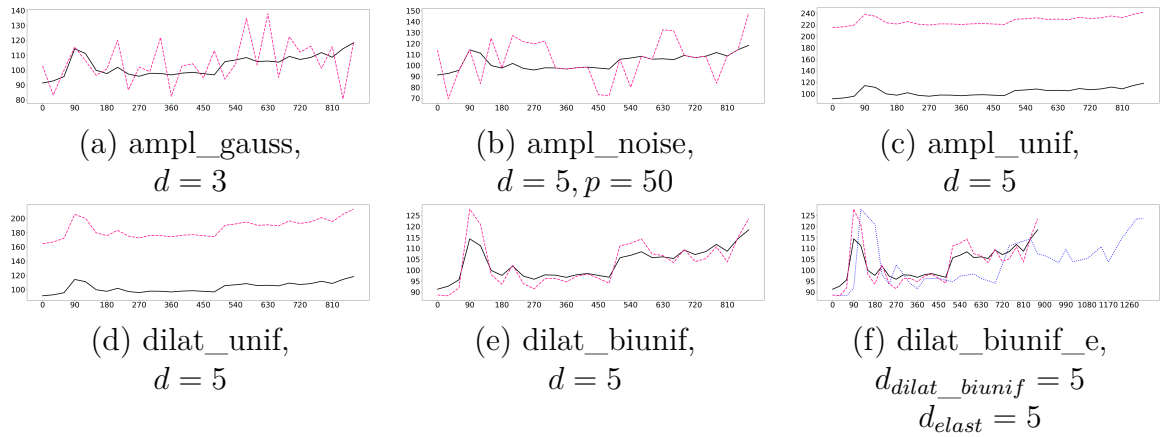


Figure 4.5 – Illustration pour les 6 perturbations considérées. Pour les sous-figures (a) à (e) sont indiquées la série temporelle initiale (en noir), et sa perturbée (en pointillé rose). Dans nos expériences, la perturbation *elast* n'est jamais utilisée seule comme le montre la sous-figure (f) avec la série temporelle initiale (en noir), la série perturbée *via dilat_biunif* (en pointillée rose) et cette série transformée *via elast* pour donner la perturbation *dilat_biunif_e* (pointillé bleu).

Le tableau 4.3 détaille ces transformations et fournit les valeurs des paramètres pour les 5 degrés de perturbation.

Perturbation	Description
<i>ampl_gauss</i>	Modifie l'amplitude à chaque horodatage $a_{pert} \sim Norm(a_{ref}, (std \times d/100)^2)$ 5 degrés : $d \in \{5, 10, 15, 20, 25\}$ <i>std</i> : écart-type de la distribution des amplitudes dans la référence Cette perturbation peut représenter le bruit éventuellement présent dans les séries temporelles et qui est propre aux différents capteurs (qui peuvent être perturbés, bouger, etc.)
<i>ampl_noise</i>	Modifie l'amplitude de $p\%$ des horodatages $a_{pert} = a_{ref} \times (1 + sign_{aleat} \times d/100)$, $sign_{aleat} \in \{+, -\}$ 5×3 degrés : $d \in \{5, 10, 15, 20, 25\}$, $p \in \{50, 75, 100\}$ Cette perturbation peut représenter le bruit éventuellement présent dans les séries temporelles et qui est propre aux différents capteurs (qui peuvent être perturbés, bouger, etc.)
<i>ampl_unif</i>	Déplace la série temporelle vers le haut, le long de l'axe des ordonnées $a_{pert} = a_{ref} + (moy \times d/100)$ 5 degrés : $d \in \{5, 10, 15, 20, 25\}$ <i>avg</i> : moyenne des amplitudes sur la référence Cette perturbation peut représenter la variation de la valeur normale du patient (pour un patient la valeur normale de la FC sera de 70 mais 80 pour un autre)
<i>dilat_unif</i>	Augmente l'amplitude à chaque horodatage, en fonction de l'amplitude initiale $a_{pert} = a_{ref} \times (1 + d/100)$ 5 degrés : $d \in \{5, 10, 15, 20, 25\}$ Cette perturbation peut représenter les effets cumulatifs d'un événement, par exemple de différentes doses (pour un médicament dose-dépendant, quand la dose augmente, l'effet augmente)
<i>dilat_biunif</i>	Augmente toutes les amplitudes qui étaient au-dessus du polynôme ajusté sur la série temporelle initiale ¹ , en fonction de l'amplitude initiale ; symétriquement, diminue toutes les amplitudes qui étaient en dessous du polynôme $a_{ref} > a_{pol} : a_{pert} = a_{ref} + ((a_{ref} - a_{pol}) \times d/100)$ $a_{ref} \leq a_{pol} : a_{pert} = a_{ref} + ((a_{pol} - a_{ref}) \times d/100)$ 5 degrés : $d \in \{20, 40, 60, 80, 100\}$ ¹ polynôme de degré 4 obtenu par régression linéaire de la référence Cette perturbation peut représenter les effets cumulatifs d'un événement, par exemple de différentes doses (pour un médicament dose-dépendant, quand la dose augmente, l'effet augmente)
<i>dilat_biunif_e</i>	Applique <i>dilat_biunif</i> puis insère $d_{elast}\%$ nouvelles valeurs à des horodatages tirés au hasard, la valeur insérée étant la moyenne entre les valeurs précédente et suivante 5×5 degrés : $d_{elast} \in \{10, 20, 30, 40, 50\}$, $d_{dilat_biunif} \in \{20, 40, 60, 80, 100\}$ Cette perturbation peut représenter un changement de cinétique métabolique qui va jouer sur la durée de l'effet d'un médicament.

Table 4.3 – Description des 6 types de perturbation considérés, et valeurs des paramètres correspondant aux degrés croissants de perturbation. A noter, dans nos expérimentations, la perturbation elast n'est jamais utilisée seule.

Le degré de perturbation est noté d . Les valeurs possibles pour d dépendent du type de perturbation. Dans la suite nous nous référons uniquement aux degrés $d \in \{1, 2, \dots, 5\}$. Par exemple, si d prend ses valeurs dans $\{5, 10, 15, 20, 25\}$, au lieu de $d = 10$, on parlera de $d = 2$ (perturbation relativement légère). a_{ref} désigne l'amplitude dans la série temporelle

à perturber (*c'est-à-dire* la référence), à un certain horodatage. a_{pert} représente l'amplitude dans la version perturbée de la référence.

Dissimilarités entre références et versions perturbées Dans la troisième étape, pour chacun des 84 jeux de données de référence S , chacune des 6 perturbations P , et chacune des 9 mesures de dissimilarité analysées D , nous avons évalué la dissimilarité entre la série de référence et chacune de ses 5 versions perturbées (Figure 4.4(c)). Un tel calcul est appelé "*Reference versus Perturbed*" (**RvP**). Ainsi, pour chaque triplet (S, P, D) , nous avons obtenu 5 distributions de valeurs de dissimilarité $\mathcal{D}_1 \cdots \mathcal{D}_5$ correspondant aux 5 degrés de perturbation.

Analyse comparative Nous avons d'abord examiné les distributions, sous la forme de diagrammes en violon, \mathcal{D}_1 à \mathcal{D}_5 , pour chaque triplet (S, P, D) (Figure 4.4(d)). Nous avons également comparé les diagrammes en violon obtenus pour $(S, P, D_1), \dots, (S, P, D_8)$ où D_i représente l'une des 9 mesures de dissimilarité comparées.

Ensuite, pour chaque triplet (S, P, D) , nous avons évalué les similitudes entre les distributions \mathcal{D}_1 et \mathcal{D}_i ($2 \leq i \leq 5$). Pour cela, nous avons utilisé deux tests statistiques, le test de Kolmogorov-Smirnov et le test de Wilcoxon/Mann-Whitney.

Enfin, nous avons répété l'expérience dans un autre cadre. Cette fois, nous avons mené l'expérience RvPO ("*Reference versus Perturbed versions of Others*"), dans laquelle les distributions $\mathcal{D}_1 \cdots \mathcal{D}_5$ sont des distributions de dissimilarités entre une série de référence et une série perturbée issue d'une autre référence. Ici, pour chaque référence et chaque degré de perturbation, nous avons tiré au hasard les versions perturbées de 10 autres références et les avons comparées à cette référence.

4.3.2 Résultats et principales conclusions

Les tendances observées pour les temps de calculs sur les séries courtes sont les mêmes pour les longueurs 10, 20 et 30. La figure 4.6 (a) souligne que sur de courtes séries, l'utilisation d'heuristiques pour calculer une DTW est toujours bénéfique en terme de temps de calcul, sauf pour la DTW multi-échelles (DTWm). Les quatre heuristiques DTW sont classées, par temps de calcul, dans la plage [1,4s - 1,9s] comme suit : DTWf, DTWsc-DTWi, DTWm. La mesure basée sur *tsfresh* présente la même complexité temporelle l'heuristique DTW la plus rapide (DTWf). L'implémentation en Cython (DTWdtai) réduit le temps de calcul à environ un vingtième du temps de calcul de la DTW classique (DTWc). MPDist

et DTWdtai ont à peu près les mêmes temps de calcul. DFT est la troisième mesure la plus rapide.

Les sous-figures 4.6(b) à (d) donnent un aperçu partiel des tendances observées pour les expériences RvP et RvPO.

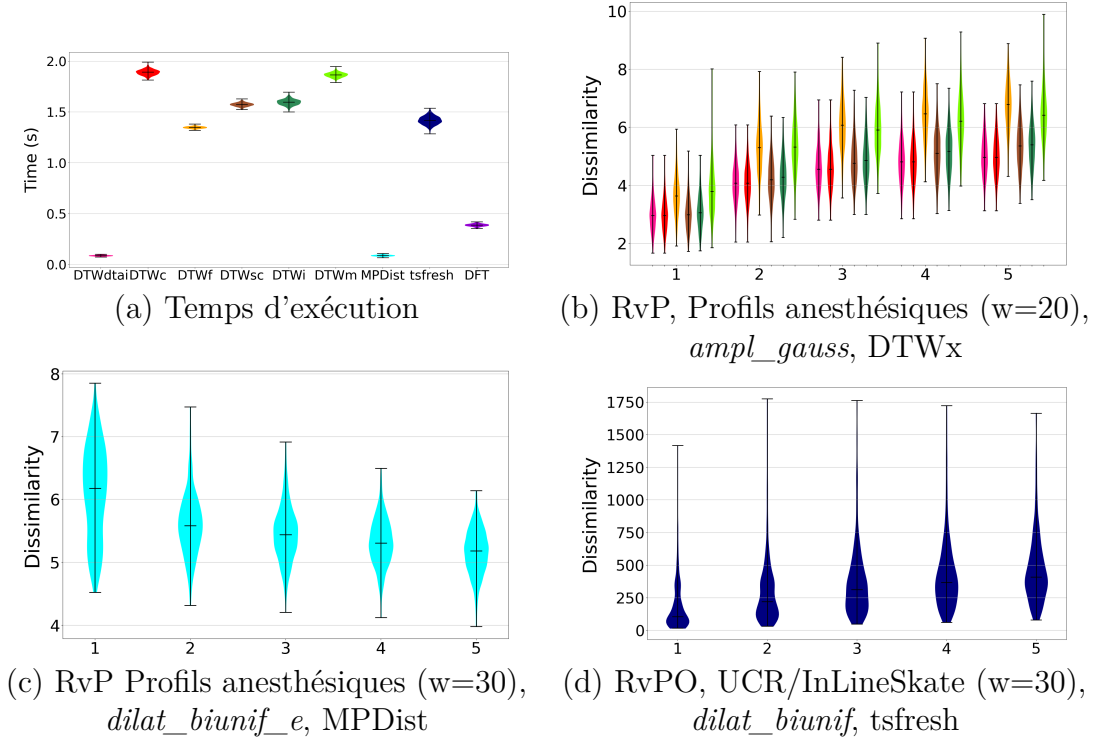


Figure 4.6 – Comparaison des 9 dissimilarité. (a) Temps d’exécution. (b) à (d) Aperçu partiel des tendances observées. En abscisses : degré croissant de perturbation. La signification des abréviations est précisée dans le tableau 4.2. DTWx : DTWdtai : rose ; DTWc : rouge ; DTWf : orange ; DTWsc : marron ; DTWi : vert clair ; DTWm : vert foncé ; MPDist : bleu clair ; tsfresh : bleu marine ; DFT : violet.

Nous définissons comme *contraste* le comportement par lequel une augmentation du degré de perturbation conduit à une augmentation de la dissimilarité entre une série et la version perturbée. Ainsi, MPDist, la mesure basée sur *tsfresh* et la DFT ont été écartées des candidats en raison soit d’un manque de contraste, d’un manque partiel de contraste (pour certains degrés seulement) ou d’un contraste inversé (le contraste diminuant à mesure que le degré augmente) pour certaines perturbations. L’expérience RvPO a également eu le mérite de disqualifier la mesure basée sur *tsfresh* qui montre un contraste dans cette expérience. Les conclusions ci-dessus s’étendent largement aux 27 autres ensembles de données de référence.

Les tests de Wilcoxon / Mann-Whitney (WMW) et Kolmogorov-Smirnov (KS) ont été exécutés sur des paires de distributions $(\mathcal{D}_1, \mathcal{D}_\ell)$, $2 \leq \ell \leq 5$. Les résultats confirment que les distributions sont différentes comme le montre la comparaison visuelle.

Un contraste visuel se traduit ainsi par des p-valeurs très faibles (dans la plage $[10^{-136}, 10^{-42}]$ (WMW), «zero» machine (KS)). Les p-valeurs comprises entre $[10^{-3} - 0,96]$ (WMW) et $[10^{-3} - 1]$ (KS) caractérisent l'absence de contraste. En absence de contraste, nous notons que les p-valeurs décroissent approximativement en fonction de ℓ .

Enfin, à partir de ces éléments nous avons sélectionné la mesure DTW la plus rapide comme mesure de dissimilarité, c'est-à-dire DTWdtai.

Conclusion

Le cadre original de nos travaux nous impose de déterminer une mesure de dissimilarité entre série temporelles multivariées. De nombreuses études comparatives existent dans la littérature, mais leur finalité ne convient pas à nos travaux.

Nous avons donc réalisé une étude comparative pour 9 mesures de dissimilarité entre séries temporelles en utilisant une grande diversité d’approches dans le choix de ces mesures. Notre étude comparative est originale en cela qu’elle cherche à montrer la capacité d’une mesure à estimer le contraste entre une série temporelle et une version perturbée de cette série temporelle. L’analyse des résultats met en avant les performances de la DTW dans son implémentation classique.

À partir de cette mesure de dissimilarité entre série temporelle univariée, nous serons en capacité de réaliser des mesures de similarité entre série temporelles multivariées comme le montreront les chapitres suivants.

SIMULATION D'UN PATIENT VIRTUEL BASÉE SUR DES DONNÉES

Introduction

L'exploration de données de santé et la construction de scénarios de simulation ne peuvent servir la formation des soignants en anesthésie que s'ils peuvent être joués par un simulateur. Ce simulateur doit permettre de simuler un patient virtuel qui soit similaire à un patient réel pour assurer le réalisme de la simulation. De plus, la simulation se doit d'être réactive aux différentes actions des apprenants, et de faire évoluer le patient virtuel en conséquence. Ce chapitre est dédié aux principes permettant la construction d'un tel simulateur.

La première partie de ce chapitre se concentrera sur un état de l'art en santé sur la modélisation jointe de séries temporelles et de séquences d'événements, les raisonnements à base de cas, et la détermination de similarités entre séries temporelles multivariées et leur prédiction.

La seconde partie présente la modalité de simulation que nous proposons, les contraintes inhérentes à la simulation réactive d'un patient virtuel ainsi que notre approche basée sur un raisonnement à base de cas.

Les parties suivantes s'intéressent aux approches et aux méthodes pour réaliser la simulation réactive d'un patient virtuel en s'appuyant sur une recherche de patients réels similaires.

5.1 État de l'art

Nos travaux se situent à l'interface de plusieurs domaines des sciences de la santé et des sciences du numérique. Le développement de la formation en santé par la simulation, aidée par le développement du numérique, a conduit de nombreux travaux à se pencher sur la notion de patient virtuel et sur la simulation de scénarios numériques.

Nos travaux s'insérant plus spécifiquement dans le domaine médical, nous focaliserons donc cet état de l'art sur les travaux qui sont liés à notre contexte. Mais nous passerons brièvement en revue les travaux en apprentissage automatique et en exploration de données qui portent sur la recherche de similarité dans les séries temporelles multivariées et leur prédiction. Nous montrerons ainsi leur inadéquation avec la simulation réactive d'un patient virtuel dans un contexte médical, et donc le caractère innovant de nos approches.

5.1.1 Simulation jointe de série temporelles et de séquences d'événements d'un patient virtuel

Le développement de la simulation pour la formation des soignants a permis le développement de scénarios jouables par les apprenants et de différents types de simulateurs, du mannequin basse fidélité à l'immersion dans une salle d'opération en réalité virtuelle.

Les simulateurs actuels se focalisent sur le réalisme des ressentis des apprenants pendant la simulation. Ainsi, les mannequins haute fidélité sont de plus en plus sophistiqués. Les plateformes utilisant la réalité virtuelle pour la formation permettent une immersion dans des situations chirurgicales qui sont enrichies de technologies de rétroaction haptiques pour former, par exemple, au maniement des outils chirurgicaux [89].

Un grand nombre d'études montre l'aspect bénéfique de l'utilisation d'un patient virtuel, dans un contexte de simulation, en termes d'intégration de connaissances et d'évaluation des compétences médicales [90]. On peut ainsi supposer que les nouvelles générations de simulateur devraient inclure un patient virtuel, et ses détails physiologiques, dans la simulation. Cependant, les tentatives de simulation de comportements physiologiques réalistes restent des exceptions. Il existe le simulateur physiologique *HumMod* [91] qui s'appuie sur la littérature pour simuler le fonctionnement des différents organes. On peut aussi noter l'existence de *Sycamore* qui est connecté au système d'information de nombreux centres de santé aux États-Unis et qui permet de simuler la réponse d'une pathologie à un traitement [92].

À notre connaissance, il n'existe pas de simulateur qui intègre, à la simulation des pratiques médicales, la simulation de la physiologie d'un patient virtuel réaliste.

L'utilisation de données de santé pour enrichir les scénarios de simulation, et donc le réalisme du patient virtuel, a fait l'objet de plusieurs travaux.

En effet, l'intelligence artificielle a beaucoup à apporter à l'extraction de connaissances à partir des données issues des interventions. Les techniques d'apprentissage automatique permettent la détection et la prédiction automatiques des activités chirurgicales, par exemple à l'aide de modèles de Markov cachés [93] ou de forêts aléatoires [94]. En particulier, l'apprentissage profond peut reconnaître divers types d'interventions chirurgicales à partir de vidéos [95]. Dans le domaine de la fouille de données, la littérature rapporte des travaux sur l'exploitation des tâches chirurgicales de bas niveau pour prédire les éventuelles tâches ultérieures des chirurgiens [96], ainsi que le temps d'intervention [97]. Des travaux appliquent des techniques de process mining à des enregistrements obtenus dans un contexte de soins, en mettant l'accent sur la découverte de processus, la vérification de conformité et l'amélioration de processus [98].

Bien que ces travaux soient proches de notre proposition, aucune approche consistant en la prédiction à court terme à la volée de séries temporelles multivariées, afin de gérer des scénarios réactifs simulant une opération chirurgicale, n'a jamais été proposée.

5.1.2 Raisonnement à base de cas

Le raisonnement à base de cas (*Case based reasoning*) est une approche fondamentalement différente des autres principales approches dans le domaine de l'intelligence artificielle. Si dans de nombreuses approches les données sont utilisées directement pour trouver une solution à un problème, le raisonnement à base de cas utilise les données pour trouver une solution de manière indirecte. Le raisonnement à base de cas résout un problème en trouvant, grâce aux données, un cas similaire déjà expérimenté et en tirant parti des connaissances sur ce cas similaire pour résoudre la nouvelle situation.

Le raisonnement à base de cas a été largement utilisé en médecine pour le diagnostic, la classification, la recommandation et la planification thérapeutique [99]. C'est une approche auto-adaptative guidée par les données, qui est bien adaptée aux problèmes dont les solutions nécessitent des connaissances difficiles à spécifier, mais pour lesquelles il existe suffisamment de données.

Cependant, il n'est pas simple de construire une nouvelle série temporelle multivariée à partir de celles de cas similaires. De plus, choisir une mesure de similarité adaptée aux

séries temporelles n'est pas aisé. Ainsi, relativement peu de travail a été fait pour exploiter les données temporelles dans le cadre d'un raisonnement à base de cas. On peut toutefois noter une application du raisonnement à base de cas pour la prédiction des taches réalisées par des chirurgiens [96].

Dans leurs travaux récents, MULAYIM *et al* utilisent le raisonnement à base de cas en grande dimension en l'appliquant à des millions de dossiers de patients. Leur solution consiste à intégrer un algorithme de k-plus proches voisins à un raisonnement à base de cas, pour éviter une évaluation inutile des voisins trop différents [100].

Notre approche de raisonnement à base de cas repose sur le concept de similarité locale entre des patients réels et un patient virtuel. De nombreux travaux rapportés dans la littérature traitent de la notion de similarité entre patients. Diverses définitions des similarités globales et locales ont donc été proposées. Elles exploitent une variété de composants des données des patients, pour les appliquer au regroupement ou la classification de patients, comme dans [101-105].

5.1.3 Score de similarité multivarié

L'exploration de séries temporelles multivariées a suscité un intérêt croissant ces dernières années. Notamment en raison du nombre croissant de domaines qui nécessitent la capacité de traiter de grands ensembles de séries temporelles.

Une grande majorité des travaux traitent de la recherche non supervisée de similarités dans les séries temporelles multivariées et étudient l'un des quatre domaines suivants :

- la réduction de dimension [106],
- l'indexation des séries temporelles [107],
- la représentation de séries temporelles [108],
- la conception de la mesure de similarité.

Il existe de nombreuses approches pour la recherche de motifs univariés dans des séries temporelles, on peut ainsi mentionner les *shapelets* [109], les approches par transformation symbolique [66] mentionnées en 4.2.3, ou plus récemment les *Matrix Profiles* [110].

Nous n'avons pu identifier qu'une seule publication traitant de la détection supervisée de motif multivarié dans des séries temporelles. Dans cette publication, LAFTCHIEV *et al* [111] cherchent à trouver des correspondances pour un motif multivarié dans de grandes séries temporelles. En particulier, ce travail accélère la recherche de séries temporelles similaires en utilisant des seuils pour éliminer les candidats peu prometteurs, dans la lignée des travaux de RAKTHANMANON [112].

Bien que le choix, ou la conception, d'une mesure de similarité entre séries temporelles soit central dans la majorité des travaux, il n'y a pas de consensus sur ce qui constitue une bonne métrique.

La méthode la plus simple consiste à moyenner les différentes dissimilarités univariées, pour obtenir une dissimilarité multivariée. Ainsi, WANG *et al* [113] obtiennent notamment une dissimilarité multivariée en combinant les dissimilarités univariées à l'aide de la méthode de vote BORDA pondérée.

De nombreux efforts pour généraliser les mesures de similarités au cas multivarié ont été consacrés à la mesure élastique la plus populaire, la *Dynamic Time Warping* (DTW) [72]. SHIFAZ *et al* [114] ont récemment généralisé cinq mesures de similarité élastiques au cas multivarié. De plus, les mesures univariées peuvent se généraliser directement aux séries temporelles multivariées, comme la mesure développée par BAYDOGAN *et al* [115], qui est basée sur une généralisation du concept d'autorégression.

Dans les travaux rapportés dans [116], une mesure de similarité temporelle est conçue pour faire face à des données de tests en laboratoire irrégulièrement mesurées, dans le but de mesurer la similarité entre des trajectoires de soins. Dans [117], les auteurs alimentent un graphe temporel avec les données temporelles des patients, pour quantifier la similarité entre patients en termes de progression de la maladie. Les données temporelles considérées dans ces deux travaux sont des événements médicaux.

5.1.4 Prédiction

De nombreux travaux ont été réalisés sur la prédiction de séries temporelles multivariées [118]. Dans notre application, les actions médicales peuvent exercer des effets synergiques ou antagonistes sur certains paramètres physiologiques du patient. Dans ces conditions, concevoir un modèle de prédiction basé sur la modélisation conjointe de séries temporelles et de séquences d'événements est très difficile.

De très nombreuses approches ont été proposées pour la prédiction de séries temporelles multivariées. Elles s'appuient sur divers modèles et techniques, notamment :

- les modèles autorégressifs [119],
- la factorisation matricielle [120],
- les modèles à facteurs dynamiques [121],
- la normalisation de flux [122].

De plus, de nombreuses approches d'apprentissage profond ont été conçues pour la prédiction de séries temporelles multivariées [123-126]. Des travaux récents se sont aussi concentrés sur des modèles hybrides [127, 128].

Nous n'avons trouvé aucune référence à des méthodes de prédiction dans lesquelles l'incorporation de connaissances (séquence d'événements) permet d'affiner la prédiction. Toutefois, [129] procède de manière symétrique à la nôtre. Pour prédire les occurrences d'événements, ces auteurs utilisent deux réseaux de neurones entrelacés : l'un pour capturer les dépendances entre les événements et l'autre pour modéliser les effets sur les séries temporelles des événements. En effet, les séries temporelles sont censées refléter l'état de l'environnement dans lequel les événements se produisent, de sorte qu'une meilleure description de cet environnement peut augmenter la précision de la prédiction des événements.

5.2 Simulation par une approche par raisonnement à base de cas

Il existe une multitude de modalités de simulation pour la formation des soignants. Comme nous l'indiquons dans la section 5.1.1, à notre connaissance il n'existe pas de simulateur de pratique médicale, qui soit réactif aux actions des apprenants et qui s'appuie sur un patient virtuel réaliste.

Dans le cadre de ces travaux de thèse nous avons fait le choix de proposer une modalité de simulation mono-utilisateur réactive basse fidélité pour la construction d'un simulateur sur ordinateur. Le fonctionnement de ce simulateur est décrit en 7.3. Cette section présente spécifiquement notre approche pour la simulation du patient virtuel

5.2.1 Approche par raisonnement à base de cas

La conception et l'entraînement d'un modèle de prédiction par apprentissage automatique est une tâche compliquée. Il faudrait pour cela d'importantes quantités de données dont l'accès est contraint, et une puissance de calcul qui n'était pas disponible au moment de la réalisation de ces travaux. De plus, les modèles produits ne seraient pas forcément adaptables aux différents objectifs pédagogiques. Nous proposons donc une approche plus simple, qui repose sur le raisonnement à base de cas.

Notre approche peut ainsi s'énoncer comme suit : Tout au long de la simulation, un patient virtuel évolue selon les actions des apprenants et de l'équipe médicale. À chaque fois qu'une action médicale est appliquée, nous identifions les caractéristiques de la situation du patient virtuel, un *cas*, et nous recherchons des instanciations connues de ce cas, pour prédire à partir de ces instanciations l'évolution à court terme du patient virtuel.

5.2.2 Définition d'un cas

Chaque *cas* correspond à une action médicale appliquée au patient virtuel, par exemple une action A, et à son *passé récent*. C'est-à-dire les actions (événements) et l'état physiologique (série temporelles), du patient virtuel qui précèdent la réalisation de l'action médicale à simuler.

La simulation est réalisée à partir d'instanciations déjà connues chez des patients réels de ce cas. Dans le cadre de notre approche, nous considérons une cohorte de patients réels qui partagent les mêmes caractéristiques que celles du patient virtuel comme un ensemble d'instanciations des différents cas.

Quand une action médicale est appliquée, nous recherchons les instanciatioins de ce cas les plus similaires au passé récent du patient virtuel, et donc les patients réels les plus similaires sur un cas précis, puis nous prédisons le futur immédiat du patient virtuel à partir des fenêtres de prédiction formées par les futurs immédiats des patients réels. La figure 5.1 présente notre approche pour un patient virtuel et 3 patients réels pour un cas.

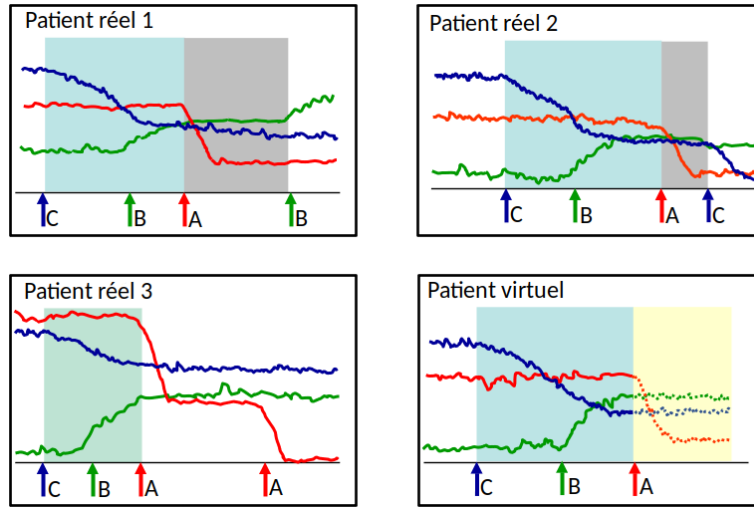


Figure 5.1 – Approche par raisonnement à base de cas pour la prédiction du futur immédiat du patient virtuel. À la survenue de l'action *A* chez le patient virtuel, on détermine son passé récent, en bleu, et on recherche chez des patients réels des passés récents similaires. Les patients réels 1 et 2 ont des passés récents similaires, mais pas le patient 3. À partir des futurs immédiats des patients réels 1 et 2, en gris, on prédit le futur immédiat du patient virtuel, en jaune.

5.2.3 Algorithmes

Pour faire évoluer le patient virtuel pendant tout un scénario, on simule sa série temporelle multivariée entre deux déclenchements d'action consécutifs, et on itère ce processus jusqu'à la fin de l'intervention. L'algorithme 3 décrit le cadre général de notre approche, son fonctionnement est présenté de manière schématique en figure 5.2.

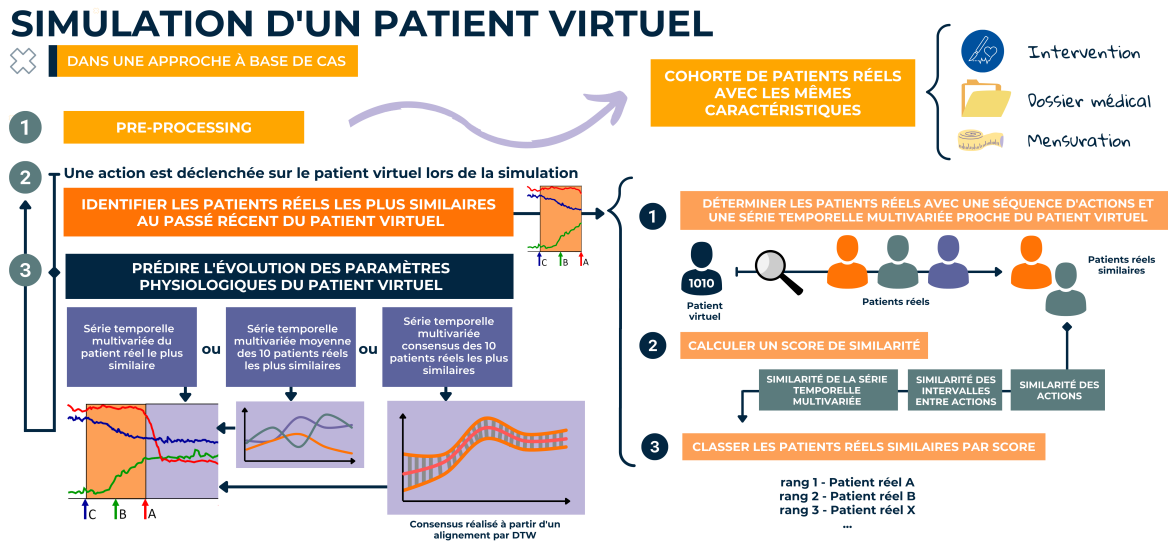


Figure 5.2 – Schéma explicatif de l'algorithme 3 pour la simulation d'un patient virtuel dans une approche à base de cas.

Notre approche repose sur deux temps, d'abord la recherche, dans une cohorte construite en amont de patients qui partagent les mêmes caractéristiques que notre patient virtuel (ligne 1, étape 1 dans la figure 5.2), de patients similaires au patient virtuel (lignes 7 à 9, étape 2 dans la figure 5.2). Puis la prédiction du futur immédiat du patient virtuel est réalisée à partir de ses patients similaires (lignes 11 à 13, étape 3 dans la figure 5.2). Chacun de ces temps pose des problématiques spécifiques en exploration de données, en détermination de similarités multivariées, et en prédiction de séries temporelles multivariées.

Algorithme 3 : Simulation du patient virtuel

Entrées :

- \mathcal{C}_{hir} : Chirurgie à simuler,
- \mathcal{C}_{ar} : Caractéristiques du patient virtuel,
- \mathcal{C}_r : Cohorte de patients réels,
- $timeout$: délai maximal permis entre le déclenchement de deux actions successives,
- n_a : Nombre d'actions dans le *passé récent*, permet de construire le contexte \mathcal{C} du patient virtuel à l'occurrence d'une action
- τ : pourcentage de relaxation pour les intervalles de temps entre actions, utilisé pour la recherche de patients similaires,
- n_s : nombre de patients réels similaires au patient virtuel,
- $strat_pred$: stratégie de prédiction utilisée.

Variables :

- \mathcal{C}_p : cohorte de patients réels ayant les caractéristiques \mathcal{C}_{ar} et subi l'intervention \mathcal{C}_{hir} ,
- \mathcal{C} : contexte du patient virtuel de longueur n_a à l'occurrence d'une action,
- \mathcal{M} : motif multivarié contextualisé du patient virtuel,
- $délai$: temps écoulé entre la dernière action et l'action juste initiée,
- $action$: action initiée par un apprenant,
- \mathcal{P}_{at_sim} : ensemble de patients similaires au contexte \mathcal{C} ,
- \mathcal{P}_{at_rank} : ensemble de patients similaires au contexte \mathcal{C} , classés par score de similarité,
- \mathcal{F} : prédiction de la série temporelle multivariée correspondant au futur immédiat du patient virtuel après l'occurrence d'une action

Sorties :

- \mathcal{E} : Séquence d'événements du patient virtuel en cours de construction,
- \mathcal{V} : Série temporelle multivariée du patient virtuel en construction

FONCTION `simulation_reactive` ($\mathcal{C}_{hir}, \mathcal{C}_{ar}, \mathcal{C}_r, timeout, n_a, \tau, n_s, strat_pred$)

```

1  $\mathcal{C}_p \leftarrow \text{Selection\_cohort}(\mathcal{C}_{hir}, \mathcal{C}_{ar}, \mathcal{C}_r)$ 
2  $(\mathcal{E}, \mathcal{V}) \leftarrow \text{Initialise\_patient\_virtuel}(n_a)$ 
3 tant que l'intervention n'est pas terminée faire
4    $(action, délai) \leftarrow \text{Attendre\_une\_action}(timeout)$ 
5   si  $action = \emptyset$  alors
6     Fin de l'intervention
7   sinon
8      $\mathcal{C} \leftarrow \text{Construire\_contexte}(\mathcal{E}, n_a, action, délai)$ 
9      $\mathcal{M} \leftarrow \text{Construire\_motif}(\mathcal{C}, \mathcal{V})$ 
10     $\mathcal{P}_{at\_sim} \leftarrow \text{Selection\_patients\_similaires}(\mathcal{C}_p, \mathcal{C}, \tau)$ 
11     $\mathcal{P}_{at\_rank} \leftarrow \text{Calcul\_score\_similarité}(\mathcal{P}_{at\_sim}, \mathcal{M}, n_s)$ 
12     $\mathcal{F} \leftarrow \text{Prédiction\_futur\_immédiat}(\mathcal{P}_{at\_rank}, strat\_pred)$ 
13     $\mathcal{V} \leftarrow \text{Ajout\_série\_temporelle\_multivariée}(\mathcal{V}, \mathcal{F})$ 
14     $\mathcal{E} \leftarrow \text{Ajout\_séquence}(\mathcal{E}, délai, action)$ 
15  fin
16 fin
```

5.3 Recherche de patients similaires

Pour chaque action réalisée par l'apprenant (ligne 4), ou l'équipe médicale, nous commençons par chercher les patients réels similaires (ligne 9) à un patient virtuel parmi la cohorte de patients dont les caractéristiques sont proches (ligne 1). Ces caractéristiques sont : le type de chirurgie, le sexe, la classe d'âge, la classe de poids et les antécédents médicaux. Les antécédents médicaux correspondent à une pathologie pré-existante à l'intervention envisagée et sans rapport avec celle-ci, ou éventuellement, à l'absence d'antécédents médicaux.

La recherche des patients réels les plus similaires au patient virtuel s'appuie sur un score de similarité multivarié (ligne 10). Toutefois, calculer un tel score sur un ensemble de patients peut être coûteux. La série temporelle multivariée d'un patient est directement déterminée par sa séquence d'événements associée. Un patient réel, qui n'a pas dans sa séquence d'événements l'action E_k venant d'être réalisée sur le patient virtuel, ne peut donc pas être considéré comme un patient réel similaire. De plus, on peut considérer *a priori* que le patient réel le plus similaire au patient virtuel ne sera pas un patient réel avec un passé récent très dissimilaire de celui du patient virtuel lors de la survenue de l'action E_k .

Notre recherche s'organise donc en deux temps : (i) la pré-sélection des patients similaires du point de vue de la séquence d'événements et de leur passé récent lors de la survenue de l'action qui vient d'être déclenchée (lignes 7 à 9) ; (ii) le calcul d'un score de similarité multivarié pour chacun de ces patients réels candidats (ligne 10). Il est ensuite possible de simuler le patient virtuel à partir de ces patients similaires (ligne 11 à 13)

5.3.1 Patient virtuel et cohorte de patients réels

Un *patient virtuel* est représenté par un ensemble de variables simulées, qui correspondent aux paramètres physiologiques, et par une séquence d'événements. La simulation du patient virtuel consiste donc à produire un profil anesthésique.

Dans la suite de ce chapitre, nous utiliserons les notations suivantes pour nous référer au patient virtuel et aux patients réels.

Ensemble de variables simulées du patient virtuel La série temporelle multivariée du patient virtuel est composée des séries temporelles univariées décrivant chacune une variable physiologique. \mathcal{V} représente l'ensemble des n_v variables dont la dynamique est simulée. Chaque variable est identifiée par son étiquette i ($1 \leq i \leq n_v$).

Séquence d'événements La séquence d'événements d'un profil anesthésique est une séquence horodatée d'actions réalisées par l'équipe médicale, y compris l'apprenant. \mathcal{E} représente l'ensemble des n_e événements anesthésiques E pour la séquence d'événements du patient virtuel pendant l'intervention simulée. Chaque événement est identifié par une étiquette k ($1 \leq k \leq n_e$).

\mathcal{E} est composé de \mathcal{E}_{act} , une séquence d'actions E , et de \mathcal{E}_{time} , la séquence d'horodatages correspondant aux temps d'occurrence des actions E de \mathcal{E}_{act} .

Cohorte de patients réels \mathcal{C}_p désigne la cohorte de n_p patients réels qui partagent les mêmes caractéristiques que le patient virtuel. Tout patient réel est identifié par son étiquette j ($1 \leq j \leq n_p$). Un *patient réel* est représenté par son profil anesthésique.

Soit $\mathcal{T}(j)$ la séquence horodatée des actes médicaux appliqués lors de l'intervention chirurgicale du patient réel j ($0 \leq j \leq n_p$). Le nombre total d'actions successives enregistrées dans $\mathcal{T}(j)$ est noté $n_t(j)$. $\mathcal{T}(j)$ est composé de $\mathcal{T}_{act}(j)$, une séquence de $n_t(j)$ actions, et de $\mathcal{T}_{time}(j)$, la séquence d'horodatages correspondant aux temps d'occurrence des actions dans $\mathcal{T}_{act}(j)$. Enfin, $\mathcal{T}_{act}(j)[m, n]$ ($1 \leq m \leq n \leq n_t(j)$) désigne un sous-vecteur de $\mathcal{T}_{act}(j)$ commençant par l'action m et se terminant par l'action n .

Pour un patient j ($0 \leq j \leq n_p$), $\mathcal{Y}(j)$ désigne l'ensemble de n_y séries temporelles univariées synchrones qui décrivent respectivement la dynamique des n_y variables. Les séries temporelles n_y ont la même longueur $\ell(j)$ spécifique au patient j . $\mathcal{Y}(j)[r, s]$ désigne la sous-série multivariée commençant à l'horodatage r et se terminant à l'horodatage s ($1 \leq r \leq s \leq \ell_j$). De plus, $\mathcal{Y}(j)[r, s][v]$ fait référence à la sous-série univariée liée à la variable v ($1 \leq v \leq n_v$).

5.3.2 Contexte et motif multivarié contextualisé

Lorsque nous prédisons l'avenir du patient virtuel après qu'une action E_k ai été initiée (ligne 4), nous savons que des actions ont été effectuées dans le *passé récent*.

Quand l'action E_k est appliquée, celle-ci n'est pas encore enregistrée dans la séquence \mathcal{E} du patient virtuel, construite jusqu'ici dans la simulation courante. Le nombre actuel d'actions enregistrées dans la séquence d'événements \mathcal{E} est de n_e actions. Le nombre d'actions, pour la durée du passé récent qui sert de contexte pour le motif multivarié contextualisé, est n_a .

Le contexte du patient virtuel \mathcal{C} est défini comme une séquence d'actions et de durées entrelacées $[E_{n_e-n_a-1}, D_{n_e-n_a-1}, \dots, E_{n_e-1}, D_{n_e-1}, E_{n_e}, D_0, E_k]$, où à une action E_x ($n_e - n_a - 2 \leq x \leq n_e$) correspond un événement $\mathcal{E}_{act}[x]$, et D_x est le temps écoulé entre les actions E_x et E_{x+1} calculé à partir de $\mathcal{E}_{time}[x]$ et $\mathcal{E}_{time}[x+1]$, horodatages respectifs de E_x et E_{x+1} . D_0 est l'intervalle de temps écoulé entre E_{n_e} , la dernière action dans \mathcal{E} , et l'action E_k tout juste appliquée. Pour la suite de cette section, la suite d'actions du contexte \mathcal{C} est notée \mathcal{C}_{act} et la suite de durées \mathcal{C}_{dur} .

Il est aussi possible de déterminer un contexte pour un patient réel en considérant la même action E_k dans sa propre séquence d'événements et les actions qui l'ont précédée. On peut ainsi noter $\mathcal{K}(j)$ le contexte d'un patient réel j pour une action E_k instanciée chez le patient virtuel.

Ainsi, si n_a est égal à 3, cela signifie que nous considérons les trois dernières actions effectuées dont l'action qui vient d'être déclenchée. Disons que ces actions sont B, C et A, alors $\langle "B", 38, "C", 86, "A" \rangle$ est un exemple de contexte pour ces actions où les intervalles de temps entre deux déclenchements successifs sont respectivement de 38 et 86 unités de temps.

Du point de vue des paramètres physiologiques, ce *passé récent* correspond à un motif multivarié, qui est une sous-série temporelle multivariée du patient virtuel en cours de construction par simulation. Comme le montre la figure 5.3, cette sous-série temporelle est délimitée à gauche par l'horodatage d'initiation de l'action $E_{n_e-n_a-1}$, et à droite par l'horodatage d'initiation de l'action en cours, E_k . Ce motif multivarié est ainsi contextualisé par les événements du passé récent et n'est donc rien d'autre qu'une troncature du profil anesthésique du patient virtuel. Ce motif multivarié contextualisé du patient virtuel est utilisé comme requête dans les profils anesthésiques des patient réels, pour identifier les motifs multivariés contextualisés réels qui sont les plus similaires à celui du patient virtuel.

Nos contraintes de simulation nous imposent de pouvoir simuler l'effet combiné de l'action en cours E_k et des n_a actions précédentes. Il faut donc que le passé récent, utilisé pour déterminer le motif multivarié contextualisé \mathcal{M} qui permet la recherche de patients réels similaires, tienne compte de cette contrainte d'effet combiné. Pour déterminer l'étendue du passé récent du patient virtuel, nous spécifions le nombre n_a d'actions passées et actuelle à prendre en compte en nous basant sur les conseils des experts en anesthésie.

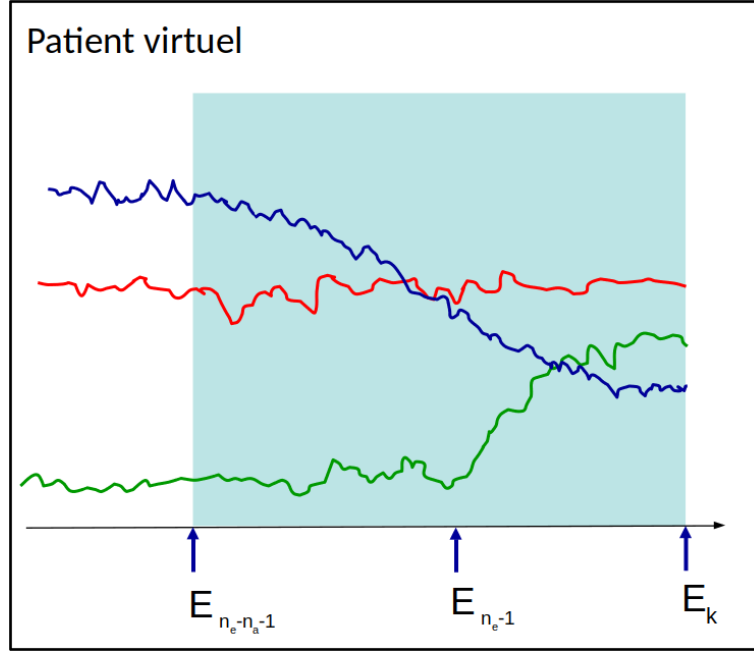


Figure 5.3 – Séquence d'événements et séries temporelles pour 3 variables d'un patient virtuel à l'occurrence de l'action E_k . Le motif multivarié contextualisé est indiqué en bleu.

Ainsi, le motif multivarié contextualisé \mathcal{M} d'un patient virtuel, à l'occurrence de l'action E_k , est défini comme :

$$\mathcal{M} = \mathcal{V}[\mathcal{E}_{time}(k - n_a), \mathcal{E}_{time}(k)]$$

où \mathcal{V} est l'ensemble des variables de la série temporelles multivariée.

Ce motif multivarié contextualisé \mathcal{M} d'un patient virtuel est relatif au contexte \mathcal{C} de ce même patient virtuel. De manière symétrique, on peut définir le motif multivarié contextualisé $\mathcal{P}(j)$ d'un patient réel j , relatif au contexte $\mathcal{K}(j)$ de ce même patient.

5.3.3 Pré-sélection des patients réels

La sélection des patients réels similaires revient à exclure de la cohorte \mathcal{C}_p les patients dont la séquence d'événements \mathcal{T} montre un passé récent pour une action E_k , c'est-à-dire un contexte $\mathcal{K}(j)$, qui est dissimilaire du contexte \mathcal{C} du patient virtuel.

Il faut donc dans un premier temps déterminer si l'action E_k est présente dans la séquence d'événements \mathcal{T} d'un patient j . Dans un second temps, il faut déterminer si le contexte \mathcal{C} du patient virtuel et le contexte $\mathcal{K}(j)$ du patient réel sont dissimilaires.

L'étendue du passé récent n_a peut évoluer selon les interventions. En s'appuyant sur les conseils des anesthésistes, nous avons dû considérer plusieurs valeurs pour n_a .

Une recherche d'occurrences multiples de \mathcal{C} de taille n_a , dans \mathcal{T} , peut aussi se montrer très coûteuse. Pour palier cet inconvénient, on peut réaliser un pré-traitement des séquences \mathcal{T} pour toute la cohorte \mathcal{C}_p .

Ce pré-traitement consiste en la construction de tous les contextes $\mathcal{K}(j)$ potentiels, et de leurs motifs $\mathcal{P}(j)$ associés, pour chaque valeur de n_a . Les paires de contextes et motifs $(\mathcal{K}(j), \mathcal{P}(j))$ caractérisés par le même $\mathcal{K}(j)$ sont ensuite groupées et assemblées dans une table d'association en utilisant $\mathcal{K}(j)$ comme clé.

De cette façon, en passant à la table d'association le contexte du patient virtuel \mathcal{C} , il est possible de récupérer tous les contextes $\mathcal{K}(j)$ identiques à \mathcal{C} et les motifs $\mathcal{P}(j)$ qui y sont liés.

Cependant, une telle approche nous contraint à trouver des occurrences exactes de \mathcal{C} . Il existe pourtant des événements différents qui ont des rôles et des impacts identiques, par exemple deux médicaments différents, mais avec la même utilisation médicale. Toutefois, grâce à l'aide des experts, nous sommes en mesure de catégoriser ces actions *synonymes*. Il est donc possible d'ajouter de la flexibilité sur la recherche de patient en substituant leurs synonymes à des événements de la sous-séquence d'événements \mathcal{C}_{act} .

Le contexte \mathcal{C} décrit les intervalles entre les différents événements. Afin d'apporter un peu de flexibilité temporelle dans la recherche des patients réels similaires nous acceptons que les intervalles de temps puissent être différents des intervalles du contexte \mathcal{C}_{dur} , et cela d'autant plus que les actions sont plus anciennes. Étant donné le pourcentage de relaxation τ , on permet un écart $\mathcal{T}_{thresh}(i)$ vis-à-vis de \mathcal{C}_{dur} pour une position i dans \mathcal{C}_{dur} . \mathcal{T}_{thresh} diminuant progressivement avec l'augmentation de la position i ($1 \leq i < n_a$) d'une action dans le contexte \mathcal{C}_{dur} , comme suit :

$$\mathcal{T}_{thresh}(i) = \tau \left(1 - \frac{i-1}{n_a-1}\right)$$

La sélection des patients est ainsi réalisée à partir du contexte \mathcal{C} , et en considérant les substitutions des événements synonymes et les relaxations temporelles sur les intervalles.

5.3.4 Score de similarité multivarié

À partir d'une pré-sélection de patients réels similaires, il faut déterminer les patients les *plus* similaires.

Pour être similaire au patient virtuel, un patient réel doit avoir une séquence d'événements la plus proche de celle du patient virtuel. De plus, pendant l'intervalle de temps défini par le passé récent, la dynamique des paramètres physiologiques devrait également être la plus proche de celle du patient virtuel.

Ainsi, pour chaque patient réel pré-sélectionné, on détermine un score \mathcal{S} défini comme suit :

$$\mathcal{S} = \lambda_1 \mathcal{S}im_{events} + \lambda_2 \mathcal{S}im_{times} + \lambda_3 \mathcal{S}im_{series}$$

où $\mathcal{S}im_{events}$ et $\mathcal{S}im_{times}$ sont deux similarités sur les sous-séquences d'événements, et $\mathcal{S}im_{series}$ une similarité sur les sous-séries temporelles multivariées. λ_1 , λ_1 et λ_1 , sont des coefficients de pondération.

5.3.4.1 Similarité entre les séquences d'événements

Pour déterminer la similarité entre le contexte \mathcal{K} d'un patient réel et le contexte \mathcal{C} du patient virtuel, on définit deux critères :

- $\mathcal{S}im_{events}$, qui représente la similarité entre \mathcal{C} et \mathcal{K} du point de vue de l'ordonnement des événements,
- $\mathcal{S}im_{times}$, qui représente la similarité entre intervalles de temps séparant les événements successifs dans \mathcal{C} et \mathcal{K} .

Pour estimer la similarité dans l'ordonnement des événements, $\mathcal{S}im_{events}$ détermine une similarité $sim()$ pour une paire d'événements présents à la même position, tel que $sim(s, s') = 1$ si $s = s'$ ou 0 sinon. Ainsi la similarité $\mathcal{S}im_{events}$ entre le patient virtuel et un patient réel j est formulée par :

$$\mathcal{S}im_{events} = \frac{1}{n_a} \sum_{i=1}^{n_a} sim(\mathcal{C}_{act}[i], \mathcal{K}_{act}(j)[i])$$

où $\mathcal{C}_{act}[i]$ et $\mathcal{K}(j)[i]$ correspondent respectivement à la $i^{ème}$ action dans le contexte du patient virtuel à la $i^{ème}$ action dans le contexte du patient réel j pour une position i ($1 \leq i < n_a$).

Au-delà de la nature et de l'ordonnement des événements, l'horodatage des actions réalisées contraint la séquence d'événements. Pour estimer la similarité des intervalles entre les actions présentes dans \mathcal{C} et $\mathcal{K}(j)$, $\mathcal{S}im_{times}$ associe deux à deux les intervalles et

calcule la somme ξ des différences entre intervalles qui est ensuite transformée en similarité sur l'ensemble des patients réels, comme suit :

$$\mathcal{S}im_{times} = \frac{\xi_{max} - \xi}{\xi_{max} - \xi_{min}}$$

$$\xi = \sum \gamma(i) | \mathcal{C}_{dur}[i] - \mathcal{K}_{dur}(j)[i] |$$

où ξ_{max} et ξ_{min} sont respectivement la somme ξ maximale et minimale des différences entre les intervalles. $\mathcal{C}_{dur}[i]$ et $\mathcal{K}_{dur}(j)[i]$ sont les intervalles entre la $i^{ème}$ action ($1 \leq i < n_a$) et la $i + 1^{ème}$ action dans les contextes \mathcal{C} et $\mathcal{K}(j)$ respectivement.

S'il est intéressant d'apporter un peu de flexibilité temporelle pour la sélection des patients réels similaires, comme indiqué en 5.3.3, il peut être pertinent d'apporter une relaxation temporelle sur le calcul de cette similarité. Nous pénalisons progressivement les écarts entre intervalles les plus récents de manière à réduire le poids des intervalles les plus anciens dans le calcul de la similarité. Ainsi, nous définissons un coefficient $\gamma(i)$ dont la valeur est une fonction de la position i ($1 \leq i < n_a$) de l'intervalle comme suit :

$$\gamma(i) = \delta \left(1 - \frac{i-1}{n_a-1} \right)$$

où δ est un coefficient de pénalité.

5.3.4.2 Similarité entre les séries temporelles multivariées

Pour sélectionner les patients réels j avec le motif multivarié contextualisé $\mathcal{P}(j)$ le plus similaire à celui du patient virtuel \mathcal{M} , il est nécessaire de calculer une similarité entre les séries temporelles multivariées.

On calcule ainsi une similarité $\mathcal{S}im_{series}$, qui repose sur une mesure de dissimilarité \bar{D} , de la façon suivante :

$$\mathcal{S}im_{series} = \frac{\bar{D}_{max} - \bar{D}}{\bar{D}_{max} - \bar{D}_{min}}$$

où \bar{D}_{max} et \bar{D}_{min} sont respectivement les valeurs maximale et minimale de la mesure de dissimilarité \bar{D} sur tous les patients réels pré-sélectionnés. Comme $\mathcal{S}im_{series}$ est un score normalisé, il est nul pour le patient réel avec la plus grande dissimilarité multivariée \bar{D}_{max} et égal à 1 pour la plus petite dissimilarité multivariée \bar{D}_{min} .

Il n'est pas aisé de déterminer une dissimilarité multivariée \bar{D}_{max} qui soit adaptée à des séries temporelles anesthésiques. De plus, le motif multivarié contextualisé \mathcal{M} est composé de séries temporelles courtes. Bien que de nombreuses mesures de dissimilarité

entre séries temporelles univariées aient été proposées dans la littérature, nous n'avons pas trouvé de publication traitant spécifiquement des séries temporelles courtes.

Nous avons donc réalisé une étude comparative pour neuf mesures de dissimilarités univariées. Cette étude est présentée dans le chapitre 4. Comme cette étude le montre, l'utilisation d'une DTW est adaptée au calcul de dissimilarité entre séries temporelles univariées anesthésiques courtes.

Ainsi, pour un patient réel j , une DTW est calculée pour toutes les sous-séries temporelles univariées présentes dans les motifs \mathcal{M} et $\mathcal{P}(j)$. La dissimilarité multivariée \bar{D} est ensuite calculée comme la moyenne des différentes dissimilarités univariées obtenues par DTW :

$$\bar{D} = \frac{1}{n_i} \sum DTW(\mathcal{V}[\mathcal{E}_{time}(k - n_a), \mathcal{E}_{time}(k)], \mathcal{V}(j)[\mathcal{E}_{time}(k - n_a), \mathcal{E}_{time}(k)])$$

Les patients pré-sélectionnés sont ensuite classés selon leurs scores de similarité \mathcal{S} .

5.4 Simulation du patient virtuel

Pour prédire l'avenir immédiat du patient virtuel, on s'appuie sur l'évolution des patients réels similaires au sein d'une fenêtre particulière, la *fenêtre de prédiction*. Pour chaque patient réel similaire, cette fenêtre est la portion de série temporelle multivariée qui est délimitée, à gauche par l'action qui vient d'être déclenchée, et à droite par le déclenchement de l'action suivante.

La prédiction de l'avenir immédiat du patient virtuel pose deux difficultés. Premièrement, nous ne savons pas sur quelle plage de temps nous devons prédire l'évolution du patient virtuel. Deuxièmement, les fenêtres de prédiction des patients les plus similaires au patient virtuel ne couvrent pas toutes le même intervalle de temps, comme le montre la figure 5.1 où les fenêtres de prédiction sont indiquées en gris.

Afin de prendre en compte ces difficultés, nous proposons trois stratégies de prédiction à court terme du patient virtuel.

5.4.1 Stratégies de prédiction

Notre algorithme de prédiction s'appuie sur la construction d'une série temporelle multivariée. Plusieurs modalités ont été envisagées pour la construction de cette série temporelle selon le nombre de patients les plus similaires utilisés et la méthode qui les associe.

Patient le plus similaire Une première stratégie consiste à prédire l'évolution du patient virtuel en fonction du patient réel le plus similaire. La fenêtre de prédiction est ainsi composée de l'avenir immédiat du patient le plus similaire. La sous-série temporelle multivariée correspondante à cet avenir immédiat est concaténée à la série temporelle multivariée du patient virtuel.

Moyenne des patients similaires Il peut arriver que le patient le plus similaire au patient virtuel offre une fenêtre de prédiction plus courte que celles des autres patients moins similaires. Il est intéressant de chercher à prédire le plus loin possible l'évolution du patient virtuel. Ainsi, on peut chercher pour cela à associer les fenêtres de prédiction des n_s ($n_s > 1$) patients les plus similaires.

Dans cet objectif, pour chaque variable v de \mathcal{V} on produit une sous-série univariée en faisant la moyenne des sous-séries univariées délimitées par les fenêtres de prédiction des patients similaires. Les différentes sous-séries univariées sont ensuite associées pour produire une nouvelle sous-série temporelle multivariée. La sous-série temporelle ainsi obtenue est concaténée à la série temporelle du patient virtuel, comme dans la première stratégie.

Consensus des patients similaires L'utilisation d'une moyenne sur les patients similaires proposée dans la deuxième stratégie est susceptible d'introduire des artefacts dans l'évolution du patient virtuel et par conséquent en réduire le réalisme. Pour une prédiction plus réaliste, une autre stratégie repose sur la création d'une série temporelle multivariée consensus, comme proposé par VAUGHAN et GABRYS [130] et présenté au chapitre 6.

Cette troisième stratégie met ainsi en oeuvre un processus hiérarchique ascendant pour produire un consensus entre les n_s patients similaires. Ce consensus est ensuite utilisé pour la prédiction du patient virtuel.

5.4.2 Évaluation de la qualité de la simulation

Pour évaluer le réalisme des scénarios générés par chacune des différentes stratégies décrites en 5.4.1 nous avons mis en oeuvre une évaluation expérimentale.

Si notre approche de simulation fonctionne correctement alors, pour des conditions d'intervention et une séquence d'événements identique à une séquence réelle donnée, la série temporelle multivariée d'une séquence simulée devrait être très similaire à la séquence réelle. Cette évaluation consiste donc à produire par simulation une série temporelles multivariée en utilisant la même séquence d'événements que la séquence d'un profil anesthésique réel, et à comparer ensuite les séries temporelles réelle et simulée.

5.4.2.1 Protocole

Notre protocole d'évaluation vise à évaluer comment les scénarios générés par chacune des stratégies s'écartent des scénarios réels.

Dans ce but, pour chacune des trois stratégies, et pour chacun des patients d'un jeu de données, nous tirons au hasard un instant t auquel une action a été initiée dans la séquence du patient réel. L'opération chirurgicale est ensuite simulée à partir de cet instant t , en utilisant la même séquence d'événements que celle du patient réel.

Pour chaque patient j , on note $\mathcal{V}_{real}(j)$ la série temporelle multivariée réelle délimitée par l'horodatage t_j , et la fin de l'opération réelle.

$\mathcal{V}_{pred}(j)[k]$ désigne la série temporelle multivariée prédite à partir de t_j jusqu'à la fin de l'opération, lorsque la stratégie k est utilisée pour la simulation.

$\bar{D}_k(j)$ est la dissimilarité \bar{D} calculée entre les séries temporelles multivariées $\mathcal{V}_{real}(j)$ et $\mathcal{V}_{pred}(j)[k]$. Un score de similarité est produit par normalisation, comme suit :

$$Sim_k(j) = \frac{\bar{D}_{k,max} - \bar{D}_k(j)}{\bar{D}_{k,max} - \bar{D}_{k,min}}$$

où $\bar{D}_{k,max}$ et $\bar{D}_{k,min}$ sont respectivement les valeur maximale et minimale de \bar{D}_k obtenues sur l'ensemble des patients.

\mathcal{D}_{Sim_k} représente la distribution des similarités normalisées calculées sur l'ensemble des patients lorsque la stratégie k est prise en compte.

Les différentes stratégies k expérimentées sont dénommées par les noms suivants :

- GS_BEST : patient le plus similaire ,
- GS_AVG_10 : moyenne des 10 patients les plus similaires,
- GS_CONS_10 : consensus des 10 patients les plus similaires.

5.4.2.2 Données expérimentales

L'utilisation de données réelles est hautement contrainte pour des raisons de protection des données de santé. Dans le cadre de cette expérimentation nous avons donc utilisé des données réalistes produites par le générateur DBLBS (voir 3.3) pour une chirurgie spécifique. Le jeu de données est ainsi composé de 1000 profils anesthésiques correspondant à la cure d'une hernie inguinale sous coelioscopie, pour un patient de sexe masculin de 30 ans et sans antécédents médicaux.

5.4.2.3 Implémentation et paramétrage

L'algorithme 3 a été implémenté en Python 3.10.0, dans le cadre du simulateur présenté en 7.3. L'apprenant ayant été remplacé par la séquence d'événements de chaque patient réel. Sur la recommandation d'un expert, n_a a été fixé à 4 pour cette intervention. La détermination du score de similarité défini en 5.3.4 implique l'utilisation de coefficients. Ces coefficients ont été fixés comme suit : $\lambda_1 = \lambda_2 = 1$, $\lambda_3 = 2$.

Il peut être difficile de trouver des patients très similaires au patient virtuel. Pour résoudre ce problème potentiel, nous avons réalisé trois recherches de similarité en parallèle,

pour trois modalités de relaxation : basse τ_b , moyenne τ_m et haute τ_h . Ces modalités sont caractérisées par les trois pourcentages de relaxation suivants : $\tau_b = 1\%$, $\tau_m = 5\%$ et $\tau_h = 10\%$.

La mesure de dissimilarité multivariée \bar{D} est obtenue comme la moyenne des dissimilarités univariées, elles-mêmes obtenues grâce à la DTW.

5.4.2.4 Résultats

Une première analyse visuelle nous permet de dire que les séries temporelles simulées sont proches de la série temporelle réelle. La figure 5.4 montre une série temporelle réelle et deux sous-séries simulées à partir du même instant t_j en utilisant différentes stratégies. La figure 5.5 présente les mêmes séries temporelles avec un focus sur la période de l'induction.

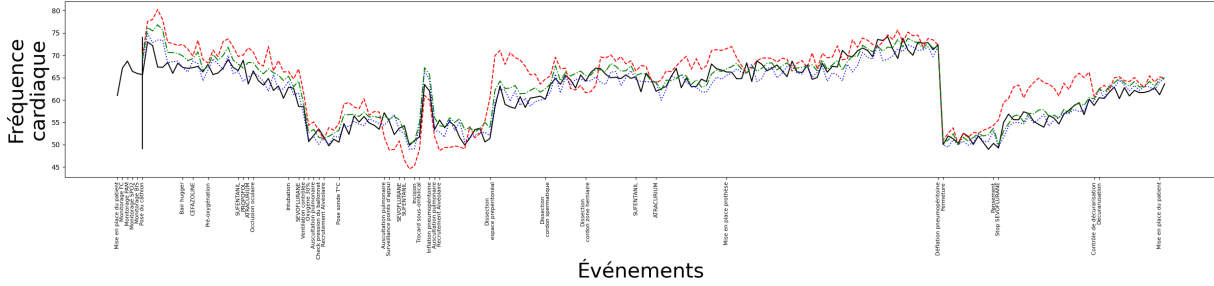


Figure 5.4 – Série temporelle de la fréquence cardiaque et les trois sous-séries simulées. La simulation a été réalisée à partir de l'action « mise en place de la voie veineuse périphérique » (trait vertical noir). Ligne noire : série temporelle réelle ; ligne rouge : stratégie GS_BEST ; ligne verte : stratégie GS_AVG_10 ; ligne bleue : stratégie GS_CONS_10.

La figure 5.6 présente sous la forme de diagrammes en violons les distributions \mathcal{D}_{Sim_k} pour les trois stratégies GS_BEST, GS_AVG_10 et GS_CONS_10. La table 5.1 donne les différentes caractéristiques de ces distributions.

Nous pouvons constater que se baser simplement sur le patient réel le plus similaire au patient virtuel pour chaque prédiction à court terme, comme le fait la stratégie GS_BEST, donne la similarité la plus faible (Moyenne = 0,694, Médiane = 0.721). La stratégie GS_CONS_10 produit les simulations les plus similaires aux patients réels (Moyenne = 0,817, Médiane = 0.820). Avec une moyenne égale à 0,783 et une médiane de 0,779, la stratégie GS_AVG_10 montre un comportement intermédiaire, mais avec une importante dispersion des valeurs sous la forme d'*outliers*. Néanmoins, chacune de ces stratégies est capable de générer des séries temporelles assez peu différentes de celles de vrais patients.

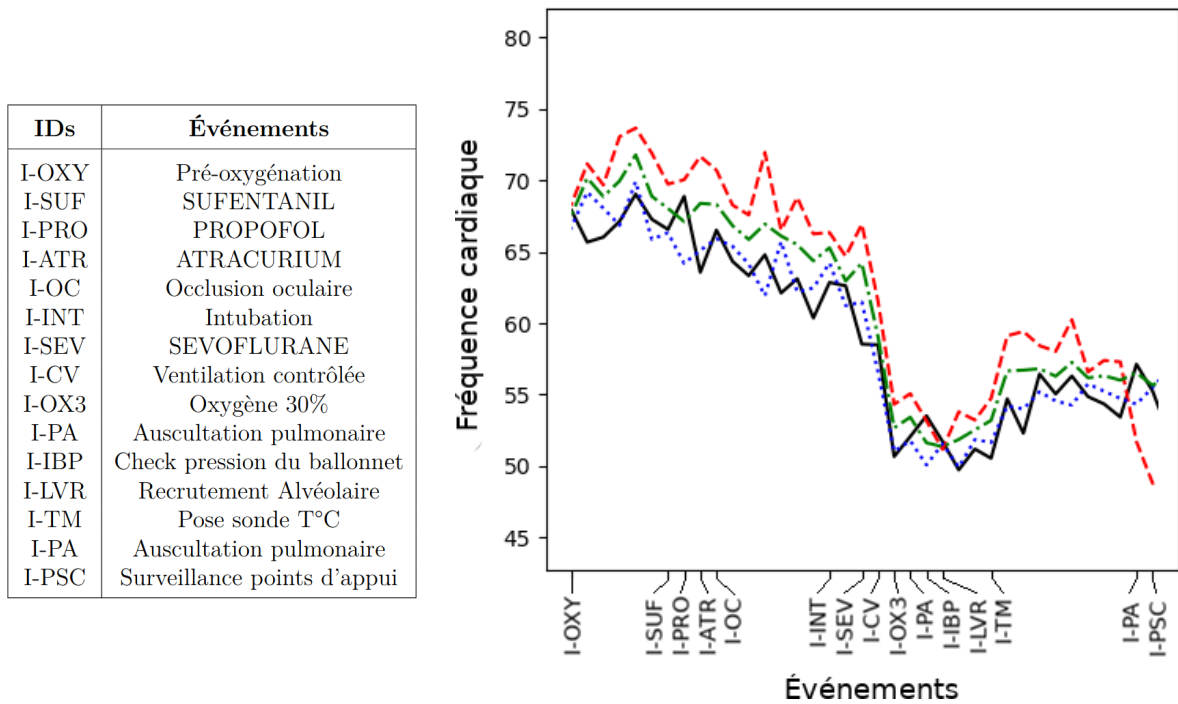


Figure 5.5 – Zoom sur la période de l’induction pour la série temporelle et les trois sous-séries simulées de la figure 5.4. Ligne noire : série temporelle réelle ; ligne rouge : stratégie GS_BEST ; ligne verte : stratégie GS_AVG_10 ; ligne bleue : stratégie GS_CONS_10.

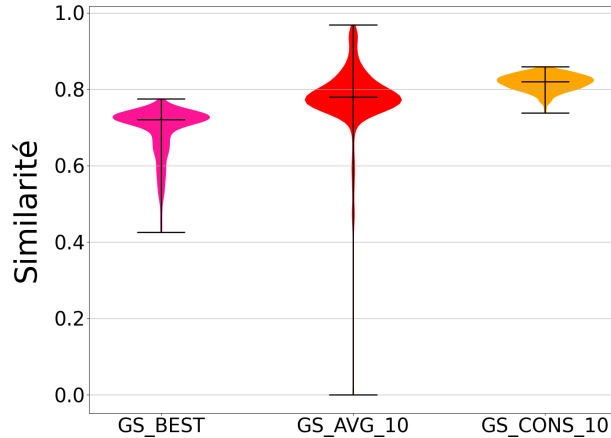


Figure 5.6 – Distribution des similarités entre séquence simulée et séquence réelle, pour les trois stratégies. Chaque violon décrit la distribution de la mesure de similarité Sim_k entre les séries temporelles multivariées réelles et simulées, pour les 1000 patients réels du jeu de données, pour une stratégie k donnée. Les barres supérieures et inférieures donnent l’étendue de la distribution, la barre centrale indique la médiane

La table 5.2 compare les distributions par couples de stratégies en présentant les valeurs des tests et p-valeurs de deux tests statistiques (Kolmogorov-Smirnov et Wilcoxon / Mann-Whitney), ainsi que la distance euclidienne.

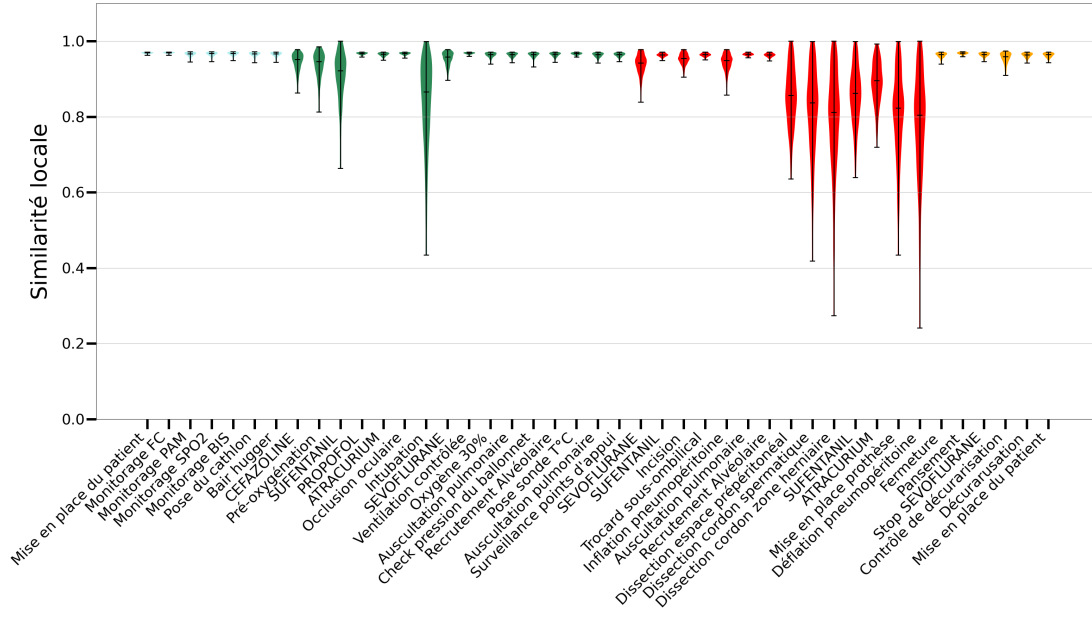
Stratégie	Moyenne	Médiane	Écart-type
GS_BEST	0,694	0.721	0.0604
GS_AVG_10	0,783	0.779	0.0697
GS_CONS_10	0,817	0.820	0.0205

Table 5.1 – Caractéristiques des différentes distributions \mathcal{D}_{Sim_k}

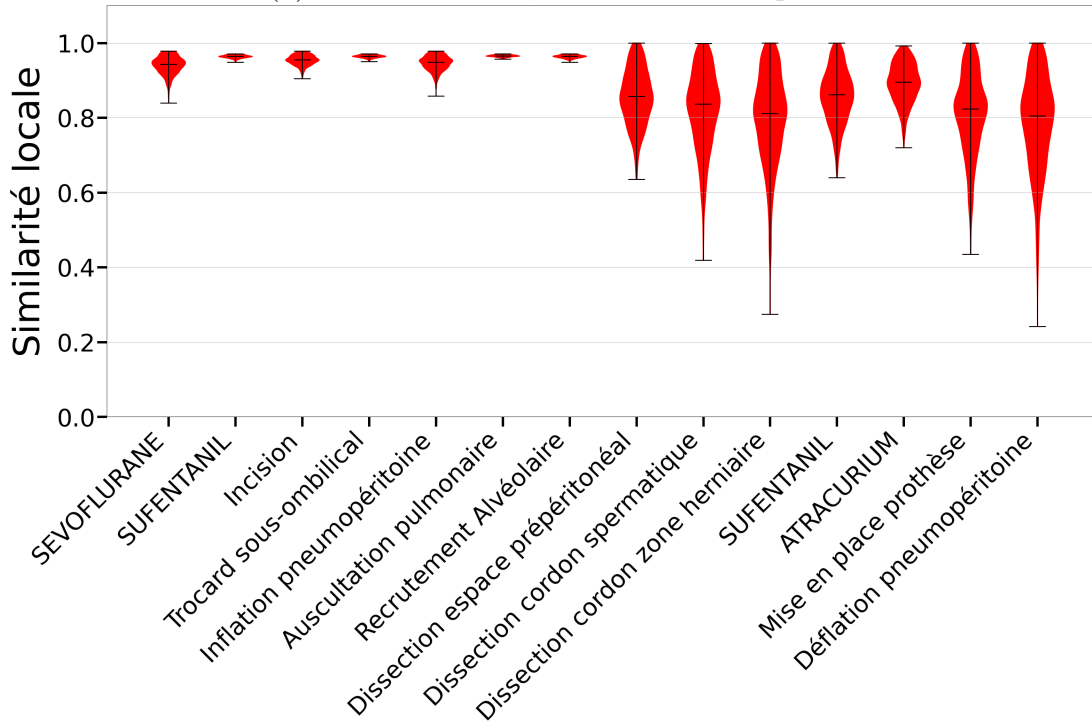
Stratégie	Stratégie	KS	WMW	DE
GS_BEST	GS_CONS_10	0.991 (0.0)	161.0 (0.0)	4.36
GS_BEST	GS_AVG_10	0.82 (0.0)	56 221.0 ($2.54e^{-245}$)	3.81
GS_AVG_10	GS_CONS_10	0.52 ($1.20e^{-125}$)	212 346.0 ($2.76e^{-95}$)	2.47

Table 5.2 – Comparaisons par paires des distributions pour les stratégies, selon deux tests statistiques et la distance euclidienne. Sont donné les valeurs des tests et entre parenthèses les p-valeurs. KS : test de Kolmogorov-Smirnov. WMW : test de Wilcoxon / Mann-Whitney. DE : distance euclidienne.

Enfin, la Figure 5.7 présente les distributions des similarités pour les sous-séries temporelles entre une action et l'action suivante pour la stratégie GS_CONS_10. Cette figure met en évidence que dans le cas de cette intervention, la cure d'une hernie inguinale sous coelioscopie, la simulation est réalisée avec un haut degré de réalisme pour la majorité de l'opération. Toutefois, des variations sont observables pour certaines actions de l'induction ou pendant l'entretien, comme le montre la figure 5.7 (b) correspondant à l'entretien de l'anesthésie. Les mêmes tendances sont observées sur toutes les stratégies et peuvent s'expliquer par la diversité dans les séries temporelles des patients réels. Les figures correspondantes sont présentées en annexe D.1.



(a) GS_CONS_10, Intervention complète



(b) GS_CONS_10, Entretien

Figure 5.7 – Distributions des similarités pour les sous-séries temporelles de chaque action pour la stratégie GS_CONS_10.

Conclusion

La construction d'un simulateur, permettant de jouer des scénarios de simulation pour la formation des soignants en anesthésie, doit s'appuyer sur une méthode pouvant simuler de manière réactive un patient virtuel qui soit similaire à un patient réel pour assurer le réalisme de la simulation. En nous appuyant sur une approche par raisonnement à base de cas, nous nous affranchissons de la nécessité d'un apprentissage de modèles complexes, imposé par les approches d'apprentissage machine, pour simuler un patient virtuel réactif aux actions d'un apprenant.

Notre approche s'appuie sur une recherche de patients similaires au patient virtuel au moment de l'occurrence d'une action anesthésique. Ces patients similaires sont ensuite utilisés pour prédire l'évolution du patient virtuel jusqu'à la prochaine action réalisée par l'apprenant.

Notre approche permet l'implémentation de plusieurs stratégies de prédiction du patient virtuel, soit par l'utilisation de la série temporelle multivariée du patient réel le plus similaire ou la construction d'une série temporelle multivariée moyenne ou consensus. L'évaluation de la qualité de la simulation produite par ces trois stratégies met en avant une bonne qualité de simulation de cette approche avec une supériorité de la stratégie par construction de séries temporelles multivariées consensus.

Cette approche, et ces stratégies, nous permettent ainsi de proposer une méthode de simulation réactive aux actions d'un apprenant qui permet l'évolution d'un patient virtuel réaliste rendant possible un simulateur de pratique médicale qui n'a jusqu'alors jamais été proposé.

Notre proposition est de plus applicable à toute situation où des séries temporelles annotées d'événements sont disponibles, pour générer des prédictions réalistes à la volée pour des données temporelles multivariées.

REPRÉSENTATIONS SYNTHÉTIQUES

Introduction

Le but des présents travaux est d'utiliser les données de santé issues des profils anesthésiques pour construire des scénarios réactifs. Il faut pouvoir donner un support à ces scénarios afin de compiler dans une seule structure de données l'ensemble des éléments nécessaires pour jouer les scénarios. Dans un chapitre précédent, nous avons présenté une formalisation informatique de l'anesthésie et sa construction. En nous appuyant sur cette formalisation, nous pouvons représenter un ensemble de séquences d'événements par un graphe orienté. Toutefois, la taille du graphe produit est un enjeu important pour la simulation, afin qu'il ne soit pas «trop grand».

La représentation synthétique des traces d'événements d'un ensemble de patients, sous la forme d'un graphe orienté, peut être le support des scénarios d'apprentissage. Cette représentation synthétique permet à la fois un contrôle pédagogique du scénario et la simulation : elle contient toutes les informations utiles pour la prédiction du patient virtuel en réaction aux actions des apprenants. L'utilisation de ces éléments par le simulateur sera présenté au chapitre suivant.

La première partie de ce chapitre s'intéresse à la construction d'un graphe représentant un ensemble de séquences d'événements.

La seconde partie présente la construction d'une cohorte de patients de taille et de diversité adaptées à une représentation par un graphe.

La partie suivante présente une méthode pour construire une représentation synthétique de profils anesthésiques et son utilisation à des fins de simulation.

La dernière partie propose une réflexion sur l'opportunité que représente une représentation synthétique pour l'anonymisation de données.

6.1 Représentation d'un ensemble de séquences d'événements par un graphe orienté

Pour pouvoir réaliser des scénarios qui s'appuient sur les données, il est nécessaire d'accéder à plusieurs profils anesthésiques. En effet, un scénario basé sur un unique profil anesthésique ne permettrait pas de proposer à un apprenant plusieurs choix d'actions, mais seulement de rejouer l'histoire du patient possédant ce profil anesthésique. Il est donc nécessaire de proposer une structure de données qui permet de contenir plusieurs profils anesthésiques, de manière structurée et adaptable, et qui puisse être utilisée comme support d'un scénario de simulation.

Un profil anesthésique est composé de deux éléments :

- une séquence d'événements, qui contient les événements ordonnancés enregistrés lors de l'intervention,
- une série temporelle multivariée qui décrit les variations des paramètres physiologiques du patient, enregistrées pendant toute l'intervention. Chacune des dimensions de la série temporelle correspond à un paramètre physiologique (par exemple, la fréquence cardiaque).

Comme le montre la section 3.1.2, toutes les anesthésies se structurent de manière similaire, ce qui se reflète dans la séquence d'événements. Pour une intervention spécifique, une part de la variabilité entre séquences d'événements est due à la variabilité temporelle : d'une séquence à l'autre, les intervalles de temps entre deux événements successifs sont différents. La variabilité entre séquences d'événements est plus rarement due à une variabilité de l'ordonnancement des événements ou aux événements eux-mêmes, ceux-ci étant peu diversifiés.

Les séquences d'événements sont définies par une suite ordonnée d'événements. Ainsi, la séquence S dans l'exemple ci-dessous est composée de plusieurs événements E :

$$T = (E_1, \dots, E_n).$$

Au sein d'une séquence d'événements, les occurrences d'un même événement sont considérées comme des événements différents. C'est par exemple le cas pour l'administration d'un médicament qui est répétée plusieurs fois pendant l'intervention.

Ancres Certains événements sont toujours, ou presque toujours, présents dans un ensemble de séquences à la même position ou à des positions proches, dans chacune des séquences de l'ensemble. Ce type particulier d'événement est dénommé *ancree*.

Ainsi la séquence S peut être composée d'événements E et d'ancres A :

$$S = (A_1, E_1, \dots, A_i, \dots, E_i, A_n, \dots, E_m).$$

Les séquences d'événements d'une intervention comportent toutes systématiquement des événements descripteurs qui indiquent le début et la fin de l'intervention. Ainsi, une séquence S composée de m événements E et de n ancres A est toujours bornée par des ancres et adopte cette forme :

$$S = (A_1, E_1, \dots, A_i, \dots, E_i, \dots, A_n).$$

Interancree La séquence des événements apparaissant entre deux ancres est dénommée *interancree*.

Grappe Un ensemble de séquences d'événements peut être représenté par un graphe orienté, où tous les événements sont des nœuds et l'ordonnancement entre ces nœuds est représenté par les arcs.

Par exemple, pour l'ensemble de séquences E_s suivant, on peut construire le graphe G de la figure 6.1.

$$E_s = (S_1, S_2, S_3)$$

$$S_1 = (A_1, E_2, E_3, A_4)$$

$$S_2 = (A_1, B_2, E_3, A_4)$$

$$S_3 = (A_1, E_2, B_3, A_4)$$

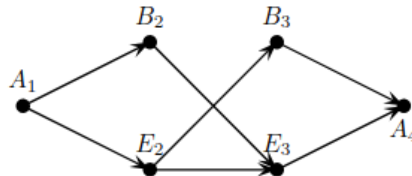


Figure 6.1 – Graphe G représentant l'ensemble de séquences E_s

Dans un contexte de simulation, l'ensemble des nœuds fils d'un nœud correspond à l'ensemble des choix qui sont offerts à l'apprenant pour cette étape précise de l'intervention.

Si l'on utilise un graphe «trop grand» comme support à la simulation d'un scénario, on risque de se trouver confronté à plusieurs difficultés. Si l'on souhaite que l'apprenant travaille une compétence spécifique dans un scénario, *via* une situation d'apprentissage particulière, mais qu'on lui offre une importante diversité de choix, alors il a la possibilité d'aller vers une autre situation d'apprentissage. Avec un graphe «trop grand», le contrôle pédagogique du scénario est difficile. De plus, d'un point de vue technique, la mémorisation d'un graphe «trop grand» peut occuper une quantité de mémoire très importante et compliquer une simulation réactive.

Dans la suite de ces travaux, le caractère «trop grand» d'un graphe, qui implique une intrication et une diversité importantes, sera qualifié de *complexe*¹.

La complexité du graphe dépend directement de la composition de l'ensemble de séquences à partir duquel il est construit. Ce graphe sera d'autant plus complexe que les séquences sont dissimilaires entre elles. Inversement, le graphe sera d'autant plus «linéaire» que les séquences sont similaires entre elles.

La diversité d'un ensemble de séquences détermine la complexité du graphe qui le représente. En amont de la construction de ce graphe, il est primordial de sélectionner un ensemble de séquences approprié pour notre objectif de simulation, et de diversité permettant de contrôler celle du graphe.

Dans cette thèse, nous souhaitons utiliser le graphe associé à un ensemble de séquences d'événements, comme support de construction de scénarios d'apprentissage. Nous souhaitons sélectionner une ou des sous-cohortes de patients dont l'ensemble de séquences soit utilisable pour notre objectif de simulation ; le contrôle de la complexité de l'ensemble de séquences associées à une telle sous-cohorte doit permettre de contrôler la complexité du graphe construit à partir de ces séquences.

1. Dans le sens de sa racine latine *complexus*, ce qui est imbriqué ensemble.

6.2 Construction de cohortes de profils anesthésiques

Pour construire un scénario basé sur les données, par l’entremise d’un graphe utilisé comme support, la première étape consiste à construire une cohorte de profils anesthésiques respectant les critères suivants. D’une part, ces profils doivent être similaires pour permettre un travail pédagogique sur un sujet précis, et permettre une simulation réaliste. D’autre part, les profils doivent être suffisamment diversifiés afin de permettre des choix à l’apprenant et de représenter la diversité des situations présentes dans une intervention.

Une première étape triviale de pré-sélection des profils anesthésiques consiste à filtrer les patients réels selon les critères suivants :

- À une époque donnée, le patient réel a subi la procédure chirurgicale qui fait l’objet de la simulation. La codification CCAM² est utilisée pour ce filtrage.
- Le patient réel présente le même sexe et les mêmes classes d’âge, de poids et de taille que le patient virtuel à simuler.
- Le patient réel présente les mêmes antécédents médicaux (obésité, diabète ou aucun antécédent médical) que le patient simulé.

La pré-sélection de patients ayant subi la même intervention (code CCAM identique) apportera de l’homogénéité à la cohorte. Une pré-sélection plus inclusive, admettant des patients ayant subi le même type d’intervention (code CCAM proche), permet plus de variabilité.

Cependant, le principal critère de sélection reste la diversité dans le déroulement des interventions. Avec une sélection basée sur des critères trop inclusifs, l’effectif de profils anesthésiques pourrait être trop important et apporter trop de diversité pour la simulation. Inversement, une sélection avec des critères trop restrictifs risque d’apporter trop peu de profil anesthésiques et donc trop peu de diversité.

Nous aimerions de plus connaître *a priori* la complexité du graphe associé à une cohorte pré-sélectionnée comme expliqué précédemment, sans construire ce graphe.

2. Les différents actes médicaux sont classés par l’assurance-maladie dans la *Classification Commune des Actes Médicaux* (CCAM) qui affecte à chaque acte une définition et un tarif de remboursement. On y retrouve les interventions qui nécessitent une anesthésie générale. Chaque acte est codé par un identifiant unique. Ainsi, le code LMMC001 correspond ainsi à la *cure bilatérale d’une hernie inguinale sous cœlioscopie avec pose de prothèse* alors que le code LMMC002 correspond à la *cure unilatérale d’une hernie inguinale sous cœlioscopie avec pose de prothèse*

La stratégie proposée dans le cadre de ces travaux enchaîne deux étapes :

- une étape de clustering de séquences d'événements générant un partitionnement des patients d'une cohorte initiale,
- une étape de regroupement ascendant hiérarchique initié à partir des clusters obtenus lors de la première étape ; notre approche contrôle le regroupement successif de deux clusters au moyen d'un critère permettant d'estimer la complexité du graphe qui serait généré par le regroupement de ces deux clusters.

Cette approche en deux temps permet de contrôler la diversité au sein des clusters. L'utilisation d'un regroupement ascendant hiérarchique fournit la possibilité, *a posteriori*, de sélectionner dans l'arbre de regroupement différents clusters de taille variable et dont la diversité est adaptée à la simulation de différents scénarios d'apprentissage.

Lorsqu'on s'élève dans les niveaux de la hiérarchie produite par le regroupement ascendant hiérarchique, la complexité du graphe associé à chaque regroupement de clusters croît. L'agglomération successive de deux clusters est autorisée jusqu'à ce qu'on atteigne un seuil minimal de complexité. Au contraire des approches de classification ascendante hiérarchique standard, nous initions notre approche en nous appuyant sur un partitionnement initial.

Nous avons donc besoin des outils suivants : (i) une première méthode de clustering, (ii) une mesure de la diversité d'un ensemble de séquences permettant d'évaluer la complexité du graphe représentatif de cet ensemble de séquences, (iii) une méthode de type clustering ascendant hiérarchique équipée d'un critère permettant d'estimer la complexité d'un graphe sans construire le graphe.

Le reste de cette section se décline ainsi : dans la partie 6.2.1, nous adaptons l'algorithme de KOHONEN [131] pour la construction de cartes auto-organisées au cas de séquences d'événements. La partie 6.2.2 présente des expérimentations dont l'objectif est de déterminer, pour un ensemble de séquences \mathcal{E}_s , quels critères C_{sq} sur les séquences et quels critères C_g sur un graphe sont corrélés, pour les couples $(C_{sq}(\mathcal{E}_s), C_g(\mathcal{E}_s))$ formés en considérant 20 candidats pour C_{sq} et 12 candidats pour C_g . La partie 6.2.3 décrit ensuite brièvement une méthode de clustering ascendant contrôlée par un critère de diversité des séquences d'événements (C_{sq}). Finalement, la partie 6.2.4 montre les résultats obtenus par application de cette stratégie à une cohorte de 1000 patients réalistes, et pour divers seuils de complexité.

6.2.1 Partitionnement d'une cohorte de profils anesthésiques

La littérature du process mining montre que la majorité des approches de clustering sont applicables aux traces, sous réserve d'avoir la bonne approche de vectorisation [132].

Dans le cadre de nos travaux, nous avons estimé qu'il était préférable d'utiliser une approche de clustering qui utilise une phase d'apprentissage, car les approches sans apprentissage présentent généralement le défaut d'être asymétriques [133] : si une classe est sur-représentée, alors les faux positifs pour cette classe sont plus nombreux. L'apprentissage profond a été exclu du fait de la quantité de données nécessaire à la phase d'apprentissage et parce que les méthodes relevant de l'apprentissage profond requièrent des temps de calculs importants. Ces caractéristiques n'étaient pas compatibles avec la puissance de calcul disponible et les effectifs de profils anesthésiques parfois faibles pour une intervention.

Nous avons donc choisi les cartes auto-organisées (SOM), qui nécessitent une phase d'apprentissage en un temps de calcul acceptable, et sont par ailleurs très bien documentées dans la littérature comme outil de clustering.

SONG *et al.* [134] ont montré la pertinence des cartes auto-organisées pour le clustering de séquences d'événements.

6.2.1.1 Repères sur le clustering par cartes auto-organisées

Les cartes auto-organisées³ sont un type de réseau de neurones artificiels utilisés en apprentissage non supervisé.

Elles permettent de produire une cartographie d'un ensemble de données dans un espace réel. Cette cartographie peut ensuite être utilisée dans un objectif de partitionnement ou de classification.

Principe des cartes auto-organisées Le principe des cartes auto-organisées a été proposé par Teuvo KOHONEN [131]. Cette méthode repose sur un réseau de neurones et met en œuvre :

- une phase d'*initialisation*, où un réseau de neurones, dont la taille est proportionnelle à la taille du jeu de données et où chaque neurone est représenté par un vecteur consensus qui est initialisé avec des valeurs aléatoires,

3. Ou cartes autoadaptatives, cartes de KOHONEN ou encore SOM (*Self-organizing map*).

- une phase d'*apprentissage* pendant laquelle les différents éléments d'un ensemble \mathcal{E}_s sont affectés successivement à un neurone de la carte dont ils sont le plus similaires. Ce neurone et ses voisins s'en voient modifiés proportionnellement.
- une phase de *partitionnement* ou de classification, où les éléments de l'ensemble à partitionner \mathcal{E}_s sont associés une dernière fois au neurone de la carte dont ils sont le plus similaires.

Le cœur de la méthode consiste à faire évoluer chaque neurone de la carte à chaque fois qu'un élément de \mathcal{E}_s lui est associé. Les activations successives d'un neurone par les divers éléments de \mathcal{E}_s permettent de mettre à jour le vecteur consensus porté par le neurone.

L'évolution des neurones est conditionnée par un *taux d'apprentissage*. L'impact sur les neurones voisins d'un neurone activé est conditionné par un *rayon*. L'ensemble \mathcal{E}_s est présenté à la carte au cours de plusieurs cycles d'apprentissage. Les valeurs du taux d'apprentissage et du rayon décroissent au fil des différents cycles d'apprentissage.

Apprentissage Les différentes séquences de l'ensemble \mathcal{E}_s sont présentées de manière aléatoire pendant des cycles d'apprentissages successifs.

Pour chacune des itérations d'un cycle d'apprentissage, on cherche pour une séquence S , issue de l'ensemble \mathcal{E}_s de taille n , le neurone k le plus similaire. Pour cela, on calcule pour chacun des neurones une similarité⁴ S_{map} , entre le vecteur v_s associé à la séquence S (défini en 6.2.1.2) et le vecteur v_i du neurone k_i (défini en 6.2.1.2). Le neurone sélectionné est celui qui maximise la similarité S_{map} .

Une fois la séquence S affectée au neurone k_{max} le plus similaire, la carte est mise à jour et chacun des autres neurones k_i de la carte évolue selon le taux d'apprentissage L_r . Les neurones à proximité du neurone k_{max} sont plus impactés que les neurones plus éloignés. Cet impact selon la distance est réglé par le rayon $R(i, k_{max})$.

$v_i = v_i + L_r().R(i, k_{max}).(v_s - v_i)$ avec v_i le vecteur associé à un neurone k_i de la carte, v_s le vecteur associé à la séquence S , $L_r()$ le taux d'apprentissage et $R(i, k_{max})$ le rayon.

Les fonctions $L_r()$ et $RR(i, k_{max})$ sont définies par des fonctions gaussiennes [131] qui dépendent des valeurs initiales définies avant l'apprentissage, $L_{initial}$ et $R_{initial}$, ainsi que du nombre h de séquences qui ont déjà été ajoutées à la carte au fil des cycles. Le rayon prend en compte les coordonnées x_i, y_i du neurone k_i mis à jour et les coordonnées x_k, y_k du neurone k_{max} activé.

4. Généralement à partir d'une distance euclidienne.

$$L_r() = \frac{L_{initial}}{1 + \frac{h}{n}}$$

$$(i, k) = e^{\frac{-1}{2} \cdot \frac{R_{initial}}{1 + \frac{h}{n}} \cdot (x_i - x_k)^2 + (y_i - y_k)^2}$$

avec $n = |\mathcal{E}_s|$.

Qualité de l'apprentissage À chaque cycle d'apprentissage, la carte évolue. Pour déterminer la qualité d'une carte, on peut utiliser l'*erreur de quantification*, qui informe sur la qualité de l'apprentissage [135, 136], et l'*erreur topologique*, qui informe sur la qualité de la projection et la conservation de la topologie de la carte [137, 138].

L'erreur de quantification Q_E est la différence moyenne entre le vecteur v_s associé à la séquence S présentée à la carte, et le vecteur v_k du neurone k_{max} activé par v_s . Cette erreur évalue la précision des données représentées dans la carte par rapport aux données d'entrée. Ainsi, une carte de bonne qualité minimise cette valeur [135]. Q_E est définie comme suit :

$$Q_E = \frac{1}{n} \sum_{i=1}^n ||v_s - v_k(S)||$$

L'erreur topologique T_E représente la distance entre le 1^{er} et le 2nd neurones, respectivement $v_1(S)$ et $v_2(S)$, les plus similaires à la séquence S présentée. Plus la distance entre ces neurones est importante, plus la topologie de la carte est déformée. Il est donc préférable que cette valeur soit minimisée [137, 138]. T_E est définie par :

$$T_E = \frac{1}{n} \sum_{i=1}^n D(v_1(S), v_2(S))$$

où $D(v_1(S), v_2(S)) = 1$ si $v_1(S)$ et $v_2(S)$ sont voisins, $D(v_1(S), v_2(S)) = 0$ sinon.

Partitionnement À l'issue de la phase d'apprentissage, les séquences de l'ensemble \mathcal{E}_s sont présentées à la carte où elles sont affectées au neurone dont elles sont le plus similaires. À l'issue de cette phase de partitionnement, chaque ensemble de séquences «présent dans un neurone» est considéré comme un cluster.

Jeux	Nombre total de séquences	Nombre de classes	Classes	Nombre de séquences
A	1000	2	A10[2-8%]A10[2-8%]A	500
			A10[92-98%]A10[92-98%]A	500
B	1500	3	A10[2-8%]A10[2-8%]A	500
			A10[47-53%]A10[47-53%]A	500
			A10[92-98%]A10[92-98%]A	500

Table 6.1 – Deux jeux expérimentaux, de deux et trois classes respectivement, utilisés pour le partitionnement d’une cohorte de profils anesthésiques par carte auto-organisée. La colonne *Classes* indique la structure des séquences d’événements qui composent les différentes classes abstraites du jeu, où A est une ancre, 10 est le nombre d’événements dans une interancre et la probabilité d’apparition des événements dans cette interancre est précisée entre crochets.

6.2.1.2 Partitionnement d’une cohorte de profils anesthésiques par cartes auto-organisées

Nous avons complété les travaux de SONG *et al.* [134] par des expérimentations sur des séquences d’événements abstraites, où nous contrôlons trois types de paramètres réglant la diversité d’un ensemble de séquences. Ces expérimentations consistent à construire des jeux de données expérimentaux composés de plusieurs classes de séquences d’événements abstraites puis à retrouver les classes d’origine par l’algorithme de KOHONEN.

Jeux de données Nous avons construit plusieurs jeux de base composés de 1000 séquences d’événements. Nous avons fait varier plusieurs paramètres :

- le nombre d’ancres (3 modalités : 3, 5 et 10 ancres),
- le nombre d’événements dans une interancre (3 modalités : 10, 15 et 20 événements),
- la probabilité de présence d’un événement dans une interancre (5 modalités : 2-8%, 24-30%, 47-53%, 70-76% et 92-98%).

À partir de chacun de ces jeux de base, un tirage aléatoire des séquences a été réalisé pour constituer 50 jeux expérimentaux composés de 2 à 3 classes dans plusieurs proportions. Deux jeux expérimentaux sont présentés en table 6.1. Une sélection de jeux expérimentaux est donnée en annexe E.1.

Adaptation de l'algorithme de Kohonen au cas de séquences d'événements

Les séquences, composées d'une suite d'événements, ne se prêtent pas facilement à la construction de vecteurs consensus, un élément central des cartes auto-organisées.

Il est possible de représenter une séquence par un vecteur contenant les similarités entre la séquence et les autres séquences de l'ensemble auquel elle appartient. Une séquence est ainsi décrite de manière relative vis-à-vis des autres séquences de l'ensemble. Une séquence S , issue d'un ensemble de n séquences \mathcal{E}_s , est ainsi représentée par un vecteur v_s défini comme :

$$v_s = (S_{im}(S, S_1), S_{im}(S, S_i), \dots, S_{im}(S, S_n)) \text{ où, } S_{im} \text{ désigne une mesure de similarité.}$$

Dans le cadre de notre expérimentation, nous avons considéré trois mesures de similarité : l'indice de JACCARD, le score d'alignement de SMITH-WATTERMANN, et la moyenne de ces deux similarités.

Vecteur représentant un neurone v_i de la carte Cette approche de vectorisation des séquences permet la création, dans chacun des neurones k de la carte, d'une séquence consensus, qui prend la forme d'un vecteur v_k dont chacune des composantes c est un coefficient correspondant à la participation relative, à cette séquence consensus, d'une séquence S_i d'un ensemble de n séquences.

$$v_k = (c(S_1), c(S_i), \dots, c(S_n)).$$

Dimensionnement de la carte Le nombre de neurones théoriques N_{th} de la carte est déterminé en fonction du nombre de séquences n dans le jeu de données, comme le recommande la littérature [139] :

$$N_{th} = 5\sqrt{n}.$$

Le nombre réel de neurones N_r est adapté pour pouvoir former une carte carrée composée de i rangs et j colonnes, comme suit :

$$\begin{aligned} i &= \lceil \sqrt{N_{th}} \rceil \\ j &= \lceil \frac{N_{th} - (N_{th} \% i)}{i} + 1 \rceil \\ N_r &= i * j. \end{aligned}$$

Chaque neurone k est composé d'un vecteur v_k qui contient n valeurs réelles, les coefficients c pour chacune des séquences du jeu de données. Avant la phase d'apprentissage, tous les coefficients c sont initialisés par des valeurs aléatoires réelles entre 0 et 1.

Paramétrage La phase d'apprentissage nécessite trois paramètres principaux : un taux d'apprentissage L_r , un rayon R , et enfin un nombre de cycles d'apprentissage.

Les valeurs utilisées dans le cadre de cette expérimentation sont présentées dans la table 6.2.

Paramètre	Valeur
$L_{initial}$	0.5
$R_{initial}$	1.5
Nombre de cycles d'apprentissage	10 000

Table 6.2 – Paramètres utilisés dans le cadre de l'expérimentation. Les valeurs ont été déterminées à partir de la littérature ou expérimentalement

Qualité de l'apprentissage Pour obtenir une carte de bonne qualité à l'issue de l'apprentissage, on cherche à minimiser l'erreur de quantification et à maintenir l'erreur topologique la plus faible possible.

L'erreur topologique augmente progressivement avec le nombre de cycles. Il est donc préférable d'utiliser une valeur relativement basse.

Nous avons déterminé que 10 000 cycles d'apprentissage sont suffisants pour réduire l'erreur de quantification tout en maintenant une erreur topologique faible. La détermination expérimentale du nombre de cycles est présentée en annexe E.2.

Qualité des clusters Dans ce mémoire, nous avons utilisé 3 critères pour évaluer la qualité des clusters. La distance intra-cluster, mesure la distance entre chacun des points x_i d'un cluster k avec le centroïde c_k de ce cluster :

$$d_{intra}(k) = \frac{\sum_{i=1}^{N_k} \|x_i - c_k\|}{N_k}.$$

La distance inter-cluster de deux clusters k et l mesure la distance entre les deux points moyens, ou centroïdes, c_k et c_l de ces clusters :

$$d_{inter}(k, l) = \|c_k - c_l\|.$$

L'indice de Davies-Bouldin BD d'un partitionnement X , composé de N_c clusters i, \dots, j , mesure la similarité moyenne entre chaque cluster et celui qui lui est le plus similaire, *via* leurs centroïdes c_i et c_j [140]. Cet indice permet d'identifier des ensembles de clusters qui sont compacts et bien séparés. Il est défini comme suit :

$$BD(X) = \frac{1}{N_c} \sum_{i=1}^{N_c} \max \left\{ \frac{d_{intra}(k) + d_{intra}(l)}{d_{inter}(k, l)} \right\}$$

Résultats du partitionnement À l'issue de la phase d'apprentissage, le partitionnement d'un jeu \mathcal{E}_s est obtenu par présentation de chacune des séquences de \mathcal{E}_s à la carte. La figure 6.2 présente le résultat de deux partitionnements, pour les jeux présentés en 6.1, en distinguant les différentes classes d'origine par des couleurs.

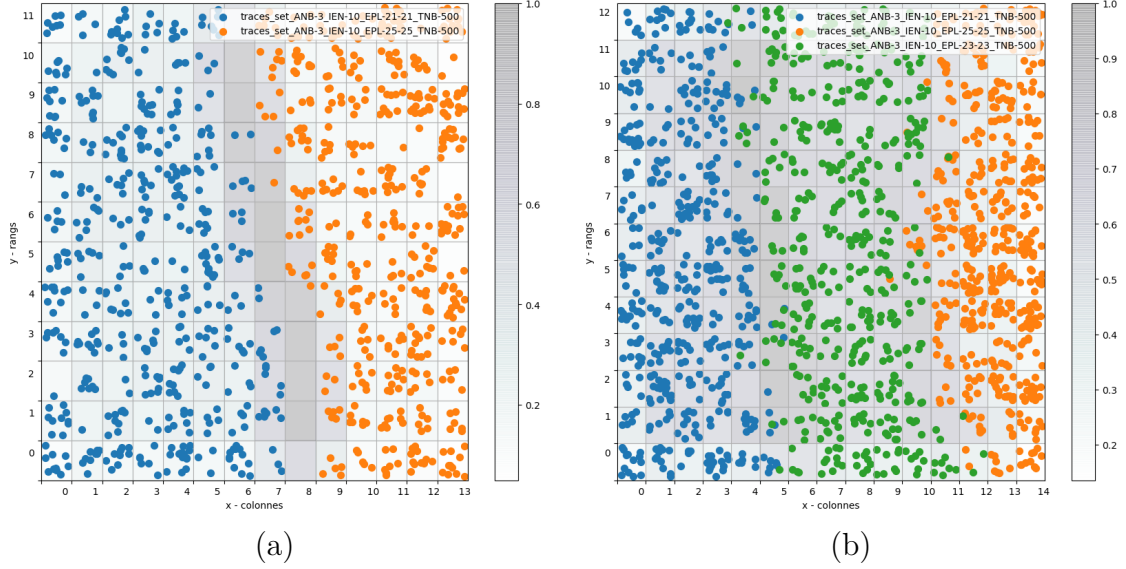


Figure 6.2 – Résultat du partitionnement pour les jeux A (a) et B (b), présentés en 6.1, et composés respectivement de deux et trois classes. Le fond de la figure présente la U-matrix⁵. Chaque case correspond à un neurone, chaque point dans une case correspond à une séquence associée à ce neurone. La couleur d'un point correspond à sa classe d'origine.

On constate visuellement que le partitionnement produit une majorité de neurones au contenu homogène et correspondant à l'une des classes d'origine. Une évaluation de la qualité des clusters est présentée dans la table 6.3. Cette évaluation vient confirmer la constatation visuelle de la qualité du partitionnement : les clusters sont bien séparés et homogènes.

Jeux	Classes	Distance intra-cluster	Distance inter-cluster	Indice de Davies-Bouldin
A	2 classes	1.194 1.209 (0.362)	7.424 7.415 (0.352)	0.631 0.641 (0.470)
B	3 classes	1.835 1.951 (0.424)	7.130 7.408 (1.024)	0.725 0.591 (0.435)

Table 6.3 – Évaluation de la qualité des clusters formés à l'issue du partitionnement pour les deux jeux de la figure 6.2. Les valeurs données sont les moyennes et médianes (en gras) calculées sur l'ensemble des clusters non vides, les écart-types sont donnés entre parenthèses

Une évaluation du partitionnement est présentée dans la table 6.4, pour les clusters

5. La U-matrix est une représentation d'une carte auto-adaptative où sont représentés, en niveaux de gris, les distances euclidiennes entre le vecteur v_k du neurone et les vecteurs v associés aux neurones voisins de la carte.

non vides. Comme le montrent les résultats de cette évaluation, les différents clusters produits sont majoritairement composés d’une seule classe.

Jeux	Classes	Nombre de clusters	Purité	Entropie	Homogénéité
A	2 classes	148	1.00	0.00	1.00
B	3 classes	191	0.798	12.37	0.62

Table 6.4 – Évaluation de la qualité des clusters formés à l’issue du partitionnement pour les deux jeux de la figure 6.2. Les valeurs données sont les moyennes calculées sur l’ensemble des clusters composés d’au moins deux séquences.

Conclusion Les expérimentations réalisées sur les différentes combinaisons de paramètres montrent que l’utilisation des cartes auto-organisées permet le partitionnement de séquences anesthésiques.

Le choix d’un nombre de neurones proportionnel au nombre de séquences dans un jeu permet de partitionner très précisément un ensemble de séquences hétérogènes. Ces cartes permettent de plus de séparer deux classes proches ou présentes dans des proportions très différentes.

Cependant, cette approche produit de très nombreux clusters, ce qui rend pertinent un regroupement destiné à produire un ensemble de cohortes exploitables chacune pour la simulation.

6.2.2 Critères estimatifs de la complexité d’un graphe

Pour pouvoir construire, à partir d’une cohorte de profils anesthésiques, un graphe de complexité adaptée, il faut pouvoir s’appuyer sur des indicateurs de cette complexité.

Il serait toutefois coûteux de construire le graphe de chacun des clusters afin d’estimer s’ils sont trop complexes ou non. Notre approche s’appuie donc sur des indicateurs de la complexité d’un graphe estimée à partir de critères et de mesures de dissimilarités entre les séquences d’événements.

Pour déterminer ces indicateurs, nous avons réalisé une expérimentation nous permettant d’identifier des corrélations entre ces indicateurs et des critères et mesures de dissimilarités entre les séquences d’événements.

6.2.2.1 Critères sur les séquences

La forme du graphe est directement liée à la similarité entre les n séquences de l’ensemble E_s à partir duquel le graphe est construit. Les critères calculés sur les séquences

se basent ainsi sur les caractéristiques des séquences ou sur des mesures de dissimilarité entre les séquences. Les mesures de dissimilarité utilisées sont présentées dans les définitions suivantes.

Indices de Jaccard La distribution des indices de JACCARD [141] nous donne une information sur la similarité entre deux populations d'événements présents dans ces séquences. L'indice de JACCARD entre deux séquences S_1 et S_2 de l'ensemble E_s se calcule comme suit :

$$Jaccard_I(S_1, S_2) = \frac{|S_1 \cap S_2|}{|S_1 \cup S_2|}.$$

Score d'alignement Le score d'alignement des séquences d'après l'algorithme de SMITH et WATERMANN [61] nous informe de la similarité de l'ordonnancement des événements dans ces séquences. Ce score a été utilisé selon deux modalités : (i) non pénalisé, où seulement les *matches* sont comptabilisés (*coût* = 1) et les éventuels *gaps* ignorés (*coût* = 0), ou bien (ii) pénalisé, où le poids des *matches* est augmenté (*coût* = 3) et les éventuels *gaps* comptabilisés (*coût* = -1).

Ainsi, nous avons sélectionné les critères suivants :

- les caractéristiques de la distribution du nombre d'événements dans l'ensemble E_s ,
- les caractéristiques de la distribution du nombre d'événements dans les interancres de l'ensemble E_s ,
- les caractéristiques de la distribution des indices de JACCARD [141] pour les paires de séquences de l'ensemble E_s ,
- les caractéristiques de la distribution d'un score d'alignement des séquences, au sens d'un alignement de Smith-Watermann [61] pour deux modalités en pénalisation et en non-pénalisation.

Les caractéristiques de ces différents critères sont présentées en table 6.5 et 6.6.

Critère	Description
<i>Critères basés sur la distribution du nombre d'événements</i>	
$Events_{Mean}$	Moyenne du nombre d'événements dans les séquences de E_s $Events_{Mean} = \frac{1}{n} \sum_{i=1}^n S_i $, où S_i est une séquence de E_s
$Events_{Range}$	Étendue de la distribution du nombre d'événements dans les séquences de E_s $Events_{Range} = \max(N_e(E_s)) - \min(N_e(E_s))$, où $N_e(E_s)$ est le nombre d'événements dans les séquences de E_s comme suit : $N_e(E_s) = \{ S_1 , \dots, S_n \}$, où S_i est une séquence de E_s
$Events_{Min}$	Nombre d'événements minimal dans les séquences de E_s $Events_{Min} = \min(N_e(E_s))$
$Events_{Max}$	Nombre d'événements maximal dans les séquences de E_s $Events_{Max} = \max(N_e(E_s))$
$Events_{Var}$	Variance de la distribution du nombre d'événements dans les séquences de E_s $Events_{Var} = \text{Variance}(N_e(E_s))$
$Events_{Std}$	Écart-type de la distribution du nombre d'événements dans les séquences de E_s $Events_{Std} = \text{EcartType}(N_e(E_s))$
<i>Critères basés sur la distribution du nombre d'événements dans les interancres</i>	
$InterA_{Mean}$	Moyenne du nombre d'événements dans les interancres des séquences de E_s $InterA_{Mean} = \frac{1}{n(N_a-1)} \sum_{i=1}^n (S_i - N_a)$, où N_a est le nombre d'ancres dans une séquence S_i de E_s
$InterA_{Range}$	Étendue de la distribution du nombre d'événements dans les interancres des séquences de E_s $InterA_{Range} = \max(N_i(E_s)) - \min(N_i(E_s))$, où $N_i(E_s)$ est le nombre d'événements dans les interancres des séquences de E_s comme suit : $N_i(E_s) = \{ S_1(a_1, a_2) , \dots, S_1(a_{N_a-1}, a_{N_a}) , \dots, S_n(a_1, a_2) , \dots, S_n(a_{N_a-1}, a_{N_a}) \}$, où S_1 est une séquence de E_s et $a_1, a_2, a_{N_a-1}, a_{N_a}$ sont respectivement les premières, secondes, pénultièmes et dernières ancres de cette séquence. $S_1(a_1, a_2)$ est la sous-séquence de S_1 comprise entre a_1 et a_2 .
$InterA_{Min}$	Nombre d'événements minimal dans les interancres des séquences de E_s $InterA_{Min} = \min(N_i(E_s))$
$InterA_{Max}$	Nombre d'événements maximal dans les interancres des séquences de E_s $InterA_{Max} = \max(N_i(E_s))$
$InterA_{Var}$	Variance de la distribution du nombre d'événements dans les interancres des séquences de E_s $InterA_{Var} = \text{Variance}(N_i(E_s))$
$InterA_{Std}$	Écart-type de la distribution du nombre d'événements dans les interancres des séquences de E_s $InterA_{Std} = \text{EcartType}(N_i(E_s))$
<i>Critères basés sur les indices de JACCARD</i>	
$JaccardI_{Mean}$	Moyenne de l'indice de JACCARD de toutes les paires de séquences de E_s $JaccardI_{Mean} = \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j=1}^n JaccardI(S_i, S_j)$, où S_i et S_j ($i \neq j$) sont des séquences de E_s
$JaccardI_{Range}$	Étendue de la distribution des indices de JACCARD de toutes les paires de séquences de E_s $JaccardI_{Range} = \max(N_J(E_s)) - \min(N_J(E_s))$, où $N_J(E_s)$ est l'ensemble des indices de JACCARD des paires de séquences de E_s comme suit : $N_J(E_s) = \{JaccardI(S_1, S_2), \dots, JaccardI(S_n, S_{n-1})\}$, où S_1, S_2, S_n, S_{n-1} sont des séquences de E_s
$JaccardI_{Var}$	Variance de la distribution des indices de JACCARD de toutes les paires de séquences de E_s $JaccardI_{Var} = \text{Variance}(N_J(E_s))$, où $N_J(E_s)$ est l'ensemble des indices de JACCARD des paires de séquences de E_s
$JaccardI_{Std}$	Écart-type de la distribution des indices de JACCARD de toutes les paires de séquences de E_s $JaccardI_{Std} = \text{EcartType}(N_J(E_s))$, où $N_J(E_s)$ est l'ensemble des indices de JACCARD des paires de séquences de E_s

Table 6.5 – Description des 16 critères basés sur la distribution du nombre d'événements, sur la distribution du nombre d'événements dans les interancres, et sur les indices de JACCARD, et calculés sur les séquences d'événements.

Critère	Description
<i>Critères basés sur le score d'alignement</i>	
$AScoreP_{Mean}$	Moyenne des scores d'alignement pénalisés de toutes les paires de séquences de E_s $AScoreP_{Mean} = \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j=1}^n AScoreP(S_i, S_j)$, où S_i et S_j ($i \neq j$) sont des séquences de E_s
$AScoreP_{Var}$	Variance de la distribution des scores d'alignement pénalisés de toutes les paires de séquences de E_s $AScoreP_{Var} = Variance(N_{SP}(E_s))$, où $N_{SP}(E_s)$ est l'ensemble des scores d'alignement pénalisés de toutes les paires de séquences de E_s
$AScoreNP_{Mean}$	Moyenne des scores d'alignement non-pénalisés de toutes les paires de séquences de E_s $AScoreNP_{Mean} = \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j=1}^n AScoreNP(S_i, S_j)$, où S_i et S_j ($i \neq j$) sont des séquences de E_s
$AScoreNP_{Var}$	Variance de la distribution des scores d'alignement non-pénalisés de toutes les paires de séquences de E_s $AScoreNP_{Var} = Variance(N_{SNP}(E_s))$, où $N_{SNP}(E_s)$ est l'ensemble des scores d'alignement non-pénalisés de toutes les paires de séquences de l'ensemble E_s

Table 6.6 – Description des 4 critères basés sur le score d'alignement, calculés sur les séquences d'événements.

6.2.2.2 Critères sur les graphes

Pour estimer la complexité d'un graphe, nous avons sélectionné des critères qui décrivent l'intrication plus ou moins importante des séquences entre elles dans le graphe. Ces critères se basent sur le nombre d'arcs et le nombre de nœuds et recourent aux définitions suivantes.

Largeur et niveau Dans l'esprit d'un parcours du graphe en largeur d'abord, on peut définir qu'un niveau L_i est l'ensemble des nœuds qui présentent la même longueur i du plus court chemin entre un nœud N_j et le premier nœud du graphe N_0 . La largeur ℓ d'un niveau L_i est le nombre de nœuds qui composent ce niveau, comme suit :

$$L_i = \{N_j | \text{Min}(\text{Distance}(N_0, N_j)) = i\}$$

$$\ell(L_i) = |L_i|$$

Ratio d'arcs Pour chaque nœud, on peut déterminer le ratio entre le nombre d'arcs entrants et le nombre d'arcs sortants. Si ce ratio est supérieur à 1, alors la largeur du graphe diminue à partir de ce nœud, à l'inverse, si ce ratio est inférieur à 1, alors la largeur du graphe augmente à partir de ce nœud.

$$Ratio = \frac{|Ae(N_i)|}{|As(N_i)|},$$

où $Ap(N_i)$ et $Ac(N_i)$ sont respectivement les arc entrants et sortants pour le nœud N_i .

Densité La densité d'un graphe est le ratio entre le nombre d'arcs et le nombre de nœuds. Pour un graphe $G = (N; A)$, où N est l'ensemble des nœuds, et A l'ensemble des arcs, la densité de $G = (N; A)$ est :

$$Density(G) = \frac{|A|}{|N|(|N|-1)}.$$

Un graphe complexe se caractérisera ainsi par une densité élevée. Toutefois, la présence d'ancres dans les séquences a pour effet de «linéariser» le graphe qui se rapproche ainsi d'une combinaison de sous-graphes liés par les ancres.

Réticulation La réticulation d'un graphe est calculée comme étant la densité du graphe, pondérée par le nombre d'ancres. Pour le graphe G il y a de multiples sous-graphes G_i entre deux ancres successives. On définit la réticulation moyenne comme étant la moyenne des densités des sous-graphes G_i . La réticulation médiane est définie de façons similaire :

$$\begin{aligned} Reticulation_{Mean} &= \text{moyenne}_{i=1, \dots, N_{sg}} (Densité(G_i)) \\ Reticulation_{Median} &= \text{médiane}_{i=1, \dots, N_{sg}} (Densité(G_i)), \end{aligned}$$

où N_{sg} représente le nombre de sous-graphes G_i entre deux ancres successives.

Ainsi, nous avons sélectionné les critères suivants :

- les caractéristiques de la distribution du ratio entre le nombre d'arcs entrants et d'arcs sortants, pour chaque nœud et pour chaque niveau du graphe,
- la largeur maximale du graphe et les caractéristiques de la distribution de la largeur de chaque niveau du graphe,
- la densité du graphe,
- les réticulations moyenne et médiane du graphe.

Les caractéristiques de ces différents critères sont présentées en table 6.7.

Critère	Description
<i>Critères basés sur la largeur du graphe</i>	
$LevelWidth_{Mean}$	Moyenne de la largeur des différents niveaux du graphe $LevelWidth_{Mean} = \frac{1}{ L } \sum_{i=1}^{ L } \ell(L_i)$, où L est l'ensemble des niveaux présents dans le graphe
$LevelWidth_{Median}$	Médiane de la largeur des différents niveaux du graphe $LevelWidth_{Median} = Mediane(E_L)$, où E_L est l'ensemble des largeurs des niveaux présents dans le graphe, comme suit : $E_L = \{\ell(L_1), \dots, \ell(L_i), \dots, \ell(L_{ L })\}$
$Width_{Max}$	Largeur maximale du graphe $Width_{Max} = Max(E_L)$
<i>Critères basés sur le ratio d'arcs entrants et d'arcs sortants</i>	
$Ratio_{Mean}$	Moyenne du ratio d'arcs entrants et d'arcs sortants pour tous les nœuds du graphe $Ratio_{Mean} = \frac{1}{ N -2} \sum_{i=2}^{ N -1} \left(\frac{ Ae(N_i) }{ As(N_i) } \right)$, où $Ae(N_i)$ et $As(N_i)$ sont respectivement les arcs entrants et sortants du nœud N_i
$Ratio_{Min}$	Ratio minimal d'arcs entrants et d'arcs sortants pour tous les nœuds du graphe $Ratio_{Min} = Min(E_R)$, où E_R est l'ensemble des ratios pour chaque nœud du graphe, comme suit : $E_R = \left\{ \left(\frac{ Ae(N_2) }{ As(N_2) } \right), \dots, \left(\frac{ Ae(N_i) }{ As(N_i) } \right), \dots, \left(\frac{ Ae(N_{ N -1})}{ As(N_{ N -1})} \right) \right\}$
$Ratio_{Max}$	Ratio maximal d'arcs entrants et d'arcs sortants pour tous les nœuds du graphe $Ratio_{Max} = Max(E_R)$, où E_R est l'ensemble des ratios pour chaque nœud du graphe
$LevelRatio_{Mean}$	Moyenne du ratio d'arcs entrants et d'arcs sortants pour tous les niveaux du graphe $LevelRatio_{Mean} = \frac{1}{ L -1} \sum_{i=1}^{ L -1} \left(\frac{ Ae(L_i) }{ As(L_i) } \right)$, où $Ae(L_i)$ et $As(L_i)$ sont respectivement les arcs entrants et sortants des nœuds du niveau L_i
$LevelRatio_{Min}$	Ratio minimal d'arcs entrants et d'arcs sortants pour tous les niveaux du graphe $LevelRatio_{Min} = Min(E_{LR})$, où E_{LR} est l'ensemble des ratios pour chaque nœud des niveaux du graphe, comme suit : $E_{LR} = \left\{ \left(\frac{ Ae(L_1) }{ As(L_1) } \right), \dots, \left(\frac{ Ae(L_i) }{ As(L_i) } \right), \dots, \left(\frac{ Ae(N_{ L })}{ As(N_{ L })} \right) \right\}$
$LevelRatio_{Max}$	Ratio maximal d'arcs entrants et d'arcs sortants pour tous les niveaux du graphe $LevelRatio_{Max} = Max(E_{LR})$, où E_{LR} est l'ensemble des ratios pour chaque nœud des niveaux du graphe
<i>Critères basés sur la densité du graphe</i>	
$Density$	Densité du graphe, le ratio entre le nombre d'arcs et le nombre de nœuds.
$Reticulation_{Mean}$	Réticulation moyenne du graphe, moyenne de la densité des sous graphe entre deux ancres.
$Reticulation_{Median}$	Réticulation mediane du graphe, médiane de la densité des sous graphe entre deux ancres.

Table 6.7 – Description des 12 critères calculés sur un graphe pour estimer sa complexité

6.2.2.3 Expérimentation

Cette expérimentation consiste à calculer les différents critères sur les séquences pour un ensemble de séquences d'événements et les différents critères descriptifs de la complexité du graphe correspondant à l'ensemble de séquences. L'objectif est de pouvoir mettre en avant d'éventuelles corrélations.

Constitution des jeux Cette expérimentation se place à un haut niveau d'abstraction ; les séquences sont représentées de la manière la plus simple, comme une série d'événements uniquement caractérisés par leurs positions. Ainsi, on trouvera dans un jeu expérimental, la séquence S composée de n événements E et ancres A .

$$S = (A_1, E_2, \dots, A_n).$$

Chaque jeu est constitué de 100 séquences et est défini par 2 paramètres :

- le nombre d'ancres dans la séquence (3 modalités : 3, 5 et 10 ancres),
- le nombre d'événements possibles (5, 10, 20 événements) dans une interancre.

Nous avons ainsi produit 9 jeux expérimentaux pour les différentes combinaisons des deux paramètres. Pour permettre de la variabilité au sein d'un jeu, et donc permettre la complexité du graphe final, du bruit est ajouté grâce à une probabilité d'apparition propre à un événement dans une interancre selon 3 modalités (12-25%, 45-55%, 75-85%).

Réplication et calculs de corrélation Chacune des combinaisons de paramètres définissant les jeux a fait l'objet d'une expérience dans laquelle un jeu est construit et l'ensemble des critères calculés. L'ensemble des critères calculés est utilisé pour calculer des corrélations entre les critères des séquences et les critères du graphe. Chaque expérience a été répliquée 100 fois.

Résultats À partir des critères calculés sur un jeu de séquences, et le graphe représentatif de cet ensemble de séquences, des corrélations de PEARSON ont été calculées afin de mettre en évidence d'éventuels liens entre critères sur les séquences et critères sur les graphes.

Les résultats pour un jeu, composé de séquences avec 3 ancres et 10 événements maximum par interancre, sont visibles dans la table 6.8 et la figure 6.3 est donnée en illustration. Les résultats pour une sélection de jeux sont présentés en annexe F.1.

Critères sur les séquences	Critères sur les graphes											
	$Density$	$Reticulation_{Median}$	$Reticulation_{Mean}$	$LevelWidth_{Mean}$	$LevelWidth_{Median}$	$Width_{Max}$	$Ratio_{Mean}$	$Ratio_{Min}$	$Ratio_{Max}$	$LevelRatio_{Mean}$	$LevelRatio_{Min}$	$LevelRatio_{Max}$
$Events_{Mean}$	-0.885 (6e-235)	-0.885 (6e-235)	-0.885 (6e-235)	-0.912 (2e-273)	0.911 (7e-271)	-0.865 (3e-212)	-0.942 (0.0)	-	-0.907 (1e-265)	-0.918 (2e-283)	-	-0.872 (8e-220)
$Events_{Min}$	-0.944 (0.0)	-0.944 (0.0)	-0.944 (0.0)	-0.875 (9e-223)	0.893 (3e-245)	-0.864 (2e-210)	-0.951 (0.0)	-	-0.925 (2e-297)	-0.861 (2e-207)	-	-0.844 (1e-191)
$Events_{Max}$	-0.760 (3e-133)	-0.760 (3e-133)	-0.760 (3e-133)	-0.888 (1e-238)	0.872 (7e-219)	-0.813 (1e-166)	-0.864 (5e-211)	-	-0.823 (5e-174)	-0.912 (4e-273)	-	-0.843 (2e-190)
$Events_{Var}$	0.333 (1e-19)	0.333 (1e-19)	0.333 (1e-19)	-0.098 (0.00)	0.041 (0.27)	0.044 (0.23)	0.117 (0.00)	-	0.155 (3e-05)	-0.197 (1e-07)	-	-0.058 (0.12)
$Events_{Std}$	0.320 (3e-18)	0.320 (3e-18)	0.320 (3e-18)	-0.121 (0.00)	0.065 (0.08)	0.024 (0.52)	0.097 (0.00)	-	0.140 (0.00)	-0.215 (8e-09)	-	-0.076 (0.04)
$Events_{Range}$	0.307 (8e-17)	0.307 (8e-17)	0.307 (8e-17)	-0.110 (0.00)	0.037 (0.32)	0.028 (0.45)	0.098 (0.00)	-	0.134 (0.00)	-0.191 (3e-07)	-	-0.075 (0.04)
$InterA_{Mean}$	-0.896 (2e-249)	-0.896 (2e-249)	-0.896 (2e-249)	-0.911 (5e-272)	0.911 (1e-271)	-0.868 (7e-215)	-0.948 (0.0)	-	-0.914 (3e-276)	-0.915 (9e-278)	-	-0.873 (2e-220)
$InterA_{Min}$	-0.914 (2e-277)	-0.914 (2e-277)	-0.914 (2e-277)	-0.754 (7e-130)	0.785 (2e-147)	-0.780 (1e-144)	-0.857 (1e-203)	-	-0.845 (1e-192)	-0.698 (1e-103)	-	-0.723 (1e-114)
$InterA_{Max}$	-0.637 (6e-81)	-0.637 (6e-81)	-0.637 (6e-81)	-0.831 (1e-180)	0.811 (2e-165)	-0.735 (6e-120)	-0.771 (4e-139)	-	-0.730 (8e-118)	-0.874 (9e-222)	-	-0.778 (3e-143)
$InterA_{Var}$	0.182 (1e-06)	0.182 (1e-06)	0.182 (1e-06)	-0.237 (1e-10)	0.181 (1e-06)	-0.094 (0.01)	-0.038 (0.30)	-	0.002 (0.95)	-0.345 (5e-21)	-	-0.199 (1e-07)
$InterA_{Std}$	0.152 (5e-05)	0.152 (5e-05)	0.152 (5e-05)	-0.279 (4e-14)	0.224 (1e-09)	-0.132 (0.00)	-0.073 (0.05)	-	-0.029 (0.44)	-0.380 (1e-25)	-	-0.232 (4e-10)
$InterA_{Range}$	0.069 (0.06)	0.069 (0.06)	0.069 (0.06)	-0.310 (3e-17)	0.259 (3e-12)	-0.171 (5e-06)	-0.144 (0.00)	-	-0.106 (0.00)	-0.414 (2e-30)	-	-0.275 (1e-13)
$JaccardI_{Mean}$	-0.985 (0.0)	-0.985 (0.0)	-0.985 (0.0)	-0.783 (1e-146)	0.810 (3e-164)	-0.810 (1e-164)	-0.929 (1e-304)	-	-0.915 (5e-278)	-0.740 (2e-122)	-	-0.772 (1e-139)
$JaccardI_{Var}$	0.810 (3e-164)	0.810 (3e-164)	0.810 (3e-164)	0.699 (7e-104)	-0.727 (1e-116)	0.717 (1e-111)	0.746 (1e-125)	-	0.746 (7e-126)	0.636 (1e-80)	-	0.644 (2e-83)
$JaccardI_{Std}$	0.798 (1e-155)	0.798 (1e-155)	0.798 (1e-155)	0.677 (4e-95)	-0.706 (7e-107)	0.700 (3e-104)	0.727 (2e-116)	-	0.729 (1e-117)	0.611 (5e-73)	-	0.622 (1e-76)
$JaccardI_{Range}$	0.706 (5e-107)	0.706 (5e-107)	0.706 (5e-107)	0.407 (2e-29)	-0.460 (5e-38)	0.489 (2e-43)	0.552 (3e-57)	-	0.577 (1e-63)	0.314 (1e-17)	-	0.396 (1e-27)
$AScoreP_{Mean}$	-0.949 (0.0)	-0.949 (0.0)	-0.949 (0.0)	-0.889 (1e-239)	0.899 (8e-254)	-0.870 (2e-217)	-0.962 (0.0)	-	-0.934 (8e-316)	-0.872 (4e-219)	-	-0.856 (1e-202)
$AScoreP_{Var}$	-0.403 (7e-29)	-0.403 (7e-29)	-0.403 (7e-29)	-0.622 (3e-76)	0.582 (5e-65)	-0.514 (1e-48)	-0.579 (5e-64)	-	-0.526 (4e-51)	-0.711 (7e-109)	-	-0.613 (1e-73)
$AScoreNP_{Mean}$	-0.911 (2e-271)	-0.911 (2e-271)	-0.911 (2e-271)	-0.909 (1e-267)	0.911 (1e-271)	-0.871 (3e-218)	-0.953 (0.0)	-	-0.921 (3e-288)	-0.907 (1e-264)	-	-0.871 (3e-218)
$AScoreNP_{Var}$	-0.078 (0.03)	-0.078 (0.03)	-0.078 (0.03)	-0.419 (2-31)	0.368 (7-24)	-0.285 (1-14)	-0.288 (7-15)	-	-0.237 (1-10)	-0.521 (4-50)	-	-0.395 (1-27)

Table 6.8 – Coefficients de corrélation et p-valeurs (entre parenthèses) calculés pour les différents critères sur les séquences et le graphe, pour un ensemble de séquences composées de 3 ancres et 10 événements au maximum dans une interancre. Les critères $Ratio_{min}$ et $LevelRatio_{Min}$ se sont révélés constants et les corrélations n'ont pas été calculées.

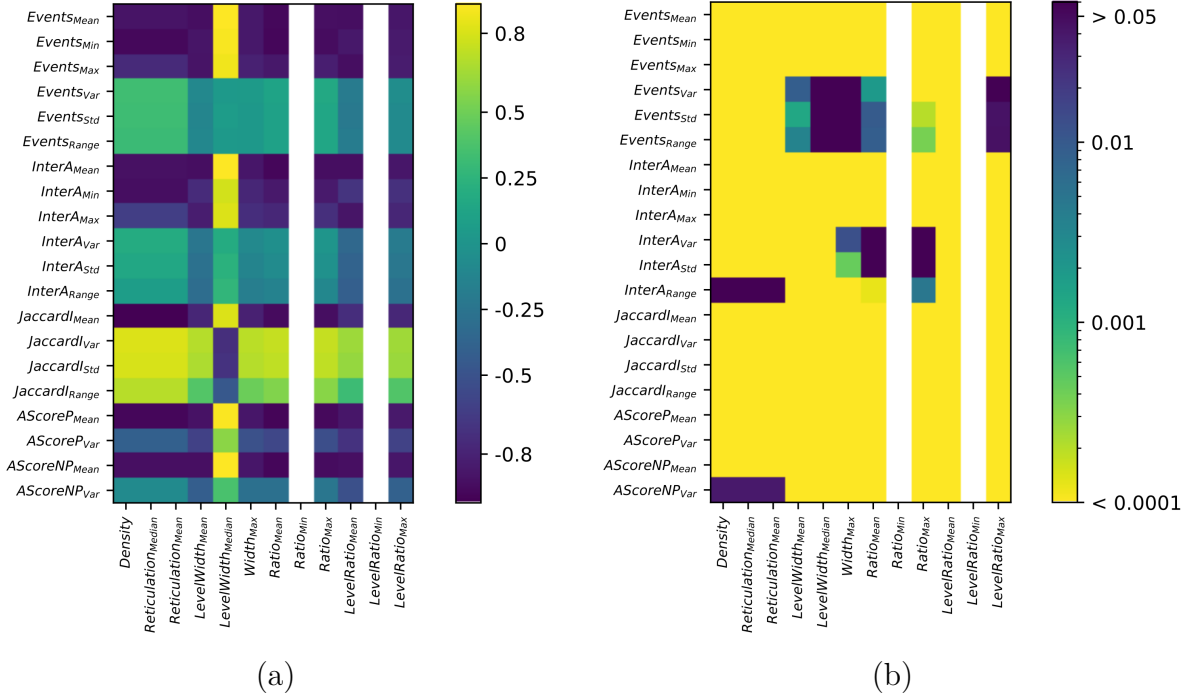


Figure 6.3 – Heatmap des coefficients de corrélation (a) et p-valeurs (b) des différents critères pour un jeu de 3 ancres et 10 interancres. Les critères calculés sur le graphe sont présentés en abscisse et les critères calculés sur les séquences sont présentés en ordonnée. Les critères $Ratio_{min}$ et $LevelRatio_{Min}$ se sont révélés constants et les corrélations n’ont pas été calculées.

Comme le montre l’ensemble des résultats, plusieurs critères sur les séquences sont corrélés (au seuil de 0.8 avec une p-valeur < 0.05) avec certains critères sur le graphe. Les deux mesures de dissimilarité utilisées, les indices de JACCARD et le score d’alignement, sont ainsi positivement corrélées avec la densité du graphe et la réticulation moyenne ou médiane de celui-ci.

Conclusion La définition de critères, et la mise en avant de corrélation entre ces critères sur les graphes et ensemble de séquences, respectivement, met en évidence le lien fort entre similarité entre les séquences et complexité du graphe construit à partir de ces séquences.

L’importante corrélation entre les mesures de dissimilarité calculées sur les séquences, les indices de JACCARD et le score d’alignement, et la densité du graphe comme la réticulation (médiane ou moyenne), nous permet de disposer de critères prédictifs de la complexité du graphe représentant un ensemble de séquences. Ces critères seront utilisés dans la suite de ces travaux.

6.2.3 Association de clusters de profils anesthésiques pour la constitution de sous-cohortes idéales

Nous rappelons que, à l'issue de l'étape de partitionnement, chacun des neurones de la carte auto-organisée décrite en 6.2.1 est représenté par un vecteur consensus de séquences d'événements appartenant à un ensemble de patients. Cet ensemble de patients est également mémorisé par ce neurone. À l'issue du partitionnement, chaque ensemble de séquences "présentes dans un neurone" est considéré comme un cluster.

On pourrait utiliser ces clusters directement comme des sous-cohortes de patients. Cependant, comme le montre la figure 6.2 (Page 139), les classes d'origine qui composent le jeu expérimental de l'expérimentation sont réparties dans plusieurs neurones (représentés sur la figure par une case).

Il est alors pertinent de réaliser un regroupement à l'aide d'un second partitionnement. Les clusters sont représentés par le vecteur consensus du neurone dont ils sont issus, et sont associés dans un processus ascendant hiérarchique naïf.

On calcule la distance pour toutes les paires de vecteurs consensus correspondant à deux clusters non vides. Pour les deux vecteurs les plus similaires, si la distance inter-cluster est plus grande que la distance intra-cluster, et si l'indice de Davies-Bouldin pour ces deux clusters est supérieure à 1, alors il est pertinent de les fusionner [142]. Si les deux clusters sont fusionnés, un nouveau vecteur consensus est produit par la moyenne des deux vecteurs consensus initiaux. Ce processus est itéré pour toutes les paires de clusters dont les vecteurs consensus sont les plus similaires.

On construit ainsi un arbre binaire de clustering pour l'ensemble des clusters, comme le montre les figures 6.4 et 6.5. Le nombre de neurones étant défini par le nombre de séquences (voir 6.2.1.2), le premier partitionnement pour le jeu A a produit 158 clusters non vides, dont 10 *singletons*⁶, sur 168 neurones, et 194 clusters non vides, dont 3 *singletons* sur 195 neurones pour le jeu B. Ces différents clusters non vides sont utilisés comme ensemble de départ pour le regroupement.

Chaque cluster formé dans la hiérarchie correspond à une cohorte de patients de plus en plus diversifiée, et donc à un graphe de plus en plus complexe. Pour tous les clusters présents dans l'arbre issu du processus ascendant hiérarchique, on calcule les critères C_{sq} ⁷ sélectionnés après l'analyse de corrélation entre critères sur l'ensemble de séquences et critères de graphe (voir tables 6.5 et 6.6).

6. Un cluster ne comportant qu'une seule trace

7. Les indices de JACCARD et le score d'alignement

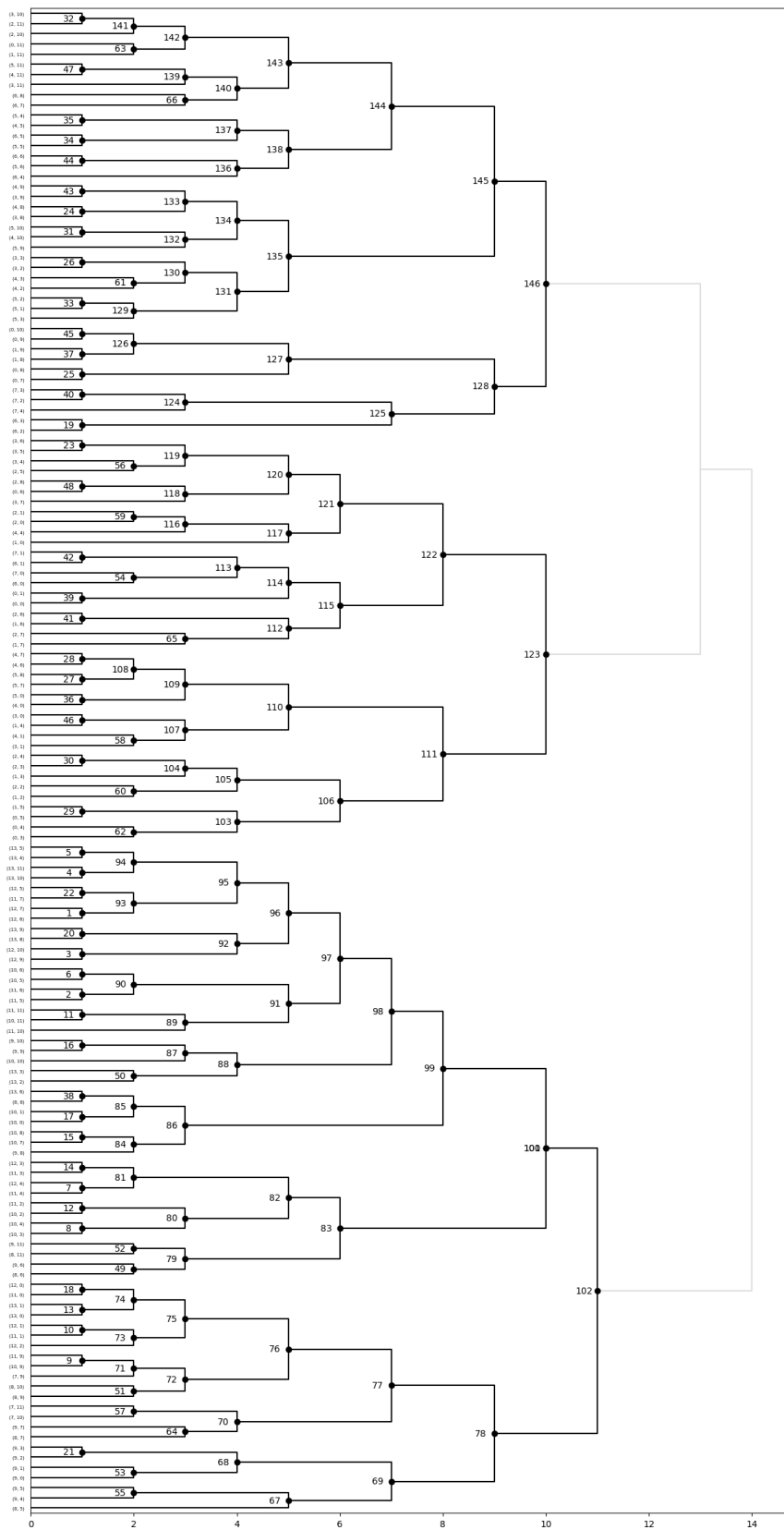


Figure 6.4 – Hiérarchie du regroupement pour le jeu A présenté en figure 6.2.

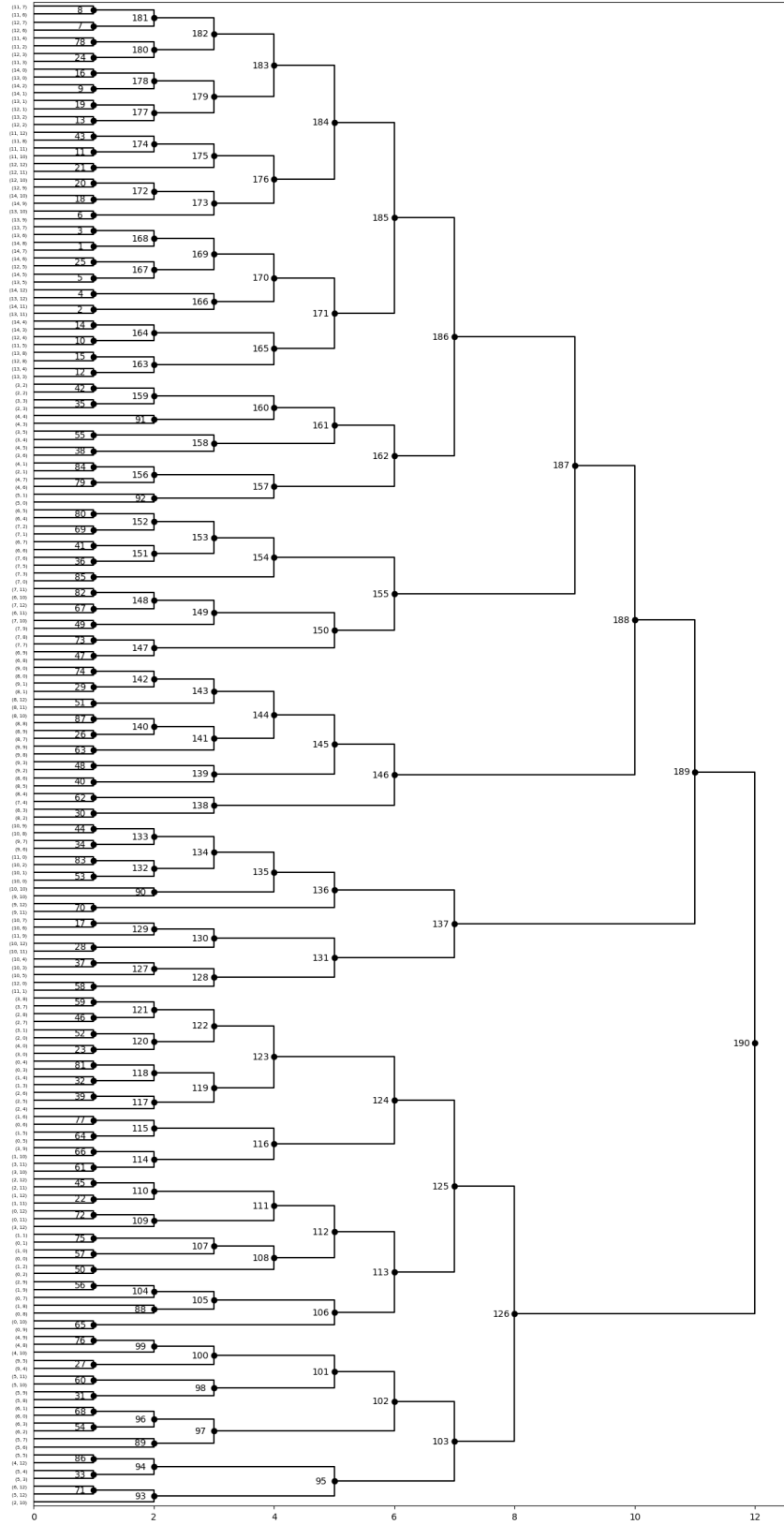


Figure 6.5 – Hiérarchie du regroupement pour le jeu B présenté en figure 6.2.

Les différents clusters finaux sont déterminés en coupant l'arbre à l'endroit qui permet une diversité adaptée à la simulation, pour chaque type de cluster.

On peut ainsi former à partir des jeux A et B, initialement composés de 2 et 3 classes respectivement, plusieurs clusters finaux selon les besoins en complexité du graphe pour la simulation. Le regroupement permet la construction de 146 clusters différents pour le jeu A et 190 pour le jeu B.

Les différents critères présentés en tables 6.9 et 6.10, correspondent à une sélection de clusters présents dans les hiérarchies présentés en 6.4 et 6.5, pour les jeux A et B respectivement. Les résultats complets pour les jeux A et B sont présentés en annexe F.2.

Clusters n°	Critères sur les séquences			Critères sur le graphe		
	$JaccardI_{Mean}$	$AScoreP_{Mean}$	$AScoreNP_{Mean}$	$Density$	$Reticulation_{Median}$	$Reticulation_{Mean}$
18	0.835	17.784	17.784	0.08	0.35	0.344
28	0.665	4.066	4.066	0.11	0.262	0.31
83	0.804	17.975	17.975	0.111	0.212	0.212
110	0.516	3.698	3.698	0.139	0.264	0.264
123	0.451	3.332	3.332	0.223	0.428	0.428
135	0.475	3.723	3.723	0.164	0.314	0.314
146	0.422	3.478	3.478	0.249	0.477	0.477

Table 6.9 – Valeurs des différents critères, sur les ensembles de séquences et sur le graphe, calculés pour différents clusters obtenus à partir du jeu A et présents dans la figure 6.4

Clusters n°	Critères sur les séquences			Critères sur le graphe		
	$JaccardI_{Mean}$	$AScoreP_{Mean}$	$AScoreNP_{Mean}$	$Density$	$Reticulation_{Median}$	$Reticulation_{Mean}$
19	0.821	17.714	17.714	0.077	0.5	0.418
101	0.419	6.554	6.554	0.184	0.352	0.352
123	0.479	3.334	3.334	0.176	0.337	0.337
127	0.623	12.714	12.714	0.105	0.304	0.341
137	0.575	11.85	11.85	0.168	0.322	0.322
145	0.478	8.741	8.741	0.192	0.367	0.367
186	0.61	13.339	13.339	0.233	0.447	0.447
190	0.431	7.99	7.99	0.261	0.5	0.5

Table 6.10 – Valeurs des différents critères, sur les ensembles de séquences et sur le graphe, calculés pour différents clusters obtenus à partir du jeu B et présents dans la figure 6.5

Ainsi, à partir de ces clusters finaux, on obtient les cohortes de patients dont la diversité est adaptée à notre objectif : utiliser un graphe représentant leurs profils anesthésiques comme support d'une simulation.

6.2.4 Application aux profils anesthésiques réalistes

Les expérimentations précédentes nous montrent que notre approche est fonctionnelle. Toutefois, pour assurer qu'elle est adaptée à notre objectif de simulation, nous avons appliqué cette approche à un ensemble de profils anesthésiques réalistes.

Nous avons utilisé une cohorte de patients réalistes, générées par le générateur DBLBS présenté au chapitre 3, et composée de 1000 patients, d'environ 30 ans et sans antécédents, ayant eu pour intervention la cure d'une hernie inguinale sous coelioscopie avec pose de prothèse.

À partir des séquences d'événements de chacun des patients de cette cohorte, nous avons construit une carte auto-organisée composée de 168 neurones, comme présenté en 6.2.1. À l'issue de l'apprentissage, les différents ensemble de séquences ont été partitionnés. Chacune des séquences a été associée à un neurone de la carte.

Ainsi, à l'issue de ce premier partitionnement, les 1000 séquences d'événements ont été assignées à 143 clusters non vides dont 8 *singletons*. Ces clusters ont ensuite fait l'objet d'un regroupement, comme présenté en 6.2.3. L'arbre résultant du partitionnement ascendant hiérarchique est présenté en figure 6.6.

On peut alors sélectionner plusieurs clusters selon les besoins de la simulation. Différents clusters présents dans la figure 6.6, ont été sélectionnés. Les critères estimatifs de la complexité des graphes issus de ces clusters sont présentés dans la table 6.11. Les résultats complets sont présentés en annexe F.3.

Clusters n°	Critères sur les séquences			Critères sur le graphe		
	$JaccardI_{Mean}$	$AScoreP_{Mean}$	$AScoreNP_{Mean}$	$Density$	$Reticulation_{Median}$	$Reticulation_{Mean}$
37	0.779	74.424	74.424	0.009	0.2	0.177
81	0.749	72.45	72.45	0.008	0.153	0.166
83	0.656	66.842	66.842	0.004	0.007	0.096
100	0.788	73.111	73.111	0.011	0.233	0.225
105	0.728	69.252	69.252	0.009	0.186	0.184
118	0.723	68.828	68.828	0.01	0.2	0.197
130	0.735	69.715	69.715	0.01	0.207	0.201
134	0.652	65.414	65.414	0.004	0.008	0.095

Table 6.11 – Valeurs des différents critères, sur les ensembles de séquences et sur le graphe, calculés pour différents clusters présents dans la figure 6.6

Toutefois, les graphes produits à partir des séquences d'événements ne contiennent qu'une partie des profils anesthésiques. Il serait pertinent de proposer une représentation dont la structure de données comporte toutes les informations de ces profils anesthésiques et pas seulement les séquences d'événements.

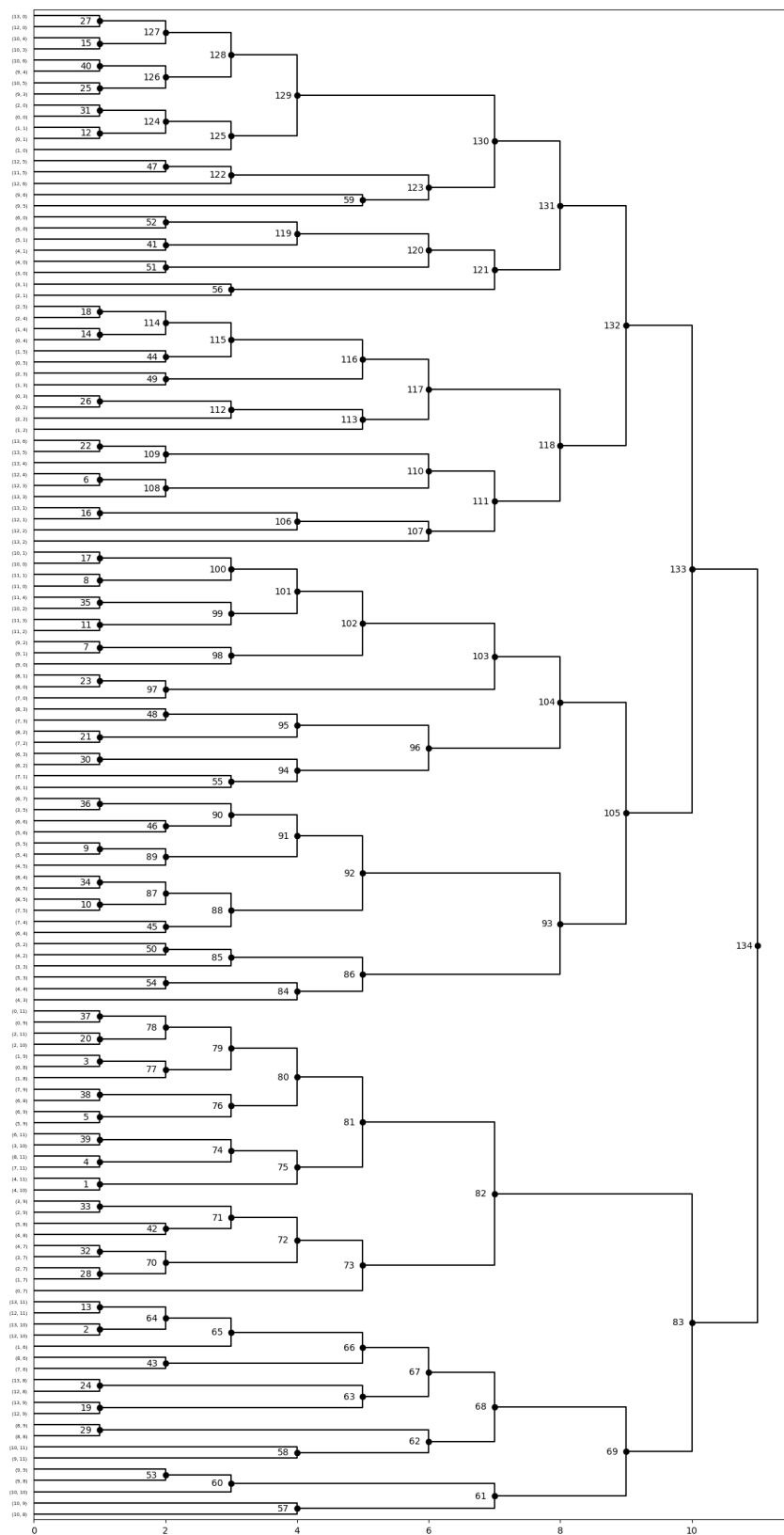


Figure 6.6 – Hiérarchie du regroupement pour une cohorte de 1000 patients de type et d'intervention identiques.

6.3 Construction d'une représentation synthétique

Les événements descripteurs (voir 3.2.2) renseignés pendant la surveillance d'une intervention sont systématiquement présents, et à la même position de l'intervention. Ce sont donc des ancres qui bornent les grandes étapes d'une anesthésie.

La structure générale d'une anesthésie est connue et formalisée. Il est ainsi possible d'utiliser la structure de la grammaire anesthésique présentée en 3.1.2, en prenant pour points de référence les ancres qui bornent les grandes étapes. L'objectif est d'aligner les séquences issues des différents profils anesthésiques et de produire un graphe comme présenté en 6.1 qui serait une *représentation* des séquences d'événements.

Une telle représentation implique que des séquences potentiellement très similaires soient toutes présentes, et implique donc une redondance. De plus, la conservation de l'intégrité des séquences sous la forme d'enchaînements nœud-arcs induit une structure de données de taille importante. Ensuite, un tel graphe contient seulement les séquences d'événements de l'ensemble de profils anesthésiques à partir duquel il est construit. Comment peut-on alors simuler un patient virtuel réactif aux actions d'un apprenant ? Il faut pouvoir simuler chaque action sur le patient virtuel, donc la série temporelle correspondante à l'effet d'un événement.

On peut pour cela ajouter en chacun des nœuds de ce graphe les informations liées à l'intervention et les portions de séries temporelles, sous la forme de consensus, entre deux événements, afin de produire une *représentation* d'un ensemble de profils anesthésiques. Il est alors possible d'utiliser la similarité entre certaines séquences pour produire une *représentation synthétique*.

Le but de cette représentation synthétique est d'associer les éléments similaires entre les profils anesthésiques, séquences d'événements et séries temporelles, tout en conservant la diversité au sein du graphe.

6.3.1 Approche

Notre approche consiste, pour chaque profil anesthésique, à réaliser plusieurs étapes séquentielles. La figure 6.7 présente de manière schématique cette approche pour la construction d'une représentation synthétique. Comme l'indique la figure, à partir de l'extraction de la séquence d'événements du profil anesthésique d'un patient (1), on réalise un «alignement» sur la grammaire anesthésique correspondant à l'intervention (2). Pour cela, on procède par affectation des différents événements de la séquence à une étape anes-

thésique de la grammaire. Cette séquence alignée est ensuite ajoutée au graphe (3). Les différents nœuds de cette séquence sont ensuite enrichis par les informations spécifiques des événements (4) et leurs sous-séries temporelles correspondantes sous la forme de séries temporelles multivariées consensus (5).

L'algorithme 4 présente la construction d'une représentation synthétique sous la forme d'un graphe $G(N, A)$, où N est l'ensemble des nœuds et A l'ensemble des arcs, à partir d'un ensemble E_s composé de n séquences S_i . La fonction *Consensus_hiérarchique()*, dans l'algorithme 4, réalise la construction des séries temporelles consensus. Son fonctionnement est présenté en algorithme 5.

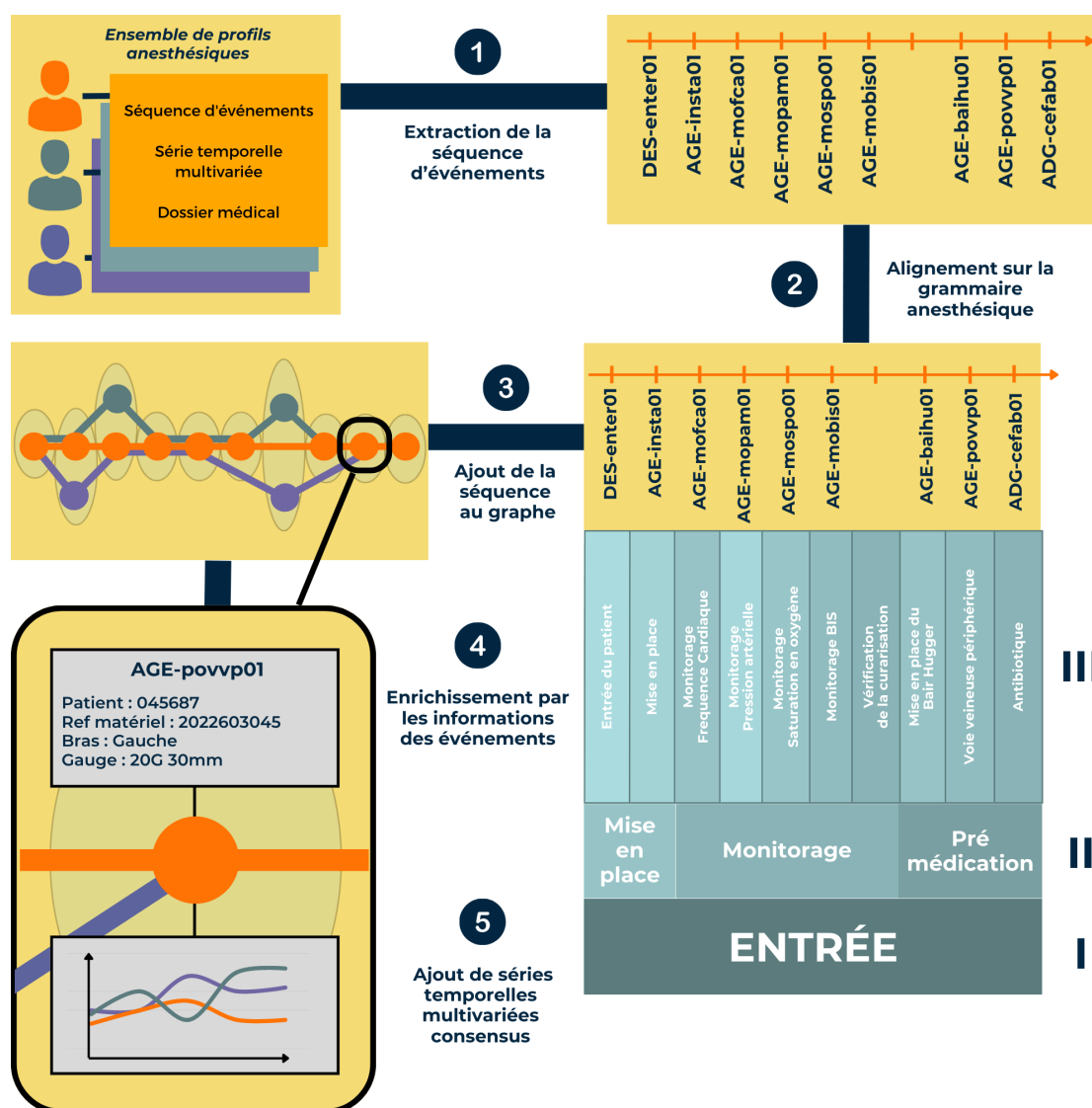


Figure 6.7 – Schéma du principe de construction d'une représentation synthétique à partir d'un ensemble de profils anesthésiques

Algorithme 4 : Construction d'une représentation synthétique**Entrées :**

- *Gramaire* : la grammaire d'une intervention
- *Ensemble_P* : un ensemble de profils anesthésiques
- *Seuil* : Seuil d'association de séries temporelles similaires pour la création de consensus

Variables :

- *Profil* : un profil anesthésique
- *Noeud* : un nœud dans le graphe
- *Séquence* : une séquence d'événements issue d'un profil anesthésique
- *Événement* : un événement d'une séquence
- *Étape* : une étape anesthésique dans la grammaire
- *Arc* : un arc entre deux nœuds
- *Arcs* : un ensemble d'arcs entre deux nœuds
- *SousSéries* : une sous série temporelle multivariées
- *SousSéries_consensus* : Sous séries temporelles consensus entre deux nœuds

Sorties :

- *Graphe* : Graphe représentant les profils anesthésiques

FONCTION *Représentation_synthétique*(*Gramaire*, *Ensemble_P*, *Seuil*)

```

1  pour Profil dans EnsembleP faire
2      Séquence ← Extraire_Séquence(Profil);
3      pour Événement dans Séquence faire
4          Étape ← position_dans_grammaire(Gramaire, Événement);
5          Metanoeud ← Existe_metanoeud(Graphe, Étape);
6          si Metanoeud = ∅ alors
7              Metanoeud ← Créer_metanoeud(Graphe, Étape);
8              Créer_noeud_dans_metanoeud(Metanoeud, Événement);
9          sinon
10             Noeud ← Noeud_similaire(Metanoeud, Événement, Séquence );
11             si Noeud = ∅ alors
12                 Noeud ← Créer_noeud_dans_metanoeud(Metanoeud, Événement);
13             sinon
14                 Fusion(Noeud, Événement);
15             fin
16         fin
17         Arc ← Créer_arc_issu_du_parent(Noeud);
18         SousSéries ← Extraction_sous_séries(Profil, Noeud, Noeud.Parent);
19         Annoter_arc(Arc, SousSéries);
20         Ajouter(Graphe, Arc);
21         Associer_arcs_similaires(Graphe, Noeud, Noeud.Parent, Seuil );
22     fin
23 fin
24 pour Paire_de_noeuds_adjacents dans Graphe faire
25     SousSérie_consensus ← Consensus_hiéarchique(Graphe, Paire_de_noeuds_adjacents, Seuil );
26     Ajouter(Graphe, SousSérie_consensus)
27 fin
28 retourner Graphe

```

Algorithme 5 : Fonctionnement de *Consensus_hiéarchique()* pour la construction de représentation synthétique

Entrées :

- *Graphe* : Graphe représentant les profils anesthésiques
- *Paire_de_noeuds_adjacents* : Paire de nœuds adjacent dans l'arc
- *Seuil* : Seuil de similarité en dessous duquel deux séries temporelles ne sont plus considérées similaires

Variables :

- *Séries_Temporelles_arcs* : Ensembles des séries temporelles associées issues des arc
- *Matrice_similarité* : matrices de similarité calculées pour un ensemble de séries
- *SSéries_temporelles* : un ensemble de séries temporelles associées
- *max_score_similarité* \leftarrow 0 : score de similarité maximal dans *Matrice_similarité*
- *score_similarité* \leftarrow 0 : score de similarité pour une paire de séries
- *Serie_i, Serie_j* : séries temporelles de la paire courante
- *Série_consensus_univariée* : série consensus issue de la combinaison des séries de la paire courante pour une variable
- *Série_consensus_multivariée* : série consensus issue de la combinaison des séries de la paire courante pour toutes les variables.
- *Séries_consensus* : Ensemble des séries consensus.

Résultat :

- *SousSéries_consensus* : ensemble de sous-séries temporelles consensus entre deux nœuds

FONCTION *Consensus_hiéarchique*(*Graphe*, *Paire_de_noeuds_adjacents*, *Seuil*)

```

1  Séries_Temporelles_arcs  $\leftarrow$  Extraction_SSérie_entre_noeuds(Graphe, Paire_de_noeuds_adjacents );
2  pour SSéries_temporelles dans Séries_Temporelles_arcs faire
3      tant que max_score_similarité > Seuil faire
4          Matrice_similarité  $\leftarrow$  Calcul_matrice_similarité_par_DTW(SSéries_temporelles);
5          max_score_similarité  $\leftarrow$  Max(Matrice_similarité);
6          tant que score_similarité > Seuil faire
7              Serie_i, Serie_j, score_similarité  $\leftarrow$  Paire_la_plus_similaire(Matrice_similarité);
8              si score_similarité > Seuil alors
9                  Ajouter(Paire_de_noeuds_adjacents.Historique, Serie_i, Serie_j);
10                 Série_consensus_multivariées  $\leftarrow$   $\emptyset$ ;
11                 pour chaque Var dans Séries_Temporelles faire
12                     Matrice_alignement  $\leftarrow$  Calcul_alignement_DTW(Serie_i[Var], Serie_j[Var]);
13                     Chemin  $\leftarrow$  Calcul_chemin_alignement_de_j_sur_i(Matrice_alignement);
14                     pour k  $\leftarrow$  0 à longueur(Serie_temporelle_i) faire
15                         Série_consensus_univariée  $\leftarrow$  Moyenne(Serie_i[Var][k], Serie_j[Var][Chemin[k]]);
16                     fin
17                 Ajouter(Série_consensus_multivariée, Série_consensus_univariée);
18             fin
19             Ajouter(Séries_consensus, Série_consensus_multivariée);
20             Mettre_à_jour(Matrice_similarité, i, j);
21         fin
22     fin
23     SSéries_temporelles  $\leftarrow$  Séries_consensus  $\cup$  Séries_dissimilaires_restantes(Matrice_similarité);
24     fin
25     Ajouter(SousSéries_consensus, SSéries_temporelles);
26 fin
27 Supprimer_Arcs(Graphe, Paire_de_noeuds_adjacents);
28 pour chaque SousSéries_consensus faire
29     Créer_nouveaux_arcs(Graphe, Paire_de_noeuds_adjacents);
30 fin
31 retourner SousSéries_consensus

```

6.3.2 Construction du graphe et association partielle des séquences similaires

La construction de la représentation synthétique d'un ensemble de profils anesthésiques consiste à réaliser dans un premier temps un alignement des différentes séquences de ces profils avec la grammaire anesthésique d'une intervention.

À titre d'exemple, nous utiliserons dans cette section la séquence d'événements présentée en table 6.12, qui correspond à la grande étape de l'entrée lors d'une intervention de *cure d'une hernie inguinale sous cœlioscopie avec pose de prothèse* et dont la séquence complète est donnée en annexe A.2.

Temps (s)	ID	Événements
0	DES-enter01	Entrée du patient
0	AGE-insta01	Mise en place du patient
30	AGE-mofca01	Monitoring FC
60	AGE-mopam01	Monitoring PAM
90	AGE-mospo01	Monitoring SPO2
120	AGE-mobis01	Monitoring BIS
150	AGE-baihu01	Bair hugger
390	AGE-povvp01	pose VVP (cathlon)
450	ADG-cefab01	CEFAZOLINE bolus

Table 6.12 – Séquence d'événements de l'entrée d'une intervention de *cure d'une hernie inguinale sous cœlioscopie avec pose de prothèse*

6.3.2.1 Grammaire squelette

La grammaire joue le rôle d'un «squelette» pour le graphe. Sa structuration en plusieurs niveaux correspond à des niveaux d'abstraction différents. On peut représenter ces différents niveaux, en dérivant les règles de production, sous la forme de graphes. Comme le montre la figure 6.8, ces graphes sont ainsi des représentations d'une intervention à différents niveaux d'abstraction. Chaque non terminal de la grammaire à un niveau précis est un nœud dans le graphe qui représente ce niveau. Ce nœud est lui-même la représentation d'un sous-graphe dans le graphe du niveau d'abstraction inférieur⁸.

Si une étape anesthésique correspond à un nœud dans le graphe du niveau II, et à un sous-graphe dans le graphe du niveau III, alors un événement de cette étape anesthésique est lui-même un nœud dans le sous-graphe du niveau III, la position de ce nœud dans le sous-graphe étant conditionnée par les règles de production de la grammaire.

8. Dans de nombreux cas, ce sous-graphe est lui-même composé d'un seul nœud.

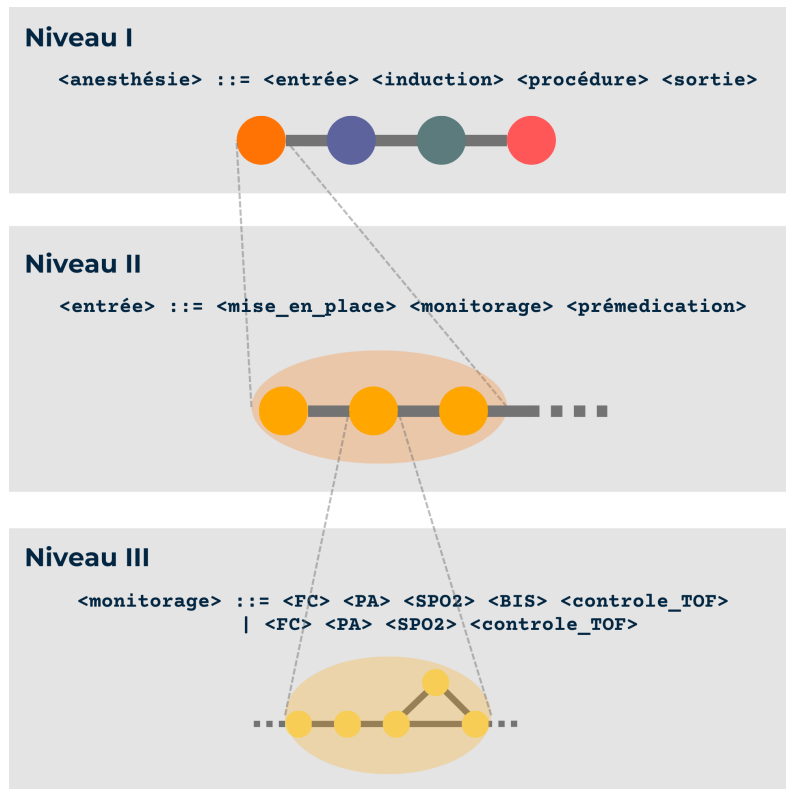


Figure 6.8 – Schéma de l’analogie entre la structure d’une grammaire anesthésique et un graphe. Dans le niveau I, le non terminal **entrée** est représenté dans le graphe par un nœud. Ce nœud correspond à un sous-graphe du graphe du niveau II. Au sein de ce sous-graphe, le nœud qui représente le non terminal **monitorage** correspond, lui aussi, à un sous-graphe au niveau III.

De plus, l’ordonnancement des événements prévu par ces règles de production nous permet d’utiliser la grammaire comme séquence de référence afin de comparer les différentes séquences issues des profils anesthésiques. L’usage de la grammaire comme référence nous affranchit alors de l’utilisation d’un algorithme d’alignement. La figure 6.8 présente ce parallélisme entre structure de la grammaire et graphe, ainsi que le principe de l’alignement prenant la grammaire comme référence.

6.3.2.2 Alignement sur la grammaire

Pour réaliser l’alignement d’une séquence sur la grammaire, il faut pouvoir associer à chaque événement de la séquence une étape anesthésique. La grammaire d’une intervention dispose déjà de cette information dans son plus bas niveau de structuration (Niveau III/IV) qui contient les identifiants des événements correspondant aux différentes actions anesthésiques.

En utilisant la grammaire d'une intervention comme une séquence de référence, il est possible de procéder à l'alignement d'une séquence avec la grammaire. Pour cela, on associe les différents événements qui composent une séquence à autant d'étapes anesthésiques.

L'association entre un événement et une étape anesthésique est ainsi réalisée par l'entremise d'un tableau associatif qui associe à chaque événement son étape anesthésique dans la grammaire. L'association événement-étape est présentée, pour l'intervention utilisée en exemple, en table 6.13. Cette séquence est présentée sous la forme d'un graphe aligné avec la grammaire en figure 6.9.

ID	I	II	Niveau	
			III	IV
DES-enter01	Entrée	Mise en place	Entrée du patient	-
AGE-insta01	Entrée	Mise en place	Entrée du patient	-
AGE-mofca01	Entrée	Monitoring	Installation monitoring FC	-
AGE-mopam01	Entrée	Monitoring	Installation monitoring PA	-
AGE-mospo01	Entrée	Monitoring	Installation monitoring SPO2	-
AGE-mobis01	Entrée	Monitoring	Installation monitoring BIS	-
AGE-baihu01	Entrée	Prémédication	Bair Hugger	-
AGE-povvp01	Entrée	Prémédication	Installation cathlon ⁹	-
ADG-cefab01	Entrée	Prémédication	Prophylaxie	-

Table 6.13 – Table d'association entre les événements de la séquence et leur étape anesthésique

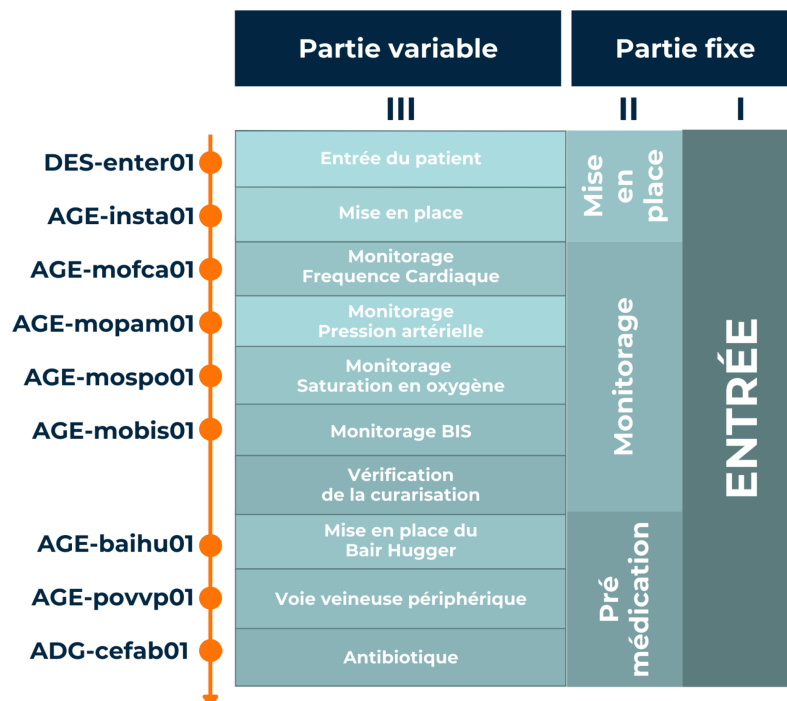


Figure 6.9 – Une séquence de l'entrée de la cure d'une hernie inguinale sous coelioscopie avec pose de prothèse alignée avec la grammaire de l'intervention. On notera l'absence dans la séquence d'un événement pour la sous-étape *Vérification de la curarisation*

6.3.2.3 Construction du graphe

Le graphe est construit de manière itérative, par présentation successive de chacune des séquences d'événements d'un ensemble de séquences. Lors de la construction du graphe, pour chacun des événements d'une séquence présentée, on crée un nœud ou on associe l'événement à un nœud pré-existant. L'ordonnancement entre les différents événements est figuré sous la forme d'arcs du graphe.

Chaque événement est caractérisé par des informations qui lui sont propres (identifiant, position dans la séquence, temps, dose, etc). Ces informations sont ajoutées au nœud sous la forme d'étiquette. Pendant le traitement de toutes les séquences, le graphe construit est ainsi enrichi d'informations sur le patient et le déroulement de l'intervention.

On pourrait chercher à associer deux séquences complètes, ce qui permettrait une première synthèse des séquences et une réduction de la taille de la représentation. Cependant, la structure de l'anesthésie, avec ses grandes étapes distinctes, a pour caractéristique une répartition hétérogène de la diversité au sein des séquences.

Ainsi, les portions de séquences concernant l'entrée et l'induction sont généralement très similaires pour une même intervention, car elles correspondent à des étapes fortement encadrées et normalisées. Pour atteindre un plus haut niveau de réduction, on peut ainsi réaliser une association partielle des séquences en associant seulement les sous-séquences similaires au lieu de séquences complètes.

Ainsi, comme le présente le fonctionnement de l'algorithme 4 en page 157, la séquence de chaque profil anesthésique (ligne 2) est ajoutée au graphe, événement par événement (ligne 3). Chaque événement est associé à une étape anesthésique (ligne 4). Au sein du graphe, les événements sont représentés par des nœuds qui portent des étiquettes décrivant l'ensemble des informations caractérisant l'événement. Pour garder la structure de la grammaire anesthésique au sein du graphe, tous les nœuds associés à une étape anesthésique sont rassemblés dans un *méta-nœud*.

Un méta-nœud contient ainsi des événements anesthésiques dans des contextes différents mais tous relatifs à la même étape anesthésique. Un méta-nœud est présenté de manière schématique en figure 6.10.

Si, lors de l'ajout de l'événement, il n'existe pas encore dans le graphe de méta-nœud pour l'étape anesthésique de l'événement, alors ce méta-nœud est créé et un nœud représentant l'événement lui est ajouté (ligne 6 à 8). Si un méta-nœud est pré-existant dans le

Similarité entre nœuds Deux nœuds dont la position dans la grammaire anesthésique et l’identifiant sont identiques ne peuvent être considérés similaires que si leurs *contextes locaux* sont similaires. La similarité du *contexte local* correspond à l’ordonnancement des événements autour d’un nœud (nœuds parents/nœuds enfants). En effet, l’ordonnement des événements anesthésiques a un effet direct sur le patient du fait de l’interaction entre les différents effets des événements sur les paramètres physiologiques du patient.

Les arcs qui correspondent à l’ordonnancement des événements dans la séquence en cours d’ajout au graphe, sont créés entre le nœud préexistant, ou nouvellement créé, et son nœud parent (ligne 13). Ces arcs sont temporairement annotés par les sous-séries temporelles entre ces deux événements avant d’être ajoutés au graphe (ligne 14 à 16).

Similarité entre les arcs Dans le graphe, les arcs représentent l’ordonnancement entre les événements. Deux arcs peuvent être considérés similaires s’ils partagent les mêmes nœuds et si les sous-séries temporelles entre ces nœuds sont elles aussi similaires.

La similarité entre les sous-séries temporelles est déterminée par le calcul d’une similarité par *Dynamic Time Warping* (DTW). La similarité entre deux séries temporelles multivariées est calculée par la moyenne des similarités des séries temporelles univariées.

Si la similarité entre les séries temporelles des arcs entre deux nœuds est suffisante à un certain seuil, alors ces arcs sont considérés similaires et sont associés (ligne 17).

Toutefois, la simple association de séries temporelles similaires ne permet pas une réduction de la quantité d’information. On pourrait choisir une série temporelle parmi le groupe de séries temporelles associées comme représentante, mais on risquerait de perdre la part de diversité présente dans les autres séries temporelles. Il est possible de conserver une part de la diversité des différentes séries temporelles associées en construisant une série temporelle consensus pour cet ensemble de séries temporelles.

Chaque ensemble de séries temporelles associées entre deux nœuds est aggloméré en une série temporelle consensus (ligne 19). Ces consensus sont ensuite associés au nœud parent de l’arc (ligne 20). Ainsi, par exemple, s’il y a 4 ensembles de séries temporelles associées entre deux nœuds¹⁰, on produit au moins 4 séries temporelles consensus.

10. Les séries temporelles entre deux nœuds sont associées entre elles uniquement si elles sont similaires. Des séries temporelles dissimilaires produisent alors plusieurs ensembles de séries temporelles associées

6.3.3 Traitement des séries temporelles, création de séries consensus et enrichissement des nœuds

La construction de séries temporelles consensus peut permettre une réduction de la taille de la représentation tout en conservant le maximum de diversité. La construction d'une série temporelle multivariée consensus revient, pour chaque variable, à réaliser une synthèse des caractéristiques des différentes séries temporelles associées entre elles.

Les différentes séries issues d'un profil anesthésique sont associées si elles présentent une forte similarité entre elles. La méthode de construction du consensus doit donc s'attacher à mettre en commun les caractéristiques communes, tout en conservant une partie des caractéristiques spécifiques de chaque séquence.

Soit deux séries temporelles, S et S' , composées respectivement de n valeurs v_i et v'_i réelles ($1 < i < n$). La méthode de construction d'une série C consensus qui combine ces deux séries est la fonction qui pour une position i des séries, à laquelle correspond les valeurs v_i et v'_i , renvoie la valeur pour la série consensus c_i .

Une approche naïve pour la formulation de cette fonction est l'utilisation de la moyenne des deux valeurs :

$$\begin{aligned} S &= [v_1, \dots, v_i, \dots, v_n] \\ S' &= [v'_1, \dots, v'_i, \dots, v'_n] \\ C &= [c_1, \dots, c_i, \dots, c_n] \\ c_i &= \frac{v_i + v'_i}{2} \end{aligned}$$

Si les valeurs sont effectivement combinées, l'utilisation de la moyenne ne permet pas la conservation des caractéristiques communes aux deux séquences. En effet, la moyenne lisse ces caractéristiques.

En revanche, un alignement par DTW est particulièrement adapté pour associer des portions de séries temporelles aux formes similaires. De plus, la DTW nous affranchit d'une contrainte forte de la moyenne : il n'est pas nécessaire que les séries temporelles aient la même longueur. Cette méthode s'appuie sur le calcul d'une matrice d'alignement qui permet d'associer à chaque valeur d'une série temporelle la valeur la plus similaire dans l'autre série.

Ainsi pour deux séries temporelles, S et S' , composées de n et m ($n > m$) valeurs réelles respectivement, un alignement par DTW associe à une valeur v_i , valeur de la séquence S à la position i , une valeur v'_j , valeur de la séquence S' à la position j .

Il est ainsi possible de combiner deux valeurs non pas en fonction de la position i des valeurs dans la série temporelle, mais de leur association par la matrice de l'alignement, comme proposé par VAUGHAN et GABRYS [130].

Tout point de S est aligné avec un point de S' et symétriquement. Ainsi, l'alignement $Al(v_i)$ définit-il une correspondance de v_i ($1 < i < n$) vers v'_j ($1 < j < m$) comme suit :

$$Al(v_i) = v'_j$$

Dans ces conditions, la construction d'une série consensus C qui combine les deux séries S et S' est l'ensemble de n ¹¹ valeurs C_i , moyenne des valeurs v_i et v'_j , comme suit :

$$\begin{aligned} S &= [v_1, \dots, v_i, \dots, v_n] \\ S' &= [v'_1, \dots, v'_j, \dots, v'_m] \\ C &= [c_1, \dots, c_i, \dots, c_n] \\ c_i &= \frac{v_i + Al(v_i)}{2} \end{aligned}$$

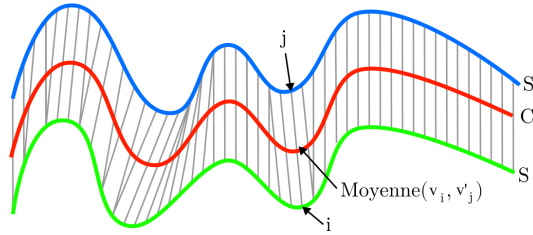


Figure 6.11 – Principe de construction d'une série temporelle consensus par l'approche de VAUGHAN et GABRYS [130]

Un alignement est spécifique aux deux séries qui sont alignées. Pour produire la série consensus de q séries temporelles ($q > 2$) il faut aligner ces séries entre elles. L'algorithme d'alignement de deux séries temporelles est un algorithme exact, de complexité $\mathcal{O}(nm)$. La complexité de l'algorithme d'alignement multiple de q séries temporelles ($q > 2$) est prohibitif. Une heuristique doit être utilisée dans ce cas. Nous avons fait le choix d'une approche itérative qui associe entre elles des paires de séries temporelles. Les figures 6.12 et 6.13 présentent cette approche de manière schématique.

¹¹. La série consensus produite est composée du même nombre de valeurs que la série temporelle la plus longue de l'alignement.

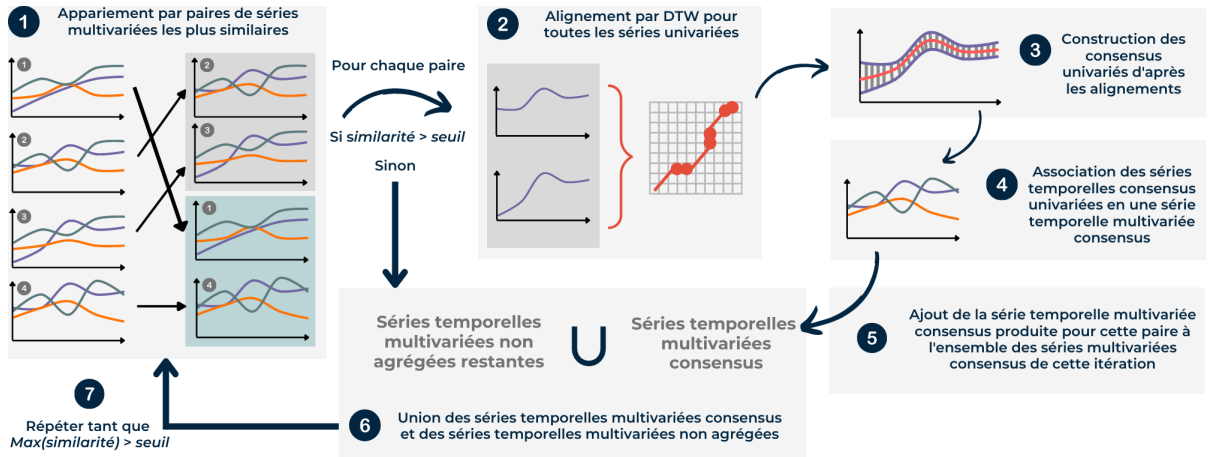


Figure 6.12 – Schéma de la création de séries temporelles consensus à partir d'alignement par DTW pour un ensemble de séries temporelles associées entre elles lors d'une itération.

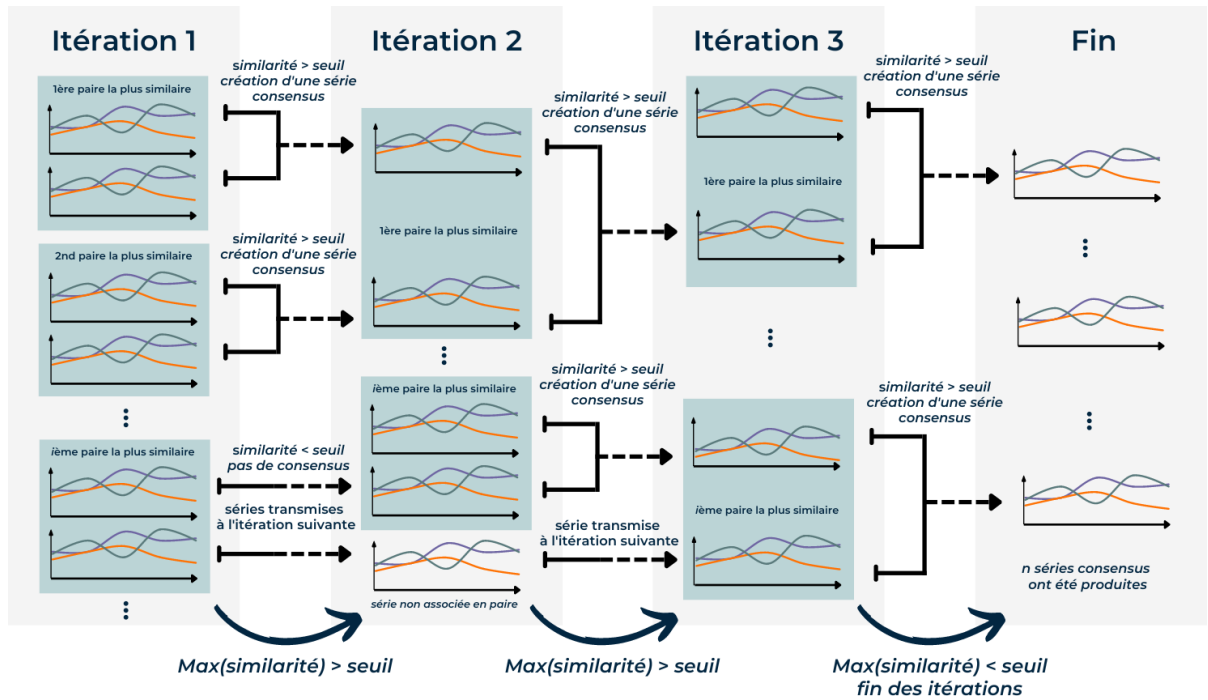


Figure 6.13 – Schéma de l'enchaînement des différentes itérations pour la création de séries temporelles consensus à partir d'alignement par DTW pour un ensemble de séries temporelles associées entre elles. Les séries consensus produites lors d'une itération sont utilisées à l'itération suivante avec les séries non agglomérées. Les itérations sont répétées tant que la similarité maximale entre deux paires de séries n'est pas inférieure au seuil fixé.

Comme le présentent l’algorithme 5 en page 158, et la figure 6.12, pour chaque ensemble de séries temporelles associées issues des arcs (ligne 1 et 2) on réalise plusieurs itérations. On commence par calculer une matrice de similarités à partir de la DTW¹² (ligne 4).

Pour chaque paire de séries temporelles les plus similaires selon la matrice de similarités (figure 6.12-1) (ligne 6), si la similarité entre ces séries est supérieure à un seuil donné (ligne 7) alors on construit leur série consensus. L’historique de construction est conservé dans les deux nœuds de l’arc (ligne 8).

Pour chaque série univariée de la paire, on réalise un alignement par DTW (figure 6.12-2) (ligne 10) et on construit une série univariée consensus (figure 6.12-3) (lignes 12 et 13). Les séries univariées consensus sont assemblées pour produire une série temporelle multivariée consensus (figure 6.12-4) (ligne 14). Cette série est ensuite ajoutée à l’ensemble des séries consensus produites pendant cette itération (figure 6.12-5) (ligne 15).

La matrice de similarité est ensuite mise à jour : les deux séries temporelles similaires sont retirées (ligne 16). Quand la similarité de la paire de séries temporelles la plus similaire est inférieure au seuil fixé, on termine cette itération (ligne 5). On réalise alors l’union de l’ensemble des séries temporelles non associées avec les séries temporelles consensus produites pendant l’itération (figure 6.12-6) (ligne 17).

On répète ces itérations de construction de séries temporelles consensus jusqu’à avoir des séries temporelles consensus qui sont trop peu similaires pour être associées entre elles (figure 6.12-7) (ligne 3).

Quand toutes les séries consensus pour les différents ensembles de séries temporelles associées ont été produites, on supprime les arcs du graphe entre les deux nœuds (ligne 18). On construit alors autant de nouveaux arcs que de séries consensus produites (ligne 19).

Une fois les séries temporelles consensus produites, elles sont ajoutées au nœud parent de la paire de nœuds auxquelles elles appartiennent.

12. La similarité entre séries temporelles est calculée comme le score d’alignement de ces deux séries (Algorithme 5, ligne 4, page 158).

6.4 Opportunités pour l'anonymisation de données

La méthode proposée pour construire une représentation synthétique d'un ensemble de profils anesthésiques se base sur le principe de l'agrégation de données, et la création de séries temporelles consensus.

Bien que ces travaux de thèse n'aient pas pour objet l'anonymisation de données, l'opportunité apportée par l'approche proposée nous conduit à réfléchir sur ce sujet, et à proposer des perspectives pour l'anonymisation de données temporelles complexes comme les profils anesthésiques.

6.4.1 Repères sur l'anonymisation de données

L'anonymisation de données personnelles consiste à faire disparaître tout élément permettant une ré-identification de la personne décrite par ces données.

Les autorités régulatrices (i.e., la CNIL¹³ en France) encadrent les pratiques de diffusion de données personnelles anonymisées et fournissent des recommandations et des règles de bonnes pratiques.

L'anonymisation est un traitement particulier appliqué aux données qui résulte d'un compromis entre conservation des informations, potentiellement identifiantes, et efficacité de l'anonymisation.

L'anonymisation de données personnelles est rendue possible grâce à plusieurs techniques, qui relèvent de deux principales approches : la randomisation et la généralisation [143, 144].

La randomisation est une approche d'anonymisation qui consiste à modifier le contenu de variables afin de briser les liens d'appartenance avec un individu [145, 146]. On peut ainsi permuter les dates de naissance de deux individus d'âges identiques (approche par substitution) [147, 148]. Il est aussi possible d'ajouter des informations différentes mais similaires au jeu de données (approche par ajout de bruit). Cette approche conserve la répartition globale des informations, mais brise potentiellement les liens entre ces informations. En réduisant la véracité des données, on affaiblit le lien entre les données et l'individu à leur origine [143].

13. La Commission Nationale de l'Informatique et des Libertés (CNIL) est une autorité indépendante chargée des questions de liberté liées au numérique.

L'anonymisation de données personnelles par généralisation modifie l'ordre de grandeur des informations. Ces informations deviennent ainsi communes à plusieurs individus. Il est ainsi possible de ne garder, pour l'âge, que les mois et années de naissance d'un individu, au lieu de la date de naissance complète (Agrégation ou k-anonymat [149]). Une approche encore plus poussée cherche à maintenir la distribution d'origine dans les diverses modalités d'une variable (l-diversité/t-proximité) [150, 151].

Toutes les techniques d'anonymisation présentent des forces et des faiblesses. L'efficacité de l'anonymisation est considérée grâce à plusieurs critères de risque : l'individualisation, la corrélation et l'inférence.

Le *risque d'individualisation* représente la possibilité d'isoler les données spécifiques d'un individu au sein du jeu de données anonymisées. Par exemple, un jeu qui contiendrait une variable, ou combinaison de variables, unique à un individu présenterait un tel risque.

Si le jeu de données permet un re-croisement des informations d'un individu, ou d'un groupe d'individus, avec lui-même ou un autre jeu, alors ce jeu ne résiste pas au *risque de corrélation*. Par exemple, un jeu contenant des adresses peut être recroisé avec un annuaire.

Le *risque d'inférence* consiste en la possibilité de déduire une information sur un individu, ou un groupe d'individus, à partir des autres informations. Si, au sein d'un jeu, tous les hommes de moins de 30 ans sont vaccinés, et si on connaît qu'un individu spécifique de moins de 30 ans est dans ce jeu, alors on en déduit qu'il est vacciné.

Les différentes méthodes d'anonymisation offrent des réponses inégales à ces différents critères d'efficacité, comme le montre la table 6.14.

Types	Sous-types	Risques		
		Individualisation	Corrélation	Inférence
Randomisation	Ajout de Bruit	Oui	Incertain	Incertain
	Substitution	Oui	Oui	Incertain
Généralisation	Agrégation / k-anonymat	Non	Oui	Oui
	L-diversité	Non	Oui	Incertain

Table 6.14 – Risques présentés par différentes méthodes d'anonymisation, d'après les recommandations du G29 [143]

Les recommandations du G29¹⁴ encouragent l'utilisation au cas par cas des techniques et la spécialisation des jeux de données anonymisés selon les objectifs recherchés [143].

Une alternative à l'anonymisation réside dans la production de données simulées réalistes, qui s'affranchissent des contraintes réglementaires des données personnelles.

6.4.2 Perspective pour des représentations synthétiques anonymisées de profils anesthésiques

L'algorithme 4 proposé en section 6.3.1 pour la construction de représentation synthétique de profils anesthésiques est focalisé sur la conservation de la diversité des séquences d'événements. Ainsi, l'association de séquences similaires se fait avec un haut niveau de conservation.

Dans le cadre d'une démarche d'anonymisation, le compromis à trouver se situe entre la conservation des informations, qui sont potentiellement identifiantes, et l'efficacité de l'anonymisation.

Si on positionne l'approche pour la construction de représentation synthétique dans le cadre de ce compromis, alors cette approche cherche à maximiser la conservation des informations sans considération pour l'effet anonymisant.

La représentation synthétique comporte plusieurs groupes de données différentes, les séries temporelles, les séquences d'événements et les éléments des dossiers médicaux des patients.

La création de consensus des séries temporelles est déjà un traitement d'anonymisation par randomisation. L'association des séquences d'événements similaires, si elle s'accompagnait d'une généralisation de certaines informations (dosage, médicaments utilisés) pourrait aussi être considérée comme un traitement d'anonymisation. Les informations des patients sont elles aussi anonymisables par des approches de k-anonymisation qui ont déjà été largement utilisées pour les jeux de données de santé [152-155].

La construction d'une représentation synthétique anonymisée implique des modifications de l'approche proposée, à la fois sur des aspects de paramétrage, et sur des aspects algorithmiques.

14. Prédécesseur du Comité Européen de la Protection des Données, un organe européen indépendant qui contribue à la collaboration et à l'application des règles en matière de protection des données au sein de l'Union européenne.

Les seuils de similarité, utilisés afin d’associer des séquences d’événements et créer des séries temporelles consensus, doivent être adaptés pour pouvoir inclure un nombre suffisant de profils anesthésiques.

Une généralisation des événements pourrait aussi être réalisée en amont de la construction d’une représentation. Toutefois, c’est prendre le risque d’associer des séquences d’événements généralisées qui ne seraient pas associées sans la généralisation. Il semble donc préférable d’intégrer aux algorithmes d’association de séquences d’événements une généralisation.

De la même manière, une k-anonymisation des informations des patients serait préférable pendant la construction d’une représentation plutôt qu’en amont, afin d’éviter le même problème de biais d’association.

Cependant, une telle approche doit se confronter aux différents risques de re-identification et faire l’objet d’une expérimentation et d’une évaluation précise de ces risques.

Conclusion

Dans le cadre de nos travaux, les scénarios d'apprentissage qui sont joués par un simulateur réactif aux actions des apprenants doivent avoir un support pour contenir les informations nécessaires à la simulation.

Un ensemble de séquences d'événements issues de profils anesthésiques peut être représenté par un graphe orienté. Une estimation de la taille de ce graphe est possible à partir d'indicateurs spécifiques calculés sur les séquences. Cela permet de prédire les caractéristiques du graphe d'un ensemble de séquences issues d'une cohorte de profils anesthésiques, sans avoir à construire ce graphe, et ainsi à en contrôler la complexité.

La structure décrite par la grammaire d'une anesthésie peut être utilisée pour construire une représentation synthétique d'un ensemble de profils anesthésiques. La construction de cette représentation s'appuie sur une association des séquences d'événement similaires et sur la création de séries temporelles multivariées consensus.

Ces représentations synthétiques, en contenant toutes les informations relatives aux patients et utilisables pour la simulation réactive d'un patient virtuel, constituent un support adapté pour la génération de scénarios d'apprentissage.

De plus, ces représentations synthétiques, par l'agglomération de profils anesthésiques, ouvrent des opportunités pour l'anonymisation de données en grandes dimensions en faisant disparaître les caractères identifiants d'un patient tout en conservant les dépendances au sein des profils anesthésiques.

GÉNÉRATION ET SIMULATION DE SCÉNARIOS D'APPRENTISSAGE CONTEXTUALISÉS

Introduction

La finalité de ces travaux est la génération automatisée de scénarios réactif contextualisés pour la formation par la simulation en anesthésie. Ces scénarios seront ensuite soumis à des apprenants afin de leur faire intégrer de nouvelles compétences. Un scénario a ainsi une finalité pédagogique qui se concrétise dans un contexte médical précis. Pour jouer un scénario, il ne suffit donc pas de simuler un patient virtuel, mais il faut guider l'apprenant vers des situations d'apprentissage spécifiques. Ce chapitre présente comment, à partir des éléments présentés dans les chapitres précédents, il est possible de produire un scénario de simulation dont on peut contrôler la finalité pédagogique et le contexte médical. Ce chapitre présente ensuite le fonctionnement d'un simulateur réactif pour jouer ces scénarios.

Dans une première partie, ce chapitre s'intéressera aux différents éléments qui caractérisent et contextualisent les scénarios, notamment le contexte médical et son éventuelle altération à des fins pédagogiques.

La seconde partie présente comment il est possible de générer des scénarios et de simuler un patient virtuel à partir d'une représentation synthétique. Il y est notamment présenté les différentes approches de modification de cette représentation, qui permettent de contrôler tant le contexte médical de la simulation que sa finalité pédagogique.

La dernière partie assemble les différents éléments présentés dans les chapitres précédents pour construire un prototype de simulateur réaliste permettant de jouer les scénarios d'apprentissage.

7.1 Les scénarios d'apprentissage contextualisés

Les scénarios d'apprentissage en simulation doivent donner l'opportunité à un apprenant de se confronter à une situation spécifique, à des fins pédagogiques, en participant à une intervention médicale simulée. Comme nous l'avons indiqué en 1.2.3.3, un scénario se caractérise par différents éléments qui correspondent à une double contextualisation : médicale et pédagogique [49, 50] .

Les éléments liés à la contextualisation médicale du scénario correspondent : (i) au contexte de prise en charge du patient, une intervention en urgence ou programmée par exemple, (ii) à l'environnement médical simulé, qui peut être un cabinet ou un bloc opératoire et (iii) à la description du patient virtuel, avec les éléments clés de son dossier médical et sa pathologie.

La composante pédagogique de la contextualisation du scénario comprend : (i) une spécialisation aux apprenants auxquels le scénario est destiné (e.g le contenu sera différent pour un interne ou un IADE), et (ii) la progression du scénario, qui conduit l'apprenant dans des situations pédagogiques choisies, avec éventuellement des éléments perturbateurs du scénario.

Les chapitres précédents ont présenté différentes méthodes permettant la simulation d'un patient virtuel réaliste et réactive aux actions d'un apprenant. Ce qui caractérise un scénario d'apprentissage vis-à-vis des simulations présentées précédemment, c'est le contrôle imposé par la contextualisation médicale et pédagogique sur la simulation.

Au-delà du gain d'expérience offert à l'apprenant par la simulation, dans un cadre de formation, les scénarios doivent lui permettre de se positionner vis-à-vis d'une référence. Dans le cadre de scénarios d'apprentissage en anesthésie, ce sont les référentiels de bonnes pratiques qui servent de référence. Pour pouvoir positionner, un apprenant vis-à-vis des référentiels de bonnes pratiques, il est nécessaire que le scénario le confronte à des situations standardisées. Ainsi, dans ce contexte pédagogique, une simulation non contrôlée correspondrait à prendre le risque que l'apprenant ne se confronte pas aux situations standardisées et passe à côté des buts de l'activité pédagogique.

Les scénarios n'ont donc pas seulement besoin d'être réactifs aux actions de l'apprenant et de permettre l'évolution d'un patient virtuel réaliste, mais ils doivent conduire l'apprenant dans des situations lui permettant de mobiliser ou de travailler de nouvelles

compétences. Pour cela, il est nécessaire de donner aux formateurs une possibilité de contrôle sur le contenu et le déroulement des scénarios. Ce contrôle se réalise sur deux aspects : en conditionnant le contexte médical du scénario, ou en altérant le déroulement du scénario pour des raisons pédagogiques.

7.1.1 Conditionnement du contexte médical

La pratique de l'anesthésie s'insère évidemment dans un contexte médical, mais surtout dans une finalité de soin. L'intervention réalisée sur un patient vient répondre à une pathologie ou à la réalisation d'un geste médical. En effet, l'éthique médicale interdit la réalisation d'une intervention non motivée par la santé du patient. Ce lien entre pathologie et soin apporte une première dépendance entre le patient et l'intervention à réaliser.

L'intervention à réaliser, qui nécessite une anesthésie, conditionne le déroulement et les choix opérés par les soignants pendant toute l'opération. Ainsi, certains médicaments ou techniques sont utilisés préférablement à d'autres pour une intervention donnée. Il y a donc une relation de dépendance entre l'intervention et l'anesthésie.

Cette double dépendance, entre patient et intervention, puis entre intervention et anesthésie, a pour conséquence que le patient utilisé à la base du scénario conditionne directement les situations auxquelles peut être confronté un apprenant. Il est donc possible de contrôler le cadre général et le contenu d'un scénario *via* le patient autour duquel il est construit.

En fonction des choix pédagogiques, on va donc choisir le type de patient et sa pathologie pour faire travailler les apprenants sur un sujet précis.

Toutefois, ce contrôle indirect ne permet pas de réorienter un apprenant dans la bonne direction en cas de mauvaise pratique, ou vers un nouveau problème à résoudre. Un contrôle direct sur le scénario est donc nécessaire par altération de son déroulement.

7.1.2 Altération du déroulement de la simulation dans un but pédagogique

Soumettre à l'apprenant des situations spécifiquement choisies correspond à le confronter à un état spécifique des paramètres physiologiques du patient virtuel et à un historique spécifique des actions médicales réalisées. Les paramètres physiologiques simulés du patient virtuel sont le résultat de l'impact des différents événements présents dans la séquence d'événements. Pour contrôler le déroulement de la simulation, il faut donc contrôler la séquence d'événements du patient virtuel et les impacts de ces événements.

7.1.2.1 Types d'altérations

Le contrôle de la séquence d'événements consiste à altérer le déroulement «normal» de la simulation du patient virtuel, en ajoutant ou supprimant des événements, dont l'impact a éventuellement été altéré, permettant de constituer les situations médicales spécifiques auxquelles on souhaite confronter l'apprenant. Les altérations peuvent prendre plusieurs formes :

- l'inclusion d'événements, ou de séquences d'événements, par exemple une réaction allergique,
- la diminution ou l'accentuation des effets d'un événement sur les paramètres physiologiques du patient virtuel,
- la modification des probabilités d'instanciation de certaines actions de l'équipe médicale ou de certaines réactions du patient virtuel.

Ces différentes altérations offrent au formateur une boîte à outils qui permet de constituer des situations d'apprentissages contrôlées.

7.1.2.2 Contrainte de crédibilité

Comme le mentionne la littérature, l'apprentissage est plus efficace si la simulation est fidèle à la réalité [34], notamment lors de l'apprentissage de tâches complexes [38].

Il serait pédagogiquement incohérent, en altérant brutalement le déroulement de la simulation, d'altérer la fidélité de la situation d'apprentissage à laquelle est confronté l'apprenant, en la rendant peu crédible. Ainsi, à l'altération du déroulement de la simulation, s'applique une contrainte de maintien de la crédibilité de la situation.

Pour ne pas briser la fidélité de la simulation, on peut interagir avec l'apprenant pour lui indiquer des éléments de contexte qui vont crédibiliser la situation d'apprentissage.

Pour cela, on peut tirer parti des événements descripteurs, qui sont présents dans la séquence d'événements du patient virtuel pour décrire le déroulement de l'intervention. En ajoutant à la séquence d'événements du patient virtuel des événements descripteurs *pédagogiques*, on peut interagir avec l'apprenant en l'informant de différents éléments concernant l'environnement, l'équipe médicale ou le dossier médical du patient. On peut ainsi par exemple crédibiliser l'apparition d'une situation imprévue dans le déroulement de l'intervention, mais souhaitée par les formateurs, en informant l'apprenant des résultats d'une analyse médicale.

7.2 Génération et simulation de scénarios à partir d'une représentation synthétique

Produire des scénarios d'apprentissage n'implique donc pas seulement de simuler un patient réaliste et réactif aux actions de l'apprenant, mais il nous est en plus nécessaire de donner aux formateurs la possibilité d'altérer la simulation.

7.2.1 Limites des modalités de simulation

Les modalités de simulation présentées en 5.4.1 et qui s'appuient uniquement sur les données d'une cohorte de patients ne permettent pas de simuler de tels scénarios d'apprentissage.

En effet, il faut pouvoir simuler un patient virtuel réaliste et réactif aux actions d'un apprenant qui s'écarterait trop des bonnes pratiques. Or, ces approches précédemment décrites ne permettent de simuler que des situations connues et donc limitées aux bonnes pratiques. Comme les situations qui ne correspondent pas aux bonnes pratiques sont particulièrement rares, si un apprenant s'écarte trop des bonnes pratiques, cela risque de mettre en échec cette modalités de simulation. C'est une limitation de l'approche par raisonnement à base de cas qui est contrainte par les données qu'elle utilise.

Un second problème est posé par l'altération du déroulement de la simulation. Ces modalités s'appuient uniquement sur les actions de l'apprenant et ne prennent pas en compte un scénario au contenu précis. De plus, ces modalités de simulation ne peuvent pas simuler l'impact d'un événement dont les effets sur les paramètres du patient virtuel auraient été modifiés pour des raisons pédagogiques. Il n'est donc pas possible pour le formateur de contrôler le déroulement de la simulation avec ces modalités.

Dans le chapitre 6, nous avons décrit une structure de données qui représente un ensemble de profils anesthésiques en s'appuyant sur la grammaire d'une intervention : une représentation synthétique.

Il est possible de palier les problèmes des modalités de simulation mentionnées ci-dessus en utilisant une représentation synthétique comme support de la simulation.

Pour utiliser une représentation synthétique comme support d'un scénario de simulation, il faut pouvoir réaliser deux choses : (i) simuler un patient virtuel en fonction des actions d'un apprenant, (ii) permettre le contrôle de la simulation par les formateurs en altérant la représentation synthétique pour produire un *scénario*.

7.2.2 Simulation d'un patient virtuel à partir d'une représentation synthétique

Il est possible d'utiliser l'algorithme 3 présenté en 5.2.3 avec une représentation synthétique comme structure de donnée support de la simulation d'un patient virtuel.

7.2.2.1 Approche à base de cas appliquée aux représentations synthétiques

Notre approche de raisonnement à base de cas consiste en deux étapes principales : (i) la recherche de patients similaires vis-à-vis du passé récent du patient virtuel en cours de simulation, (ii) la prédiction de l'avenir du patient virtuel basé sur les patients les plus similaires trouvés.

La représentation synthétique contient, sous la forme d'un chemin dans le graphe et des informations portées par les nœuds, l'ordonnancement des événements horodatés des différents profils anesthésiques. Il est donc possible de calculer pour chaque patient représenté dans la représentation synthétique les deux critères sur les séquences d'événements du score de similarité multivarié présenté en 5.3.4. Ces critères sont la similarité entre les ordonnancements des événements, et la similarité entre les intervalles entre les événements.

Toutefois, la série temporelle multivariée d'un patient n'est plus présente dans la représentation synthétique, elle a été associée à celles d'autres patients dont les séries temporelles étaient très similaires, pour constituer une série temporelle consensus. On compare alors la série temporelle multivariée du patient virtuel avec les séries consensus. Cela représente toutefois une altération de l'approche à base de cas, la série consensus n'étant plus relative à un cas, mais à un ensemble de cas similaires.

Il est donc possible de retrouver, dans la représentation synthétique, des patients similaires vis-à-vis du passé récent du patient virtuel en cours de simulation et de réaliser un classement de ces patients comme le décrit l'algorithme 3.

Nous avons présenté en 5.4.1 plusieurs modalités de simulation. La modalité produisant la simulation la plus réaliste est la modalité utilisant un consensus des patients les plus similaires. Les séries temporelles de nos patients réels sont elles aussi associées, dans la représentation synthétique, sous la forme d'un consensus. En simulant le futur immédiat du patient virtuel à partir de la série temporelle consensus la plus similaire au patient virtuel, on simule en fait le consensus de plusieurs patients réels très similaires.

7.2.2.2 Qualité de la simulation et performances

La qualité de la simulation avec cette nouvelle modalité a été évaluée de la même manière que les autres modalités de simulation (voir 5.4.2). Une séquence d'événements réel a été utilisée avec une représentation synthétique non altérée. Comme le montre la figure 7.1 et la table 7.1, la simulation avec une représentation synthétique est de qualité équivalente (RS : Moyenne = 0.823, Médiane = 0.824) à la modalité basée sur des consensus (GS_CONS_10 : Moyenne = 0.817, Médiane = 0.820). De plus, la dispersion des valeurs (RS : 0.017) est plus réduite que la modalité basée sur des consensus (GS_CONS_10 : 0.020). En effet, cette modalité construit un consensus à partir des 10 meilleurs patients réels, indépendamment de la similarité entre leurs séries temporelles, alors que les consensus contenus dans la représentation synthétique sont constitués uniquement à partir de séries temporelles similaires (selon un seuil défini, voir chapitre 6).

De plus, la modalité basée sur une représentation synthétique a montré de meilleures performances (RS : 0.320 s) que les autres modalités de simulation qui associent plusieurs patients (GS_AVG_10 : 0.356 s et GS_CONS_10 : 0.396 s), comme le montre la figure 7.1 et la table 7.2. En effet, la représentation synthétique comportant déjà en son sein les séries temporelles consensus, il n'est pas nécessaire de les générer à la volée pendant la simulation. Cette modalité a donc une performance (RS : 0.320 s) proche de la performance de la modalité utilisant le patient réel le plus similaire (GS_BEST : 0.311 s).

Stratégie	Similarité (%)		
	Moyenne	Médiane	Écart-type
GS_BEST	0,694	0.721	0.0604
GS_AVG_10	0,783	0.776	0.0697
GS_CONS_10	0,817	0.820	0.0205
RS	0,823	0.824	0.017

Table 7.1 – Caractéristiques des différentes distributions \mathcal{D}_{Sim_k}

Stratégie	Temps de recherche (s)		
	Moyenne	Médiane	Écart-type
GS_BEST	0.311	0.311	0.00051
GS_AVG_10	0.356	0.356	0.00050
GS_CONS_10	0.396	0.396	0.00077
RS	0.320	0.320	0.00060

Table 7.2 – Caractéristiques des différentes distributions des temps de recherche des patients les plus similaires et de construction de la série temporelles multivariée avant simulation.

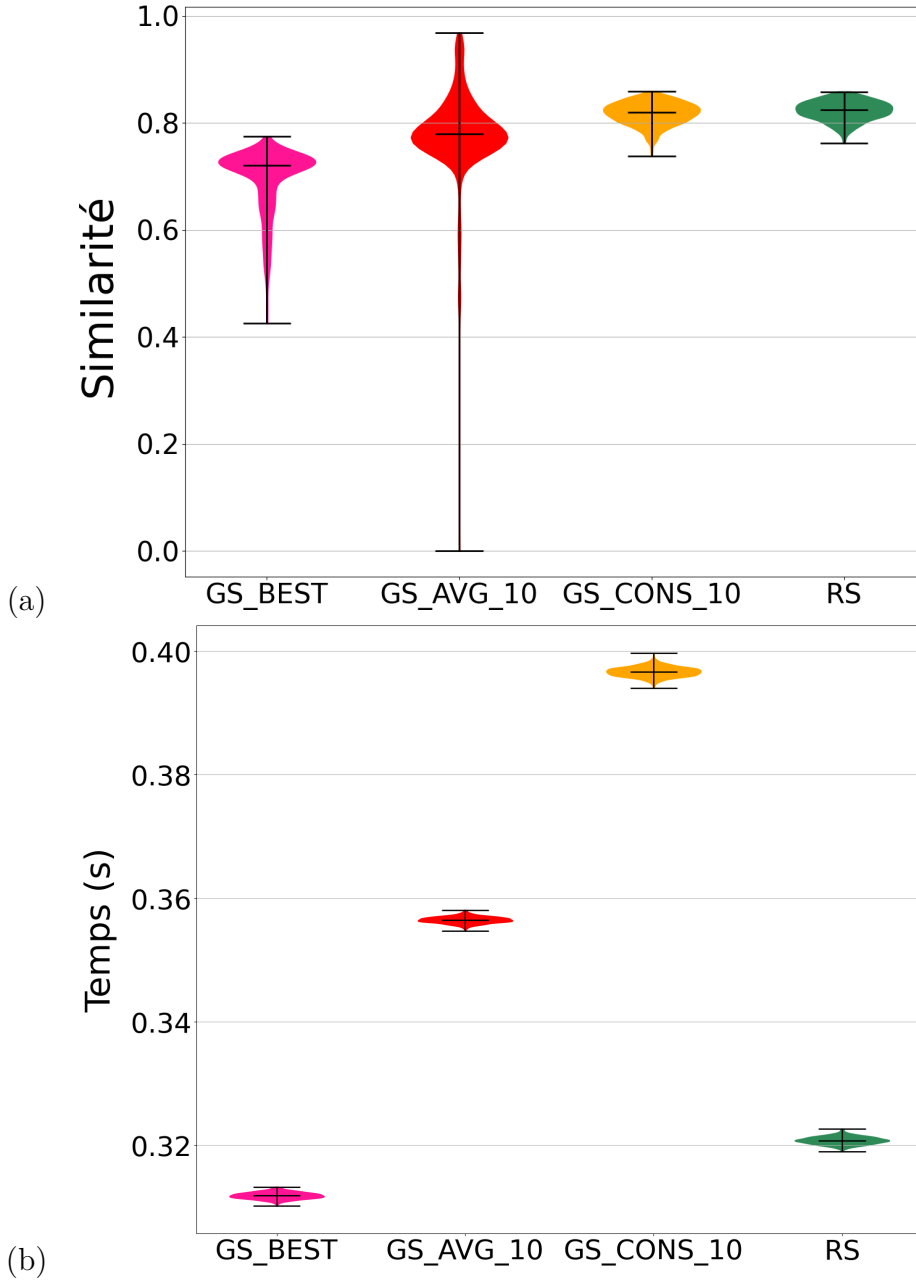


Figure 7.1 – (a) Distribution des similarités \mathcal{D}_{Sim_k} entre séquences simulées et séquences réelles, pour les trois modalités présentées en 5.4.2 et la modalité utilisant une représentation synthétique. Les barres supérieures et inférieures donnent l'étendue de la distribution, la barre centrale indique la médiane. GS_BEST : patient le plus similaire, GS_AVG_10 : moyenne des 10 patients les plus similaires, GS_CONS_10 : consensus des 10 patients les plus similaires, RS : Simulation à partir d'une représentation synthétique.

(b) Distributions des temps de recherche des patients les plus similaires et de construction de la série temporelle multivariée avant simulation.

7.2.3 Altération d'une représentation synthétique

Au-delà de pouvoir simuler un patient virtuel avec une représentation synthétique, pour pouvoir produire des scénarios d'apprentissage contextualisé, il est nécessaire de permettre le contrôle de la simulation par les formateurs en altérant la représentation synthétique.

Une représentation synthétique est constituée d'un graphe dont les nœuds représentent les événements survenus pendant une intervention. Ces nœuds sont enrichis d'informations concernant l'événement qu'ils représentent et des séries temporelles consensus. Ces éléments sont ensuite utilisés pour simuler le patient virtuel comme présenté plus haut.

Pour altérer le déroulement de la simulation du patient virtuel, il suffit que les informations contenues dans les nœuds soient cohérentes avec la simulation souhaitée. On peut pour cela modifier le graphe de la représentation synthétique en ajoutant de nouveaux nœuds. Ces *nœuds artificiels* peuvent avoir été créés par les formateurs ou être des nœuds de la représentation, dont les informations qu'ils portent ont été modifiées. Ces modifications sont réalisées grâce aux altérations présentées en 7.1.2.1. Une représentation synthétique altérée peut ainsi être le support d'un scénario d'apprentissage.

7.2.3.1 Modification des effets d'un événement

L'effet d'un événement sur les paramètres physiologiques du patient virtuel se concrétise par la variation des variables de la série temporelle multivariée du patient.

Notre approche par raisonnement à base de cas s'appuie sur des séries temporelles consensus pour simuler le patient virtuel. Pour modifier les effets d'un événement, il faut donc remplacer les séries temporelles consensus associées à cet événement par d'autres.

Pour produire les séries temporelles consensus de remplacement appropriées, deux approches s'offrent à nous.

Dans une première approche, il est possible de transformer les séries temporelles consensus du nœud à l'aide de l'une des transformations sur les séries temporelles présentées dans la table 4.3 du chapitre 4. La perturbation *ampl_unif* permet aux formateurs de déplacer la série temporelle vers le haut, pour représenter une augmentation des valeurs d'un paramètre physiologique. Les perturbations *dilat_unif* et *dilat_biunif* augmentent l'amplitude à chaque point de la série, selon d'autres modalités, ce qui permet aux formateurs de représenter un effet plus important de l'impact d'un événement. Il est aussi

possible d'utiliser la perturbation *dilat_biunif_e*, qui modifie l'amplitude dans les deux dimensions de la série temporelle que sont les valeurs du paramètre et la durée de la série temporelle.

Une seconde approche peut être de créer un événement artificiel. Pour modéliser l'impact artificiel de cet événement, on peut avoir recours à une modélisation du type de celle mise en oeuvre dans le générateur de profils anesthésiques DBLBS présenté en 3.2.2. On rappelle que DBLBS n'utilise pas de données, mais des connaissances expertes. On peut alors réaliser la simulation de cet impact d'événement artificiel pour un unique patient, utilisé comme consensus, dans la représentation synthétique. On peut aussi générer des sous-séries temporelles pour plusieurs patients, dont on réalise ensuite le consensus.

Ces deux approches nous permettent de produire un nœud artificiel à intégrer dans la représentation synthétique.

7.2.3.2 Inclusion d'événements et de déclencheurs

Les événements appartiennent à différentes catégories (voir 3.2.2), dont :

- Des réactions du patient, déclenchée par le simulateur
- Des actions, déclenchées par l'équipe médicale simulée, ou bien par l'apprenant.

Pour inclure des événements dans la séquence d'événements du patient virtuel, il faut ajouter un nœud dans la représentation synthétique. Au sein du graphe de la représentation synthétique, les nœuds sont contenus dans des métanœuds relatifs à une étape anesthésique. L'inclusion d'un événement E dans le graphe $G(M, A)$ consiste à associer cet événement au métanœud correspondant à son étape anesthésique, et à construire les arcs adaptés à l'ordonnancement souhaité par les formateurs.

Pendant la simulation d'un patient virtuel, l'apprenant réalise les différentes actions médicales en fonction de celles qui lui sont proposées. À chaque étape de l'intervention, les actions qui sont proposées à l'apprenant sont les nœuds enfants du dernier nœud instancié par l'apprenant.

Si on souhaite qu'un apprenant se voit proposer une action médicale supplémentaire, il faut alors ajouter un arc sortant dans le graphe, à partir du nœud où on souhaite plus de choix. Cet aspect peut être réalisé, en mode *off-line* ou au contraire, dynamiquement pendant la simulation.

Si un apprenant présente une maîtrise suffisante lors de la simulation, le formateur peut

souhaiter le confronter à une nouvelle situation d'apprentissage. Le formateur prévoira alors le déclenchement d'un événement, ou d'une séquence d'événements, permettant de conduire l'apprenant vers cette nouvelle situation. Il en conditionne le déclenchement à l'aide d'indicateurs du niveau de maîtrise de l'apprenant, par exemple le temps passé à réaliser une précédente étape anesthésique.

Pour ce cas précis, un formateur peut prévoir dans la représentation synthétique le déclenchement d'une réaction, par exemple l'apparition d'un pneumothorax¹, ce qui permet ainsi de conduire l'apprenant vers une nouvelle situation d'apprentissage.

Le formateur peut aussi vouloir interdire le choix d'une action médicale à l'apprenant. Par exemple, pour forcer l'apprenant à utiliser d'un médicament dont les conditions d'utilisation sont particulières, le formateur peut interdire le choix d'autres médicaments du même type.

Ainsi, tous les nœuds artificiels comportent un champ «déclencheur» qui contient différentes conditions d'exécution, par exemple l'existence d'un événement dans le passé du patient virtuel, la durée de simulation d'une étape, ou un niveau de valeur pour l'un des paramètres physiologiques. Si les conditions de déclenchement ne sont pas satisfaites, alors l'événement n'est pas instancié ou n'est pas proposé à l'apprenant.

7.2.3.3 Modification des probabilités d'instanciation

Les événements qui impactent le patient virtuel sont soit des actions médicales, soit des réactions du patient (voir 3.2.2). Les actions médicales sont réalisées par l'apprenant, qui décide de l'action à réaliser sur le patient virtuel, ou bien par l'équipe médicale.

Quand, pour une étape anesthésique, il y a la possibilité que l'équipe médicale déclenche une action, ou que le patient ait une réaction, alors ces événements sont instanciés selon leur probabilité d'instanciation. Cette probabilité est calculée comme le ratio du nombre de patients de la représentation synthétique dont la séquence d'événements présente cet événement, sur le nombre total de patients ayant contribué à la représentation synthétique.

Un formateur peut modifier cette probabilité d'instanciation s'il souhaite que l'événement soit instancié plus fréquemment, ou à chaque simulation. Ainsi, par exemple, le formateur peut s'assurer de confronter un apprenant à une réaction allergique dans le cadre d'un scénario, bien que cette réaction soit rare en situation réelle.

1. Un pneumothorax est la présence d'air entre les deux couches de la plèvre (la membrane qui recouvre les poumons et qui tapisse aussi l'intérieur de la paroi thoracique), entraînant des difficultés respiratoires

7.3 Simulateur réactif

Notre contribution dans le cadre de ces travaux se positionne dans les champs de l'exploration de données, de la prédiction et du traitement de séquence d'événements et de séries temporelles. Les mannequins haute fidélité se positionnent dans le domaine de la robotique et les environnements virtuels dans celui de la réalité virtuelle, deux domaines qui ne sont pas l'enjeu de ces travaux. Le choix des caractéristiques du simulateur pour lequel nous avons opté s'est donc concentré sur ses capacités de simulation et non sur son interface. Nous avons ainsi fait le choix d'un simulateur mono-utilisateur basse fidélité².

Ce simulateur apporte plusieurs avantages qui répondent bien aux objectifs de ces travaux et à la durée d'une thèse :

- la facilité de conception, il n'est pas nécessaire de construire une interface complexe, soit en réalité virtuelle ou par l'entremise d'un mannequin de simulation.
- l'absence de matériel spécifique, car ce simulateur ne nécessite pas de casque de réalité virtuelle, d'accessoire spécifiques ou de mannequin de simulation.
- la capacité d'évolution, le simulateur peut facilement évoluer vers les interfaces de haute fidélité que nous avons mentionnés au chapitre 1.

De plus, si on peut penser que ce simulateur ne serait que transitoire, en attendant une évolution en simulateur haute fidélité, il propose pourtant une modalité mono-utilisateur, qui est un dispositif de simulation qui est rarement offert dans les centres de simulation.

7.3.1 Contraintes

Le cadre pédagogique dans lequel sera utilisé le simulateur impose différentes contraintes. Premièrement, nous nous plaçons dans un contexte où seules les bonnes pratiques médicales doivent être appliquées ; il n'est donc pas possible de proposer une infinité d'actions à l'apprenant mais uniquement des actions cohérentes médicalement avec le déroulement de l'intervention. Ainsi, on ne propose pas à un apprenant d'arrêter le respirateur pendant la procédure chirurgicale, ou bien d'intuber le patient sans qu'il ne soit endormi. Deuxièmement, nous sommes dans le cas d'un support numérique pour la formation en anesthésie, c'est donc un soignant en formation qui réalise tous les actes liés à l'anesthésie. Toutefois, l'apprenant se place dans un contexte réaliste et il est donc accompagné d'une équipe

2. Un simulateur basse fidélité correspond à un dispositif de simulation dont l'*aspect* est de faible fidélité par rapport à la réalité. Cette caractéristique de basse fidélité ne s'applique pas aux scénarios qui sont joués sur ce simulateur

médicale virtuelle qui peut interagir avec le patient. La survenue des actions de l'équipe médicale est réalisée par le simulateur en suivant le scénario, alors que le simulateur doit réagir aux actions qui sont décidées par l'apprenant. Le simulateur doit donc permettre une interaction contrôlée et réactive avec l'apprenant.

Des différentes contraintes pédagogiques, découlent des contraintes de simulation pour permettre la génération de scénarios réactifs et réalistes. Ces contraintes s'appliquent aussi, pour partie, aux scénarios d'apprentissage.

Contrainte de réactivité Il est nécessaire que, pour toute action de l'équipe médicale (y compris l'apprenant), les paramètres physiologiques du patient virtuel évoluent en réaction à l'action réalisée.

Contrainte de réalisme L'enchaînement d'actions de l'équipe médicale entraîne éventuellement une combinaison d'effets synergiques ou antagonistes sur certains paramètres du patient virtuel. Ces effets, synergiques ou antagonistes, doivent être simulés et réalistes.

Contrainte de cohérence médicale Les choix offerts à l'apprenant doivent être médicalement cohérents avec la situation du patient virtuel, dans un cadre de bonnes pratiques.

Contrainte d'adaptabilité Selon les objectifs pédagogiques, pour faire travailler les apprenants sur certains gestes ou procédures, un scénario peut être modifié pour s'écarter du réalisme médical (allongement du temps, non-décès d'un patient).

Contrainte de performance La prédiction du patient virtuel ne doit pas perturber le bon déroulement du scénario par un temps de calcul trop long, après chaque action de l'équipe médicale ou de l'apprenant.

Les différents éléments présentés dans les chapitres précédents et plus haut, apportent des solutions à ces différentes contraintes. Ainsi, la simulation d'un patient virtuel réactif aux actions d'un apprenant proposée par l'algorithme 3 permet de répondre aux contraintes de réactivité et de réalisme. L'utilisation d'une représentation synthétique comme support du scénario d'apprentissage et comme support de la simulation permet de satisfaire les contraintes de cohérence médicale et d'adaptabilité. Enfin, comme nous l'avons montré plus haut, ces éléments s'exécutent avec des performances adaptées.

7.3.2 Construction du simulateur

Le prototype de simulateur construit pour jouer un scénario d'apprentissage a été réalisé en Python 3.10. Les représentations synthétiques altérées, qui servent de support au scénario, ont été implémentées sous la forme de fichiers *JSON*.

Le choix de la représentation altérée conditionne l'intervention et le type de patient auxquels sera confronté l'apprenant pendant le scénario.

7.3.2.1 Aspects fonctionnels

Le simulateur fonctionne de manière similaire à la première étape du générateur de profils anesthésiques DBLBS décrit en 3.3.1.

Au début du scénario, la représentation synthétique est instanciée sous la forme d'un graphe orienté. La génération de la séquence d'événements du patient virtuel est réalisée en se déplaçant de nœud en nœud dans ce graphe, en fonction des choix de l'apprenant ou des probabilités d'instanciations des réactions du patient ou des actions de l'équipe médicale.

Quand un événement est instancié, les paramètres physiologiques du patient virtuel sont simulés à partir des séries consensus de la représentation, comme présenté en 7.2.2.

En fonction des événements qui sont instanciés, la représentation synthétique évolue par neutralisation de certains arcs, ou par modification de certaines probabilités d'instanciation. Ainsi, si un événement se caractérise par son unicité, il ne peut plus être instancié une nouvelle fois.

De plus, cette évolution dynamique de la représentation synthétique donne la possibilité aux formateurs d'une altération en temps réel, et donc un contrôle en temps réel du scénario.

7.3.2.2 Interface

Le simulateur ne doit pas seulement être capable de simuler un patient virtuel, mais il doit interagir avec l'apprenant. Cette interaction comprend deux aspects : (i) transmettre à l'apprenant les informations sur le déroulement de l'intervention, sur l'état du patient virtuel, et les prochaines actions qu'il est possible de réaliser ; (ii) permettre à l'apprenant de choisir l'action qu'il souhaite réaliser.

Le design d'interface n'étant pas l'enjeu de ces travaux, l'interface réalisée pour permettre ces interactions avec l'apprenant est basique. Toutefois, elle prend en compte le comportement des soignants lors d'une intervention et leurs usages des différents appareils.

En effet, les échanges avec les experts nous ont appris qu'au cours d'une intervention, toutes les variables physiologiques du patient n'ont pas toute la même pertinence. Ainsi, la température n'est pas scrutée par les soignants en permanence, au contraire de la fréquence cardiaque. De plus, les différentes variables sont observées par les soignants de 3 principales manières :

- par le scope³, qui présente le monitoring de certaines variables hémodynamiques (FC, PAM, PAS, PAD) et respiratoires (SPO2, FR) ainsi que la température,
- par le respirateur, qui présente les variables respiratoires,
- par constatation ou information *via* l'équipe médicale, d'un élément concernant le patient. Cela peut être un mouvement du patient ou les résultats d'une analyse.

Ainsi, l'interface réalisée se décompose en deux unités : Le scope avec les principaux paramètres vitaux, et l'interface principale composée d'un journal d'événements, des actions proposées à l'apprenant et des variables du respirateur.

Le scope produit par le simulateur est présenté en figure 7.2 (a) ; l'interface principale est présentée en sous figure (b).

3. Le scope (ou moniteur) est un écran qui permet de suivre en permanence les paramètres vitaux du patient qui sont monitorés par différents capteurs installés au début de l'intervention.

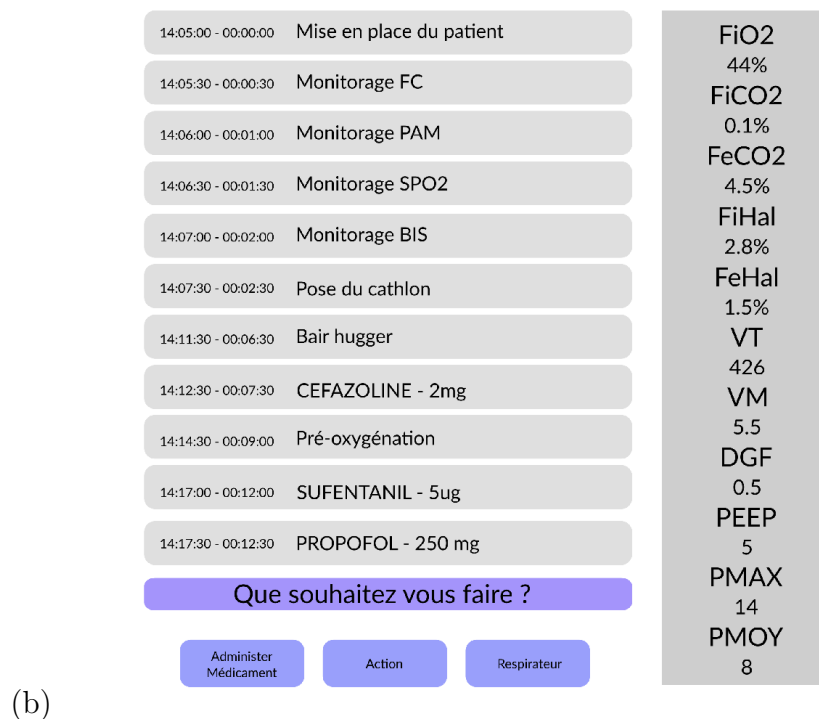
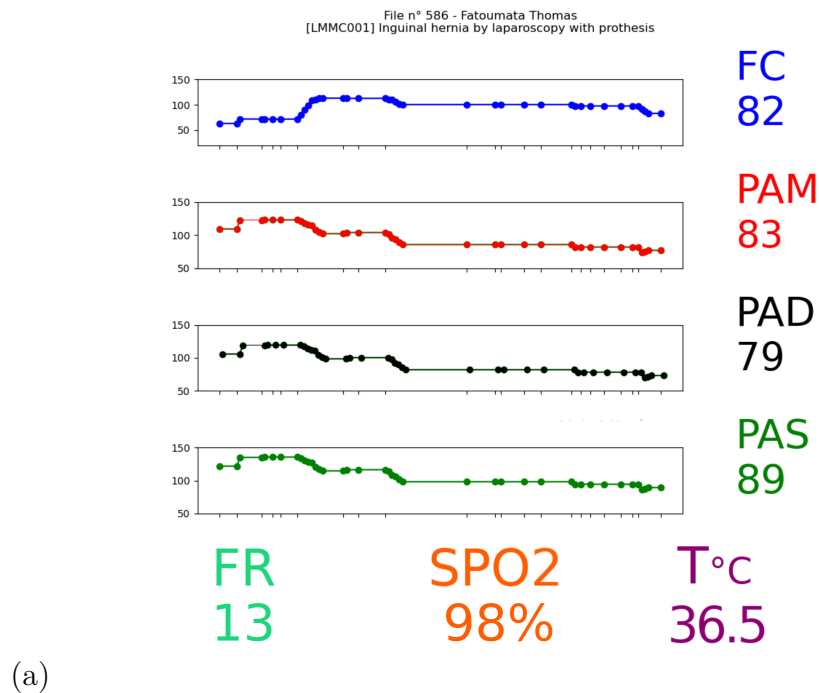


Figure 7.2 – (a) Scope présentant à l'apprenant les principaux paramètres physiologiques du patient virtuel de manière similaire à celle du bloc opératoire (b) Journal d'événement et interface d'interaction entre l'apprenant et le simulateur. Le cadre de droite présente les variables respiratoires du patient virtuel.

Conclusion

Si les modalités de simulation basées sur une approche à base de cas permettent la simulation d'un patient virtuel, les différentes caractéristiques d'un scénario d'apprentissage : sa double contextualisation médicale et pédagogique, la nécessité de contrôle, et la contrainte de crédibilité du scénario, font que ces modalités de simulation ne sont pas adaptées à la simulation de scénarios d'apprentissage.

La représentation synthétique d'un ensemble de profils anesthésiques confère une structure unique et cohérente aux données anesthésiques. Elle permet le fonctionnement de notre approche de simulation d'un patient virtuel par raisonnement à base de cas tout en rendant possible son altération pour des raisons pédagogiques.

Cette simulation est de qualité et de performances légèrement supérieure aux autres modalités. Ainsi, une représentation synthétique est à la fois un support pour la simulation du patient virtuel et un support pour le scénario d'apprentissage.

La représentation synthétique contient tous les éléments nécessaires à la simulation par le simulateur. Notre simulateur permet l'interaction avec l'apprenant et la simulation du patient virtuel en s'appuyant sur les éléments de la représentation synthétique. Il ne contient pas d'élément spécifique à l'anesthésie.

Son utilisation n'est ainsi pas limitée à l'anesthésie, mais peut s'adapter à toutes les pratiques de simulation impliquant de simuler les séries temporelles d'un système, qui est dans notre cas un patient virtuel.

L'utilisation d'une représentation synthétique comme support d'un scénario d'apprentissage permet, au contraire d'autres modalités de simulation, de simuler un patient virtuel.

CONCLUSION

Le développement continu de l'anesthésie en fait aujourd'hui une pratique courante, légalement et techniquement encadrée et ainsi de plus en plus sûre. La surveillance médico-légale des anesthésies génère des données qui couvrent toute la procédure d'une anesthésie et qui sont numérisées depuis 2004 au CHU de Nantes.

L'évolution des pratiques des centres de formation en santé, ainsi que le développement professionnel continu des soignants, ont permis le développement en France de la pratique de la simulation avec l'encouragement de la Haute Autorité de Santé. Les centres de simulation se sont ainsi développés de manière importante sur la dernière décennie.

La science des données peut apporter des solutions pour répondre aux besoins de diversité des scénarios et de pilotage automatique des séances de simulation. La pratique de la simulation peut alors s'enrichir des données de santé accumulées dans les centres de soins, sous la forme de profils anesthésiques, pour produire des scénarios d'apprentissage réactifs basés sur les données.

L'ensemble de ces travaux de thèse apporte des solutions, théoriques et techniques, à la création automatisée de scénarios d'apprentissage réactifs afin de consolider l'expertise en anesthésie grâce à un simulateur, et cela, en s'appuyant sur les données réelles.

Éléments clefs

Les contraintes légales et techniques imposées par l'exploitation de données réelles contenues dans les profils anesthésiques, nous ont conduits, dans le cadre de ces travaux, à la création d'un générateur de données réalistes. Le réalisme des données générées a été évalué avec succès grâce à une analyse de similarité et une analyse statistique menée sur des données réelles et simulées.

Ce générateur de données réalistes a été construit sur la base des connaissances expertes des professionnels de l'anesthésie. Pour cela, nous avons proposé une formalisation informatique d'une anesthésie, ce qui n'a jamais été réalisé jusqu'à présent. Sa construction a nécessité de réaliser des modélisations des effets des événements anesthésiques sur le patient à partir de l'expertise des soignants.

Le souhait d'utiliser les données réelles comme base de la simulation a conduit à la conception et au développement d'une nouvelle méthode de simulation, inspirée du raisonnement à base de cas, mais prenant en compte, pour la première fois dans la littérature, des données temporelles multivariées, de divers types, et interdépendantes.

La méthode de simulation que nous proposons s'appuie sur une mesure de dissimilarité entre profils anesthésiques, construite spécifiquement pour notre approche. Cette mesure utilise une mesure de dissimilarité entre séries temporelles univariées. Nous avons montré la pertinence du choix de cette mesure au cours d'expérimentations intensives dédiées à la comparaison de neuf mesures de dissimilarité candidates.

Le réalisme de la simulation obtenu par cette méthode a été évalué grâce à la comparaison des séries temporelles de patients simulés et réels soumis aux mêmes séquences d'actions médicales.

L'utilisation de données réelles en grande dimension pour la simulation a aussi posé des enjeux de structure de données et de contrôle de ces données utilisées : nous avons proposé la construction de représentations synthétiques, représentant les profils anesthésiques d'une cohorte de patients, pour permettre le bon fonctionnement des méthodes de simulation envisagées.

Pour la construction de cette représentation synthétique, nous avons proposé, et expérimenté, une approche de partitionnement de séquences d'événements combinant un partitionnement par cartes auto-organisées et un partitionnement ascendant hiérarchique.

Nous avons aussi déterminé, par expérimentation, des critères estimatifs de la complexité d'une représentation synthétique à partir de l'ensemble des séquences d'événements à son origine. Ces éléments permettent la production d'une représentation dont la *complexité* est contrôlée.

Ces différents éléments nous ont permis de construire des scénarios d'apprentissage médicalement et pédagogiquement contextualisés à partir d'un ensemble de données anesthésiques. Nous avons construit un simulateur réactif prototype pour pouvoir jouer ces scénarios et permettre l'interaction avec l'apprenant. Pour cela, le prototype de simulateur développé utilise diverses fonctionnalités de contrôle qui sont offertes aux formateurs pour adapter précisément le contexte médical et la finalité pédagogique des scénarios.

Perspectives

Bien que notre contribution soit axée sur un objectif de formation en anesthésie par la simulation assistée par le numérique, son intérêt dépasse le cadre de la simulation en milieu médical et ouvre des possibilités en médecine personnalisée, dans le développement de jumeaux numériques pour la prédiction au bloc opératoire ou en anonymisation de données.

Médecine personnalisée

La médecine personnalisée cherche à améliorer la prise en charge des patients en adaptant le parcours de soins et la spécificité des actes médicaux aux patients. Elle ambitionne d'être une médecine prédictive, préventive, personnalisée et participative.

Notre contribution, en proposant des approches de simulation d'un patient virtuel, nous a conduits à réaliser des prédictions de l'avenir de ce patient virtuel.

En choisissant les caractéristiques du patient virtuel pour qu'elles soient identiques à un patient réel, alors notre patient virtuel est personnalisé et devient un jumeau numérique de ce patient réel.

La simulation d'un jumeau numérique, en amont d'une intervention peut nous permettre d'identifier toutes les potentielles difficultés qui seront rencontrées pendant l'intervention. On s'appuie pour cela sur les données anesthésiques des interventions similaires connues. Une simulation permettrait ainsi une action préventive avant une intervention.

Ces données réelles sont issues de milliers de patients ce qui donne à nos approches un caractère participatif. L'intervention d'un patient permet, avec celle de nombreux autres, de contribuer à la formation des futurs soignants.

Nos travaux contribuent ainsi à la médecine personnalisée et permettraient d'améliorer la prise en charge de patients en utilisant notre simulateur comme un outil pour le soin.

Outil de prédiction au bloc opératoire

Nos travaux nous ont conduits à proposer des méthodes pour simuler un patient virtuel dans le cadre de scénario d'apprentissage. Ces méthodes ne sont pas restreintes à leur contexte pédagogique et peuvent permettre de simuler un patient virtuel dans de nombreux autres cas.

Notre simulateur est réactif aux actions d'un utilisateur qui, dans un contexte pédagogique est un soignant en formation. Il simule à la volée le patient virtuel en prédisant son avenir immédiat en fonction des choix de l'utilisateur. On peut sortir de notre cadre pédagogique, et se placer dans le cadre de la pratique de soin, en remplaçant l'apprenant par un soignant. Cela n'aurait pas d'impact sur le fonctionnement du simulateur.

En utilisant un patient virtuel comme le jumeau numérique d'un patient qui fait l'objet d'une intervention, et un soignant comme utilisateur, alors notre simulateur prédirait, à chaque action du soignant, l'avenir immédiat du patient *via* son jumeau numérique.

En s'interfaçant directement aux dispositifs d'enregistrement médico-légal d'un bloc opératoire, notre simulateur pourrait être utilisé comme un outil de prédiction au bloc opératoire. Il permettrait alors de faire bénéficier les soignants de l'expérience de toutes les interventions similaires dont on dispose des données.

Données de santé anonymisées

Nos travaux s'appuient sur des données de santé réelles. Ces données sont soumises à un cadre légal contraignant qui ne permet pas une exploitation large de ces données. L'anonymisation est alors une solution pour faciliter l'exploitation de ces données.

Nous avons proposé une représentation pour un ensemble de patients. Cette représentation produit une synthèse des patients les plus similaires. Les méthodes proposées pour la construction de cette représentation synthétique pourraient permettre, avec quelques altérations, la construction d'une représentation synthétique anonymisée. Une telle représentation permettrait de faciliter l'exploitation des données de santé réelles.

De plus, nous avons construit un générateur de profils anesthésiques à partir de connaissances expert, et montré le caractère réaliste des profils anesthésiques qu'il produit. Ces profils anesthésiques réalistes ne sont pas liés à un patient réel, et ne sont donc pas soumis au cadre légal des données de santé.

Notre générateur donne ainsi la possibilité d'utiliser des jeux de données réalistes en remplacement des jeux de données réels.

BIBLIOGRAPHIE

1. GRANRY, J. & MOLL, M., Rapport de mission : État de l'art (national et international) en matière de pratiques de simulation dans le domaine de la santé, *Haute Autorité en Santé* (2012).
2. DADURE, C., MARIE, A., SEGURET, F. & CAPDEVILA, X., One year of anaesthesia in France : A comprehensive survey based on the national medical information (PMSI) database. Part 1 : In-hospital patients, *Anaesthesia Critical Care & Pain Medicine* **34**, 191-197 (2015).
3. BOISAUBERT, H., *Repères statistiques de la base EXAN, archives médico-légales des interventions anesthésiques au CHU de Nantes de 2000 à 2019* 2022.
4. CLERGUE, F. *et al.*, L'anesthésie en France en 1996. Résultats d'une enquête de la Société française d'anesthésie et de réanimation (Sfar), *Chirurgie* **124**, 115-121 (1999).
5. LIENHART, A. *et al.*, Premiers résultats de l'enquête SFAR-INSERM sur la mortalité imputable à l'anesthésie en France : réduction par 10 du taux de ces décès en 20 ans, *Bulletin de l'Académie Nationale de Médecine* **188**, 1429-1441, ISSN : 0001-4079, <https://www.sciencedirect.com/science/article/pii/S0001407919336660> (2004).
6. SFAR, *Information médicale sur l'anesthésie - Qu'est ce que l'anesthésie ?* <https://sfar.org/pour-le-grand-public/information-medicales-sur-lanesthesie/>.
7. *Décret n°94-1050 du 5 décembre 1994 relatif aux conditions techniques du fonctionnement des établissements de santé en ce qui concerne la pratique de l'anesthésie et modifiant le code de la santé publique* Journal Officiel de la République Française, 1994.
8. MALLAMPATI, S. R. *et al.*, A clinical sign to predict difficult tracheal intubation ; a prospective study, *Canadian Anaesthetists' Society Journal* **32**, 429-434 (1985).
9. SAMSOON, G. & YOUNG, J., Difficult tracheal intubation : a retrospective study, *Anaesthesia* **42**, 487-490 (1987).

-
10. SFAR, Prévention du risque allergique peranesthésique, *Ann Fr Anesth Reanim* **20**, f56-69 (2001).
 11. ASO, ASA physical status classification system, *ASA House of Delegates* (2014).
 12. BAILLARD, C., BOURGAIN, J., BOUROCHE, G. *et al.*, *Muscle relaxants and reversal in anaesthesia. Guidelines from the French society of anaesthesia & intensive Care medicine* 2018.
 13. DÉPARTEMENT D'ANESTHÉSIE RÉANIMATION DE BICÈTRE, *Protocole d'anesthésie réanimation* 15^e éd. (MAPAR, 78, rue du général leclerc 94275 Le kremelin bicêtre, 2019).
 14. LE SIMU, *Laboratoire Expérimental de Simulation de Médecine Intensive de Nantes Université* <https://lesimudenantes.univ-nantes.fr/>.
 15. CHINIARA, G., *Simulation médicale pour acquisition des compétences en anesthésie. Congrès national d'anesthésie et de réanimation. Conférences d'actualisation* 2007.
 16. LANE, J. L., SLAVIN, S. & ZIV, A., Simulation in medical education : A review, *Simulation & Gaming* **32**, 297-314 (2001).
 17. *Décret n°2001-486 du 6 juin 2001 portant publication de la Convention européenne sur la protection des animaux vertébrés utilisés à des fins expérimentales ou à d'autres fins scientifiques, adoptée à Strasbourg le 18 mars 1986 et signée par la France le 2 septembre 1987* Journal officiel, 2001.
 18. *Décret n°2001-464 du 29 mai 2001 modifiant le décret n° 87-848 du 19 octobre 1987 pris pour l'application de l'article 454 du Code pénal et du troisième alinéa de l'article 276 du Code rural et relatif aux expériences pratiquées sur les animaux.* Journal officiel, 2001.
 19. GELIS, A. *et al.*, Peer role-play for training communication skills in medical students : a systematic review, *Simulation in Healthcare* **15**, 106-111 (2020).
 20. HUWENDIEK, S. *et al.*, Towards a typology of virtual patients, *Medical teacher* **31**, 743-748 (2009).
 21. TAG, B., Utilisation de cadavres et de parties de cadavres dans la recherche médicale et la formation prégraduée, postgraduée et continue : recommandations de l'Académie suisse des sciences médicales (ASSM), *BULLETIN DES MEDECINS SUISSES* **90**, 102-107 (2009).

-
22. MELLER, G., A typology of simulators for medical education, *Journal of digital imaging* **10**, 194-196 (1997).
 23. BOET, S., GRANRY, J.-C. & SAVOLDELLI, G., *La simulation en santé : De la théorie à la pratique* (Springer, 2013).
 24. GELBART, N. R., in *The King's Midwife* (University of California Press, 1998).
 25. BARROWS, H. S. *et al.*, An overview of the uses of standardized patients for teaching and evaluating clinical skills, *Academic Medicine-Philadelphia*- **68**, 443-443 (1993).
 26. HOFFMAN, K. I. & ABRAHAMSON, S., The 'cost-effectiveness' of Sim One, *Academic Medicine* **50**, 1127-1128 (1975).
 27. PHILIP, J., *A computer program to teach theoretical and practical aspects of inhalational anesthetic pharmacokinetics using pictorial and graphical displays and continuous user interaction* in *Proc of Microcomputer in Anesthesia IV* (Ashville NC, 1983).
 28. ROSEN, K. R., The history of medical simulation, *Journal of critical care* **23**, 157-166 (2008).
 29. KOHN, L. T., CORRIGAN, J. M., DONALDSON, M. S. *et al.*, *To err is human : Building a safer health system. Institute of Medicine, Committee on Quality of Health Care in America* 2000.
 30. HAUTE AUTORITÉ DE SANTÉ, Guide de bonnes pratiques en matière de simulation en santé, *Haute Autorité en Santé* **97** (2012).
 31. COMITÉ PÉDAGOGIQUE DE LA SOFRASIMS, *État des lieux de l'emploi des jeux sérieux pour la formation des professionnels de santé en France en 2019* 2021.
 32. HAS - SERVICE ÉVALUATION ET OUTILS POUR LA QUALITÉ ET LA SÉCURITÉ DES SOINS, *Guide pour l'évaluation des infrastructures de simulation en santé* 2015.
 33. REHMANN, A. J., MITMAN, R. D. & REYNOLDS, M. C., *A Handbook of Flight Simulation Fidelity Requirements for Human Factors Research*. rapp. tech. (Crew System Ergonomics Information Analysis Center Wright-Patterson AFB OH, 1995).
 34. NORMAN, G., DORE, K. & GRIERSON, L., The minimal relationship between simulation fidelity and transfer of learning, *Medical education* **46**, 636-647 (2012).

-
35. TETERIS, E., FRASER, K., WRIGHT, B. & MCCLAUGHLIN, K., Does training learners on simulators benefit real patients ?, *Advances in health sciences education* **17**, 137-144 (2012).
 36. PATEL, V. L. & KAUFMAN, D. R., in *Biomedical informatics* 109-148 (Springer, 2014).
 37. ALESSI, S. M., Fidelity in the design of instructional simulations. *Journal of computer-based instruction* (1988).
 38. MARAN, N. J. & GLAVIN, R. J., Low-to high-fidelity simulation—a continuum of medical education ?, *Medical education* **37**, 22-28 (2003).
 39. HENNEMAN, E. A., LEE, J. L. & COHEN, J. I., Collaboration : a concept analysis, *Journal of advanced Nursing* **21**, 103-109 (1995).
 40. COOK, D. A. *et al.*, Technology-enhanced simulation for health professions education : a systematic review and meta-analysis, *Jama* **306**, 978-988 (2011).
 41. ABRAHAMSON, S., DENSON, J. S. & WOLF, R. M., Effectiveness of a simulator in training anesthesiology residents, *Academic Medicine* **44**, 515-9 (1969).
 42. SIMON, R., RAEMER, D. & RUDOLPH, J., Debriefing assessment for simulation in healthcare : rater version, *Cambridge, MA : Center for Medical Simulation* (2009).
 43. BRUPPACHER, H. R., CHEN, R. P. & LACHAPELLE, K., First, do no harm : using simulated patient death to enhance learning ?, *Medical education* **45**, 317-318 (2011).
 44. ANDRÉ, B. & DEKONINCK, J., *Motiver pour enseigner-Analyse transactionnelle et pédagogie : Analyse transactionnelle et pédagogie* (Hachette éducation, 2014).
 45. BARRY ISSENBERG, S., MCGAGHIE, W. C., PETRUSA, E. R., LEE GORDON, D. & SCALESE, R. J., Features and uses of high-fidelity medical simulations that lead to effective learning : a BEME systematic review, *Medical teacher* **27**, 10-28 (2005).
 46. MCGAGHIE, W. C., ISSENBERG, S. B., PETRUSA, E. R. & SCALESE, R. J., A critical review of simulation-based medical education research : 2003–2009, *Medical education* **44**, 50-63 (2010).
 47. BOET, S. *et al.*, *Within-team debriefing versus instructor-led debriefing for simulation-based education : a randomized controlled trial* 2013.
 48. BOET, S. *et al.*, Looking in the mirror : self-debriefing versus instructor debriefing for simulated crises, *Critical Care Medicine* **39**, 1377-1381 (2011).

-
49. GABA, D. M., The future vision of simulation in health care, *BMJ Quality & Safety* **13**, i2-i10 (2004).
 50. SPRAWLS, P., Evolving models for medical physics education and training : a global perspective, *Biomedical imaging and intervention journal* **4** (2008).
 51. OFFICIEL, J., *loi relative à l'informatique, aux fichiers et aux libertés du 6 janvier 1978* 1978.
 52. PARLEMENT EUROPÉEN, *Directive 95/46/CE sur la protection des données personnelles* 1995.
 53. PARLEMENT EUROPÉEN, *Règlement du Parlement européen et du Conseil relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données* 2016.
 54. CNIL, *Méthodologie de référence MR-003 - Recherches dans le domaine de la santé sans recueil du consentement - Délibération n° 2018-154 du 3 mai 2018* 2018.
 55. SACKETT, D. L., ROSENBERG, W. M., GRAY, J. M., HAYNES, R. B. & RICHARDSON, W. S., *Evidence based medicine : what it is and what it isn't* 1996.
 56. CNEAR, *Référentiel d'anesthésie-réanimation et de médecine péri-opératoire* (CNEAR, 2020).
 57. MILLER, R. D. *et al.*, *Miller. Anesthesia. Chapter 1 to 55* (Elsevier, 2016).
 58. ROBINSON, J. D., LUPKIEWICZ, S. M., PALENIK, L., LOPEZ, L. M. & ARIET, M., Determination of ideal body weight for drug dosage calculations. *American journal of hospital pharmacy* **40**, 1016-1019 (1983).
 59. KAMMERER, M. R., PORTER, M. M., BEEKLEY, A. C. & TICHANSKY, D. S., Ideal body weight calculation in the bariatric surgical population, *Journal of Gastrointestinal Surgery* **19**, 1758-1762 (2015).
 60. MCCARRON, M. M. & DEVINE, B. J., Clinical pharmacy : case studies : case number 25 gentamicin therapy, *Drug Intelligence & Clinical Pharmacy* **8**, 650-655 (1974).
 61. SMITH, T. F., WATERMAN, M. S. *et al.*, Identification of common molecular subsequences, *Journal of molecular biology* **147**, 195-197 (1981).

-
62. WANG, X. *et al.*, Experimental comparison of representation methods and distance measures for time series data, *Data Mining and Knowledge Discovery* **26**, 275-309 (2013).
 63. GIUSTI, R. & BATISTA, G. E., *An empirical comparison of dissimilarity measures for time series classification* in *2013 Brazilian Conference on Intelligent Systems* (2013), 82-88.
 64. LHERMITTE, S., VERBESSELT, J., VERSTRAETEN, W. W. & COPPIN, P., A comparison of time series similarity measures for classification and change detection of ecosystem dynamics, *Remote sensing of environment* **115**, 3129-3152 (2011).
 65. KIANIMAJD, A. *et al.*, Comparison of different methods of measuring similarity in physiologic time series, *IFAC-PapersOnLine* **50**, 11005-11010 (2017).
 66. DING, H., TRAJCEVSKI, G., SCHEUERMANN, P., WANG, X. & KEOGH, E., Querying and mining of time series data : experimental comparison of representations and distance measures, *Proceedings of the VLDB Endowment* **1**, 1542-1552 (2008).
 67. CASSISI, C., MONTALTO, P., ALIOTTA, M., CANNATA, A. & PULVIRENTI, A., Similarity measures and dimensionality reduction techniques for time series data mining, *Advances in data mining knowledge discovery and applications'(InTech, Rijeka, Croatia, 2012*, 71-96 (2012).
 68. FINK, E. & PRATT, K. B., in *Data mining in time series databases* 43-65 (World Scientific, 2004).
 69. KEOGH, E. & RATANAMAHATANA, C. A., Exact indexing of dynamic time warping, *Knowledge and information systems* **7**, 358-386 (2005).
 70. VINTSYUK, T. K., Speech discrimination by dynamic programming, *Cybernetics* **4**, 52-57 (1968).
 71. CORMEN, T. H., LEISERSON, C. E., RIVEST, R. L. & STEIN, C., Introduction to algorithms mit press, *Cambridge MA* (1990).
 72. SAKOE, H. & CHIBA, S., Dynamic programming algorithm optimization for spoken word recognition, *IEEE transactions on acoustics, speech, and signal processing* **26**, 43-49 (1978).
 73. ITAKURA, F., Minimum prediction residual principle applied to speech recognition, *IEEE Transactions on acoustics, speech, and signal processing* **23**, 67-72 (1975).

-
74. ŁUCZAK, M., Hierarchical clustering of time series data with parametric derivative dynamic time warping, *Expert Systems with Applications* **62**, 116-130 (2016).
 75. VLACHOS, M., KOLLIOS, G. & GUNOPULOS, D., *Discovering similar multidimensional trajectories* in *Proceedings 18th international conference on data engineering* (2002), 673-684.
 76. AGRAWAL, R., FALOUTSOS, C. & SWAMI, A., *Efficient similarity search in sequence databases* in *International conference on foundations of data organization and algorithms* (1993), 69-84.
 77. CAI, Y. & NG, R., *Indexing spatio-temporal trajectories with chebyshev polynomials* in *Proceedings of the 2004 ACM SIGMOD international conference on Management of data* (2004), 599-610.
 78. RALANAMAHATANA, C. A. *et al.*, in *Data mining and knowledge discovery handbook* 1069-1103 (Springer, 2005).
 79. CHAN, K.-P. & FU, A. W.-C., *Efficient time series matching by wavelets* in *Proceedings 15th International Conference on Data Engineering (Cat. No. 99CB36337)* (1999), 126-133.
 80. KEOGH, E., CHAKRABARTI, K., PAZZANI, M. & MEHROTRA, S., Dimensionality reduction for fast similarity search in large time series databases, *Knowledge and information Systems* **3**, 263-286 (2001).
 81. LIN, J., KEOGH, E., WEI, L. & LONARDI, S., Experiencing SAX : a novel symbolic representation of time series, *Data Mining and knowledge discovery* **15**, 107-144 (2007).
 82. ABANDA, A., MORI, U. & LOZANO, J. A., A review on distance based time series classification, *Data Mining and Knowledge Discovery* **33**, 378-412 (2019).
 83. SALVADOR, S. & CHAN, P., Fastdtw : Toward accurate dynamic time, *Warping in Linear Time and Space* (2007).
 84. MÜLLER, M., MATTES, H. & KURTH, F., *An efficient multiscale approach to audio synchronization.* in *ISMIR* **546** (2006), 192-197.
 85. GHARGHABI, S., IMANI, S., BAGNALL, A., DARVISHZADEH, A. & KEOGH, E., *Matrix profile xii : Mpdist : a novel time series distance measure to allow data mining in more challenging scenarios* in *2018 IEEE International Conference on Data Mining (ICDM)* (2018), 965-970.

-
86. CHRIST, M., BRAUN, N., NEUFFER, J. & KEMPA-LIEHR, A. W., Time series feature extraction on basis of scalable hypothesis tests (tsfresh—a python package), *Neurocomputing* **307**, 72-77 (2018).
 87. DAU, H. A. *et al.*, The UCR time series archive, *IEEE/CAA Journal of Automatica Sinica* **6**, 1293-1305 (2019).
 88. DAMA, F. & SINOQUET, C., Partially Hidden Markov Chain Linear Autoregressive model : inference and forecasting, *arXiv preprint arXiv :2102.12584* (2021).
 89. KLEVEN, N. F. *et al.*, Virtual operating room for collaborative training of surgical nurses in *CYTED-RITOS International Workshop on Groupware* (2014), 223-238.
 90. ISAZA-RESTREPO, A., GÓMEZ, M. T., CIFUENTES, G. & ARGÜELLO, A., The virtual patient as a learning tool : a mixed quantitative qualitative study, *BMC Medical Education* **18**, 1-10 (2018).
 91. HESTER, R. L. *et al.*, HumMod : a modeling environment for the simulation of integrative human physiology, *Frontiers in physiology* **2**, 12 (2011).
 92. HESTER, R., PRUETT, W., CLEMMER, J. & RUCKDESCHEL, A., Simulation of integrative physiology for medical education, *Morphologie* **103**, 187-193 (2019).
 93. MEISSNER, C., MEIXENSBERGER, J., PRETSCHNER, A. & NEUMUTH, T., Sensor-based surgical activity recognition in unconstrained environments, *Minimally Invasive Therapy & Allied Technologies* **23**, 198-205 (2014).
 94. STAUDER, R. *et al.*, Random forests for phase detection in surgical workflow analysis in *International Conference on Information Processing in Computer-Assisted Interventions* (2014), 148-157.
 95. KHALID, S., GOLDENBERG, M., GRANTCHAROV, T., TAATI, B. & RUDZICZ, F., Evaluation of deep learning models for identifying surgical actions and measuring performance, *JAMA network open* **3**, e201664-e201664 (2020).
 96. FORESTIER, G., PETITJEAN, F., RIFFAUD, L. & JANNIN, P., Automatic matching of surgeries to predict surgeons' next actions, *Artificial intelligence in medicine* **81**, 3-11 (2017).
 97. FRANKE, S., MEIXENSBERGER, J. & NEUMUTH, T., Intervention time prediction from surgical low-level tasks, *Journal of biomedical informatics* **46**, 152-159 (2013).

-
98. GURGEN ERDOGAN, T. & TARHAN, A., A goal-driven evaluation method based on process mining for healthcare processes, *Applied Sciences* **8**, 894 (2018).
 99. CHOUDHURY, N. & BEGUM, S. A., A survey on case-based reasoning in medicine, *International Journal of Advanced Computer Science and Applications* **7** (2016).
 100. MÜLÂYIM, M. O. & ARCOS, J. L., Fast anytime retrieval with confidence in large-scale temporal case bases, *Knowledge-Based Systems* **206**, 106374 (2020).
 101. NG, K., SUN, J., HU, J. & WANG, F., Personalized predictive modeling and risk factor identification using patient similarity, *AMIA Summits on Translational Science Proceedings* **2015**, 132 (2015).
 102. BROWN, S.-A., Patient similarity : emerging concepts in systems and precision medicine, *Frontiers in physiology* **7**, 561 (2016).
 103. PARIMBELLI, E., MARINI, S., SACCHI, L. & BELLAZZI, R., Patient similarity for precision medicine : A systematic review, *Journal of biomedical informatics* **83**, 87-96 (2018).
 104. WANG, N. *et al.*, Measurement and application of patient similarity in personalized predictive modeling based on electronic medical records, *Biomedical engineering online* **18**, 1-15 (2019).
 105. FANG, H. S. A. *et al.*, Patient similarity analytics for explainable clinical risk prediction, *BMC medical informatics and decision making* **21**, 1-12 (2021).
 106. KANE, A. & SHIRI, N., *Multivariate time series representation and similarity search using pca* in *Industrial Conference on Data Mining* (2017), 122-136.
 107. KALE, D. C. *et al.*, *An examination of multivariate time series hashing with applications to health care* in *2014 IEEE international conference on data mining* (2014), 260-269.
 108. RIBEIRO, M. X. & de ANDRADE, C. G., Similarity search in multidimensional time series using the coulomb's law, *Journal of Information and Data Management* **5**, 74-74 (2014).
 109. YE, L. & KEOGH, E., *Time series shapelets : a new primitive for data mining* in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining* (2009), 947-956.

-
110. YEH, C.-C. M., KAVANTZAS, N. & KEOGH, E., *Matrix Profile VI : Meaningful Multidimensional Motif Discovery* in *2017 IEEE International Conference on Data Mining (ICDM)* (2017), 565-574.
 111. LAFTCHIEV, E. & LIU, Y., *Finding multidimensional patterns in multidimensional time series* in *KDD workshop on MiLeTS '18, London* (2018).
 112. RAKTHANMANON, T. *et al.*, *Searching and mining trillions of time series subsequences under dynamic time warping* in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining* (2012), 262-270.
 113. WANG, J., ZHU, Y., LI, S., WAN, D. & ZHANG, P., Multivariate time series similarity searching, *The Scientific World Journal* **2014** (2014).
 114. SHIFAZ, A., PELLETIER, C., PETITJEAN, F. & WEBB, G. I., Elastic similarity measures for multivariate time series classification, *arXiv preprint arXiv :2102.10231* (2021).
 115. BAYDOGAN, M. G. & RUNGER, G., Time series representation and similarity based on local autopatterns, *Data Mining and Knowledge Discovery* **30**, 476-509 (2016).
 116. SHA, Y., VENUGOPALAN, J. & WANG, M. D., *A novel temporal similarity measure for patients based on irregularly measured data in electronic health records* in *Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics* (2016), 337-344.
 117. GANZINGER, M., SCHRODT, J. & KNAUP, P., *A Concept for Graph-Based Temporal Similarity of Patient Data.* in *MedInfo* (2019), 138-142.
 118. LIU, Z., ZHU, Z., GAO, J. & XU, C., Forecast methods for time series data : a survey, *IEEE Access* **9**, 91896-91912 (2021).
 119. SIMS, C. A., in *Time Series and Statistics* (éd. EATWELL, J., MILGATE, M. & NEWMAN, P.) 175-186 (Palgrave Macmillan UK, London, 1990), ISBN : 978-1-349-20865-4, https://doi.org/10.1007/978-1-349-20865-4_23.
 120. YU, H.-F., RAO, N. & DHILLON, I. S., Temporal regularized matrix factorization for high-dimensional time series prediction, *Advances in neural information processing systems* **29** (2016).
 121. RANGAPURAM, S. S. *et al.*, Deep state space models for time series forecasting, *Advances in neural information processing systems* **31** (2018).

-
122. RASUL, K., SHEIKH, A.-S., SCHUSTER, I., BERGMANN, U. & VOLLGRAF, R., Multi-variate probabilistic time series forecasting via conditioned normalizing flows, *arXiv preprint arXiv :2002.06103* (2020).
 123. SALINAS, D., FLUNKERT, V., GASTHAUS, J. & JANUSCHOWSKI, T., DeepAR : Probabilistic forecasting with autoregressive recurrent networks, *International Journal of Forecasting* **36**, 1181-1191 (2020).
 124. LIM, B. & ZOHREN, S., Time-series forecasting with deep learning : a survey, *Philosophical Transactions of the Royal Society A* **379**, 20200209 (2021).
 125. SEZER, O. B., GUDELEK, M. U. & OZBAYOGLU, A. M., Financial time series forecasting with deep learning : A systematic literature review : 2005–2019, *Applied soft computing* **90**, 106181 (2020).
 126. TEALAB, A., Time series forecasting using artificial neural networks methodologies : A systematic review, *Future Computing and Informatics Journal* **3**, 334-340 (2018).
 127. LAI, G., CHANG, W.-C., YANG, Y. & LIU, H., *Modeling long-and short-term temporal patterns with deep neural networks* in *The 41st international ACM SIGIR conference on research & development in information retrieval* (2018), 95-104.
 128. SEN, R., YU, H.-F. & DHILLON, I. S., Think globally, act locally : A deep neural network approach to high-dimensional time series forecasting, *Advances in neural information processing systems* **32** (2019).
 129. XIAO, S. *et al.*, Learning time series associated event sequences with recurrent point process networks, *IEEE transactions on neural networks and learning systems* **30**, 3124-3136 (2019).
 130. VAUGHAN, N. & GABRYS, B., Comparing and combining time series trajectories using dynamic time warping, *Procedia Computer Science* **96**, 465-474 (2016).
 131. KOHONEN, T. & HONKELA, T., Kohonen network, *Scholarpedia* **2**, 1568 (2007).
 132. SONG, M., YANG, H., SIADAT, S. H. & PECHENIZKIY, M., A comparative study of dimensionality reduction techniques to enhance trace clustering performances, *Expert Systems with Applications* **40**, 3722-3737 (2013).
 133. JIN, C. & MALTHOUSE, E. C., On the bias and inconsistency of K-means clustering (2016).

-
134. SONG, M., GÜNTHER, C. W. & VAN DER AALST, W. M., *Trace clustering in process mining* in *International conference on business process management* (2008), 109-120.
 135. KOHONEN, T., *Self-organizing maps* (Springer Science & Business Media, 2012).
 136. BERGLUND, E. & SITTE, J., The parameterless self-organizing map algorithm, *IEEE Transactions on neural networks* **17**, 305-316 (2006).
 137. BAUER, H.-U., HERRMANN, M. & VILLMANN, T., Neural maps and topographic vector quantization, *Neural networks* **12**, 659-676 (1999).
 138. KIVILUOTO, K., *Topology preservation in self-organizing maps* in *Proceedings of International Conference on Neural Networks (ICNN'96)* **1** (1996), 294-299.
 139. TIAN, J., AZARIAN, M. H. & PECHT, M., *Anomaly detection using self-organizing maps-based k-nearest neighbor algorithm* in *PHM Society European Conference* **2** (2014).
 140. DAVIES, D. L. & BOULDIN, D. W., A cluster separation measure, *IEEE transactions on pattern analysis and machine intelligence*, 224-227 (1979).
 141. JACCARD, P., Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines, *Bull Soc Vaudoise Sci Nat* **37**, 241-272 (1901).
 142. VESANTO, J. & ALHONIEMI, E., Clustering of the self-organizing map, *IEEE Transactions on neural networks* **11**, 586-600 (2000).
 143. G29, *Avis 05/2014 du groupe de travail « article 29 » du 10 avril 2014 sur les Techniques d'anonymisation* 2014.
 144. VERYKIOS, V. S. *et al.*, State-of-the-art in privacy preserving data mining, *ACM Sigmod Record* **33**, 50-57 (2004).
 145. LIEW, C. K., CHOI, U. J. & LIEW, C. J., A data distortion by probability distribution, *ACM Transactions on Database Systems (TODS)* **10**, 395-411 (1985).
 146. WARNER, S. L., Randomized response : A survey technique for eliminating evasive answer bias, *Journal of the American Statistical Association* **60**, 63-69 (1965).
 147. FIENBERG, S. E. & MCINTYRE, J., *Data swapping : Variations on a theme by dalenius and reiss* in *International Workshop on Privacy in Statistical Databases* (2004), 14-29.

-
148. DALENIUS, T. & REISS, S. P., Data-swapping : A technique for disclosure control, *Journal of statistical planning and inference* **6**, 73-85 (1982).
 149. SWEENEY, L., Weaving technology and policy together to maintain confidentiality, *The Journal of Law, Medicine & Ethics* **25**, 98-110 (1997).
 150. MACHANAVAJJHALA, A., KIFER, D., GEHRKE, J. & VENKITASUBRAMANIAM, M., l-diversity : Privacy beyond k-anonymity, *ACM Transactions on Knowledge Discovery from Data (TKDD)* **1**, 3-es (2007).
 151. LI, N., LI, T. & VENKATASUBRAMANIAN, S., *t-closeness : Privacy beyond k-anonymity and l-diversity* in *2007 IEEE 23rd international conference on data engineering* (2006), 106-115.
 152. KUSHIDA, C. A. *et al.*, Strategies for de-identification and anonymization of electronic health record data for use in multicenter research studies, *Medical care* **50**, S82 (2012).
 153. PURI, V., SACHDEVA, S. & KAUR, P., Privacy preserving publication of relational and transaction data : Survey on the anonymization of patient data, *Computer Science Review* **32**, 45-61 (2019).
 154. EL EMAM, K., RODGERS, S. & MALIN, B., Anonymising and sharing individual patient data, *bmj* **350** (2015).
 155. MURTHY, S., BAKAR, A. A., RAHIM, F. A. & RAMLI, R., *A comparative study of data anonymization techniques* in *2019 IEEE 5th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing,(HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS)* (2019), 306-309.
 156. NEEDLEMAN, S. B. & WUNSCH, C. D., A general method applicable to the search for similarities in the amino acid sequence of two proteins, *Journal of molecular biology* **48**, 443-453 (1970).
 157. HATTON, F., TIRET, L., MAUJOL, L., N'DOYE, P. & VOURC'H, G., *Enquête épidémiologique sur les anesthésies : premiers résultats* in *Annales françaises d'anesthésie et de réanimation* **2** (1983), 331-385.
 158. SFAR, *Recommandations Formalisées d'Experts - Prise en charge anesthésique des patients en hospitalisation ambulatoire* in *Annales Françaises d'Anesthésie et de Réanimation* **29** (2010), 67-72.

-
159. VAN DER AALST, W., *Process mining : discovery, conformance and enhancement of business processes* (Springer, 2011).
 160. VAN DER AALST, W. M. *et al.*, Business process mining : An industrial application, *Information Systems* **32**, 713-732 (2007).
 161. VAN DER AALST, W. *et al.*, *Process mining manifesto* in *International Conference on Business Process Management* (2011), 169-194.
 162. VAN DER AALST, W. M. & WEIJTERS, A. J., *Process mining : a research agenda* 2004.
 163. SRIKANT, R. & AGRAWAL, R., *Mining sequential patterns : Generalizations and performance improvements* in *International Conference on Extending Database Technology* (1996), 1-17.
 164. AGRAWAL, R. & SRIKANT, R., *Mining sequential patterns* in *Proceedings of the eleventh international conference on data engineering* (1995), 3-14.
 165. GRECO, G., GUZZO, A., PONTIERI, L. & SACCA, D., Discovering expressive process models by clustering log traces, *IEEE Transactions on Knowledge and Data Engineering* **18**, 1010-1027 (2006).
 166. DE MEDEIROS, A. K. A. *et al.*, *Process mining based on clustering : A quest for precision* in *International Conference on Business Process Management* (2007), 17-29.
 167. RISTAD, E. S. & YIANILOS, P. N., Learning string-edit distance, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20**, 522-532 (1998).
 168. BOSE, R. J. C. & VAN DER AALST, W. M., *Context aware trace clustering : Towards improving process mining results* in *Proceedings of the 2009 SIAM International Conference on Data Mining* (2009), 401-412.
 169. DE WEERDT, J., VANDEN BROUCKE, S., VANTHIENEN, J. & BAESENS, B., Active trace clustering for improved process discovery, *IEEE Transactions on Knowledge and Data Engineering* **25**, 2708-2720 (2013).
 170. BOSE, R. J. C. & van der AALST, W. M., *Trace clustering based on conserved patterns : Towards achieving better process models* in *International Conference on Business Process Management* (2009), 170-181.

-
171. DELIAS, P., DOUMPOS, M., GRIGOROUDIS, E. & MATSATSINIS, N., A non-compensatory approach for trace clustering, *International Transactions in Operational Research* **26**, 1828-1846 (2019).
 172. LEOPOLD, L. B., WOLMAN, M. G., MILLER, J. P. & WOHL, E., *Fluvial processes in geomorphology* (Courier Dover Publications, 2020).
 173. ROY, B., *Paradigms and challenges*. J. Figueira, S. Greco, M. Enrgott, eds., *Multiple Criteria Decision Analysis : State of the Art Surveys* 2005.
 174. LI, H., Multivariate time series clustering based on common principal component analysis, *Neurocomputing* **349**, 239-247 (2019).
 175. VAN DER AALST, W., in *Process mining* 3-23 (Springer, 2016).
 176. VAN DER AALST, W. M., *Process mining : discovering and improving Spaghetti and Lasagna processes* in *2011 IEEE symposium on computational intelligence and data mining (CIDM)* (2011), 1-7.
 177. VAN DER AALST, W., WEIJTERS, T. & MARUSTER, L., Workflow mining : Discovering process models from event logs, *IEEE transactions on knowledge and data engineering* **16**, 1128-1142 (2004).
 178. LEEMANS, S. J., FAHLAND, D. & VAN DER AALST, W. M., *Discovering block-structured process models from event logs-a constructive approach* in *International conference on applications and theory of Petri nets and concurrency* (2013), 311-329.
 179. WEIJTERS, A. & RIBEIRO, J. T. S., *Flexible heuristics miner (FHM)* in *2011 IEEE symposium on computational intelligence and data mining (CIDM)* (2011), 310-317.
 180. CHATAIN, T., CARMONA, J. & DONGEN, B. v., *Alignment-based trace clustering* in *International Conference on Conceptual Modeling* (2017), 295-308.
 181. BOLTENHAGEN, M., CHATAIN, T. & CARMONA, J., *Generalized alignment-based trace clustering of process behavior* in *International Conference on Applications and Theory of Petri Nets and Concurrency* (2019), 237-257.
 182. RENDÓN, E., ABUNDEZ, I., ARIZMENDI, A. & QUIROZ, E. M., Internal versus external cluster validation indexes, *International Journal of computers and communications* **5**, 27-34 (2011).

-
183. VON STORCH, H. & ZWIERS, F. W., *Statistical analysis in climate research* (Cambridge university press, 2001).
 184. DUDA, R., Hart. P., E., and Stork, DG,” Pattern Classification”, *Nueva York : John Wiley and Sons* (2001).
 185. PERNG, C.-S., WANG, H., ZHANG, S. R. & PARKER, D. S., *Landmarks : a new model for similarity-based pattern querying in time series databases* in *Proceedings of 16th International Conference on Data Engineering (Cat. No. 00CB37073)* (2000), 33-42.
 186. CHRIST, M., KEMPA-LIEHR, A. W. & FEINDT, M., Distributed and parallel time series feature extraction for industrial big data applications, *arXiv preprint arXiv :1610.07717* (2016).
 187. BAHRAMI, M., HADADGAR, A. & FULADVANDI, M., Designing virtual patients for education of nursing students in cancer course, *Iranian journal of nursing and midwifery research* **26**, 133 (2021).
 188. KOTSIFAKOS, A. & PAPAPETROU, P., *Model-based time series classification* in *International Symposium on Intelligent Data Analysis* (2014), 179-191.
 189. ZHANG, D., ZUO, W., ZHANG, D. & ZHANG, H., *Time series classification using support vector machine with Gaussian elastic metric kernel* in *2010 20th International Conference on Pattern Recognition* (2010), 29-32.
 190. XI, X., KEOGH, E., SHELTON, C., WEI, L. & RATANAMAHATANA, C. A., *Fast time series classification using numerosity reduction* in *Proceedings of the 23rd international conference on Machine learning* (2006), 1033-1040.

Annexes

PROFILS ANESTHÉSQUES

A.1 Exemple de dossier médical

Le dossier médical présenté ci-dessous correspond à un patient de type «normal» sans caractéristiques ou pathologies pouvant impacter défavorablement le déroulement d’une intervention. Les éléments d’un dossier médical sont des données personnelles de santé. En dehors des soignants tenus au secret médical, ces données ne sont consultables que de manière pseudonymisée, sans variables directement identifiantes. Ainsi, le dossier médical présenté ci-dessous est présenté lui aussi de manière pseudonymisée.

Le composant dossier médical d’un profil anesthésique comprend plusieurs éléments :

- les informations d’état civil du patient, présentées en table A.1,
- les antécédents médicaux du patient, présentés en table A.2,
- les détails de la consultation et de la visite pré-anesthésique, présentés en table A.3, les variables de l’examen clinique présentés en A.4 et les facteurs de risques du patient présentés en A.5. On y apprend notamment que le patient a été précédemment opéré d’une sténose du pylore¹ à l’âge d’un mois.

Variable	Exemple	Description
SEXE	masculin	Sexe biologique
MOIS_NAISSANCE	01	Mois de naissance
ANNEE_NAISSANCE	1992	Année de naissance
MOIS_DECES	-	Mois de décès, vide si vivant
ANNEE_DECES	-	Année de décès, vide si vivant
CP	44000	Code postal

Table A.1 – Variables d’état civil présentes dans le dossier médical d’un patient

1. La sténose du pylore est une pathologie infantile consistant en une hypertrophie (épaississement) du muscle pylorique, le muscle qui se trouve entre l’estomac et l’intestin. Cet épaississement empêche le passage des aliments dans l’intestin et entraîne des vomissements ; par conséquent l’enfant ne peut pas s’alimenter.

Variable	Exemple	Description
HISTOIRE_MALADIE	Hernie suite à marathon	Information sur l'origine et l'évolution de la pathologie
TRANSFUSE_JAMAIS	Vrai	Vrai si le patient n'a jamais été transfusé.
TRANSFUSE_PLUS_100J	Faux	Vrai si le patient a été transfusé il y a plus de 100 jours.
TRANSFUSE_MOINS_100J	Faux	Vrai si le patient a été transfusé il y a moins de 100 jours.
TECHNIQUES_ANESTHESIE_LAST	Anesthésie générale avec intubation	Technique anesthésique utilisée lors de la dernière intervention (vide si pas de précédente intervention)
ATCD_NB_INTERVENTION	1	Nombre d'interventions précédentes
ATCD_SYNTHESE	Sténose du pylore ¹ (1 mois), Ancien fumeur (sevré en 2015)	Historique des pathologies et antécédents du patient .

Table A.2 – Variables d'antécédents et d'historique de santé dans le dossier médical du patient présenté dans la table A.1.

Variable	Exemple	Description
PROGRAMATION LIBELLE_INTERVENTION	Programmée Hernie inguinale sous coelioscopie avec prothèse	Indique si l'intervention est programmée ou Non Libellé de l'intervention prévue
DATE_INTERVENTION	10/08/2022	Date prévue pour l'intervention
PAS	122	Valeur de la PAS lors de la consultation
PAD	73	Valeur de la PAD lors de la consultation
FC	73	Valeur de la FC lors de la consultation
POIDS	80	Poids du patient lors de la consultation
TAILLE	182	Taille du patient lors de la consultation
EXAMENS		Examens complémentaires prescrits lors de la consultation
BILAN_PREOPERATOIRE	Bilan à faire : Numération globulaire, Plaquettes, Groupe Sanguin	Bilans à réaliser en amont de l'intervention
DATE_CS	01/08/2022	Date de la consultation
HEURE_CS	14:02:05	Heure de la consultation
INFO_PATIENT	Vrai	Vrai si le patient a bien été informé des caractéristiques et risques de l'anesthésie.
DATE_INFO_PATIENT	01/08/2022	Date de l'information du patient sur les risques anesthésiques
CONCLUSION_CS	Non accompagné. Information sur l'anesthésie générale avec intubation (risque de bris dentaire), risque hémorragique et transfusionnel	Commentaires de conclusion de la consultation
SYNTHESE	Auscultation cardio-pulmonaire sans particularité. Rhinite en cours	Commentaire de synthèse de la consultations concernant l'examen clinique

Table A.3 – Variables de la consultation anesthésique et de la visite pré-anesthésique dans le dossier médical d'un patient

Variable	Exemple	Description
DENTITION	-	Commentaires éventuels sur la dentition du patient
APP_CARDI	Faux	Vrai si présence d'une pathologie liée à l'appareil cardiaque.
APP_RESPI	Faux	Vrai si présence d'une pathologie liée à l'appareil respiratoire.
SYST_NERV	Faux	Vrai si présence d'une pathologie liée au système nerveux.
PSYCHISME	Faux	Vrai si présence d'une pathologie psychiatrique.
APP_GASTRO	Faux	Vrai si présence d'une pathologie liée à l'appareil gastro-intestinal.
APP_URO	Faux	Vrai si présence d'une pathologie liée à l'appareil urologique.
TETE_COU	Faux	Vrai si présence d'une pathologie liée à la tête ou au cou.
APP_LOCO_MOTEUR	Faux	Vrai si présence d'une pathologie liée à l'appareil loco-moteur.
AIRES_GG	Faux	Vrai si présence d'une pathologie liée aux aires ganglionnaires.
PEAU	Faux	Vrai si présence d'une pathologie liée à la peau.
SYST_ENDOCRINIEN	Faux	Vrai si présence d'une pathologie liée au système endocrinien.
NUQUE_RAIDE	Faux	Vrai si présence d'une nuque raide
SIGNE_PRIEUR	Faux	Vrai si le patient réalise un signe du prier partiel.
DENT_ABIMEE	Faux	Vrai si présence de dents abimées.
DENTIER_FIXE	Faux	Vrai si présence d'un dentier fixe.
DENTIER_AMOVIBLE	Faux	Vrai si présence d'un dentier amovible.
LENTILLES_CONTACT	Faux	Vrai si présence de lentilles de contact.
OUVERTURE_BOUCHE	> 3,5 cm	Mesure de l'ouverture de la bouche
DISTANCE_MENTON	> 65 mm	Mesure de la distance menton - os hyoïde
MALAMPATTI	1	Score de Malampatti

Table A.4 – Variables de l'examen clinique, réalisé pendant la consultation anesthésique et la visite pré-anesthésique, dans le dossier médical d'un patient

Variable	Exemple	Description
GRAND_AGE	Faux	Vrai si l'âge du patient est supérieur à 65 ans.
DIABETE	Faux	Vrai si le patient présente un diabète.
OBESE	Faux	Vrai si le patient est obèse (IMC >30).
I_REIN	Faux	Vrai si le patient présente une insuffisance des reins.
I_HEPATIQUE	Faux	Vrai si le patient présente une insuffisance hépatique.
I_RESPIRATOIRE	Faux	Vrai si le patient présente une insuffisance respiratoire.
I_SURRENALE	Faux	Vrai si le patient présente une insuffisance surrénale.
I_CARDIAQUE	Faux	Vrai si le patient présente une insuffisance cardiaque.
POLYARTERIEL	Faux	Vrai si le patient présente une pathologie polyartérielle.
HTA	Faux	Vrai si le patient présente une hypertension artérielle (HTA) .
HTAP	Faux	Vrai si le patient présente une hypertension artérielle pulmonaire (HTAP).
IDM_6MOIS	Faux	Vrai si le patient a eu un infarctus du myocarde (IDM).
ANGOR	Faux	Vrai si le patient a une angine de poitrine (appelé aussi angor).
VALVULAIRE	Faux	Vrai si le patient présente une pathologie valvulaire.
NYHA	1	La classification NYHA (New-York Heart Association) permet de classer la gravité des insuffisances cardiaques.
MALDIE_NEUROMUSC	Faux	Vrai si le patient présente une pathologie neuromusculaire.
VIRUS	Faux	Vrai si le patient présente une infection virale.
ALLERGIE	Faux	Vrai si le patient présente des allergies.
INTUB_DIFFICILE	Faux	Vrai si le patient présente un risque d'intubation difficile.
RACHIS_DIFFICILE	Faux	Vrai si le patient présente un rachis raide ou peu accessible.
VEINE_DIFFICILE	Faux	Vrai si le patient présente un capital veineux pauvre ou peu accessible.
RAI_PLUS	Faux	Vrai si le patient présente des agglutinines irrégulières qui empêcheraient une transfusion.
URGENCE	Faux	Vrai si l'intervention est réalisée en urgence.
ETAT_INFECTIEUX	Faux	Vrai si le patient présente un état infectieux.
ASA	1	Score ASA du patient

Table A.5 – Variables des facteurs de risque, déterminés pendant la consultation anesthésique, dans le dossier médical d'un patient

A.2 Exemple d'une séquence d'événements

Cette séquence d'événements correspond à une intervention pour la cure d'une hernie inguinale sous coelioscopie avec pose de prothèse réalisée chez le patient décrit en A.1.

Temps (s)	ID	Événements
0	AGE-insta01	Mise en place du patient
30	AGE-mofca01	Monitorage FC
60	AGE-mopam01	Monitorage PAM
90	AGE-mospo01	Monitorage SPO2
120	AGE-mobis01	Monitorage BIS
150	AGE-povvp01	Pose VVP (cathlon)
390	AGE-baihu01	Bair hugger
450	ADG-cefab01	CEFAZOLINE bolus
600	AGE-preox01	pré-oxygénation
780	ADG-sufei01	SUFENTANIL induction
780	ADG-ketai01	KETAMINE induction
810	ADG-propoi01	PROPOFOL induction
840	ADG-atrai01	ATRACURIUM induction
870	AGE-occoc01	Occlusion oculaire
1230	AGE-intub01	Intubation
1500	ADG-sevoc01	SEVOFLURANE continu
1530	AGE-venco01	Ventilation contrôlé intubée
1560	AGE-oxyge01	Oxygène 30%
1590	AGE-auspu01	Auscultation pulmonaire
1620	AGE-supeb01	Surveillance pression ballonnet
1650	AGE-marea01	Recrutement alvéolaire
1740	AGE-motem01	Pose sonde T°C
2010	AGE-auspu01	Auscultation pulmonaire
2040	AGE-supoa01	Surveillance points d'appui
2190	ADG-sevoc01	SEVOFLURANE continu
2220	ADG-sufec01	SUFENTANIL continu
2280	AGE-insci01	Incision
2760	AGE-troca01	Trocard sous-ombilical
2820	AGE-inpne01	Inflation pneumopéritoine
3270	ADG-atrac01	ATRACURIUM continu
3300	AGE-auspu01	Auscultation pulmonaire
3330	AGE-marea01	Manoeuvre Recrutement Alvéolaire
3630	AGE-disep01	Dissection espace préperitonéal
3960	AGE-discs01	Dissection cordon spermatique
4200	AGE-diszh01	Dissection cordon zone herniaire
4500	ADG-sufec01	SUFENTANIL continu
5670	ADG-atrac01	ATRACURIUM continu
6390	AGE-mppro01	Mise en place prothèse
8490	AGE-depne01	Déflation pneumopéritoine
8550	AGE-ferme01	Fermeture
8910	AGE-panse01	Pansement
9540	AGE-cdecu01	Contrôle de décurarisation
9720	AGE-decur01	Décurarisation
10020	AGE-insta01	Mise en place du patient

Table A.6 – Séquence d'événements d'un profil anesthésique pour la cure d'une hernie inguinale sous coelioscopie avec pose de prothèse sur un homme de 30 ans en bonne santé et sans antécédents médicaux

A.3 Paramètres physiologiques enregistrés

Nom	Abréviation	Unité	Fréquence	Enregistrement
Fréquence cardiaque	FC	Battements minutes	150 secondes	Systématique
Pression artérielle systolique	PAS	Millimètres de mercure	150 secondes	Systématique
Pression artérielle moyenne	PAM	Millimètres de mercure	150 secondes	Systématique
Pression artérielle diastolique	PAD	Millimètres de mercure	150 secondes	Systématique
Fraction inspirée en CO₂	FeCO ₂	Pourcentage	150 secondes	Systématique
Fraction inspirée en gaz halogéné	FiHal	Pourcentage	150 secondes	Systématique
Fraction expirée en gaz halogéné	FeHal	Pourcentage	150 secondes	Systématique
Température	T	Degré Centigrade	150 secondes	Systématique
Fréquence respiratoire	FR	Cycles par minute	150 secondes	Systématique
Volume de ventilation	VT	Litres	150 secondes	Systématique
Volume moyen	VM	Litres	150 secondes	Systématique
Debit de Gaz Frais	-	Litres par minute	150 secondes	Systématique
Fraction inspirée en CO₂	FiCO ₂	Pourcentage	150 secondes	Systématique
Fraction inspirée en O₂	FiO ₂	Pourcentage	150 secondes	Systématique
Fraction inspirée en NO ₂	FiN ₂ O	Pourcentage	150 secondes	Ponctuel
Pression maximale	PMAX	Millimètres de mercure	150 secondes	Systématique
Pression expiratoire positive	PEEP	Millimètres de mercure	150 secondes	Systématique
Pression moyenne	PMOY	Millimètres de mercure	150 secondes	Systématique
Index AAI	AAI	Sans unité	150 secondes	Rare
Hématrocrite	-	Pourcentage	Variable	Ponctuel
Glycémie	-	Grammes par litres	Variable	Ponctuel
Coagulation	-	Pourcentage	Variable	Ponctuel
Hémoglobine	-	Grammes par litres	Variable	Ponctuel
Pression artérielle pulmonaire systolique	PAPS	Millimètres de mercure	150 secondes	Ponctuel
Pression artérielle pulmonaire moyenne	PAPM	Millimètres de mercure	150 secondes	Ponctuel
Pression artérielle pulmonaire diastolique	PAPD	Millimètres de mercure	150 secondes	Ponctuel
Pression veineuse centrale	PVC	Millimètres de mercure	150 secondes	Ponctuel
Pression artérielle pulmonaire d'occlusion	PCPW	Millimètres de mercure	150 secondes	Ponctuel
O ₂ dans le Sang Veineux	SVO ₂	Pourcentage	150 secondes	Ponctuel
Index cardiaque	IC	Pourcentage	150 secondes	Ponctuel
Segment ST1	ST1	Millivolts	150 secondes	Ponctuel
Segment ST2	ST2	Millivolts	150 secondes	Ponctuel
Dérivation ST1	-	Millivolts	150 secondes	Ponctuel
Dérivation ST2	-	Millivolts	150 secondes	Ponctuel
<i>Train of Four</i>	TOF	Seconde	Variable	Ponctuel
Ratio T ₁ /T ₄	T ₄ T ₁	Sans unité	Variable	Systématique
Index Bi-spectral (BIS)	BIS	Sans unité	150 secondes	Ponctuel
Qualité du BIS	-	Sans unité	150 secondes	Ponctuel
Pression intra crânienne	PIC	Millimètres de mercure	150 secondes	Rare

Table A.7 – Paramètres physiologiques du patient et de l'intervention enregistrés par Pégase sous la forme de séries temporelles. Les variables en gras sont celles simulées par le générateur de profils anesthésiques présenté au chapitre 3.

A.4 Exemple d'une série temporelle multivariée

Cette série temporelle multivariée correspond à la cure d'une hernie inguinale sous coelioscopie avec pose de prothèse réalisée chez le patient décrit en A.1 avec la séquence d'événements décrite en A.2.

Les différentes variables hemodynamiques sont présentées sous la forme de courbes, en figure A.1 et A.2.

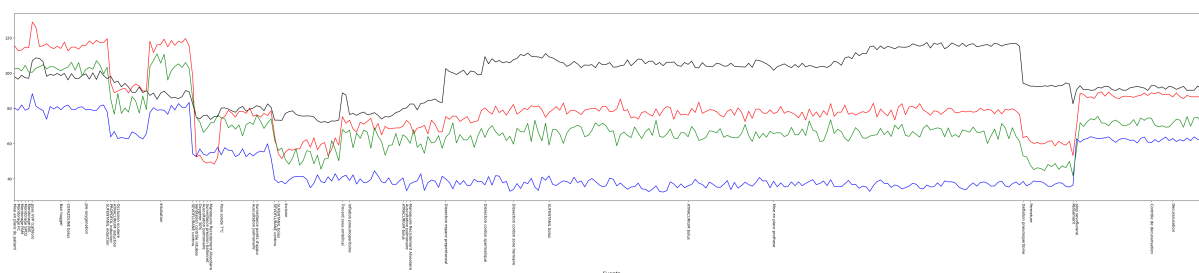


Figure A.1 – Variables hémodynamiques pendant la cure d'une hernie inguinale sous coelioscopie du patient décrit en A.1

La Figure A.2 présente les premières valeurs des séries temporelles des variables hémodynamiques, correspondant aux étapes de l'entrée et de l'induction. Ces valeurs sont données dans la table A.8.

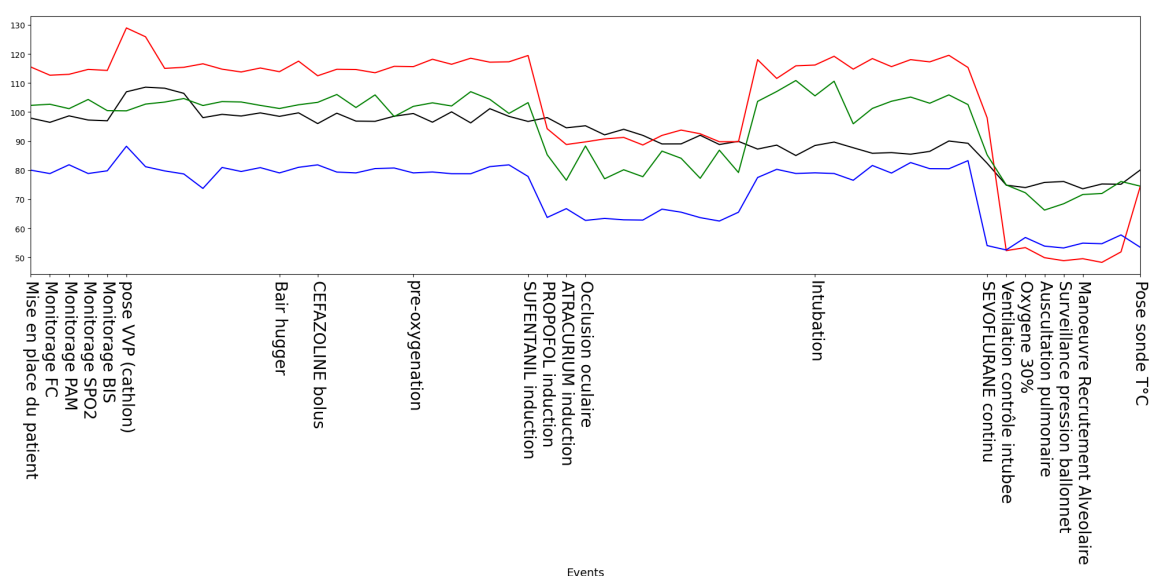


Figure A.2 – Variables hémodynamiques pendant la cure d'une hernie inguinale sous coelioscopie du patient décrit en A.1, pour les étapes de l'entrée et de l'induction

Temps (s)	ID	Événements	FC	PAS	PAD	PAM
0	AGE-insta01	Mise en place du patient	89	93	105	80
30	AGE-mofca01	Monitoring FC	98	102	116	80
60	AGE-mopam01	Monitoring PAM	96	103	113	80
90	AGE-mospo01	Monitoring SPO2	99	101	113	80
120	AGE-mobis01	Monitoring BIS	97	104	115	80
150	AGE-povvp01	Pose VVP (cathlon)	97	100	114	80
180	-		107	100	129	80
210			109	103	126	80
240			108	103	115	80
270			106	105	115	80
300			98	102	117	80
330			99	104	115	80
360			99	103	114	80
390	AGE-baihu01	Bair hugger	100	102	115	80
420			99	101	114	80
450	ADG-cefab01	CEFAZOLINE bolus	100	102	118	80
480			96	103	112	82
510			100	106	115	79
540			97	102	115	79
570			97	106	114	81
600	AGE-preox01	Pré-oxygénation	99	98	116	81
630			99	102	116	79
660			97	103	118	79
690			100	102	116	79
720			96	107	119	79
750			101	104	117	81
780	ADG-sufei01	SUFENTANIL induction	99	100	117	82
810	ADG-ketai01	KETAMINE induction	97	103	119	78
840	ADG-propoi01	PROPOFOL induction	98	85	94	64
870	ADG-atrai01	ATRACURIUM induction	95	77	89	67
900	AGE-occoc01	Occlusion oculaire	95	88	90	63
930			92	77	91	63
960			94	80	91	63
990			92	78	89	63
1020			89	87	92	67
1050			89	84	94	66
1080			92	77	93	64
1110			89	87	90	62
1140			90	79	90	66
1170			87	104	118	77
1200			89	107	112	80
1230	AGE-intub01	Intubation	85	111	116	79
1260			88	106	116	79
1290			90	111	119	79
1320			88	96	115	77
1350			86	101	118	82
1380			86	104	116	79
1410			85	105	118	83
1440			86	103	117	81
1470			90	106	120	80
1500	ADG-sevoc01	SEVOFLURANE continu	89	103	115	83
1530	AGE-venco01	Ventilation contrôlé intubée	82	85	98	54
1560	AGE-oxyge01	Oxygène 30%	75	75	52	53
1590	AGE-auspu01	Auscultation pulmonaire	74	72	53	57
1620	AGE-supeb01	Auscultation pulmonaire	76	66	50	54
1650	AGE-marea01	Check pression ballonnet	76	68	49	53
1680		Recrutement alvéolaire	74	72	50	55
1710			75	72	48	55
1740	AGE-motem01	Pose sonde T°C	75	76	52	58
			80	75	74	53

Table A.8 – Série temporelle des variables hémodynamiques pour les étapes de l’entrée et de l’induction de la cure d’une hernie inguinale sous coelioscopie avec pose de prothèse

A.5 Caractéristiques d’une intervention chirurgicale et d’une anesthésie

La table A.9 présente les caractéristiques d’une intervention, et la table A.10 celles d’une anesthésie, enregistrées par Pégase et présentes dans un profil anesthésique. Ces caractéristiques sont fournies pour exemple.

Variable	Exemple	Commentaire
LIEN	12345678901111213	Clé étrangère pour le système d’information du CHU
INTITULE	Cure unilatérale d’une hernie de l’aîne avec pose de prothèse, par vidéochirurgie	D’après la CCAM ¹
CODE	LMMC002	D’après la CCAM
LATERALITE	-	Si pertinent
DATE_DEBUT	10/08/2022	
HEURE_DEBUT	18 : 39 : 07	
HEURE_FIN	19 : 32 : 12	
TOPOGRAPHIE_1	Orthopedie générale	Grand appareil anatomique
TOPOGRAPHIE_2	Paroi abdominale	Organe ou zone
VERSION	CCAM	Version de la classification
IDENTIFIANT	01231456789	Identifiant unique du patient

Table A.9 – Variables associées aux caractéristiques d’une intervention enregistrées par Pégase.

¹ CCAM : Classification commune des actes médicaux

Variable	Exemple	Commentaire
ARRIVEE_SALLE	18 : 39 : 07	Entrée du patient dans le bloc opératoire
INDUCTION	18 : 50 : 46	Début de l’anesthésie, le patient n’est plus conscient
INCISION	18 : 52 : 20	Début de la procédure chirurgicale
FIN_CHIRURGIE	19 : 29 : 20	Fin de la procédure chirurgicale
SORTIE_SALLE	19 : 32 : 12	Sortie du bloc et entrée en Salle de Surveillance Post-Interventionnelle
SORTIE_SSPI	20 : 43 : 14	Sortie de la salle de surveillance post-interventionnelle
DUREE_ANESTHESIE	0 : 39 : 52	De l’induction à la sortie du bloc
DUREE_SSPI	01 : 11 : 02	De la sortie du bloc à la sortie de la salle de surveillance post-interventionnelle, le patient est conscient
DUREE_CHIRURGIE	00 : 39 : 00	De l’incision à la fin de la chirurgie
DUREE_PREPA	00 : 13 : 13	De l’entrée au bloc à l’incision
DUREE_SALLE	00 : 53 : 05	Temps de présence au bloc
ORIGINE	Chirurgie Digestive	Service d’origine du patient
SORTIE	Chirurgie Digestive	Service de sortie du patient
PROGRAMMATION	Programmée	Programmation de l’intervention parmi : Programmée, Ambulatoire, Semi-urgent ou Urgent
JEUNE	Vrai	Vrai si l’âge du patient < 65 ans, siNon faux
INTUBATION_FACILE	Facile	Facile, si l’intubation s’est déroulée sans problème, siNon vide
TYPE_INTUBATION	Oro-trachéale	Technique utilisée pour l’intubation parmi : masque facial, masque laryngé, oro-trachéale
TOTAL_URINES	0	Urines récoltées pendant l’intervention, en millilitres
TOTAL_PERFS	500	Perfusion administré au patient pendant l’intervention, en millilitres
TOTAL_PERTES	0	Pertes totales en liquide pendant l’intervention, en millilitres
DUREE_CEC	00 : 00 : 00	Durée de la circulation extracorporelle, si pertinent
DUREE_CLAMPAGE	00 : 00 : 00	Durée de la du clampage, si pertinent
DUREE_GARROT	00 : 00 : 00	Durée de la du garrot, si pertinent
DUREE_INSUFLATION	00 : 32 : 30	Durée de l’insuflation, si pertinent
TRANSFU_CG_NB	0	Nombre de transfusion de concentrés de globules rouges (CG)
TRANSFU_PFC_NB	0	Nombre de transfusion de plasma frais congelé (PFC)
TRANSFU_CP_NB	0	Nombre de transfusion de concentrés plaquettaires (CP)

Table A.10 – Variables associées aux caractéristiques d’une anesthésie enregistrées par Pégase

GRAMMAIRE ANESTHÉSIQUE ET TYPOLOGIE DES ÉVÉNEMENTS

La grammaire d'une anesthésie se structure en deux niveaux supérieurs, fixes et commun à toutes les anesthésies générales, et en un troisième niveau, variable selon les interventions. Un quatrième niveau peut éventuellement venir préciser le niveau trois dans certains cas.

B.1 Grammaire anesthésique - Niveau 1

Notation de Backus-Naur

```
<anesthésie> ::= <entrée> <induction> <procédure> <sortie>
```

B.2 Grammaire anesthésique - Niveau 2

Notation de Backus-Naur

```
<entrée> ::= <mise_en_place> <monitorage> <prémedication>  
<induction> ::= <préoxygénation> <médication> <intubation> <controle_fin_induction>  
<procédure> ::= <préparation> <procédure>  
<sortie> ::= <décurarisation> <sortie>
```

B.3 Grammaire - Hernie inguinale

Représentation par un graphe

Notation de Backus-Naur

```
<mise_en_place> ::= <entrée> <mise_en_place>
<monitorage> ::= <FC> <PA> <SP02> <BIS> <controle_TOF>
| <FC> <PA> <SP02> <controle_TOF>
<prémedication> ::= <bair_hugger> <vvp> <prophylaxie> <rv>
| <vvp> <prophylaxie> <rv>
| <vvp> <prophylaxie>
<préoxygénation> ::= <masque_facial> <préox_admin> <morphiniques> <analgesie>
<médication> ::= <hypnotiques>
<intubation> ::= <curare> <ventilation> <fermeture des yeux> <intub_procedure>
| <curare> <fermeture des yeux> <ventilation> <intub_procedure>
| <curare> <ventilation> <intub_procedure>
<controle_fin_induction> ::= <ventil_control> <entretien_anesth> <O2_30%> <Auscul_pul>
| <ballonet> <recrut_alvéolaire> <temp> <fermeture des yeux>
| <ventil_control> <entretien_anesth> <O2_30%> <Auscul_pul>
| <ballonet> <recrut_alvéolaire> <temp>
| <ventil_control> <entretien_anesth> <O2_30%> <Auscul_pul>
| <ballonet> <temp>
<préparation> ::= <mise_en_place_chir> <Auscul_pul> <points_support>
| <mise_à_disposition>
| <mise_en_place_chir> <Auscul_pul>
| <mise_à_disposition>
<procédure> ::= <incision> <trocart_ombilical> <inflation> <Auscul_pul> <recrut_alvéolaire>
| <dissec_espace_prépéritonéal> <dissec_cordon_spermatique> <dissec_zone_herniaire>
| <prothèse> <déflation> <fermeture> <pansement>
| <incision> <trocart_ombilical> <inflation> <Auscul_pul>
| <dissec_espace_prépéritonéal> <dissec_cordon_spermatique> <dissec_zone_herniaire>
| <prothèse> <déflation> <fermeture> <pansement>
| <incision> <trocart_ombilical> <inflation> <Auscul_pul> <recrut_alvéolaire>
| <dissec_espace_prépéritonéal> <dissec_cordon_spermatique> <dissec_zone_herniaire>
| <déflation> <fermeture> <pansement>
| <incision> <trocart_ombilical> <inflation> <Auscul_pul>
| <dissec_espace_prépéritonéal> <dissec_cordon_spermatique> <dissec_zone_herniaire>
| <déflation> <fermeture> <pansement>
<décurarisation> ::= <controle_TOF> <attente_décurarisation>
| <controle_TOF>
| <controle_TOF> <antidote_décurarisation>
<sortie> ::= <mise_en_place> <sortie>
```

B.4 Grammaire - Prothèse totale de genou

Notation de Backus-Naur

```
<mise_en_place> ::= <entrée> <mise_en_place>
<monitorage> ::= <FC> <PA> <SP02> <BIS> <controle_TOF>
| <FC> <PA> <SP02> <controle_TOF>
<prémedication> ::= <bair_hugger> <vvp> <prophylaxie> <rv>
| <vvp> <prophylaxie> <rv>
| <vvp> <prophylaxie>
<préoxygénation> ::= <masque_facial> <préox_admin> <morphiniques> <analgesie>
<médication> ::= <hypnotiques>
<intubation> ::= <curare> <ventilation> <fermeture_des_yeux> <intub_procedure>
| <curare> <fermeture_des_yeux> <ventilation> <intub_procedure>
| <curare> <ventilation> <intub_procedure>
<controle_fin_induction> ::= <ventil_control> <entretien_anesth> <02_30\%> <Auscult_pul>
| <ballonet> <recrut_alvéolaire> <temp> <fermeture_des_yeux>
| <ventil_control> <entretien_anesth> <02_30\%> <Auscult_pul>
| <ballonet> <recrut_alvéolaire> <temp>
| <ventil_control> <entretien_anesth> <02_30\%> <Auscult_pul>
| <ballonet> <temp>
<préparation> ::= <mise_en_place_chir> <garrot> <points_support>
| <mise_à_disposition>
| <mise_en_place_chir> <garrot> <mise_à_disposition>
| <mise_en_place_chir> <mise_à_disposition>
| <mise_en_place_chir> <points_support>
| <mise_à_disposition>
<procédure> ::= <gonflage_garrot> <incision> <incision_parapatellaire_interne>
| <luxation_genou> <incision_osseuse> <excision_surfaces_articulaires>
| <taille_condyles> <préparation_ciment> <choix_implants>
| <condyles_prothétiques_fémoral> <condyles_prothétiques_tibial>
| <condyles_prothétiques_rotuliennes> <réduction_luxation>
| <analyse_ligamentaire> <équilibre_ligamentaire> <contrôle_laxité>
| <relâchement_garrot> <fermeture> <pansement>
| <incision> <incision_parapatellaire_interne> <luxation_genou>
| <incision_osseuse> <excision_surfaces_articulaires>
| <taille_condyles> <préparation_ciment> <choix_implants>
| <condyles_prothétiques_fémoral> <condyles_prothétiques_tibial>
| <condyles_prothétiques_rotuliennes> <réduction_luxation>
| <analyse_ligamentaire> <équilibre_ligamentaire> <contrôle_laxité>
| <fermeture> <pansement>
<décurarisation> ::= <controle_TOF> <attente_décurarisation>
| <controle_TOF>
| <controle_TOF> <antidote_décurarisation>
<sortie> ::= <mise_en_place> <sortie>
```

B.5 Typologie des événements anesthésiques

Médicaments

ID	Type	Médicament	Type patient	Poids à utiliser	Allergie
alfe-01	Morphinique	ALFENTANIL	Standard	Poids réel	Oui
sufe-01	Morphinique	SUFENTANIL	Standard	Poids réel	Oui
remi-01	Morphinique	REMIFENTANIL	Standard	Poids idéal	Oui
lido-01	Anesthésique local	LIDOCAINE	Standard	Poids réel	Oui
sevo-02	Anesthésique	SEVOFLURANE	Standard	Poids réel	Non
keta-01	Anesthésique	KETAMINE	Standard	Poids réel	Oui
prop-01	Anesthésique	PROPOFOL	Standard	Poids réel	Oui
prop-02	Anesthésique	PROPOFOL	Sportif	Poids réel	Oui
etom-01	Hypnotique	ETOMIDATE	Standard	Poids réel	Non
thio-01	Hypnotique	THIOPENTAL	Standard	Poids réel	Oui
ephe-01	Constricteur vasculaire	EPHEDRINE	Standard	Poids réel	Non
atro-01	Antispasmodique	ATROPINE	Standard	Poids idéal	Non
dexa-01	Anti-inflammatoire	DEXAMETHASONE	Standard	Poids réel	Oui
keto-01	Anti-inflammatoire	KETOPROFENE	Standard	Poids idéal	Non
nepo-01	Antalgique	NEFOPAM	Standard	Poids réel	Non
cefa-01	Antibiotique	CEFAZOLINE	Standard	Poids ajusté ($f = 0,3$)	Oui
adre-01	Sympathomimétique	ADRENALINE	Standard	Poids idéal	Non
nora-01	Constricteur vasculaire	NORADRENALINE	Standard	Poids idéal	Non
rocu-01	Curare	ROCURIUM	Standard	Poids réel	Oui
cisa-01	Curare	CISATRACURIUM	Standard	Poids réel	Oui
atra-01	Curare	ATRACURIUM	Standard	Poids idéal	Oui
atra-02	Curare	ATRACURIUM	Sportif	Poids idéal	Oui
celo-01	Curare	CÉLOCURINE	Standard	Poids réel	Oui

Table B.1 – Typologie des médicaments utilisés lors des deux interventions étudiées, la cure d’une hernie inguinale sous coelioscopie avec pose de prothèse et la mise en place d’une prothèse totale de genou

Descripteurs

ID	Label	Commentaire
enter-01	Entrée en salle	Indique l’entrée du patient au bloc opératoire
induc-01	Induction	Indique le début de l’induction et donc de la perte de conscience du patient
occoc-01	Occlusion oculaire	Fermeture des yeux du patient
miadc-01	Mise à disposition du chirurgien	Le patient est stage et prêt pour commencer la procédure chirurgicale ou médicale
insta-01	Installation	Le patient est mis en place conformément aux besoins de l’intervention
dproc-01	Début de procédure	Début de la procédure chirurgicale ou médicale
fproc-01	Fin de procédure	DFin de la procédure chirurgicale ou médicale
dinsu-01	Début d’insufflation	Début de l’insufflation de l’abdomen du patient, spécifique à la coelioscopie
finsu-01	Fin d’insufflation	Fin de l’insufflation de l’abdomen du patient, spécifique à la coelioscopie
panse-01	Pansement	Réalisation du pansement
exits-01	Sortie de Salle	Sortie du bloc vers la salle de surveillance post-intervention

Table B.2 – Typologie des événements descripteurs d’une intervention

Gestes médicaux

ID	Action	Patient
stimn-01	Stimulation nociceptive mineure	Standard
stimn-02	Stimulation nociceptive moyenne	Standard
stimn-03	Stimulation nociceptive majeure	Standard
preox-01	Pré-oxygénation	Standard
intub-01	Intubation	Standard
pomala-01	Pose masque laryngé	Standard
pexpo-01	Pression expiratoire positive (PEP)	Standard
pexpo-01	Pression expiratoire positive (PEP)	Sportif
povvp-01	Pose VVP (cathlon)	État veineux normal
povvp-02	Pose VVP (cathlon)	État veineux pauvre
inpne-01	Inflation pneumopéritoine	Standard
depne-01	Déflation pneumopéritoine	Standard
panse-01	Pansement	Standard
posas-01	Position assise	Standard
decur-01	Décurarisation	Standard
marea-01	Manoeuvre Recrutement Alvéolaire	Standard
baihu-01	Bair hugger	Standard
baihu-02	Maintien Bair hugger	Standard
venco-01	Ventilation contrôlé intubée	Standard
venco-01	Ventilation contrôlé intubée	Sportif
venma-01	Ventilation au masque facial	Standard
venma-01	Ventilation au masque facial	Sportif
insta-01	Mise en place du patient	Standard
ferme-01	Fermeture	Standard
insci-01	Incision	Standard
troca-01	Trocard sous-ombilical	Standard
disep-01	Dissection espace préperitonéal	Standard
discs-01	Dissection cordon spermatique	Standard
diszh-01	Dissection cordon zone herniaire	Standard
mppro-01	Mise en place prothèse	Standard
mofca-01	Monitoring FC	Standard
mopam-01	Monitoring PAM	Standard
mospo-01	Monitoring SPO2	Standard
mobis-01	Monitoring BIS	Standard
occoc-01	Occlusion oculaire	Standard
auspu-01	Auscultation pulmonaire	Standard
supeb-01	Surveillance pression ballonnet	Standard
motem-01	Pose sonde T°C	Standard
cdecu-01	Contrôle de décurarisation	Standard
supoa-01	Surveillance points d'appui	Standard
oxyge-01	Oxygène 30%	Standard

Table B.3 – Typologie des gestes médicaux réalisés par l'équipe médicale lors de la cure d'une hernie inguinale sous coelioscopie

ÉTUDE COMPARATIVE DE NEUF DISSIMILARITÉS ENTRE SÉRIES TEMPORELLES UNIVARIÉES

C.1 Description des jeux de données

Les 28 jeux de données ont été produits pour cette étude comparative. Ils se répartissent comme suit :

- 26 jeux issus de la *UCR Time Series Classification Archive* [87],
- 1 jeu produit par un modèle *Partially Hidden Markov Chain Linear AutoRegressive* (PHMC-LAR) [88],
- 1 jeu produit par notre générateur de profils anesthésiques pour la cure d’une hernie inguinale sous coelioscopie (voir Chapitre 3)

C.1.1 Jeux de profils anesthésiques simulés

Nous avons également utilisé dans le cadre de cette étude 2 ensembles de données de profils anesthésiques simulés par notre générateur (voir Chapitre 3) ou le modèle PHMC-LAR. La table C.1 présente les principales caractéristiques de ces jeux.

Nom	Effectif	Longueur (points)
PHMC-LAR	100	1000
Générateur DBLBS	1000	Variable

Table C.1 – Caractéristiques des 2 jeux de profils anesthésiques simulés utilisés lors de l’expérimentation

C.1.2 Jeux de la *UCR archive*

Ces jeux ont été sélectionné en écartant les jeux de données manifestement différent des séries temporelles observées lors du suivi des patients au bloc opératoire. Nous avons ainsi exclu les jeux de données avec des séries temporelles cycliques ou saisonnières, ou encore ceux montrant des croissances ou décroissances fréquentes. La table C.2 présente les principales caractéristiques de ces jeux.

Nom du jeu	Effectif	Longueur
Beef	60	470
Coffee	56	286
DistalPhalanxOutlineAgeGroup	539	80
DistalPhalanxOutlineCorrect	876	80
DistalPhalanxTW	539	80
ECG200	200	96
ECG5000	5000	140
InlineSkate	600	1882
Mallat	2400	1024
NonInvasiveFetalECGThorax1	3765	750
NonInvasiveFetalECGThorax2	3765	750
SyntheticControl	600	60
TwoLeadECG	1162	82
UWaveGestureLibraryAll	4478	945
AllGestureWiimoteX	1000	Variable
AllGestureWiimoteY	1000	Variable
AllGestureWiimoteZ	1000	Variable
BME	180	12
Crop	24 000	46
EOGHorizontalSignal	724	1250
EOGVerticalSignal	724	1250
GestureMidAirD1	338	Variable
GestureMidAirD2	338	Variable
GestureMidAirD3	338	Variable
GesturePebbleZ1	304	Variable
GesturePebbleZ2	304	Variable

Table C.2 – Caractéristiques des 26 jeux de la *UCR archive* utilisés lors de l’expérimentation

C.2 Résultats complémentaires

C.2.1 Jeux de profils anesthésiques

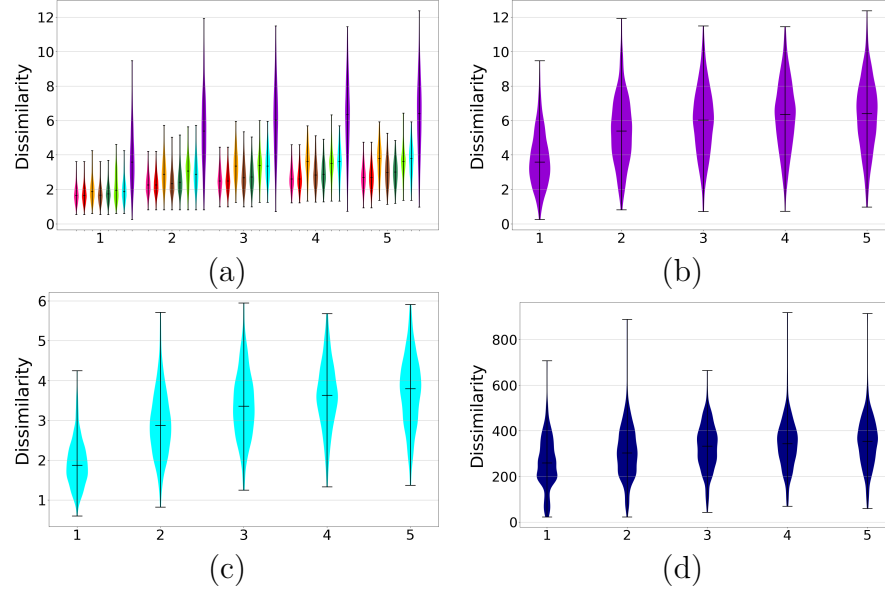


Figure C.1 – Comparaison des 9 dissimilarités sur le jeu DBLBS ($w=10$) et la perturbation *ampl_gauss*. En abscisses : degré croissant de perturbation. DTWx : DTWdtai : rose ; DTWc : rouge ; DTWf : orange ; DTWsc : marron ; DTWi : vert clair ; DTWm : vert foncé ; MPDist : bleu clair ; tsfresh : bleu marine ; DFT : violet.

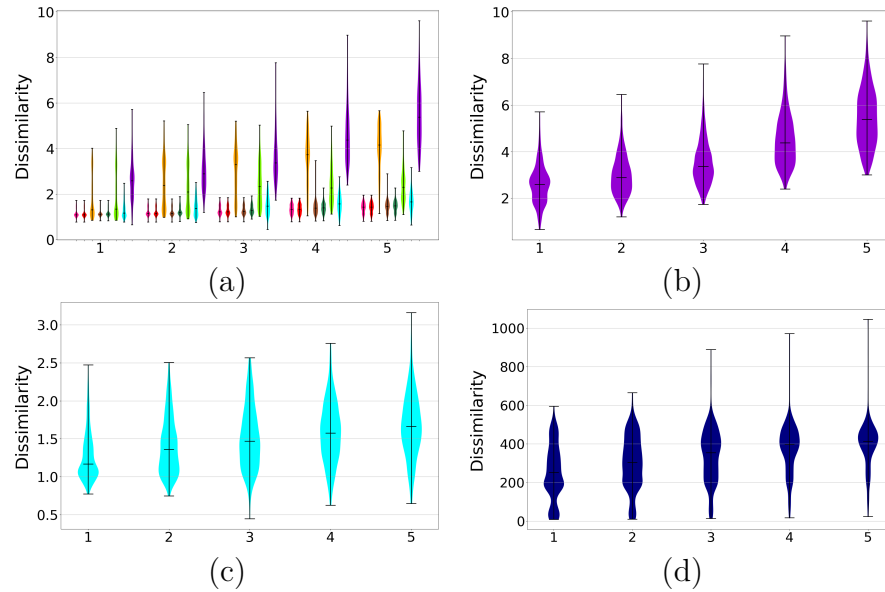


Figure C.2 – Comparaison des 9 dissimilarités sur le jeu DBLBS ($w=10$) et la perturbation *dilat_biunif_e*. En abscisses : degré croissant de perturbation. DTWx : DTWdtai : rose ; DTWc : rouge ; DTWf : orange ; DTWsc : marron ; DTWi : vert clair ; DTWm : vert foncé ; MPDist : bleu clair ; tsfresh : bleu marine ; DFT : violet.

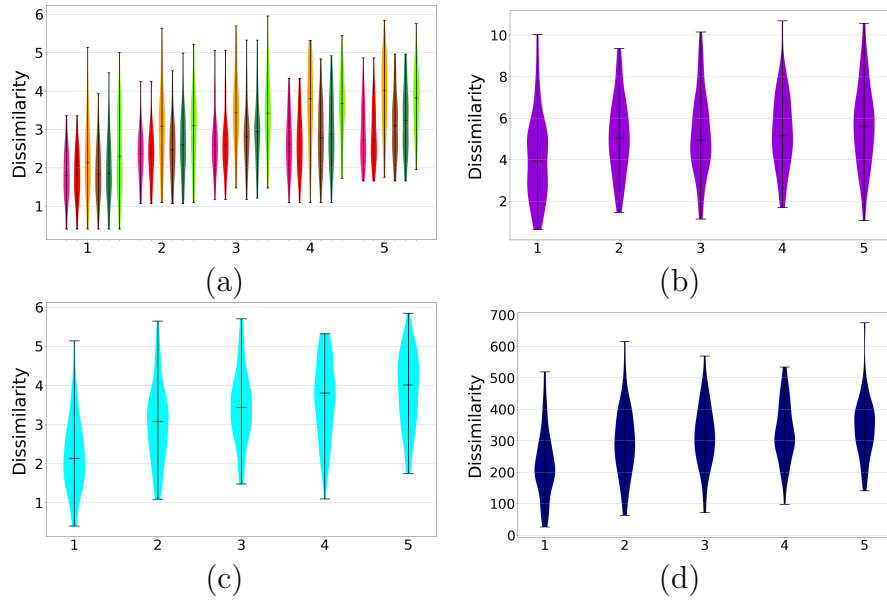


Figure C.3 – Comparaison des 9 dissimilarités sur le jeu PHMC-LAR ($w=10$) et la perturbation *ampl_gauss*. En abscisses : degré croissant de perturbation. DTWx : DTWdtai : rose ; DTWc : rouge ; DTWf : orange ; DTWsc : marron ; DTWi : vert clair ; DTWm : vert foncé ; MPDist : bleu clair ; tsfresh : bleu marine ; DFT : violet.

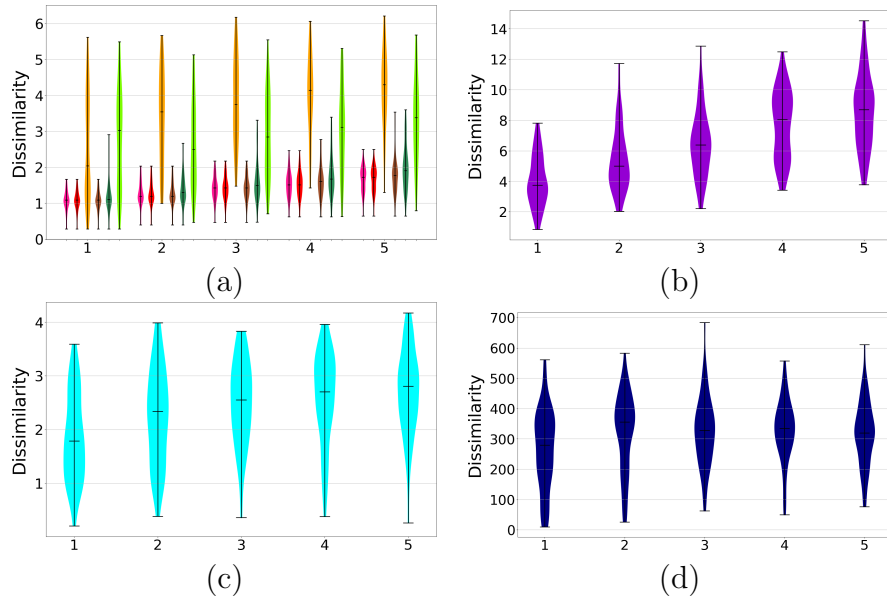


Figure C.4 – Comparaison des 9 dissimilarités sur le jeu PHMC-LAR ($w=10$) et la perturbation *dilat_biunif_e*. En abscisses : degré croissant de perturbation. DTWx : DTWdtai : rose ; DTWc : rouge ; DTWf : orange ; DTWsc : marron ; DTWi : vert clair ; DTWm : vert foncé ; MPDist : bleu clair ; tsfresh : bleu marine ; DFT : violet.

C.2.2 Jeu *inlineskate* de la *UCR archive*

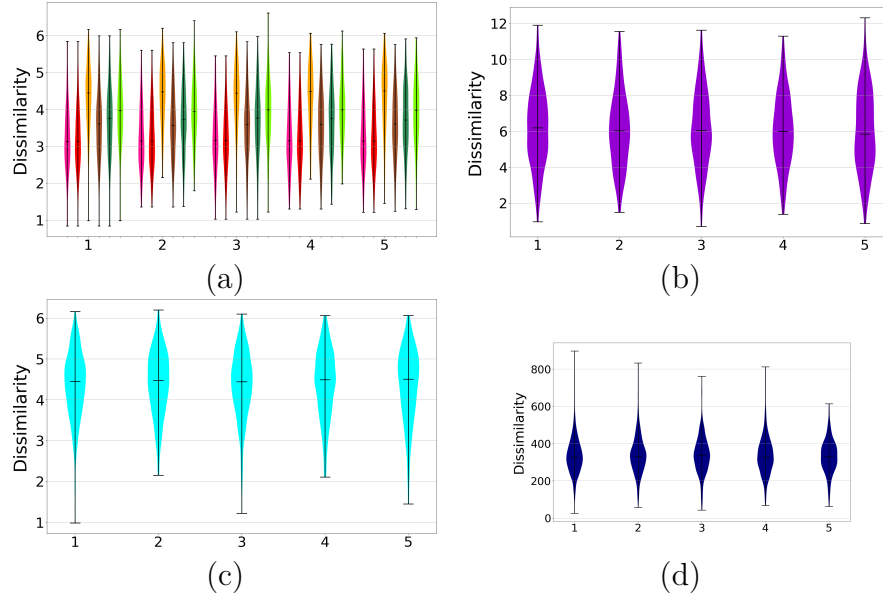


Figure C.5 – Comparaison des 9 dissimilarités sur le jeu UCR inlineskate ($w=10$) et la perturbation *ampl_gauss*. DTWx : DTWdtai : rose ; DTWc : rouge ; DTWf : orange ; DTWsc : marron ; DTWi : vert clair ; DTWm : vert foncé ; MPDist : bleu clair ; tsfresh : bleu marine ; DFT : violet.

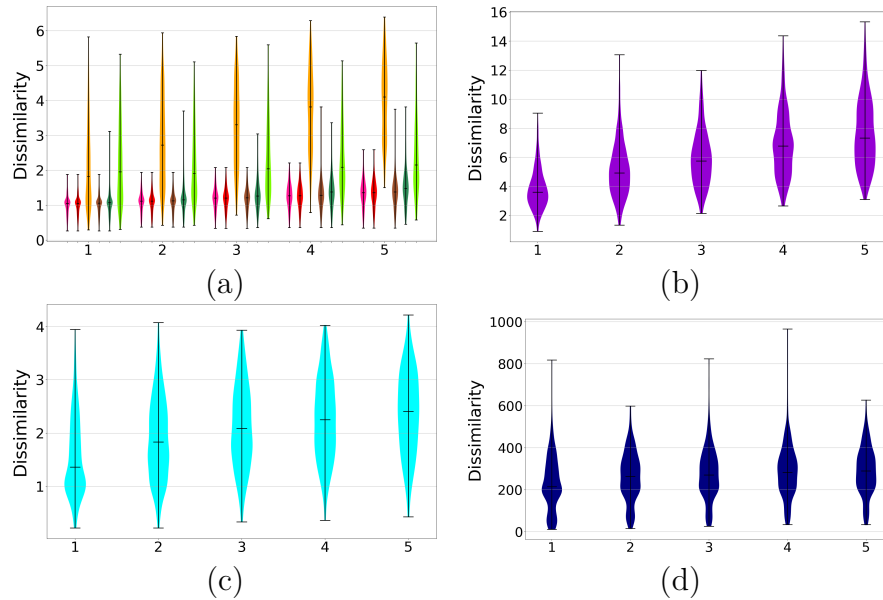


Figure C.6 – Comparaison des 9 dissimilarités sur le jeu UCR inlineskate ($w=10$) et la perturbation *dilat_biunif_e*. DTWx : DTWdtai : rose ; DTWc : rouge ; DTWf : orange ; DTWsc : marron ; DTWi : vert clair ; DTWm : vert foncé ; MPDist : bleu clair ; tsfresh : bleu marine ; DFT : violet.

SIMULATION DU PATIENT VIRTUEL

D.1 Qualité de la simulation

Stratégie GS_BEST

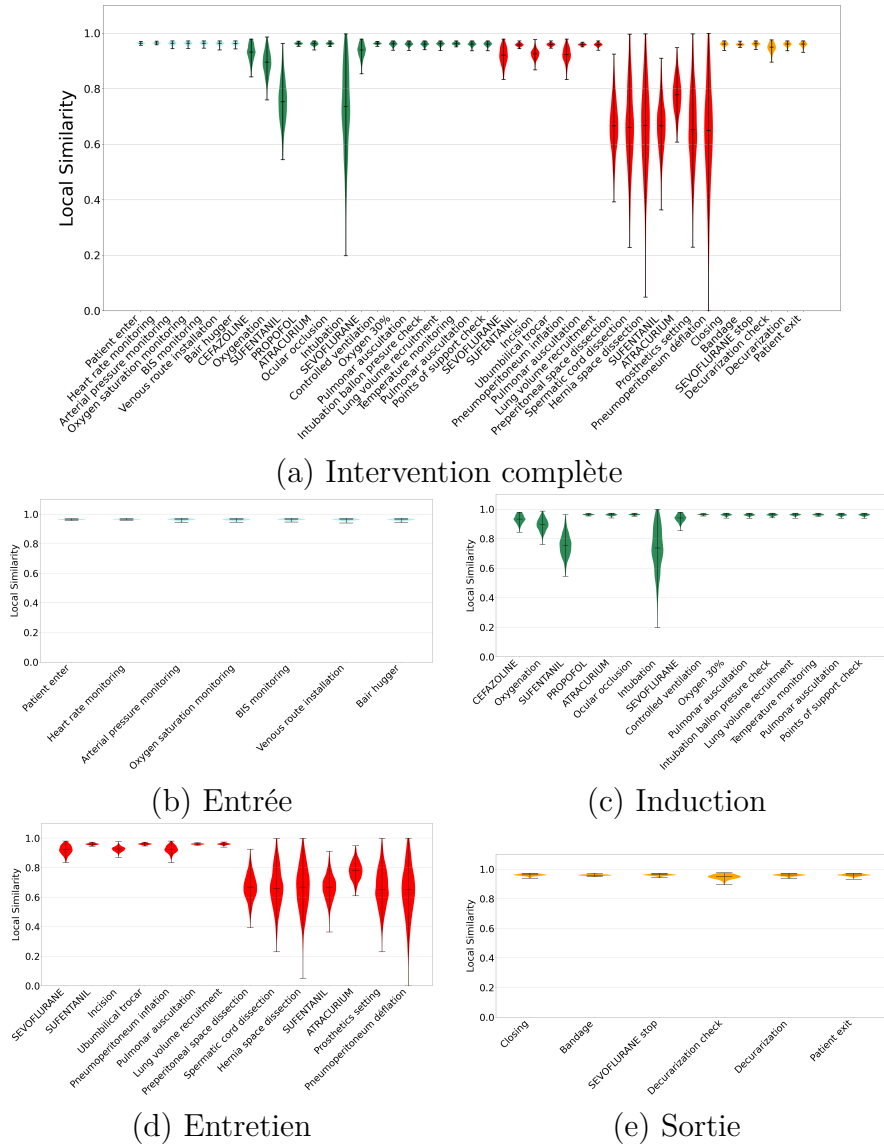


Figure D.1 – Distributions des similarités pour les sous-séries temporelles de chaque action pour la stratégie GS_BEST.

Stratégie GS_AVG_10

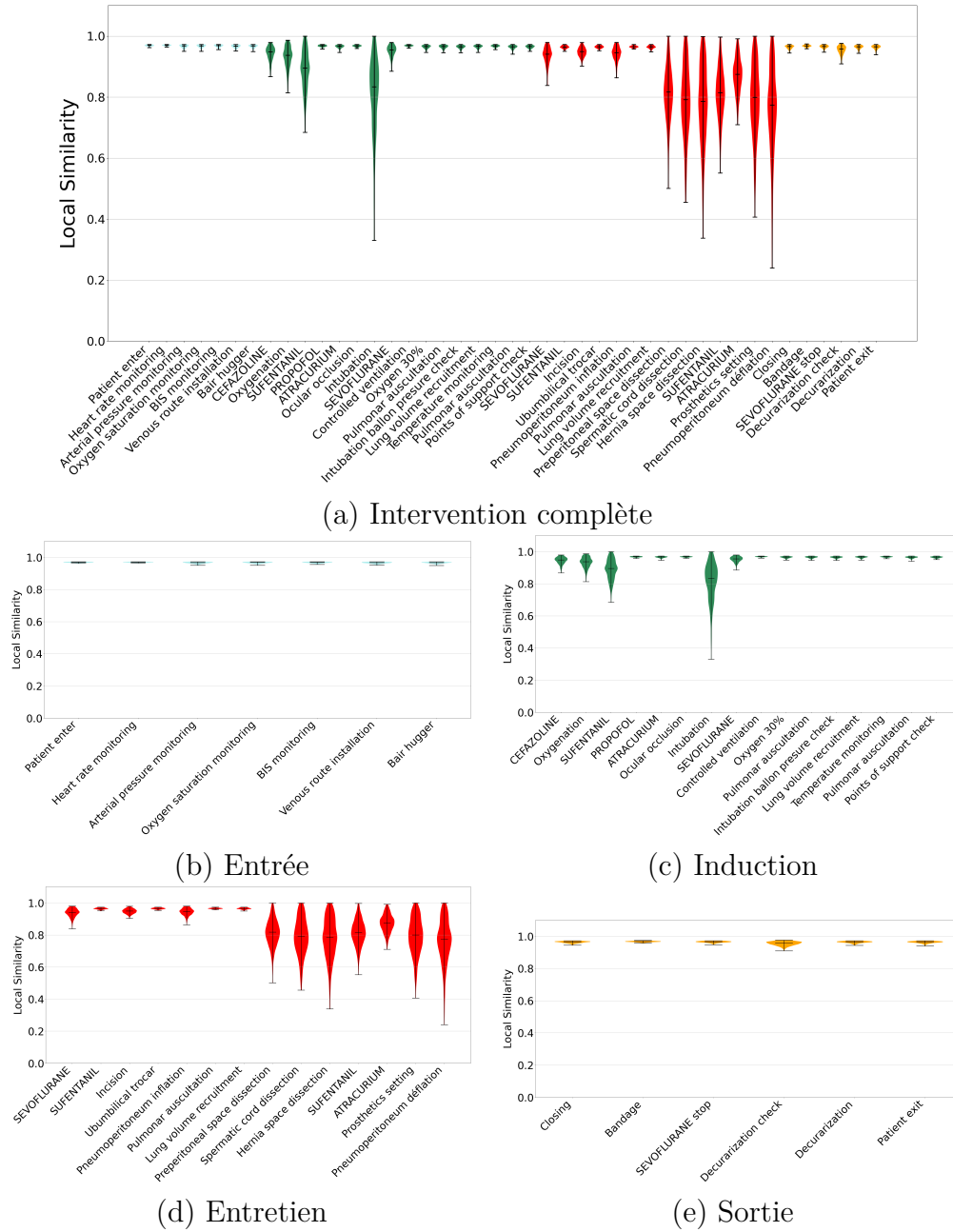


Figure D.2 – Distributions des similarités pour les sous-séries temporelles de chaque action pour la stratégie GS_AVG_10.

Stratégie GS_CONS_10

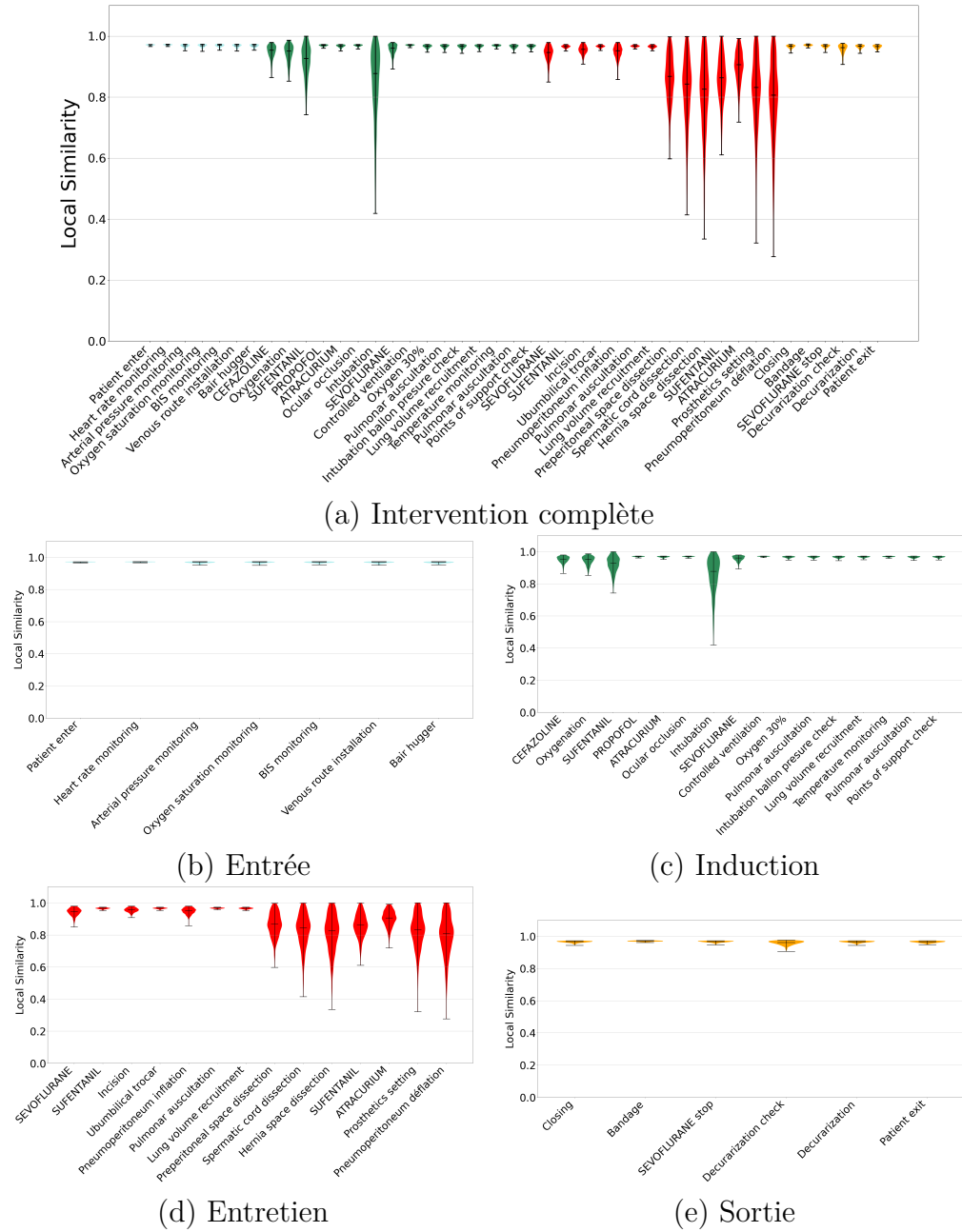


Figure D.3 – Distributions des similarités pour les sous-séries temporelles de chaque action pour la stratégie GS_CONS_10.

PARTITIONNEMENT D'UNE COHORTE DE PROFILS ANESTHÉSQUES PAR CARTES AUTO-ORGANISÉES

E.1 Jeux expérimentaux

Nous avons construit plusieurs jeux de données de base composés de 1000 séquences d'événements. Nous avons fait varier plusieurs paramètres :

- le nombre d'ancres (3 modalités : 3, 5 et 10 ancres),
- le nombre d'événements dans une interancre (3 modalités : 10, 15 et 20 événements),
- la probabilité de présence d'un événement dans une interancre (5 modalités : 2-8%, 24-30%, 47-53%, 70-76% et 92-98%).

Nous avons cherché à constituer des jeux permettant de mesurer la capacité des cartes auto-organisées à partitionner des jeux dont les séquences d'événements sont très proches et dans des proportions différentes.

Ainsi, à partir de chacun de ces jeux de base, un tirage aléatoire des séquences a été réalisé pour constituer des 50 jeux expérimentaux composés de 2 à 4 classes dans plusieurs proportions. Une sélection de jeux expérimentaux est présentée en table E.1

Les résultats des partitionnements sont présentés dans les figures E.1 et suivantes. La qualité des clusters formés est donnée dans la table E.2.

Jeux	Nombre de séquences	Nombre de classes	Classes	Nombre de séquences
<i>Deux classes très différentes en proportions égales</i>				
A	1000	2	A[2-8%]A[2-8%]A A[92-98%]A[92-98%]A	500 500
<i>Trois classes différentes en proportions égales</i>				
B	1500	3	A[2-8%]A[2-8%]A A[47-53%]A[47-53%]A A[92-98%]A[92-98%]A	500 500 500
<i>Deux classes peu différentes en proportions égales</i>				
C	1000	2	A[24-30%]A[24-30%]A A[92-98%]A[92-98%]A	500 500
<i>Deux classes très peu différentes en proportions non égales</i>				
D	1500	2	A[70-76%]A[70-76%]A[70-76%]A[70-76%]A A[92-98%]A[92-98%]A[92-98%]A[92-98%]A	500 1000
<i>Trois classes très peu différentes en proportions non égales</i>				
E	1500	3	A[47-53%]A[47-53%]A[47-53%]A[47-53%]A A[92-98%]A[70-76%]A[70-76%]A[70-76%]A A[92-98%]A[92-98%]A[92-98%]A[92-98%]A	500 500 1000
<i>Quatre classes différentes en proportions non égales</i>				
E	2500	4	A[24-30%]A[24-30%]A A[2-8%]A[2-8%]A A[70-76%]A[70-76%]A[70-76%]A[70-76%]A A[92-98%]A[92-98%]A[92-98%]A[92-98%]A	500 500 500 1000

Table E.1 – Sélection de 6 jeux expérimentaux utilisés pour le partitionnement d'une cohorte de profils anesthésiques par carte auto-organisée

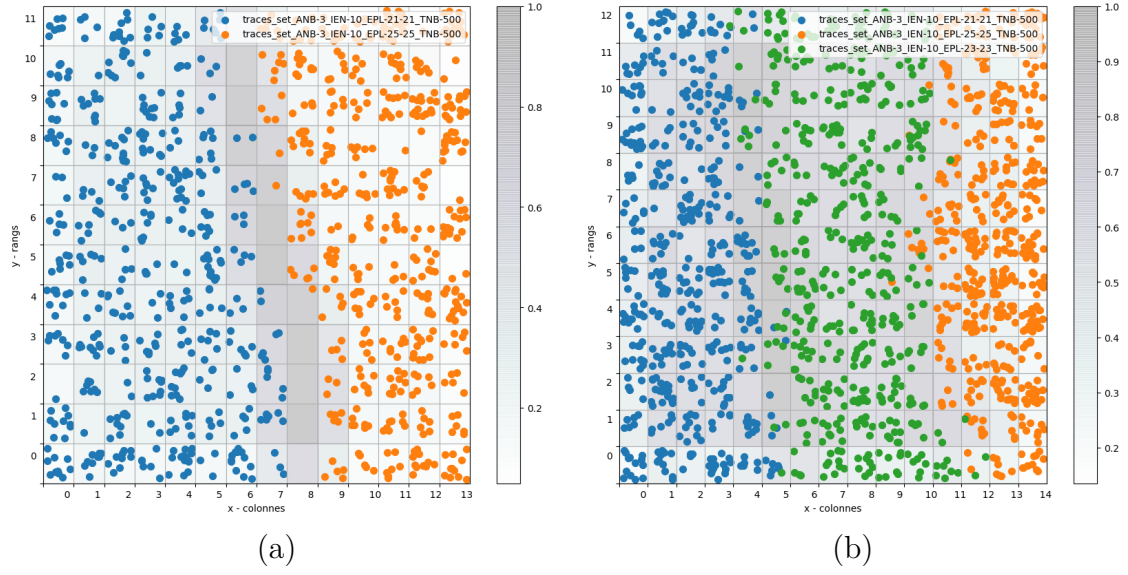


Figure E.1 – Résultat du partitionnement pour les jeux A (a) et B (b), présenté en E.1, et composés respectivement de deux et trois classes. Le fond de la figure présente la U-matrix

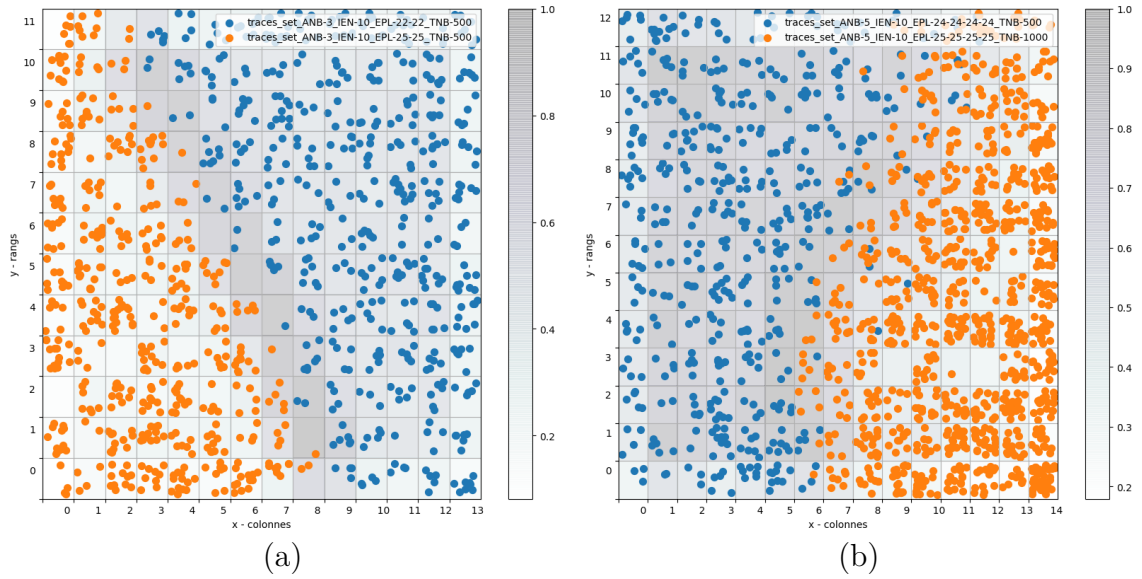


Figure E.2 – Résultat du partitionnement pour les jeux C (a) et D (b), présenté en E.1, et composés de deux classes. Le fond de la figure présente la U-matrix

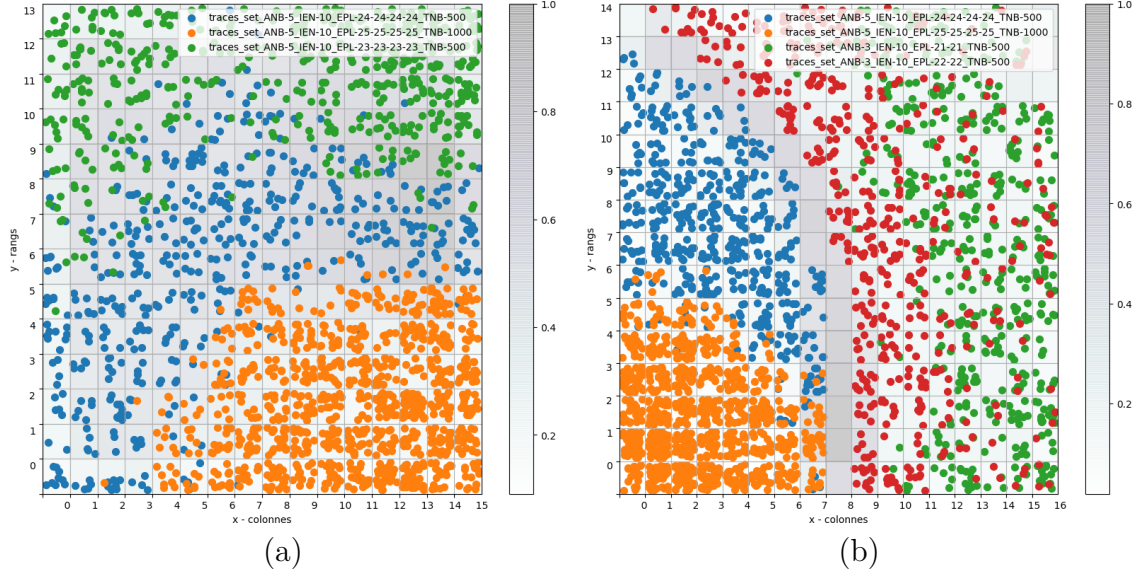


Figure E.3 – Résultat du partitionnement pour les jeux E (a) et F (b), présenté en E.1, et composés de trois et quatre classes. Le fond de la figure présente la U-matrix

Jeux	Classes	Distance intra-cluster	Distance inter-cluster	Indice de Davies-Bouldin
A	2 classes	1.208 1.244 (0.331)	7.354 6.974 (0.758)	0.632 0.723 (0.464)
B	3 classes	1.872 1.954 (0.373)	6.942 7.066 (1.051)	0.735 0.610 (0.420)
C	2 classes	1.398 1.516 (0.401)	6.218 6.288 (0.611)	0.702 0.562 (0.457)
D	2 classes	0.971 0.974 (0.249)	3.690 3.486 (0.755)	0.784 0.623 (0.501)
E	3 classes	1.457 1.528 (0.368)	5.747 5.522 (1.259)	0.785 0.586 (0.547)
F	4 classes	1.753 1.789 (0.504)	13.440 12.502 (2.207)	0.600 0.339 (0.539)

Table E.2 – Évaluation de la qualité des clusters formés à l'issue du partitionnement pour les deux jeux de la table E.1. Les valeurs données sont les moyennes des clusters non vides, les médianes sont données en gras et les écart-types sont donnés entre parenthèse

E.2 Estimation du nombre de cycles d'apprentissage

Pour déterminer le nombre de cycles nécessaire, on cherche à faire converger l'erreur de quantification et l'erreur topologique vers une valeur stable.

Plusieurs expériences, consistant en l'apprentissage de cartes avec des nombres de séquences variant dans 200, 400, 1000 pendant 1000, 10 000 et 100 000 cycles nous ont permis de déterminer le nombre de cycles d'apprentissage nécessaires pour un minimiser l'erreur de quantification et à maintenir l'erreur topologique la plus faible possible.

Comme le montre la figure E.4 l'erreur de quantification diminue rapidement et l'erreur topologique croît progressivement avec des profils similaires pour les 3 valeurs du nombre de cycles. On remarque cependant qu'à partir de 100 000 cycles l'erreur topologique est très proche de l'erreur initiale, correspondant à une carte initialisée avec des valeurs réelles aléatoires, nous avons donc choisi un seuil plus faible.

Pour cela nous avons comparé les partitionnements réalisés par les cartes après 1000 et 10 000 cycles. Comme le montre les figures E.5 et E.6, avec seulement 1000 cycles, bien que le partitionnement semble correct, la carte n'est pas totalement explorée.

Nous avons donc choisi de réaliser 10 000 cycles d'apprentissage pour chacune de nos expérimentations.

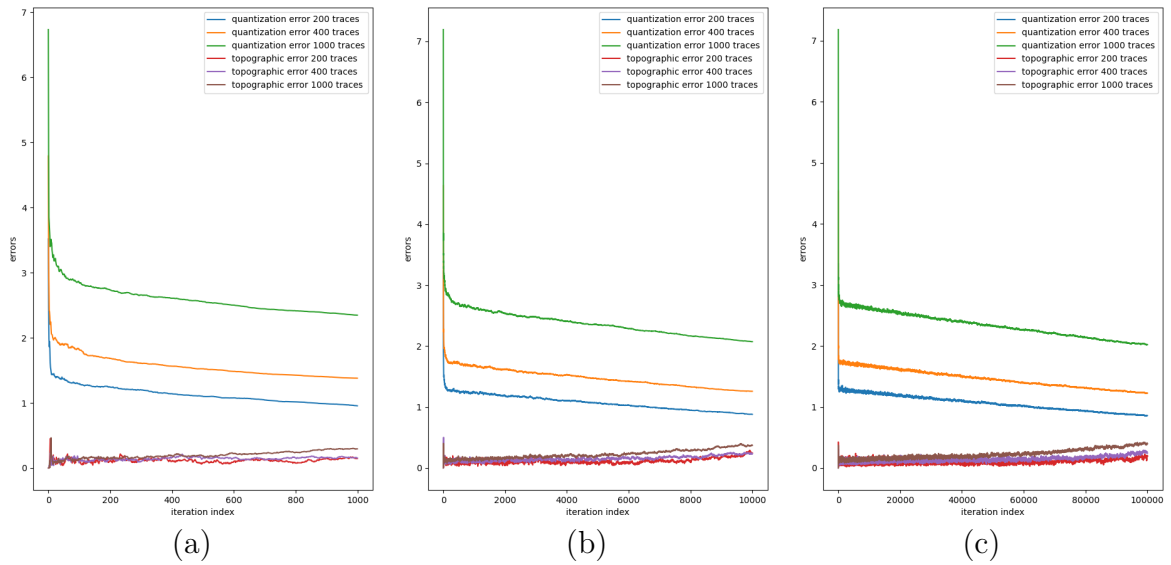


Figure E.4 – Erreur de quantification et erreur topologiques en fonction de 1000 (a), 10 000 (b) et 100 000 (c) cycles pour différents nombre de séquences 200, 400, 1000

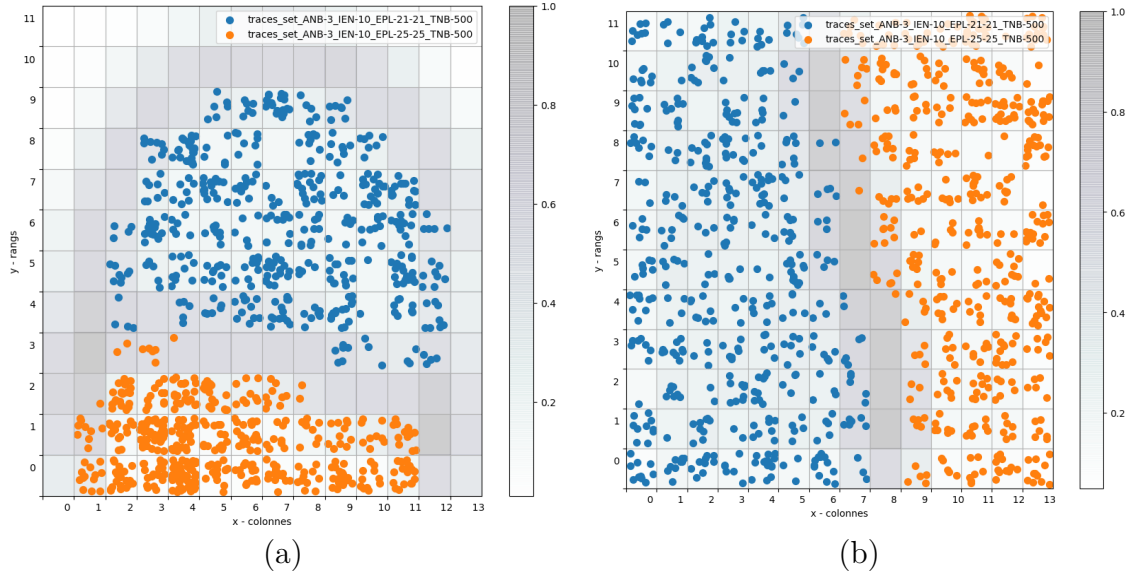


Figure E.5 – Résultat du partitionnement pour le jeu A, présenté en E.1, pour 1000 (a) et 10 000 (b) cycles. Le fond de la figure présente la U-matrix

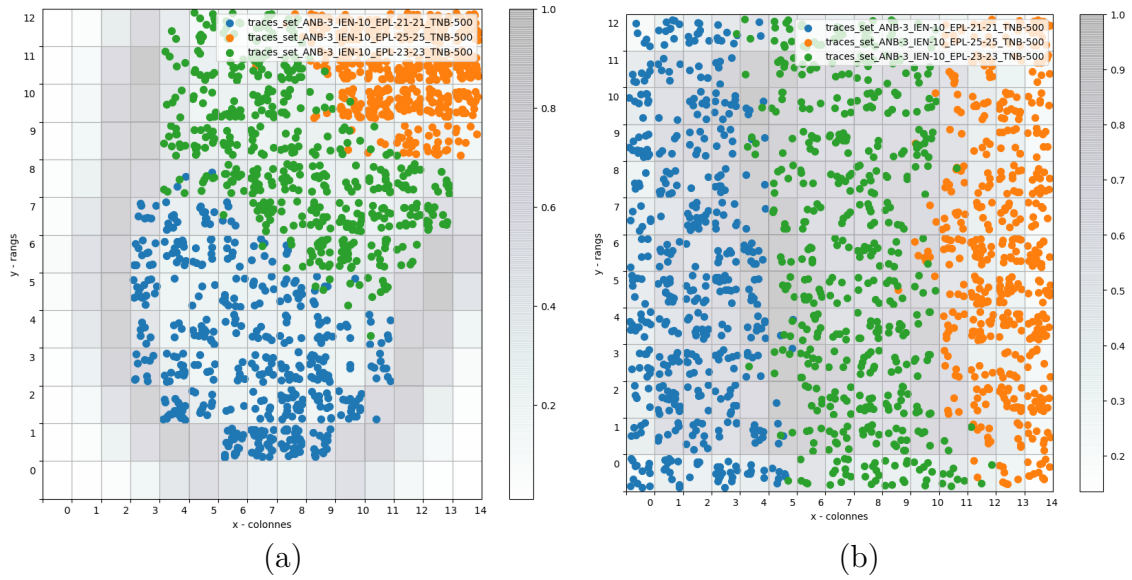


Figure E.6 – Résultat du partitionnement pour le jeu B, présenté en E.1, pour 1000 (a) et 10 000 (b) cycles. Le fond de la figure présente la U-matrix

CRITÈRES ESTIMATIFS DE LA COMPLEXITÉ D'UN GRAPHE

F.1 Corrélations entre critères sur les séquences et critères sur les graphes

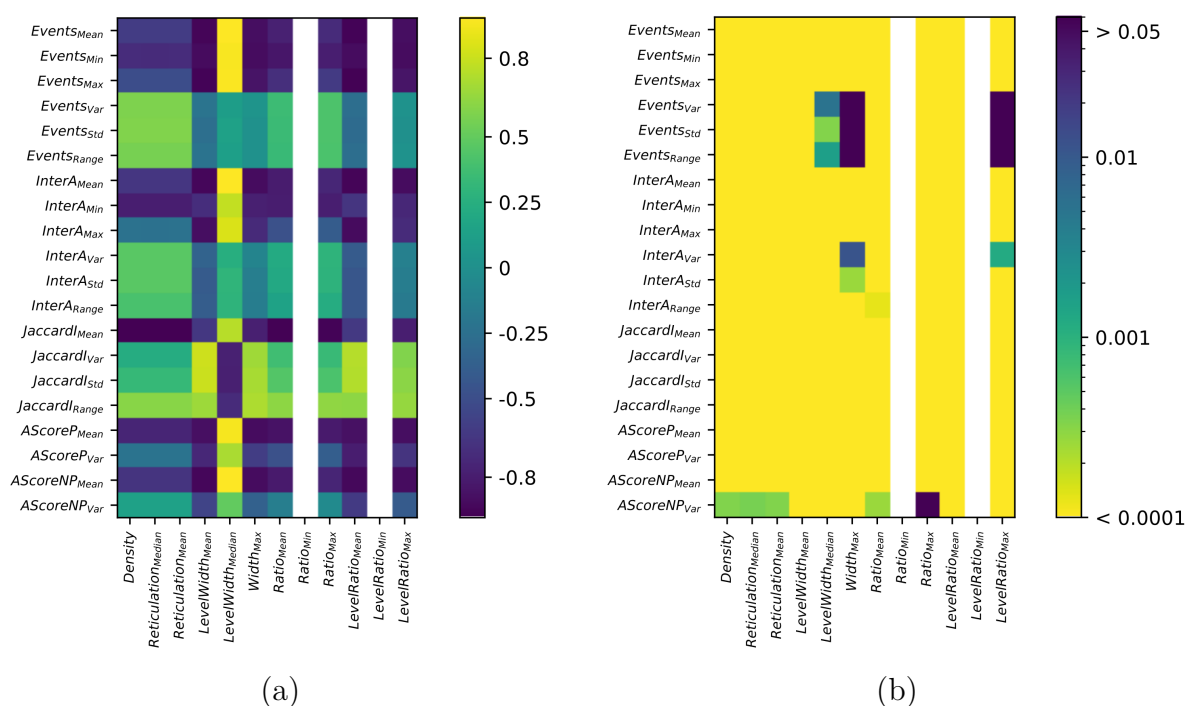


Figure F.1 – Heatmap des coefficients de corrélation (a) et p-value (b) des différents critères pour un jeu de 5 ancres et 10 interancres. Les critères calculés sur le graphe sont présentés en abscisse et les critères calculés sur les séquences sont présentés en ordonnée. Les critères $Ratio_{min}$ et $LevelRatio_{Min}$ se sont révélés constant et les corrélations n'ont pas été calculées

Critères sur les séquences	Critères sur les graphes											
	$Density$	$Reticulation_{Median}$	$Reticulation_{Mean}$	$LevelWidth_{Mean}$	$LevelWidth_{Median}$	$Width_{Max}$	$Ratio_{Mean}$	$Ratio_{Min}$	$Ratio_{Max}$	$LevelRatio_{Mean}$	$LevelRatio_{Min}$	$LevelRatio_{Max}$
$Events_{Mean}$	-0.617 (1e-74)	-0.618 (4e-75)	-0.617 (1e-74)	-0.931 (1e-308)	0.953 (0.0)	-0.886 (5e-236)	-0.780 (1e-144)	-	-0.719 (7e-113)	-0.932 (1e-311)	-	-0.883 (3e-232)
$Events_{Min}$	-0.715 (6e-111)	-0.716 (2e-111)	-0.715 (6e-111)	-0.896 (1e-249)	0.938 (0.0)	-0.892 (3e-243)	-0.841 (1e-188)	-	-0.792 (7e-152)	-0.892 (6e-243)	-	-0.889 (1e-239)
$Events_{Max}$	-0.504 (2e-46)	-0.505 (1e-46)	-0.504 (2e-46)	-0.935 (6e-318)	0.939 (0.0)	-0.852 (1e-198)	-0.698 (1e-103)	-	-0.627 (5e-78)	-0.942 (0.0)	-	-0.848 (4e-195)
$Events_{Var}$	0.585 (1e-65)	0.584 (1e-65)	0.585 (1e-65)	-0.218 (4e-09)	0.105 (0.00)	0.031 (0.40)	0.350 (1e-21)	-	0.419 (3e-31)	-0.263 (1e-12)	-	0.018 (0.61)
$Events_{Std}$	0.594 (3e-68)	0.593 (9e-68)	0.594 (3e-68)	-0.254 (9e-12)	0.135 (0.00)	0.009 (0.79)	0.349 (1e-21)	-	0.422 (1e-31)	-0.287 (9e-15)	-	0.005 (0.88)
$Events_{Range}$	0.564 (4e-60)	0.562 (1e-59)	0.564 (4e-60)	-0.230 (7e-10)	0.119 (0.00)	0.013 (0.71)	0.335 (7e-20)	-	0.409 (9e-30)	-0.265 (9e-13)	-	0.016 (0.65)
$InterA_{Mean}$	-0.653 (1e-86)	-0.654 (5e-87)	-0.653 (1e-86)	-0.923 (1e-293)	0.952 (0.0)	-0.892 (6e-244)	-0.807 (3e-162)	-	-0.748 (7e-127)	-0.925 (5e-296)	-	-0.892 (6e-243)
$InterA_{Min}$	-0.791 (4e-151)	-0.791 (2e-151)	-0.791 (4e-151)	-0.703 (1e-105)	0.775 (9e-142)	-0.784 (3e-147)	-0.800 (1e-157)	-	-0.794 (4e-153)	-0.659 (1e-88)	-	-0.745 (4e-125)
$InterA_{Max}$	-0.238 (1e-10)	-0.239 (1e-10)	-0.238 (1e-10)	-0.881 (1e-229)	0.843 (1e-190)	-0.726 (6e-116)	-0.478 (2e-41)	-	-0.400 (2e-28)	-0.895 (1e-247)	-	-0.721 (1e-113)
$InterA_{Var}$	0.466 (4e-39)	0.464 (7e-39)	0.466 (4e-39)	-0.346 (3e-21)	0.239 (1e-10)	-0.095 (0.01)	0.209 (2e-08)	-	0.284 (1e-14)	-0.405 (3e-29)	-	-0.122 (0.00)
$InterA_{Std}$	0.465 (5e-39)	0.464 (9e-39)	0.465 (5e-39)	-0.403 (7e-29)	0.292 (3e-15)	-0.137 (0.00)	0.200 (8e-08)	-	0.277 (7e-14)	-0.445 (1e-35)	-	-0.149 (7e-05)
$InterA_{Range}$	0.404 (5e-29)	0.403 (9e-29)	0.404 (5e-29)	-0.389 (9e-27)	0.285 (1e-14)	-0.147 (9e-05)	0.144 (0.00)	-	0.225 (1e-09)	-0.442 (6e-35)	-	-0.175 (3e-06)
$JaccardI_{Mean}$	-0.954 (0.0)	-0.954 (0.0)	-0.954 (0.0)	-0.643 (3e-83)	0.745 (5e-125)	-0.786 (2e-148)	-0.946 (0.0)	-	-0.932 (3e-311)	-0.638 (2e-81)	-	-0.797 (1e-155)
$JaccardI_{Var}$	0.228 (1e-09)	0.230 (6e-10)	0.228 (1e-09)	0.808 (4e-163)	-0.778 (2e-143)	0.679 (4e-96)	0.371 (2e-24)	-	0.340 (1e-20)	0.739 (2e-122)	-	0.594 (3e-68)
$JaccardI_{Std}$	0.324 (1e-18)	0.326 (7e-19)	0.324 (1e-18)	0.800 (3e-157)	-0.788 (2e-149)	0.706 (5e-107)	0.439 (1e-34)	-	0.415 (1e-30)	0.730 (6e-118)	-	0.621 (5e-76)
$JaccardI_{Range}$	0.613 (9e-74)	0.615 (3e-74)	0.613 (9e-74)	0.671 (5e-93)	-0.723 (2e-114)	0.717 (6e-112)	0.629 (2e-78)	-	0.633 (9e-80)	0.632 (1e-79)	-	0.650 (1e-85)
$AScoreP_{Mean}$	-0.757 (2e-131)	-0.758 (5e-132)	-0.757 (2e-131)	-0.877 (8e-225)	0.927 (7e-300)	-0.894 (9e-246)	-0.863 (6e-210)	-	-0.819 (7e-171)	-0.866 (2e-212)	-	-0.886 (6e-236)
$AScoreP_{Var}$	-0.226 (1e-09)	-0.227 (1e-09)	-0.226 (1e-09)	-0.740 (1e-122)	0.710 (1e-108)	-0.606 (1e-71)	-0.482 (3e-42)	-	-0.385 (3e-26)	-0.801 (2e-158)	-	-0.666 (5e-91)
$AScoreNP_{Mean}$	-0.668 (9e-92)	-0.669 (3e-92)	-0.668 (9e-92)	-0.919 (1e-284)	0.950 (0.0)	-0.894 (1e-246)	-0.813 (1e-166)	-	-0.757 (1e-131)	-0.916 (2e-279)	-	-0.890 (1e-240)
$AScoreNP_{Var}$	0.135 (0.00)	0.134 (0.00)	0.135 (0.00)	-0.566 (9e-61)	0.494 (1e-44)	-0.365 (1e-23)	-0.137 (0.00)	-	-0.043 (0.25)	-0.629 (1e-78)	-	-0.409 (9e-30)

Table F.1 – Coefficients de corrélation et p -value (entre parenthèses) calculés pour les différents critères sur les séquences et le graphe, pour un ensemble de séquences composées de 5 ancrés et 10 événements au maximum dans une interancré

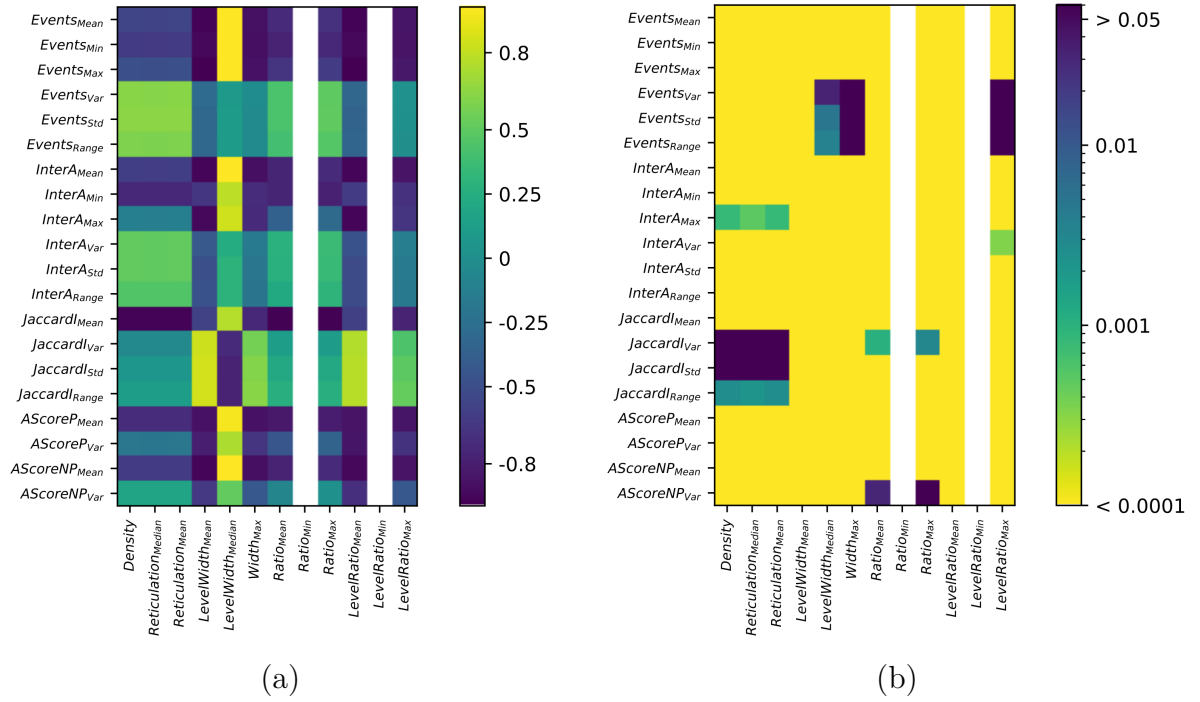


Figure F.2 – *Heatmap* des coefficients de corrélation (a) et p-value (b) des différents critères pour un jeu de 10 ancres et 10 interancres. Les critères calculés sur le graphe sont présentés en abscisse et les critères calculés sur les séquences sont présentés en ordonnée. Les critères $Ratio_{min}$ et $LevelRatio_{Min}$ se sont révélés constant et les corrélations n'ont pas été calculées

Critères sur les séquences	Critères sur les graphes											
	$Density$	$Reticulation_{Median}$	$Reticulation_{Mean}$	$LevelWidth_{Mean}$	$LevelWidth_{Median}$	$Width_{Max}$	$Ratio_{Mean}$	$Ratio_{Min}$	$Ratio_{Max}$	$LevelRatio_{Mean}$	$LevelRatio_{Min}$	$LevelRatio_{Max}$
$Events_{Mean}$	-0.561 (2e-59)	-0.564 (4e-60)	-0.561 (2e-59)	-0.945 (0.0)	0.977 (0.0)	-0.879 (8e-227)	-0.728 (9e-117)	-	-0.686 (1e-98)	-0.947 (0.0)	-	-0.853 (5e-200)
$Events_{Min}$	-0.628 (3e-78)	-0.630 (6e-79)	-0.628 (3e-78)	-0.919 (1e-285)	0.972 (0.0)	-0.881 (1e-229)	-0.775 (2e-141)	-	-0.740 (1e-122)	-0.919 (4e-285)	-	-0.859 (2e-205)
$Events_{Max}$	-0.490 (1e-43)	-0.493 (2e-44)	-0.490 (1e-43)	-0.959 (0.0)	0.971 (0.0)	-0.867 (1e-213)	-0.675 (2e-94)	-	-0.626 (1e-77)	-0.963 (0.0)	-	-0.838 (4e-186)
$Events_{Var}$	0.631 (4e-79)	0.626 (1e-77)	0.631 (4e-79)	-0.282 (2e-14)	0.080 (0.03)	-0.016 (0.65)	0.425 (3e-32)	-	0.497 (4e-45)	-0.320 (3e-18)	-	0.017 (0.65)
$Events_{Std}$	0.643 (3e-83)	0.639 (1e-81)	0.643 (3e-83)	-0.316 (9e-18)	0.106 (0.00)	-0.032 (0.38)	0.430 (6e-33)	-	0.500 (1e-45)	-0.342 (1e-20)	-	0.011 (0.75)
$Events_{Range}$	0.599 (1e-69)	0.594 (4e-68)	0.599 (1e-69)	-0.307 (8e-17)	0.110 (0.00)	-0.037 (0.31)	0.394 (1e-27)	-	0.468 (1e-39)	-0.328 (4e-19)	-	-0.001 (0.96)
$InterA_{Mean}$	-0.605 (3e-71)	-0.607 (5e-72)	-0.605 (3e-71)	-0.931 (4e-309)	0.976 (0.0)	-0.883 (4e-232)	-0.762 (2e-134)	-	-0.722 (4e-114)	-0.935 (1e-316)	-	-0.863 (4e-210)
$InterA_{Min}$	-0.751 (6e-128)	-0.750 (1e-127)	-0.751 (6e-128)	-0.656 (2e-87)	0.780 (1e-144)	-0.718 (3e-112)	-0.766 (3e-136)	-	-0.784 (9e-147)	-0.616 (1e-74)	-	-0.684 (4e-98)
$InterA_{Max}$	-0.126 (0.00)	-0.131 (0.00)	-0.126 (0.00)	-0.920 (5e-287)	0.840 (5e-188)	-0.724 (4e-115)	-0.357 (1e-22)	-	-0.296 (1e-15)	-0.926 (3e-298)	-	-0.666 (6e-91)
$InterA_{Var}$	0.509 (1e-47)	0.503 (2e-46)	0.509 (1e-47)	-0.423 (7e-32)	0.235 (2e-10)	-0.163 (1e-05)	0.276 (9e-14)	-	0.358 (1e-22)	-0.475 (1e-40)	-	-0.135 (0.00)
$InterA_{Std}$	0.512 (3e-48)	0.506 (6e-47)	0.512 (3e-48)	-0.488 (2e-43)	0.292 (2e-15)	-0.203 (5e-08)	0.271 (2e-13)	-	0.348 (2e-21)	-0.520 (8e-50)	-	-0.155 (3e-05)
$InterA_{Range}$	0.453 (8e-37)	0.447 (8e-36)	0.453 (8e-37)	-0.477 (4e-41)	0.293 (2e-15)	-0.217 (6e-09)	0.216 (7e-09)	-	0.295 (1e-15)	-0.514 (1e-48)	-	-0.180 (1e-06)
$JaccardI_{Mean}$	-0.942 (0.0)	-0.942 (0.0)	-0.942 (0.0)	-0.588 (1e-66)	0.763 (1e-134)	-0.742 (1e-123)	-0.953 (0.0)	-	-0.950 (0.0)	-0.596 (8e-69)	-	-0.776 (4e-142)
$JaccardI_{Var}$	-0.051 (0.17)	-0.048 (0.20)	-0.051 (0.17)	0.832 (4e-181)	-0.725 (1e-115)	0.569 (2e-61)	0.123 (0.00)	-	0.111 (0.00)	0.754 (1e-129)	-	0.437 (4e-34)
$JaccardI_{Std}$	0.052 (0.16)	0.055 (0.14)	0.052 (0.16)	0.845 (8e-193)	-0.766 (1e-136)	0.614 (9e-74)	0.209 (2e-08)	-	0.205 (4e-08)	0.766 (2e-136)	-	0.485 (9e-43)
$JaccardI_{Range}$	0.113 (0.00)	0.115 (0.00)	0.113 (0.00)	0.842 (1e-189)	-0.781 (4e-145)	0.631 (3e-79)	0.260 (2e-12)	-	0.257 (4e-12)	0.770 (1e-138)	-	0.516 (5e-49)
$AScoreP_{Mean}$	-0.715 (9e-111)	-0.717 (1e-111)	-0.715 (9e-111)	-0.869 (2e-216)	0.952 (0.0)	-0.870 (7e-217)	-0.829 (7e-179)	-	-0.805 (4e-161)	-0.862 (1e-208)	-	-0.854 (2e-200)
$AScoreP_{Var}$	-0.192 (2e-07)	-0.198 (1e-07)	-0.192 (2e-07)	-0.798 (9e-156)	0.741 (4e-123)	-0.661 (1e-89)	-0.440 (1e-34)	-	-0.344 (5e-21)	-0.856 (2e-202)	-	-0.679 (6e-96)
$AScoreNP_{Mean}$	-0.616 (1e-74)	-0.619 (2e-75)	-0.616 (1e-74)	-0.926 (3e-298)	0.975 (0.0)	-0.881 (3e-230)	-0.767 (8e-137)	-	-0.731 (4e-118)	-0.925 (4e-296)	-	-0.859 (1e-205)
$AScoreNP_{Var}$	0.173 (4e-06)	0.167 (8e-06)	0.173 (4e-06)	-0.641 (2e-82)	0.511 (7e-48)	-0.430 (5e-33)	-0.082 (0.02)	-	0.013 (0.71)	-0.698 (9e-104)	-	-0.428 (1e-32)

Table F.2 – Coefficients de corrélation et p -value (entre parenthèses) calculés pour les différents critères sur les séquences et le graphe, pour un ensemble de séquences composées de 10 ancres et 10 événements au maximum dans une interancree

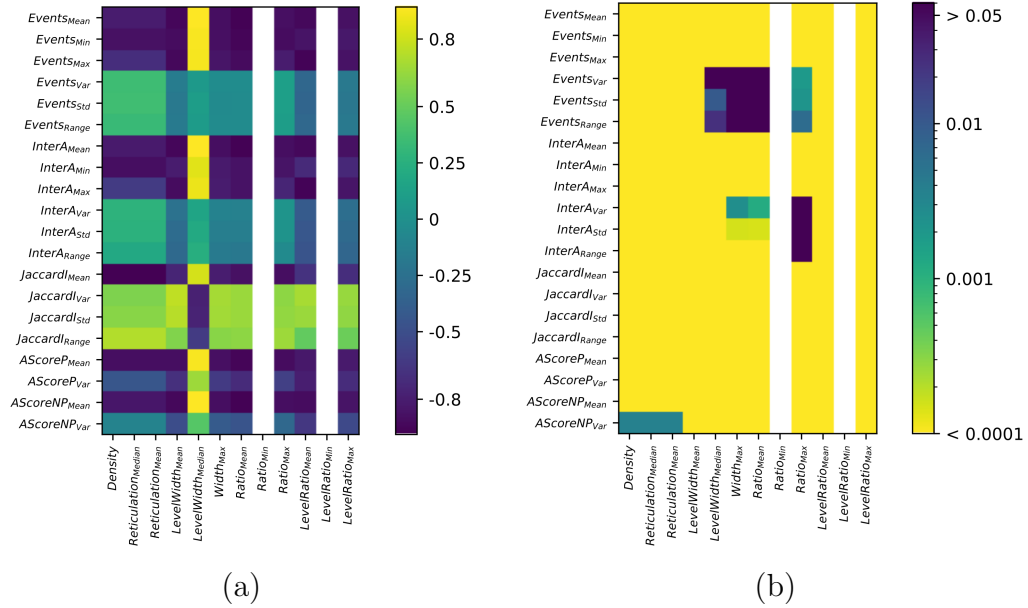


Figure F.3 – *Heatmap* des coefficients de corrélation (a) et p-value (b) des différents critères pour un jeu de 3 ancrés et 15 interancres. Les critères calculés sur le graphe sont présentés en abscisse et les critères calculés sur les séquences sont présentés en ordonnée. Les critères $Ratio_{min}$ et $LevelRatio_{Min}$ se sont révélés constant et les corrélations n'ont pas été calculées

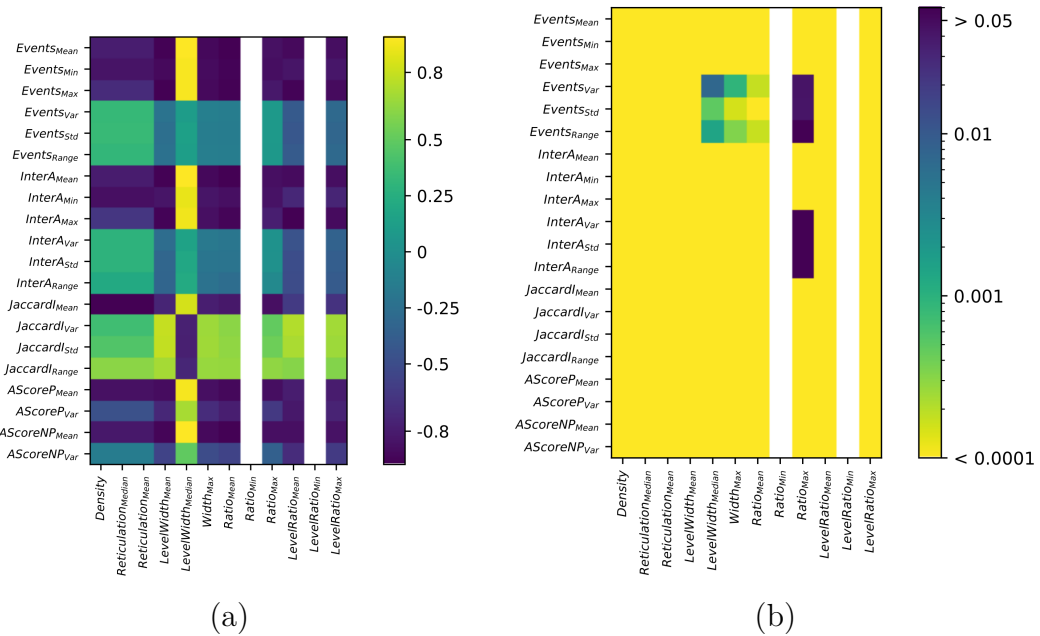


Figure F.4 – *Heatmap* des coefficients de corrélation (a) et p-value (b) des différents critères pour un jeu de 3 ancrés et 20 interancres. Les critères calculés sur le graphe sont présentés en abscisse et les critères calculés sur les séquences sont présentés en ordonnée. Les critères $Ratio_{min}$ et $LevelRatio_{Min}$ se sont révélés constant et les corrélations n'ont pas été calculées

Critères sur les séquences	Critères sur les graphes											
	$Density$	$Reticulation_{Median}$	$Reticulation_{Mean}$	$LevelWidth_{Mean}$	$LevelWidth_{Median}$	$Width_{Max}$	$Ratio_{Mean}$	$Ratio_{Min}$	$Ratio_{Max}$	$LevelRatio_{Mean}$	$LevelRatio_{Min}$	$LevelRatio_{Max}$
$Events_{Mean}$	-0.810 (1e-164)	-0.810 (1e-164)	-0.810 (1e-164)	-0.925 (1e-296)	0.942 (0.0)	-0.886 (6e-236)	-0.941 (0.0)	-	-0.876 (3e-224)	-0.916 (1e-280)	-	-0.871 (5e-218)
$Events_{Min}$	-0.873 (5e-220)	-0.873 (5e-220)	-0.873 (5e-220)	-0.892 (1e-243)	0.926 (2e-299)	-0.878 (1e-226)	-0.933 (1e-313)	-	-0.895 (1e-247)	-0.856 (2e-202)	-	-0.834 (4e-183)
$Events_{Max}$	-0.705 (2e-106)	-0.705 (2e-106)	-0.705 (2e-106)	-0.919 (9e-285)	0.916 (2e-279)	-0.852 (1e-198)	-0.904 (1e-260)	-	-0.813 (9e-167)	-0.937 (6e-323)	-	-0.866 (3e-213)
$Events_{Var}$	0.354 (3e-22)	0.354 (3e-22)	0.354 (3e-22)	-0.162 (1e-05)	0.069 (0.06)	-0.032 (0.39)	-0.030 (0.41)	-	0.117 (0.00)	-0.323 (1e-18)	-	-0.189 (4e-07)
$Events_{Std}$	0.358 (1e-22)	0.358 (1e-22)	0.358 (1e-22)	-0.188 (5e-07)	0.097 (0.00)	-0.045 (0.22)	-0.038 (0.31)	-	0.115 (0.00)	-0.340 (1e-20)	-	-0.202 (6e-08)
$Events_{Range}$	0.330 (2e-19)	0.330 (2e-19)	0.330 (2e-19)	-0.178 (2e-06)	0.084 (0.02)	-0.037 (0.32)	-0.037 (0.31)	-	0.103 (0.00)	-0.316 (8e-18)	-	-0.185 (7e-07)
$InterA_{Mean}$	-0.821 (2e-172)	-0.821 (2e-172)	-0.821 (2e-172)	-0.924 (1e-293)	0.942 (0.0)	-0.888 (3e-238)	-0.944 (0.0)	-	-0.882 (4e-231)	-0.912 (6e-274)	-	-0.870 (9e-217)
$InterA_{Min}$	-0.887 (2e-237)	-0.887 (2e-237)	-0.887 (2e-237)	-0.809 (2e-163)	0.859 (8e-206)	-0.817 (2e-169)	-0.862 (3e-209)	-	-0.857 (6e-204)	-0.730 (8e-118)	-	-0.737 (6e-121)
$InterA_{Max}$	-0.622 (1e-76)	-0.622 (1e-76)	-0.622 (1e-76)	-0.894 (1e-245)	0.885 (1e-234)	-0.812 (8e-166)	-0.865 (1e-211)	-	-0.756 (4e-131)	-0.936 (1e-319)	-	-0.847 (2e-194)
$InterA_{Var}$	0.270 (3e-13)	0.270 (3e-13)	0.270 (3e-13)	-0.244 (5e-11)	0.150 (6e-05)	-0.113 (0.00)	-0.122 (0.00)	-	0.029 (0.43)	-0.408 (1e-29)	-	-0.273 (1e-13)
$InterA_{Std}$	0.263 (1e-12)	0.263 (1e-12)	0.263 (1e-12)	-0.287 (8e-15)	0.196 (1e-07)	-0.142 (0.00)	-0.143 (0.00)	-	0.015 (0.68)	-0.441 (1e-34)	-	-0.300 (4e-16)
$InterA_{Range}$	0.187 (5e-07)	0.187 (5e-07)	0.187 (5e-07)	-0.306 (1e-16)	0.230 (6e-10)	-0.176 (2e-06)	-0.197 (1e-07)	-	-0.045 (0.23)	-0.464 (8e-39)	-	-0.327 (6e-19)
$JaccardI_{Mean}$	-0.956 (0.0)	-0.956 (0.0)	-0.956 (0.0)	-0.759 (1e-132)	0.823 (6e-174)	-0.804 (2e-160)	-0.869 (1e-215)	-	-0.888 (7e-238)	-0.676 (9e-95)	-	-0.708 (1e-107)
$JaccardI_{Var}$	0.573 (1e-62)	0.573 (1e-62)	0.573 (1e-62)	0.761 (1e-133)	-0.783 (3e-146)	0.688 (1e-99)	0.660 (6e-89)	-	0.622 (3e-76)	0.693 (1e-101)	-	0.646 (3e-84)
$JaccardI_{Std}$	0.607 (1e-71)	0.607 (1e-71)	0.607 (1e-71)	0.737 (7e-121)	-0.766 (2e-136)	0.681 (1e-96)	0.655 (2e-87)	-	0.633 (1e-79)	0.659 (1e-88)	-	0.625 (2e-77)
$JaccardI_{Range}$	0.723 (1e-114)	0.723 (1e-114)	0.723 (1e-114)	0.584 (2e-65)	-0.629 (1e-78)	0.607 (8e-72)	0.618 (3e-75)	-	0.660 (4e-89)	0.489 (2e-43)	-	0.518 (1e-49)
$AScoreP_{Mean}$	-0.886 (2e-235)	-0.886 (2e-235)	-0.886 (2e-235)	-0.885 (6e-234)	0.923 (2e-292)	-0.875 (1e-222)	-0.928 (3e-302)	-	-0.897 (7e-250)	-0.837 (2e-185)	-	-0.822 (6e-173)
$AScoreP_{Var}$	-0.447 (8e-36)	-0.447 (8e-36)	-0.447 (8e-36)	-0.694 (4e-102)	0.661 (2e-89)	-0.635 (2e-80)	-0.721 (1e-113)	-	-0.599 (2e-69)	-0.797 (1e-155)	-	-0.714 (1e-110)
$AScoreNP_{Mean}$	-0.840 (2e-188)	-0.840 (2e-188)	-0.840 (2e-188)	-0.917 (2e-281)	0.941 (0.0)	-0.887 (1e-237)	-0.942 (0.0)	-	-0.888 (1e-238)	-0.895 (1e-247)	-	-0.859 (5e-206)
$AScoreNP_{Var}$	-0.109 (0.00)	-0.109 (0.00)	-0.109 (0.00)	-0.507 (4-47)	0.442 (7-35)	-0.409 (1-29)	-0.463 (1-38)	-	-0.317 (7-18)	-0.650 (2-85)	-	-0.535 (4-53)

Table F.3 – Coefficients de corrélation et p -value (entre parenthèses) calculés pour les différents critères sur les séquences et le graphe, pour un ensemble de séquences composées de 3 ancrés et 15 événements au maximum dans une interancré

Critères sur les séquences	Critères sur les graphes											
	$Density$	$Reticulation_{Median}$	$Reticulation_{Mean}$	$LevelWidth_{Mean}$	$LevelWidth_{Median}$	$Width_{Max}$	$Ratio_{Mean}$	$Ratio_{Min}$	$Ratio_{Max}$	$LevelRatio_{Mean}$	$LevelRatio_{Min}$	$LevelRatio_{Max}$
$Events_{Mean}$	-0.793 (2e-152)	-0.793 (2e-152)	-0.793 (2e-152)	-0.936 (4e-319)	0.957 (0.0)	-0.906 (3e-263)	-0.947 (0.0)	-	-0.863 (3e-210)	-0.897 (1e-250)	-	-0.877 (7e-225)
$Events_{Min}$	-0.847 (1e-194)	-0.847 (1e-194)	-0.847 (1e-194)	-0.902 (2e-258)	0.942 (0.0)	-0.889 (2e-239)	-0.929 (8e-305)	-	-0.880 (3e-228)	-0.838 (6e-186)	-	-0.833 (3e-182)
$Events_{Max}$	-0.715 (1e-110)	-0.715 (1e-110)	-0.715 (1e-110)	-0.942 (0.0)	0.942 (0.0)	-0.896 (4e-249)	-0.937 (1e-321)	-	-0.822 (1e-173)	-0.930 (2e-307)	-	-0.895 (4e-247)
$Events_{Var}$	0.327 (6e-19)	0.327 (6e-19)	0.327 (6e-19)	-0.221 (2e-09)	0.101 (0.00)	-0.124 (0.00)	-0.141 (0.00)	-	0.076 (0.04)	-0.405 (3e-29)	-	-0.296 (1e-15)
$Events_{Std}$	0.335 (7e-20)	0.335 (7e-20)	0.335 (7e-20)	-0.254 (8e-12)	0.131 (0.00)	-0.142 (0.00)	-0.151 (5e-05)	-	0.076 (0.04)	-0.429 (7e-33)	-	-0.317 (7e-18)
$Events_{Range}$	0.313 (1e-17)	0.313 (1e-17)	0.313 (1e-17)	-0.237 (1e-10)	0.120 (0.00)	-0.135 (0.00)	-0.141 (0.00)	-	0.070 (0.06)	-0.399 (3e-28)	-	-0.299 (5e-16)
$InterA_{Mean}$	-0.802 (1e-158)	-0.802 (1e-158)	-0.802 (1e-158)	-0.933 (3e-314)	0.957 (0.0)	-0.906 (7e-264)	-0.948 (0.0)	-	-0.868 (1e-215)	-0.893 (7e-245)	-	-0.875 (4e-222)
$InterA_{Min}$	-0.872 (1e-219)	-0.872 (1e-219)	-0.872 (1e-219)	-0.837 (1e-185)	0.897 (2e-250)	-0.839 (2e-187)	-0.869 (4e-216)	-	-0.858 (1e-204)	-0.738 (2e-121)	-	-0.751 (3e-128)
$InterA_{Max}$	-0.648 (7e-85)	-0.648 (7e-85)	-0.648 (7e-85)	-0.930 (1e-305)	0.918 (9e-283)	-0.873 (1e-220)	-0.920 (1e-286)	-	-0.786 (3e-148)	-0.944 (0.0)	-	-0.895 (5e-248)
$InterA_{Var}$	0.284 (1e-14)	0.284 (1e-14)	0.284 (1e-14)	-0.269 (4e-13)	0.149 (6e-05)	-0.174 (3e-06)	-0.197 (1e-07)	-	0.022 (0.56)	-0.457 (1e-37)	-	-0.349 (1e-21)
$InterA_{Std}$	0.283 (2e-14)	0.283 (2e-14)	0.283 (2e-14)	-0.317 (7e-18)	0.195 (1e-07)	-0.206 (3e-08)	-0.218 (5e-09)	-	0.013 (0.71)	-0.494 (2e-44)	-	-0.381 (1e-25)
$InterA_{Range}$	0.207 (3e-08)	0.207 (3e-08)	0.207 (3e-08)	-0.329 (3e-19)	0.218 (4e-09)	-0.229 (7e-10)	-0.264 (1e-12)	-	-0.051 (0.17)	-0.505 (1e-46)	-	-0.401 (1e-28)
$JaccardI_{Mean}$	-0.941 (0.0)	-0.941 (0.0)	-0.941 (0.0)	-0.749 (3e-127)	0.832 (1e-181)	-0.785 (1e-147)	-0.828 (4e-178)	-	-0.870 (1e-216)	-0.626 (1e-77)	-	-0.666 (5e-91)
$JaccardI_{Var}$	0.376 (5e-25)	0.376 (5e-25)	0.376 (5e-25)	0.790 (8e-151)	-0.775 (1e-141)	0.679 (5e-96)	0.629 (1e-78)	-	0.501 (7e-46)	0.741 (4e-123)	-	0.694 (7e-102)
$JaccardI_{Std}$	0.443 (4e-35)	0.443 (4e-35)	0.443 (4e-35)	0.781 (6e-145)	-0.781 (4e-145)	0.686 (1e-98)	0.638 (1e-81)	-	0.538 (8e-54)	0.712 (1e-109)	-	0.678 (1e-95)
$JaccardI_{Range}$	0.626 (1e-77)	0.626 (1e-77)	0.626 (1e-77)	0.701 (9e-105)	-0.742 (1e-123)	0.666 (5e-91)	0.653 (1e-86)	-	0.641 (2e-82)	0.610 (1e-72)	-	0.607 (7e-72)
$AScoreP_{Mean}$	-0.865 (2e-211)	-0.865 (2e-211)	-0.865 (2e-211)	-0.881 (3e-230)	0.931 (8e-309)	-0.872 (3e-219)	-0.906 (3e-263)	-	-0.874 (8e-222)	-0.796 (1e-154)	-	-0.801 (5e-158)
$AScoreP_{Var}$	-0.478 (2e-41)	-0.478 (2e-41)	-0.478 (2e-41)	-0.739 (6e-122)	0.708 (1e-107)	-0.704 (4e-106)	-0.785 (6e-148)	-	-0.631 (3e-79)	-0.822 (5e-173)	-	-0.762 (5e-134)
$AScoreNP_{Mean}$	-0.822 (4e-173)	-0.822 (4e-173)	-0.822 (4e-173)	-0.923 (3e-293)	0.955 (0.0)	-0.900 (1e-254)	-0.939 (0.0)	-	-0.873 (7e-220)	-0.869 (9e-216)	-	-0.857 (1e-203)
$AScoreNP_{Var}$	-0.150 (6e-05)	-0.150 (6e-05)	-0.150 (6e-05)	-0.568 (4e-61)	0.492 (5e-44)	-0.503 (2e-46)	-0.564 (3e-60)	-	-0.365 (1e-23)	-0.709 (2e-108)	-	-0.620 (7e-76)

Table F.4 – Coefficients de corrélation et p -value (entre parenthèses) calculés pour les différents critères sur les séquences et le graphe, pour un ensemble de séquences composées de 3 ancres et 20 événements au maximum dans une interancree

F.2 Critères calculés sur les clusters constitués après regroupement ascendant hiérarchique pour les jeux expérimentaux

Clusters	Critères sur les séquences			Critères sur le graphe		
	$JaccardI_{Mean}$	$AScoreP_{Mean}$	$AScoreNP_{Mean}$	$Density$	$Reticulation_{Median}$	$Reticulation_{Mean}$
1	0.885	19.711	19.711	0.063	0.5	0.453
2	0.909	20.0	20.0	0.071	0.425	0.425
3	0.905	20.723	20.723	0.075	0.354	0.354
4	0.914	20.105	20.105	0.084	0.208	0.266
5	0.901	19.897	19.897	0.077	0.417	0.364
6	0.847	18.333	18.333	0.069	0.417	0.43
7	0.876	19.462	19.462	0.061	0.5	0.472
8	0.851	18.374	18.374	0.071	0.375	0.388
9	0.897	19.6	19.6	0.071	0.5	0.447
10	0.836	17.813	17.813	0.075	0.35	0.37
11	0.894	19.597	19.597	0.084	0.189	0.287
12	0.842	18.217	18.217	0.077	0.35	0.387
13	0.845	18.257	18.257	0.074	0.35	0.39
14	0.873	18.93	18.93	0.074	0.5	0.439
15	0.847	18.385	18.385	0.071	0.417	0.419
16	0.836	18.121	18.121	0.065	0.5	0.471
17	0.839	18.091	18.091	0.074	0.5	0.437
18	0.835	17.784	17.784	0.08	0.35	0.344
19	0.65	4.933	4.933	0.126	0.266	0.266
20	0.872	19.552	19.552	0.081	0.417	0.37
21	0.763	16.19	16.19	0.071	0.417	0.41
22	0.867	19.448	19.448	0.069	0.45	0.433
23	0.678	4.091	4.091	0.144	0.35	0.383
24	0.667	4.038	4.038	0.124	0.256	0.256
25	0.663	4.822	4.822	0.152	0.281	0.281
26	0.617	4.208	4.208	0.137	0.286	0.286
27	0.581	4.636	4.636	0.124	0.3	0.334
28	0.665	4.066	4.066	0.11	0.262	0.31
29	0.633	4.321	4.321	0.132	0.4	0.377
30	0.601	4.333	4.333	0.143	0.333	0.378
31	0.638	5.19	5.19	0.128	0.35	0.372
32	0.64	4.636	4.636	0.119	0.333	0.349
33	0.648	4.944	4.944	0.121	0.35	0.355
34	0.666	4.075	4.075	0.117	0.241	0.241
35	0.666	4.524	4.524	0.145	0.308	0.308
36	0.669	4.037	4.037	0.101	0.262	0.296
37	0.668	4.045	4.045	0.119	0.3	0.344
38	0.877	19.116	19.116	0.082	0.5	0.4
39	0.626	4.195	4.195	0.133	0.248	0.248

Table F.5 – Valeurs des différents critères, sur les ensembles de séquences et sur le graphe, calculés pour différents clusters obtenus à partir du jeu A et présents dans l’arbre de la figure 6.4 page 150

Clusters	Critères sur les séquences			Critères sur le graphe		
	$JaccardI_{Mean}$	$AScoreP_{Mean}$	$AScoreNP_{Mean}$	$Density$	$Reticulation_{Median}$	$Reticulation_{Mean}$
40	0.649	4.822	4.822	0.141	0.3	0.3
41	0.613	4.511	4.511	0.147	0.417	0.395
42	0.627	4.509	4.509	0.152	0.317	0.317
43	0.586	4.44	4.44	0.154	0.332	0.332
44	0.622	4.533	4.533	0.108	0.223	0.223
45	0.629	4.486	4.486	0.107	0.35	0.319
46	0.658	4.644	4.644	0.173	0.363	0.363
47	0.402	4.182	4.182	0.11	0.209	0.209
48	0.576	3.858	3.858	0.16	0.298	0.298
49	0.891	18.933	18.933	0.063	0.5	0.472
50	0.851	18.455	18.455	0.061	0.5	0.479
51	0.842	17.987	17.987	0.074	0.458	0.433
52	0.847	18.217	18.217	0.073	0.216	0.216
53	0.787	16.622	16.622	0.076	0.458	0.428
54	0.684	4.0	4.0	0.101	0.211	0.211
55	0.809	17.385	17.385	0.077	0.458	0.431
56	0.625	4.238	4.238	0.145	0.327	0.327
57	0.732	15.879	15.879	0.085	0.417	0.397
58	0.68	4.667	4.667	0.128	0.269	0.269
59	0.642	4.176	4.176	0.148	0.302	0.302
60	0.696	4.0	4.0	0.144	0.35	0.383
61	0.606	4.473	4.473	0.121	0.249	0.249
62	0.634	4.567	4.567	0.117	0.3	0.322
63	0.643	4.509	4.509	0.121	0.417	0.349
64	0.846	18.167	18.167	0.069	0.417	0.433
65	0.652	4.244	4.244	0.173	0.417	0.428
66	0.445	4.733	4.733	0.094	0.262	0.285
67	0.781	16.784	16.784	0.085	0.4	0.387
68	0.745	16.044	16.044	0.087	0.417	0.375
69	0.74	16.134	16.134	0.103	0.197	0.197
70	0.755	16.598	16.598	0.093	0.367	0.358
71	0.836	18.439	18.439	0.077	0.5	0.443
72	0.811	17.921	17.921	0.087	0.351	0.353
73	0.824	17.789	17.789	0.083	0.31	0.365
74	0.824	17.824	17.824	0.091	0.286	0.267
75	0.798	17.489	17.489	0.099	0.189	0.189
76	0.778	17.347	17.347	0.111	0.212	0.212
77	0.766	17.102	17.102	0.115	0.22	0.22
78	0.752	16.765	16.765	0.125	0.239	0.239
79	0.817	17.92	17.92	0.085	0.189	0.288
80	0.832	18.129	18.129	0.091	0.3	0.264
81	0.853	18.898	18.898	0.075	0.5	0.444
82	0.825	18.311	18.311	0.101	0.193	0.193
83	0.804	17.975	17.975	0.111	0.212	0.212
84	0.834	18.175	18.175	0.075	0.35	0.335
85	0.832	18.378	18.378	0.093	0.178	0.178
86	0.81	18.029	18.029	0.097	0.186	0.186
87	0.827	18.037	18.037	0.073	0.417	0.431
88	0.799	17.741	17.741	0.079	0.35	0.362
89	0.868	19.343	19.343	0.083	0.189	0.285
90	0.879	19.216	19.216	0.08	0.35	0.37
91	0.847	18.995	18.995	0.091	0.174	0.174
92	0.877	19.946	19.946	0.083	0.314	0.334

Table F.6 – Suite du tableau précédent

Clusters	Critères sur les séquences			Critères sur le graphe		
	$JaccardI_{Mean}$	$AScoreP_{Mean}$	$AScoreNP_{Mean}$	$Density$	$Reticulation_{Median}$	$Reticulation_{Mean}$
93	0.858	19.357	19.357	0.071	0.45	0.442
94	0.872	19.575	19.575	0.091	0.208	0.274
95	0.851	19.297	19.297	0.095	0.236	0.283
96	0.856	19.503	19.503	0.099	0.236	0.288
97	0.845	19.236	19.236	0.107	0.205	0.205
98	0.835	18.983	18.983	0.109	0.208	0.208
99	0.824	18.713	18.713	0.113	0.216	0.216
101	0.814	18.459	18.459	0.125	0.239	0.239
102	0.793	17.923	17.923	0.134	0.258	0.258
103	0.55	4.049	4.049	0.107	0.206	0.206
104	0.589	4.277	4.277	0.129	0.321	0.341
105	0.611	4.17	4.17	0.117	0.292	0.319
106	0.531	3.873	3.873	0.136	0.263	0.263
107	0.596	4.321	4.321	0.138	0.255	0.255
108	0.603	4.172	4.172	0.105	0.25	0.301
109	0.543	3.741	3.741	0.124	0.232	0.232
110	0.516	3.698	3.698	0.139	0.264	0.264
111	0.474	3.538	3.538	0.168	0.322	0.322
112	0.597	4.226	4.226	0.129	0.367	0.349
113	0.634	4.117	4.117	0.093	0.19	0.19
114	0.542	3.747	3.747	0.119	0.231	0.231
115	0.505	3.629	3.629	0.13	0.25	0.25
116	0.567	3.787	3.787	0.157	0.295	0.295
117	0.551	3.647	3.647	0.147	0.278	0.278
118	0.56	3.474	3.474	0.143	0.268	0.268
119	0.643	4.098	4.098	0.117	0.262	0.337
120	0.527	3.441	3.441	0.121	0.23	0.23
121	0.504	3.385	3.385	0.132	0.253	0.253
122	0.475	3.362	3.362	0.172	0.33	0.33
123	0.451	3.332	3.332	0.223	0.428	0.428
124	0.585	4.636	4.636	0.125	0.283	0.283
125	0.567	4.541	4.541	0.126	0.239	0.239
126	0.618	4.179	4.179	0.1	0.3	0.282
127	0.586	4.149	4.149	0.104	0.198	0.198
128	0.488	3.803	3.803	0.142	0.273	0.273
129	0.567	4.381	4.381	0.121	0.229	0.229
130	0.578	4.186	4.186	0.113	0.232	0.232
131	0.529	4.019	4.019	0.121	0.231	0.231
132	0.594	5.0	5.0	0.124	0.271	0.271
133	0.604	4.168	4.168	0.117	0.24	0.24
134	0.57	4.222	4.222	0.121	0.246	0.246
135	0.475	3.723	3.723	0.164	0.314	0.314
136	0.581	4.449	4.449	0.11	0.227	0.227
137	0.637	4.095	4.095	0.105	0.215	0.215
138	0.596	4.144	4.144	0.1	0.205	0.205
139	0.394	4.066	4.066	0.119	0.229	0.229
140	0.378	4.02	4.02	0.134	0.258	0.258
141	0.634	4.252	4.252	0.121	0.367	0.338
142	0.605	4.204	4.204	0.093	0.3	0.28
143	0.462	3.892	3.892	0.154	0.295	0.295
144	0.453	3.682	3.682	0.186	0.356	0.356
145	0.429	3.521	3.521	0.231	0.443	0.443
146	0.422	3.478	3.478	0.249	0.477	0.477

Table F.7 – Suite du tableau précédent

Clusters	Critères sur les séquences			Critères sur le graphe		
	$JaccardI_{Mean}$	$AScoreP_{Mean}$	$AScoreNP_{Mean}$	$Density$	$Reticulation_{Median}$	$Reticulation_{Mean}$
1	0.868	18.933	18.933	0.075	0.333	0.37
2	0.871	18.945	18.945	0.063	0.417	0.427
3	0.874	19.609	19.609	0.067	0.5	0.467
4	0.884	19.367	19.367	0.082	0.3	0.277
5	0.882	19.968	19.968	0.073	0.417	0.41
6	0.886	19.353	19.353	0.084	0.381	0.357
7	0.89	19.483	19.483	0.082	0.3	0.306
8	0.838	18.076	18.076	0.076	0.5	0.444
9	0.854	18.538	18.538	0.076	0.375	0.384
10	0.89	19.515	19.515	0.082	0.3	0.322
11	0.78	16.528	16.528	0.069	0.417	0.396
12	0.873	19.281	19.281	0.071	0.417	0.403
13	0.841	18.402	18.402	0.077	0.417	0.395
14	0.883	19.393	19.393	0.075	0.417	0.381
15	0.844	18.647	18.647	0.067	0.383	0.375
16	0.832	17.647	17.647	0.082	0.417	0.431
17	0.809	17.055	17.055	0.071	0.5	0.45
18	0.859	18.417	18.417	0.076	0.35	0.365
19	0.821	17.714	17.714	0.077	0.5	0.418
20	0.838	18.088	18.088	0.073	0.363	0.372
21	0.839	18.108	18.108	0.08	0.417	0.397
22	0.648	4.192	4.192	0.143	0.35	0.379
23	0.64	4.095	4.095	0.111	0.268	0.33
24	0.827	17.673	17.673	0.069	0.417	0.45
25	0.871	19.479	19.479	0.083	0.317	0.322
26	0.534	9.267	9.267	0.097	0.264	0.324
27	0.587	10.838	10.838	0.109	0.208	0.208
28	0.668	13.306	13.306	0.081	0.417	0.377
29	0.535	9.235	9.235	0.113	0.333	0.311
30	0.558	9.542	9.542	0.115	0.22	0.22
31	0.468	6.321	6.321	0.127	0.247	0.247
32	0.602	3.717	3.717	0.2	0.393	0.393
33	0.486	5.422	5.422	0.114	0.218	0.218
34	0.581	11.212	11.212	0.097	0.4	0.374
35	0.639	4.124	4.124	0.138	0.417	0.383
36	0.552	8.689	8.689	0.102	0.225	0.225
37	0.613	12.356	12.356	0.097	0.286	0.328
38	0.584	4.381	4.381	0.138	0.45	0.395
39	0.634	3.615	3.615	0.147	0.286	0.286
40	0.555	9.934	9.934	0.117	0.223	0.223
41	0.532	7.91	7.91	0.105	0.216	0.216
42	0.599	4.35	4.35	0.143	0.278	0.278
43	0.763	16.463	16.463	0.089	0.343	0.342
44	0.646	12.708	12.708	0.095	0.182	0.182
45	0.613	4.407	4.407	0.154	0.286	0.286
46	0.667	4.026	4.026	0.099	0.262	0.278
47	0.519	8.073	8.073	0.108	0.209	0.209
48	0.547	9.917	9.917	0.119	0.227	0.227
49	0.53	8.992	8.992	0.119	0.229	0.229

Table F.8 – Valeurs des différents critères, sur les ensembles de séquences et sur le graphe, calculés pour différents clusters obtenus à partir du jeu B et présents dans l’arbre de la figure 6.5 page 151

Clusters	Critères sur les séquences			Critères sur le graphe		
	$JaccardI_{Mean}$	$AScoreP_{Mean}$	$AScoreNP_{Mean}$	$Density$	$Reticulation_{Median}$	$Reticulation_{Mean}$
50	0.624	4.156	4.156	0.125	0.304	0.335
51	0.557	9.725	9.725	0.117	0.223	0.223
52	0.642	4.489	4.489	0.141	0.417	0.389
53	0.56	10.564	10.564	0.105	0.201	0.201
54	0.478	7.058	7.058	0.117	0.223	0.223
55	0.67	4.042	4.042	0.11	0.262	0.328
56	0.67	4.712	4.712	0.132	0.35	0.367
57	0.616	4.156	4.156	0.129	0.304	0.357
58	0.654	13.8	13.8	0.077	0.458	0.415
59	0.612	4.417	4.417	0.138	0.4	0.37
60	0.466	6.773	6.773	0.105	0.201	0.201
61	0.63	4.543	4.543	0.129	0.267	0.267
62	0.546	8.886	8.886	0.105	0.357	0.327
63	0.591	10.524	10.524	0.084	0.393	0.361
64	0.636	4.216	4.216	0.124	0.256	0.256
65	0.635	4.154	4.154	0.118	0.31	0.325
66	0.509	3.964	3.964	0.126	0.238	0.238
67	0.496	8.0	8.0	0.11	0.21	0.21
68	0.437	5.806	5.806	0.1	0.19	0.19
69	0.494	7.778	7.778	0.119	0.227	0.227
70	0.587	11.212	11.212	0.095	0.182	0.182
71	0.496	6.85	6.85	0.128	0.243	0.243
72	0.601	4.846	4.846	0.129	0.278	0.278
73	0.521	8.47	8.47	0.101	0.304	0.286
74	0.554	9.125	9.125	0.119	0.226	0.226
75	0.64	4.636	4.636	0.132	0.3	0.362
76	0.48	5.536	5.536	0.111	0.207	0.207
77	0.645	4.209	4.209	0.141	0.35	0.367
78	0.781	16.885	16.885	0.079	0.305	0.322
79	0.603	4.91	4.91	0.114	0.243	0.243
80	0.462	7.117	7.117	0.132	0.254	0.254
81	0.569	3.688	3.688	0.133	0.267	0.267
82	0.496	8.036	8.036	0.111	0.219	0.219
83	0.621	12.192	12.192	0.095	0.211	0.303
84	0.601	4.556	4.556	0.125	0.278	0.347
85	0.509	7.78	7.78	0.119	0.375	0.337
86	0.455	5.306	5.306	0.111	0.211	0.211
87	0.507	8.949	8.949	0.117	0.223	0.223
88	0.66	4.05	4.05	0.11	0.25	0.327
89	0.51	6.607	6.607	0.129	0.243	0.243
90	0.632	12.015	12.015	0.097	0.417	0.403
91	0.575	4.592	4.592	0.1	0.286	0.274
92	0.539	4.923	4.923	0.105	0.286	0.304
93	0.427	5.153	5.153	0.142	0.273	0.273
94	0.421	4.971	4.971	0.126	0.241	0.241
95	0.4	4.883	4.883	0.184	0.352	0.352
96	0.437	6.36	6.36	0.152	0.292	0.292
97	0.418	6.062	6.062	0.178	0.341	0.341
98	0.42	6.111	6.111	0.146	0.28	0.28
99	0.462	5.051	5.051	0.132	0.248	0.248

Table F.9 – Suite du tableau précédent

Clusters	Critères sur les séquences			Critères sur le graphe		
	$JaccardI_{Mean}$	$AScoreP_{Mean}$	$AScoreNP_{Mean}$	$Density$	$Reticulation_{Median}$	$Reticulation_{Mean}$
100	0.428	6.947	6.947	0.15	0.288	0.288
101	0.419	6.554	6.554	0.184	0.352	0.352
102	0.401	6.171	6.171	0.221	0.424	0.424
103	0.385	5.571	5.571	0.237	0.455	0.455
104	0.611	4.497	4.497	0.125	0.278	0.347
105	0.608	4.192	4.192	0.116	0.25	0.332
106	0.529	3.753	3.753	0.143	0.273	0.273
107	0.592	4.171	4.171	0.11	0.292	0.331
108	0.522	3.77	3.77	0.13	0.247	0.247
109	0.488	4.214	4.214	0.137	0.262	0.262
110	0.595	4.129	4.129	0.12	0.228	0.228
111	0.519	4.012	4.012	0.139	0.268	0.268
112	0.471	3.63	3.63	0.176	0.337	0.337
113	0.456	3.497	3.497	0.2	0.383	0.383
114	0.535	4.083	4.083	0.118	0.224	0.224
115	0.562	3.867	3.867	0.127	0.236	0.236
116	0.484	3.619	3.619	0.152	0.283	0.283
117	0.563	3.463	3.463	0.127	0.243	0.243
118	0.538	3.514	3.514	0.167	0.324	0.324
119	0.516	3.364	3.364	0.173	0.324	0.324
120	0.615	4.116	4.116	0.097	0.211	0.307
121	0.619	4.13	4.13	0.107	0.286	0.299
122	0.521	3.667	3.667	0.14	0.269	0.269
123	0.479	3.334	3.334	0.176	0.337	0.337
124	0.464	3.342	3.342	0.206	0.394	0.394
125	0.444	3.336	3.336	0.245	0.47	0.47
126	0.401	3.701	3.701	0.257	0.492	0.492
127	0.623	12.714	12.714	0.105	0.304	0.341
128	0.617	12.833	12.833	0.115	0.236	0.236
129	0.79	16.756	16.756	0.075	0.5	0.443
130	0.688	14.675	14.675	0.097	0.186	0.186
131	0.631	13.488	13.488	0.128	0.246	0.246
132	0.571	11.17	11.17	0.126	0.242	0.242
133	0.594	11.765	11.765	0.121	0.231	0.231
134	0.562	11.225	11.225	0.15	0.288	0.288
135	0.556	11.131	11.131	0.156	0.299	0.299
136	0.552	11.033	11.033	0.158	0.303	0.303
137	0.575	11.85	11.85	0.168	0.322	0.322
138	0.528	8.985	8.985	0.136	0.261	0.261
139	0.518	9.528	9.528	0.146	0.28	0.28
140	0.515	9.066	9.066	0.138	0.265	0.265
141	0.52	9.241	9.241	0.144	0.277	0.277
142	0.513	8.828	8.828	0.148	0.284	0.284
143	0.497	8.785	8.785	0.164	0.314	0.314
144	0.481	8.661	8.661	0.186	0.356	0.356
145	0.478	8.741	8.741	0.192	0.367	0.367
146	0.472	8.594	8.594	0.194	0.371	0.371
147	0.474	7.767	7.767	0.138	0.265	0.265
148	0.475	7.789	7.789	0.139	0.266	0.266
149	0.476	8.048	8.048	0.154	0.295	0.295

Table F.10 – Suite du tableau précédent

Clusters	Critères sur les séquences			Critères sur le graphe		
	$JaccardI_{Mean}$	$AScoreP_{Mean}$	$AScoreNP_{Mean}$	$Density$	$Reticulation_{Median}$	$Reticulation_{Mean}$
150	0.46	7.766	7.766	0.18	0.345	0.345
151	0.501	7.838	7.838	0.142	0.273	0.273
152	0.44	7.03	7.03	0.158	0.303	0.303
153	0.436	7.031	7.031	0.184	0.352	0.352
154	0.431	6.97	6.97	0.198	0.379	0.379
155	0.429	7.151	7.151	0.215	0.413	0.413
156	0.505	4.161	4.161	0.129	0.245	0.245
157	0.468	4.098	4.098	0.142	0.273	0.273
158	0.61	4.14	4.14	0.105	0.268	0.322
159	0.588	4.089	4.089	0.14	0.259	0.259
160	0.559	4.128	4.128	0.125	0.239	0.239
161	0.502	3.755	3.755	0.148	0.284	0.284
162	0.452	3.67	3.67	0.2	0.383	0.383
163	0.841	18.77	18.77	0.077	0.5	0.414
164	0.853	19.052	19.052	0.087	0.305	0.33
165	0.831	18.717	18.717	0.093	0.333	0.347
166	0.851	18.889	18.889	0.089	0.17	0.17
167	0.862	19.538	19.538	0.087	0.333	0.333
168	0.847	19.009	19.009	0.085	0.392	0.364
169	0.848	19.217	19.217	0.095	0.333	0.308
170	0.838	19.004	19.004	0.109	0.208	0.208
171	0.828	18.813	18.813	0.117	0.223	0.223
172	0.823	17.954	17.954	0.093	0.178	0.178
173	0.847	18.531	18.531	0.099	0.189	0.189
174	0.742	16.151	16.151	0.097	0.35	0.345
175	0.765	16.785	16.785	0.101	0.362	0.355
176	0.782	17.385	17.385	0.119	0.227	0.227
177	0.806	17.807	17.807	0.091	0.25	0.288
178	0.801	17.62	17.62	0.089	0.433	0.387
179	0.785	17.485	17.485	0.107	0.205	0.205
180	0.788	17.091	17.091	0.089	0.351	0.348
181	0.862	18.826	18.826	0.089	0.3	0.267
182	0.809	17.885	17.885	0.099	0.189	0.189
183	0.78	17.474	17.474	0.115	0.22	0.22
184	0.771	17.315	17.315	0.132	0.254	0.254
185	0.796	18.0	18.0	0.132	0.254	0.254
186	0.61	13.339	13.339	0.233	0.447	0.447
187	0.562	12.01	12.01	0.253	0.485	0.485
188	0.541	11.352	11.352	0.257	0.492	0.492
189	0.543	11.4	11.4	0.257	0.492	0.492
190	0.431	7.99	7.99	0.261	0.5	0.5

Table F.11 – Suite du tableau précédent

F.3 Critères calculés sur les clusters constitués après regroupement ascendant hiérarchique pour un ensemble de profils anesthésiques

Clusters n°	Critères sur les séquences			Critères sur le graphe		
	$JaccardI_{Mean}$	$AScoreP_{Mean}$	$AScoreNP_{Mean}$	$Density$	$Reticulation_{Median}$	$Reticulation_{Mean}$
1	0.789	74.951	74.951	0.01	0.214	0.201
2	0.616	72.533	72.533	0.009	0.02	0.093
3	0.808	75.267	75.267	0.013	0.236	0.204
4	0.789	74.971	74.971	0.01	0.214	0.202
5	0.796	75.123	75.123	0.011	0.214	0.186
6	0.8	73.758	73.758	0.012	0.242	0.229
7	0.81	74.258	74.258	0.012	0.233	0.229
8	0.771	72.238	72.238	0.012	0.25	0.238
9	0.785	72.962	72.962	0.011	0.224	0.216
10	0.799	73.689	73.689	0.012	0.238	0.23
11	0.788	73.156	73.156	0.012	0.25	0.235
12	0.779	72.67	72.67	0.011	0.238	0.228
13	0.597	71.617	71.617	0.007	0.013	0.092
14	0.795	73.288	73.288	0.012	0.214	0.22
15	0.791	73.258	73.258	0.011	0.233	0.22
16	0.801	73.75	73.75	0.012	0.25	0.243
17	0.816	74.582	74.582	0.012	0.238	0.224
18	0.791	73.272	73.272	0.011	0.238	0.228
19	0.651	75.879	75.879	0.008	0.022	0.118
20	0.795	74.91	74.91	0.011	0.194	0.189
21	0.793	73.362	73.362	0.012	0.236	0.223
22	0.786	73.03	73.03	0.012	0.238	0.22
23	0.805	74.0	74.0	0.013	0.238	0.238
24	0.639	72.648	72.648	0.008	0.021	0.11
25	0.781	72.762	72.762	0.012	0.244	0.24
26	0.784	72.922	72.922	0.011	0.236	0.226
27	0.775	72.41	72.41	0.012	0.25	0.23
28	0.781	73.824	73.824	0.011	0.176	0.169
29	0.775	75.918	75.918	0.011	0.233	0.202
30	0.793	73.045	73.045	0.012	0.238	0.227
31	0.792	73.199	73.199	0.011	0.214	0.213
32	0.78	74.181	74.181	0.011	0.194	0.172
33	0.766	73.564	73.564	0.011	0.2	0.194
34	0.777	72.556	72.556	0.011	0.25	0.233
35	0.778	72.59	72.59	0.011	0.238	0.228
36	0.771	72.231	72.231	0.011	0.236	0.215
37	0.779	74.424	74.424	0.009	0.2	0.177
38	0.785	74.176	74.176	0.01	0.178	0.182
39	0.785	73.5	73.5	0.011	0.189	0.181

Table F.12 – Valeurs des différents critères, sur les ensembles de séquences et sur le graphe, calculés pour différents clusters obtenus à partir de 1000 profils anesthésiques pour des patients de type et d'intervention identiques

Clusters n°	Critères sur les séquences			Critères sur le graphe		
	$JaccardI_{Mean}$	$AScoreP_{Mean}$	$AScoreNP_{Mean}$	$Density$	$Reticulation_{Median}$	$Reticulation_{Mean}$
40	0.763	71.179	71.179	0.012	0.238	0.221
41	0.811	74.321	74.321	0.012	0.25	0.238
42	0.796	75.314	75.314	0.011	0.2	0.199
43	0.789	73.163	73.163	0.011	0.214	0.223
44	0.795	73.475	73.475	0.011	0.233	0.225
45	0.786	72.989	72.989	0.011	0.244	0.228
46	0.796	73.3	73.3	0.011	0.224	0.221
47	0.795	73.067	73.067	0.013	0.25	0.247
48	0.77	72.197	72.197	0.011	0.25	0.229
49	0.779	71.905	71.905	0.013	0.25	0.247
50	0.799	73.718	73.718	0.012	0.25	0.237
51	0.779	72.649	72.649	0.011	0.233	0.219
52	0.796	73.556	73.556	0.012	0.244	0.231
53	0.745	72.857	72.857	0.011	0.2	0.171
54	0.759	71.081	71.081	0.011	0.207	0.219
55	0.786	72.619	72.619	0.013	0.224	0.224
56	0.789	73.179	73.179	0.012	0.242	0.224
57	0.753	74.059	74.059	0.01	0.2	0.166
58	0.789	76.576	76.576	0.011	0.214	0.203
59	0.786	73.022	73.022	0.012	0.25	0.232
60	0.741	72.658	72.658	0.01	0.2	0.166
61	0.727	71.774	71.774	0.009	0.2	0.154
62	0.768	75.49	75.49	0.01	0.214	0.191
63	0.632	72.989	72.989	0.007	0.018	0.099
64	0.578	68.985	68.985	0.006	0.012	0.086
65	0.556	65.952	65.952	0.006	0.012	0.092
66	0.558	62.949	62.949	0.006	0.011	0.092
67	0.55	63.598	63.598	0.005	0.01	0.098
68	0.571	63.138	63.138	0.005	0.01	0.098
69	0.588	63.571	63.571	0.004	0.008	0.097
70	0.765	73.177	73.177	0.01	0.112	0.151
71	0.755	73.005	73.005	0.011	0.2	0.179
72	0.745	72.267	72.267	0.01	0.112	0.151
73	0.744	72.272	72.272	0.01	0.112	0.151
74	0.779	73.89	73.89	0.009	0.165	0.164
75	0.767	73.442	73.442	0.009	0.165	0.162
76	0.785	74.357	74.357	0.009	0.178	0.176
77	0.801	74.382	74.382	0.012	0.233	0.2
78	0.773	73.934	73.934	0.009	0.165	0.167
79	0.77	73.53	73.53	0.009	0.158	0.166
80	0.755	72.75	72.75	0.008	0.153	0.166
81	0.749	72.45	72.45	0.008	0.153	0.166
82	0.733	71.243	71.243	0.007	0.141	0.135
83	0.656	66.842	66.842	0.004	0.007	0.096
84	0.764	71.554	71.554	0.011	0.2	0.205
85	0.792	73.362	73.362	0.012	0.25	0.231
86	0.764	71.232	71.232	0.011	0.2	0.203
87	0.783	72.884	72.884	0.011	0.25	0.229
88	0.773	72.308	72.308	0.01	0.233	0.218
89	0.785	72.841	72.841	0.011	0.189	0.192

Table F.13 – Suite du tableau précédent

Clusters n°	Critères sur les séquences			Critères sur le graphe		
	$JaccardI_{Mean}$	$AScoreP_{Mean}$	$AScoreNP_{Mean}$	$Density$	$Reticulation_{Median}$	$Reticulation_{Mean}$
90	0.776	72.367	72.367	0.011	0.2	0.203
91	0.763	71.699	71.699	0.011	0.178	0.19
92	0.747	70.878	70.878	0.01	0.197	0.197
93	0.741	70.289	70.289	0.009	0.197	0.195
94	0.779	72.31	72.31	0.011	0.214	0.217
95	0.758	70.903	70.903	0.011	0.233	0.218
96	0.752	70.666	70.666	0.01	0.214	0.21
97	0.789	73.167	73.167	0.012	0.25	0.235
98	0.805	73.983	73.983	0.012	0.233	0.221
99	0.774	72.375	72.375	0.011	0.233	0.219
100	0.788	73.111	73.111	0.011	0.233	0.225
101	0.772	72.272	72.272	0.011	0.233	0.226
102	0.769	72.123	72.123	0.01	0.233	0.222
103	0.763	71.792	71.792	0.01	0.233	0.219
104	0.75	70.832	70.832	0.01	0.214	0.205
105	0.728	69.252	69.252	0.009	0.186	0.184
106	0.785	72.439	72.439	0.012	0.238	0.229
107	0.779	72.31	72.31	0.012	0.214	0.221
108	0.795	73.493	73.493	0.012	0.242	0.226
109	0.77	72.029	72.029	0.011	0.233	0.222
110	0.769	72.007	72.007	0.011	0.233	0.216
111	0.76	71.463	71.463	0.011	0.214	0.213
112	0.771	72.25	72.25	0.011	0.233	0.226
113	0.767	72.027	72.027	0.011	0.242	0.23
114	0.778	72.525	72.525	0.011	0.214	0.215
115	0.773	72.262	72.262	0.011	0.214	0.214
116	0.768	71.795	71.795	0.01	0.214	0.211
117	0.758	70.896	70.896	0.01	0.207	0.208
118	0.723	68.828	68.828	0.01	0.2	0.197
119	0.785	72.957	72.957	0.011	0.244	0.224
120	0.77	72.189	72.189	0.011	0.236	0.218
121	0.763	71.788	71.788	0.01	0.236	0.216
122	0.783	72.673	72.673	0.012	0.238	0.228
123	0.755	70.705	70.705	0.011	0.214	0.21
124	0.783	72.781	72.781	0.011	0.214	0.21
125	0.775	72.055	72.055	0.01	0.2	0.198
126	0.757	71.038	71.038	0.011	0.214	0.208
127	0.743	70.037	70.037	0.011	0.214	0.209
128	0.741	69.974	69.974	0.01	0.214	0.203
129	0.738	69.905	69.905	0.01	0.207	0.202
130	0.735	69.715	69.715	0.01	0.207	0.201
131	0.739	70.075	70.075	0.01	0.207	0.201
132	0.726	69.071	69.071	0.009	0.2	0.197
133	0.721	68.767	68.767	0.009	0.186	0.184
134	0.652	65.414	65.414	0.004	0.008	0.095

Table F.14 – Suite du tableau précédent

Titre : Conception automatisée de scénarios d'apprentissage réalistes et variés, pour l'acquisition et la consolidation d'expertise en anesthésie, assistées par le numérique.

Mot clés : Anesthésie, Séries temporelles, Patient virtuel, Fouille de données

Résumé : Ces travaux portent sur la simulation de l'évolution des paramètres physiologiques d'un patient virtuel au bloc opératoire, en réaction aux actions d'un apprenant et d'une équipe médicale virtuelle. Pour cela, nous avons étendu l'approche du raisonnement à partir de cas, à des cas décrits par des traces d'événements et des séries temporelles multivariées, et où la prédiction est aussi une série temporelle multivariée. Le test de trois instantiations de cette approche nous a permis de conclure au réalisme de la simulation dans chacun des cas. Par ailleurs, nous avons proposé un modèle de représentation synthétique du déroulement d'une chirurgie, à partir des données d'une cohorte de patients. Pour obtenir de telles cohortes, nous proposons une stratégie combinant deux approches de clustering et utilisant un critère de com-

plexité. Ce critère a été validé au cours d'expérimentations préalables sur des données simulées. Ce modèle de représentation synthétique original, annoté par des séries temporelles, permet un contrôle de sa complexité ainsi que la simulation d'un patient virtuel avec un réalisme similaire aux précédents types de simulation. Cette représentation synthétique annotée représente de plus un intérêt pour l'anonymisation de ces données temporelles complexes. La construction automatisée de ces simulations repose sur l'exploitation des données d'anesthésie collectées lors de chirurgies. L'exploitation de ces données étant réglementée, nous avons construit un simulateur de données réalistes. Nous avons montré que cette approche inédite, basée sur l'expertise de soignants en anesthésie, fournit bien des séries temporelles réalistes.

Title: Automated design of realistic and varied learning scenarios, for the acquisition and consolidation of expertise in anesthesia, assisted by digital technologies.

Keywords: Anesthesia, Time series, Virtual patient, Data mining

Abstract: This work concerns the simulation of the evolution of the physiological parameters of a virtual patient in the operating room, in response to the actions of a learner and a virtual medical team. For this, we have extended the case-based reasoning approach to cases described by event traces and multivariate time series, and where the prediction is also of a temporal nature (multivariate time series). The test of three instantiations of this approach allowed us to conclude that the simulation was realistic in each case. In addition, we have proposed a synthetic representation model of the course of a surgery, based on data from a cohort of patients. To obtain such cohorts, we propose a strategy combining two clustering approaches and us-

ing a complexity criterion. This criterion was validated during prior experiments on simulated data. This original synthetic representation model, annotated by time series, allows a control of its complexity as well as the simulation of a virtual patient with a realism similar to the previous types of simulation. This annotated synthetic representation is also of interest for the anonymization of these complex temporal data. The automated construction of these simulations is based on the exploitation of anesthesia data collected during surgeries. The exploitation of this data being regulated, we have built a realistic data simulator. We have shown that this novel approach, based on the expertise of anaesthesia caregivers, does provide realistic time series.