

UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

**MODÉLISATION DE LA CINÉTIQUE DE PÉDALAGE ET DU
RECRUTEMENT MUSCULAIRE EN CYCLISME À PARTIR DE LA
CINÉMATIQUE AU MOYEN DE TECHNIQUES D'APPRENTISSAGE
AUTOMATIQUE**

**THÈSE PRÉSENTÉ
COMME EXIGENCE PARTIELLE
DOCTORAT EN GÉNIE ÉLECTRIQUE**

**PAR
ANDRÉS TORRES VELASQUEZ**

30 juillet 2025

Université du Québec à Trois-Rivières

Service de la bibliothèque

Avertissement

L'auteur de ce mémoire, de cette thèse ou de cet essai a autorisé l'Université du Québec à Trois-Rivières à diffuser, à des fins non lucratives, une copie de son mémoire, de sa thèse ou de son essai.

Cette diffusion n'entraîne pas une renonciation de la part de l'auteur à ses droits de propriété intellectuelle, incluant le droit d'auteur, sur ce mémoire, cette thèse ou cet essai. Notamment, la reproduction ou la publication de la totalité ou d'une partie importante de ce mémoire, de cette thèse et de son essai requiert son autorisation.

UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES
DOCTORAT EN GÉNIE ÉLECTRIQUE (Ph. D.)

Direction de recherche :

Frédéric Domingue directeur de recherche
Université du Québec à Trois-Rivières

François Nougarou codirecteur de recherche
Université du Québec à Trois-Rivières

Jury d'évaluation :

Frédéric Domingue directeur de recherche
Université du Québec à Trois-Rivières

François Nougarou codirecteur de recherche
Université du Québec à Trois-Rivières

Maude Josée-Blondin
Université de Sherbrooke évaluateur externe

Sébastien Duc évaluateur externe
Université de Reims Champagne-Ardenne

Thèse soutenue le 5 juin 2025

Avant-propos

Le cyclisme est une activité que j'ai voulu étudier en tant qu'athlète et ingénieur et à laquelle je souhaite contribuer par la conception d'applications technologiques innovantes afin de comprendre les phénomènes de mouvement qui se produisent dans le geste de pédalage. Le mouvement des membres inférieurs qui permettent la production de puissance en cyclisme combine de multiples facteurs physiologiques et physiques, nécessitant ces développements technologiques pour comprendre comment ils sont liés et faciliter leur analyse. Grâce au soutien et à l'accompagnement du Professeur Frédéric Domingue, qui m'a guidé par son expérience, ses connaissances et sa compréhension de cette problématique, ainsi qu'à la codirection du Professeur François Nougarou et aux conseils du Postdoctorant Geoffrey Millour, j'ai réalisé plusieurs projets dans le cadre de ma formation et qui sont présentés dans ce travail. Aujourd'hui, je vois un avenir prometteur pour tous les outils proposés qui permettent une compréhension et optimisation du mouvement de pédalage et j'espère apporter des réponses aux intervenants, athlètes et chercheurs qui s'intéressent à ce beau sport.

Dédicace

À Dieu, à Paulina, Laura, qui par leur amour et leur lumière m'ont accompagné dans ce chemin...à Santiago, Izha, Adriana Maria, Laura et la famille Gomez Fernandez, qui, avec leur amour, m'ont continuellement conseillé et encouragé.

Résumé

La production de force en cyclisme est un phénomène qui résulte de différents facteurs biomécaniques associés au mouvement des membres inférieurs. Dans ce contexte, il est intéressant d'étudier ces facteurs afin d'établir certains paramètres d'optimisation qui peuvent être utiles aux entraîneurs et aux cyclistes. Il existe aujourd'hui de nombreuses technologies qui permettent de capturer les données liées au mouvement ainsi que des variables physiologiques de l'athlète. Grâce aux informations obtenues par ces technologies, il est possible de connaître la cinématique du cycliste, c'est-à-dire les angles, les vitesses et les accélérations du membre inférieur. Aussi, il est également possible de connaître les forces appliquées sur les pédales, ce qui permet d'évaluer l'efficience mécanique du pédalage. Finalement, il est possible de connaître l'activité musculaire et d'évaluer le recrutement musculaire du cycliste à l'aide de l'électromyographie de surface (sEMG). Bien que ces données soient disponibles, même aujourd'hui, des questions subsistent quant aux facteurs biomécaniques qui permettent de transmettre efficacement les forces aux pédales. Pour interpréter la grande quantité d'informations disponibles lors d'une analyse biomécanique, les méthodes d'apprentissage automatique (« machine-learning »), impliquées en l'intelligence artificielle, s'avèrent intéressantes afin de développer des modèles mathématiques qui permettent de prédire ou même d'interpréter les variables. L'apprentissage automatique offre une grande variété d'algorithmes qui peuvent être optimisés, fournissant ainsi un outil puissant pour modéliser n'importe quel phénomène. De manière innovante, plusieurs modèles basés sur l'apprentissage

automatique sont présentés pour permettre l'estimation de métriques destinées à évaluer la technique de pédalage à partir de la cinématique du membre inférieur. En outre, en utilisant également ces données, un modèle est présenté pour prédire les forces exercées sur la pédale à l'aide d'une technique d'apprentissage automatique. Et finalement, un modèle permettant d'estimer le recrutement musculaire de certains muscles des membres inférieurs sans qu'il soit nécessaire d'utiliser des capteurs fixés aux muscles est également présenté. En conséquence, ce projet présente des modèles basés sur l'apprentissage automatique, qui permettent d'analyser les principaux facteurs liés à la production de force en cyclisme.

Mots-clés :

Biomécanique, apprentissage automatique, forces de pédalage, indice d'efficacité, proportion d'impulsion positive, électromyographie de surface

Table des Matières

Avant-propos	iv
Dédicace	v
Résumé	vi
Mots-clés :	vii
Liste des Figures	xiv
Liste des Tableaux.....	xx
Liste des Abréviations.....	xxii
CHAPITRE 1 - Introduction	25
1.1 Mise en contexte	26
1.2 Problématique de recherche	29
1.3 Proposition méthodologique	32
1.4 Objectives de la recherche et contributions	33
1.4.1 Objectives de la recherche	33
1.4.2 Principales contributions scientifiques	34
1.5 Originalité de la recherche	38
1.6 Organisation de la thèse	39
CHAPITRE 2 - État de l'art.....	42
2.1 Biomécanique du cyclisme	42
2.1.1 Mécanisme de propulsion du vélo	42
2.1.2 Géométrie du cadre	44
2.1.3 Cinétique du pédalage.....	46

2.1.4 Cinématique du pédalage.....	52
2.1.5 Recrutement musculaire pendant le pédalage.....	56
2.2 Facteurs qui affectent la production de force.....	63
2.2.1 Facteurs liés à la géométrie et aux composants du vélo	63
2.2.2 Facteurs biomécaniques	66
2.2.3 Facteurs physiologiques.....	67
2.3 La modélisation en biomécanique	68
2.3.1 Modèle de dynamique directe	68
2.3.2 Modèle de dynamique inverse	69
2.3.3 Modèles biomécaniques basés sur techniques d'apprentissage automatique.....	74
2.4 Concepts de base de l'apprentissage automatique.....	76
2.4.1 Régression linéaire multiple	80
2.4.2 Arbre de décision et forêt aléatoires	81
2.4.3 Régression LASSO	83
2.4.4 Régression rigide	83
2.4.5 Régression en composantes principales (PCR)	84
2.4.6 Régression partielle du dernier carré (PLSR).....	84
2.4.7 Réseaux neuronaux artificiels	85
2.4.8 Méthodes de sélection des variables	91
2.4.9 Validation du modèle.....	95

2.4.10 Critères d'information, métriques de validation et indicateurs d'erreur	96
CHAPITRE 3 - Infrastructure de capture des données	100
3.1 Système de capture de mouvement.....	100
3.2 Système de mesure des forces de pédalage.....	103
3.3 Système de capture de l'activité musculaire.....	104
3.4 Système de contrôle et mesure de cadence et puissance.....	105
CHAPITRE 4 - Étude 1, prédition de l'indice d'efficacité et proportion d'impulsions positives	106
4.1 Méthodologie	106
4.2 Protocole d'acquisition des données	107
4.3 Participants.....	108
4.4 Post-traitement des données.....	109
4.5 Implémentation des modèles.....	110
4.5.1 Modélisation par régression linéaire.....	111
4.5.2 Modélisation par arbre de régression.....	113
4.5.3 Modélisation par réseaux de neurones.....	114
4.6 Résultats	116
4.6.1 Résultats sélection de prédicteurs	116
4.6.2 Résultats des régressions linéaires multiples	124
4.6.3 Résultats des arbres de régression	129
4.6.4 Résultats des modèles de forêts aléatoires	132

4.6.5 Résultats réseaux neuronaux.....	134
4.6.6 Résultats validation croisée	136
4.7 Analyse des résultats.....	139
4.7.1 Analyse des méthodes de sélection des prédicteurs.....	139
4.7.2 Évaluation des hypothèses de régressions linéaires multiples.....	142
4.7.3 Analyse de validation croisée des régressions linéaires multiples ..	147
4.7.4 Analyse des modèles de régression	150
4.7.5 Analyse et comparaison de tous les modèles.....	152
4.7.6 Conclusion générale chapitre 4.....	157
CHAPITRE 5 – Étude 2, prédiction des forces de pédalage et du recrutement musculaire	158
5.1 Protocole d’acquisition des données	159
5.2 Participants.....	160
5.3 Post-traitement des données.....	160
5.4 Implémentation des modèles.....	161
5.4.1 Réseau neuronal pour prédire les forces de pédalage	162
5.4.2 Réseau neuronal pour prédire la courbe enveloppe RMS EMG.....	165
5.5 Résultats	168
5.5.1 Résultats réseau neuronal prédiction forces de pédalage.....	168
5.5.2 Résultats réseau neuronal prédiction courbe enveloppe EMG RMS178	
5.6 Analyse des résultats.....	184
5.6.1 Analyse des résultats prédiction forces du pédalage	184

5.6.2 Analyse des résultats prédition courbe enveloppe EMG RMS.....	195
5.6.3 Conclusion générale chapitre 5	201
CHAPITRE 6 – Conclusions, limitations et recommandations	202
6.1 Conclusions générales.....	202
6.2 Limitations	203
6.3 Perspectives et recommandations	205
Bibliographie.....	208
ANNEXE A - Contributions Affiche.....	218
ANNEXE B - Autres Contributions.....	219
ANNEXE C - Analyse de la dynamique inverse en pédalant.....	221
ANNEXE D - Équations cinématiques et cinétiques.....	234
ANNEXE E - Certificat d'éthique étude 1	235
ANNEXE F - Certificat d'éthique étude 2	236
ANNEXE G - Prédition de la force de pédalage Fx à 100 W	237
ANNEXE H - Prédition de la force de pédalage Fz à 100 W	240
ANNEXE I - Prédition de la force de pédalage Fx à 150 W.....	243
ANNEXE J - Prédition de la force de pédalage Fz à 150 W.....	246
ANNEXE K - Prédition de la force de pédalage Fx à 200 W	249
ANNEXE L - Prédition de la force de pédalage Fz à 200 W.....	252
ANNEXE M - Prédition courbe enveloppe EMG RMS 100 W.....	255
ANNEXE N - Prédition courbe enveloppe EMG RMS 150 W	261
ANNEXE O - Prédition courbe enveloppe EMG RMS 200 W	267

ANNEXE P - Publication soumise	273
ANNEXE Q - Publication à soumettre	274

Liste des Figures

<i>Figure 1.1</i>	<i>Contexte général de la recherche</i>	28
<i>Figure 2.1</i>	<i>Vélo d'Henry John Lawson 1879</i>	43
<i>Figure 2.2</i>	<i>Géométrie du vélo des années 1892, adapté de [16]</i>	44
<i>Figure 2.3</i>	<i>Schéma de la force appliquée sur la pédale</i>	45
<i>Figure 2.4</i>	<i>Définition des forces F_{eff}, F_R, F_L (a), et décompositions trigonométriques des forces F_x et F_z (b)</i>	46
<i>Figure 2.5</i>	<i>Définition des forces appliquées à la manivelle en isométrique</i>	47
<i>Figure 2.6</i>	<i>Profils des forces appliquées à la manivelle</i>	48
<i>Figure 2.7</i>	<i>Représentation des différentes phases du cycle de pédalage</i>	50
<i>Figure 2.8</i>	<i>Environnement de capture de mouvement</i>	53
<i>Figure 2.9</i>	<i>Modèle biomécanique du segment de liaison, (a) représentation informatique de la capture de mouvement, (b) modèle biomécanique du segment de liaison</i>	54
<i>Figure 2.10</i>	<i>Modèle biomécanique bidimensionnelle</i>	55
<i>Figure 2.11</i>	<i>Muscles impliqués pendant le mouvement de pédalage. Extrait du livre: Optimiser son positionnement pour performer en Cyclisme, L-Tips</i>	59

<i>Figure 2.12 Étapes du traitement du signal sEMG.....</i>	59
<i>Figure 2.13 Recrutement musculaire d'un cycliste typique à 250 W et 80 rpm.....</i>	61
<i>Figure 2.14 Facteurs géométriques influençant la production de force de pédalage</i>	64
<i>Figure 2.15 Modèle bidimensionnel à 5 barres du membre inférieur.....</i>	71
<i>Figure 2.16 Profils de puissance des articulations du membre inférieur.....</i>	72
<i>Figure 2.17 Flux de travail apprentissage automatique.....</i>	79
<i>Figure 2.18 Architecture d'un arbre de décision.....</i>	82
<i>Figure 2.19 Architecture générale d'un réseau de neurones.....</i>	86
<i>Figure 2.20 Architecture des réseaux neuronaux multicouches.....</i>	87
<i>Figure 2.21 Architecture d'un réseau LSTM, adapté de [84].....</i>	89
<i>Figure 3.1 Protocole biomécanique « Bike-fitting left side & floor ».....</i>	101
<i>Figure 3.2 Schème du protocole cinématique.....</i>	102
<i>Figure 3.3 Système de pédales Forped.....</i>	103

<i>Figure 3.4 Système d'électromyographie Delsys.....</i>	104
<i>Figure 3.5 Environnement de mesure.....</i>	105
<i>Figure 4.1 Méthodologie appliquée</i>	106
<i>Figure 4.2 Schéma du protocole d'étude I</i>	107
<i>Figure 4.3 Diagramme de la méthodologie modèles de régression.....</i>	111
<i>Figure 4.4 Flux de travail sélection des prédicteurs.....</i>	112
<i>Figure 4.5 Diagramme méthodologie des modèles basé sur arbres de régression</i>	113
<i>Figure 4.6 Diagramme méthodologie modèles basés sur réseaux neuronaux.....</i>	114
<i>Figure 4.7 Architecture des deux réseaux neuronaux étude I</i>	116
<i>Figure 4.8 Graphiques de régression, (a) IE, (c) PIP et Bland-Altman (b) IE, (d) PIP, régression linéaire multiple 11 prédicteurs</i>	126
<i>Figure 4.9 Graphiques de régression, (a) IE, (c) PIP et Bland-Altman, (b) IE, (d) PIP, régression LASSO.....</i>	127
<i>Figure 4.10 Graphiques de régression, (a) IE, (c) PIP et Bland-Altman, (b) IE, (d) PIP, régression par moindres carrés partiels PLSR</i>	128
<i>Figure 4.11 Arbre de régression optimisé pour prédire indice d'efficacité.....</i>	130
<i>Figure 4.12 Importance prédicteurs arbre de régression indice d'efficacité.....</i>	131

<i>Figure 4.13 Arbre de régression optimisé pour prédire la proportion d'impulsions positives</i>	131
<i>Figure 4.14 Importance prédicteurs arbre de régression proportion d'impulsions positives</i>	132
<i>Figure 4.15 Graphiques de régression, (a) IE, (c) PIP et Bland-Altman, (b) IE, (d) PIP, forêt aléatoire</i>	133
<i>Figure 4.16 Importance des prédicteurs modèle forêts aléatoires pour prédire l'indice d'efficacité</i>	134
<i>Figure 4.17 Importance des prédicteurs modèle forêts aléatoires pour prédire la proportion d'impulsions positives</i>	134
<i>Figure 4.18 Graphiques de régression, (a) IE, (c) PIP et Bland-Altman, (b) IE, (d) PIP, des réseaux neuronaux</i>	135
<i>Figure 4.19 Importance des prédicteurs algorithme Garson, réseau de neurones indice d'efficacité</i>	136
<i>Figure 4.20 Importance des prédicteurs algorithme Garson, réseau de neurones proportion d'impulsions positives</i>	136
<i>Figure 4.21 Évaluations des hypothèses modèle de régression linéaire multiple pour prédire l'IE, (a) linéarité, (b) normalité, (c) homoscédasticité</i>	143
<i>Figure 4.22 Matrice de corrélation (a) et calcul du VIF (b) modèle de régression linéaire multiple pour la prédiction de l'IE</i>	144
<i>Figure 4.23 Évaluations des hypothèses modèle de régression linéaire multiple pour prédire le PIP, (a) linéarité, (b) normalité, (c) homoscédasticité</i>	145
<i>Figure 4.24 Matrice de corrélation (a) et calcul du VIF (b) modèle de régression linéaire multiple pour la prédiction de PIP</i>	146
<i>Figure 5.1 Schéma du protocole d'étude 2</i>	160

<i>Figure 5.2</i>	<i>Diagramme méthodologie réseau neuronal pour prédire forces de pédalage</i>	163
<i>Figure 5.3</i>	<i>Architecture réseau neuronal pour prédire forces de pédalage</i>	165
<i>Figure 5.4</i>	<i>Diagramme méthodologie réseau neuronal LSTM pour prédire la courbe enveloppe EMG RMS</i>	166
<i>Figure 5.5</i>	<i>Architecture du réseau de neurones LSTM pour prédire la courbe enveloppe EMG RMS</i>	168
<i>Figure 5.6</i>	<i>Résultats prédiction des forces F_x et F_z à 100 W, R^2 (a), RMSE (b)</i>	170
<i>Figure 5.7</i>	<i>Résultats validation croisée à 100 W, R^2 moyenne (a), RMSE moyenne (b)</i>	171
<i>Figure 5.8</i>	<i>Résultats prédiction des forces F_x et F_z à 150 W, R^2 (a), RMSE (b)</i>	172
<i>Figure 5.9</i>	<i>Résultats validation croisée à 150 W, R^2 moyenne (a), RMSE moyenne (b)</i>	173
<i>Figure 5.10</i>	<i>Résultats prédiction des forces F_x et F_z à 200 W, R^2 (a), RMSE (b)</i>	174
<i>Figure 5.11</i>	<i>Résultats validation croisée à 200 W, R^2 moyenne (a), RMSE moyenne (b)</i>	175
<i>Figure 5.12</i>	<i>R^2 (a), RMSE (b) de la prédiction des forces F_x et F_z à puissance variable, et R^2 (c), RMSE (d) de la prédiction des forces F_x et F_z intra-sujets regroupant toutes les puissances</i>	176
<i>Figure 5.13</i>	<i>R^2 (a) et RMSE (b) moyenne et écart-type des forces prédictes F_x et F_z dans les gammes de puissance évaluées</i>	177
<i>Figure 5.14</i>	<i>R^2 (a) et RMSE (b) moyenne et écart-type de la validation croisée dans les gammes de puissance évaluées</i>	178

<i>Figure 5.15 R² (a) et RMSE (b) prédictions de la courbe linéaire enveloppe EMG RMS avec le réseau neuronal LSTM à 100 W.....</i>	179
<i>Figure 5.16 R² (a) et RMSE (b) prédictions de la courbe linéaire enveloppe EMG RMS avec le réseau neuronal LSTM à 150 W.....</i>	180
<i>Figure 5.17 R²(a) et RMSE (b) prédictions de la courbe linéaire enveloppe EMG RMS avec le réseau neuronal LSTM à 200 W.....</i>	181
<i>Figure 5.18 R² (a) et RMSE (b) prédictions de la courbe linéaire enveloppe EMG RMS avec le réseau neuronal LSTM à puissance variable.....</i>	182
<i>Figure 5.19 R² par puissance (a) 100 W, (b) 150 W, (c) 200 W et (d) à puissance variable pour chaque muscle analysé.....</i>	183
<i>Figure 5.20 RMSE par puissance (a) 100 W, (b) 150 W, (c) 200 W à puissance variable pour chaque muscle analysé.....</i>	183
<i>Figure 5.21 Caractéristiques des profils de force Fx et Fz à différentes puissances chez les individus 2 et 8 (ligne bleue mesurée, ligne rouge prédictive).....</i>	185
<i>Figure 5.22 Graphiques Bland-Altman des forces prédictes (a) 100 W Fx, (b) 100 W Fz, (c) 150 W Fx, (d) 150 W Fz, (e) 200 W Fx, (f) 200 W Fz.....</i>	189

Liste des Tableaux

<i>Tableau 2.1</i>	<i>Caractéristiques des principaux muscles impliqués dans le pédalage</i>	62
<i>Tableau 4.1</i>	<i>Caractéristiques moyennes des cyclistes évalués</i>	108
<i>Tableau 4.2</i>	<i>Sélection de prédicteurs, indice d'efficacité « best subset selection »</i>	118
<i>Tableau 4.3</i>	<i>Sélection de prédicteurs, indice d'efficacité « forward selection », « backward selection »</i>	119
<i>Tableau 4.4</i>	<i>Sélection de prédicteurs, indice d'efficacité « hybrid selection », LASSO, Rigide</i>	120
<i>Tableau 4.5</i>	<i>Sélection de prédicteurs, proportion d'impulsions positives « best subset selection »</i>	121
<i>Tableau 4.6</i>	<i>Sélection de prédicteurs, proportion d'impulsions positives « forward selection », « backward selection »</i>	122
<i>Tableau 4.7</i>	<i>Sélection de prédicteurs, proportion d'impulsions positives « hybrid selection », LASSO, Rigide</i>	123
<i>Tableau 4.8</i>	<i>Coefficients régressions linéaires multiples 11 prédicteurs. A. Indice d'efficacité, B. Proportion d'impulsions positives</i>	125
<i>Tableau 4.9</i>	<i>Résultats méthodes régression de composantes principales PCR et régression partielle du dernier carré PLSR</i>	128
<i>Tableau 4.10</i>	<i>Sélection de prédicteurs, méthode élimination récursive des caractéristiques pour prédire l'indice d'efficacité et la proportion d'impulsions positives</i>	129

<i>Tableau 4.11 Caractéristiques des modèles basés sur des arbres de régression...</i>	130
<i>Tableau 4.12 Caractéristiques des forêts aléatoires optimisés</i>	132
<i>Tableau 4.13 Caractéristiques des réseaux neuronaux étude 1</i>	135
<i>Tableau 4.14 Validation croisée étude 1, régression linéaire multiple, régression LASSO.....</i>	137
<i>Tableau 4.15 Validation croisée l'étude 1, régression PLSR et forêts aléatoires ...</i>	138
<i>Tableau 4.16 Validation croisée étude 1, réseau neuronal</i>	139
<i>Tableau 4.17 Synthèse des modèles développés, étude 1</i>	156
<i>Tableau 5.1 Indice d'efficacité des individus évalués.....</i>	186
<i>Tableau 5.2 Moyenne des coefficients, R^2 et de RMSE des forces prédites.....</i>	190
<i>Tableau 5.3 Moyenne R^2 et RMSE validation croisée des forces prédites.....</i>	193
<i>Tableau 5.4 Moyenne des coefficients R^2 et RMSE des muscles évalués.....</i>	198

Liste des Abréviations

Symboles	Description
ADAM	: « Adaptive Moment Estimation »
AIC	: Critère d'information d'Akaike, « Akaike information criterion »
ANN	: Réseaux de neurones artificiel, « Artificial neural network »
BIC	: Critère d'information bayésien, « Bayesian information criterion »
CMV	: Contraction maximale volontaire
sEMG	: Électromyographie de surface, « Surface electromyography »
FFNN	: « Feedforward Neural Network »
IA	: Intelligence artificielle
IMU	: Centrales inertielles, « Inertial measurements units »
IE	: Indice d'efficacité, « index of effectiveness »
ISB	: « International Society of Biomechanics »
KOPS	: « Knee Over Pedal Spindle »

LASSO	:	Opérateur de sélection et de retrait le plus absolu « least absolute shrinkage and selection operator »
LSTM	:	Mémoire à long terme, « Long Short-Term Memory »
L-TIPS	:	Laboratoire de technologies & d'innovation pour la performance sportive
MAE	:	« Mean absolute error »
MSE	:	Erreur quadratique moyenne, « Mean Squared Error »
NARX	:	Modèles autorégressifs non linéaires avec entrée exogène, « Nonlinear autoregressive models with exogenous input »
OMC	:	« Optical motion capture »
PCA	:	« Principal component analysis »
PIP	:	Proportion d'impulsions positives, « Positive impulse proportion »
PLSR	:	« Partial least squares regression »
PMB	:	Point mort bas
PMH	:	Point mort haut

r	:	Coefficient de corrélation de Pearson
RFE	:	« Recursive feature elimination »
ReLU	:	« Rectified linear unit »
RNN	:	Réseau neuronal récurrent, « Recurrent neural network »
RMS	:	Valeur moyenne quadratique, « Root mean square »
RMSE	:	Erreur quadratique moyenne, « Root mean squared error »
RMSProp	:	« Root Mean Square Propagation »
rpm	:	Révolutions par minute, « Revolutions per minute »
RSE	:	Erreur standard résiduelle, « Residual standard error »
SENIAM	:	ElectroMyoGraphie de surface pour l'évaluation non invasive des muscles, « Surface ElectroMyoGraphy for the Non-Invasive Assessment of Muscles »
SVM	:	« Support vector machines »
VIF	:	« Variance inflation factor »
XAI	:	« Explainable Artificial Intelligence »

CHAPITRE 1 - Introduction

Pour comprendre comment les humains produisent une force par l'intermédiaire de leurs membres supérieurs ou inférieurs, indépendamment de l'activité physique effectuée, une analyse complète du mouvement des segments du corps et des muscles impliqués qui produisent ces forces est indispensable. Le domaine scientifique qui permet l'analyse du mouvement humain est la biomécanique, qui se définit comme l'application de la mécanique classique à divers problèmes biologiques [1]. Le rôle de la biomécanique est de créer des modèles mathématiques qui nous permettent de comprendre comment le mouvement est généré, et quels sont ses effets sur les interfaces physiques avec lesquelles le corps humain est en contact. La biomécanique s'intéresse également à l'étude des structures du système musculosquelettique, telles que les articulations, les tendons et les muscles. En biomécanique, il existe le champ de la cinématique, qui consiste à décrire les aspects géométriques et temporels du mouvement, sans se préoccuper des forces qui en sont la cause [1]. D'autre part, l'analyse des forces internes ou externes qui provoquent le mouvement est connue sous le terme de cinétique [2]. Par conséquent, il est de grand intérêt d'appliquer ces approches au cyclisme, sport de grande expansion au niveau professionnel et amateur, parce que la production de la force de pédalage est le facteur le plus important dans la génération de puissance sur les manivelles du vélo. Plus précisément, il est important d'étudier l'interface pied-pédale où sont appliquées les forces produites par les membres inférieurs. À partir de ces forces, il est possible de définir des paramètres pour évaluer la technique de pédalage, tels que l'indice d'efficacité (IE) et la proportion d'impulsion positive (PIP), entre autres [3]. Toutefois, ces approximations conduisent à se poser plusieurs questions, par exemple si la vitesse angulaire et

l'accélération angulaire sont liées à ce phénomène. En particulier, il serait important de déterminer quels points et valeurs cinématiques lors de la rotation de la manivelle, sont les plus importants dans le phénomène de production de force. Il est possible de collecter des données pour créer des modèles mathématiques permettant de comprendre comment la force de pédalage est produite. En conséquence, dans ce projet de recherche, l'objectif est de présenter les méthodologies innovantes qui peuvent intégrer toutes les sources d'information lors de l'analyse du mouvement d'un cycliste, afin de faire l'estimation des variables et comprendre le phénomène de production de force en cyclisme.

1.1 Mise en contexte

Depuis plusieurs années, les chercheurs, les entraîneurs et les professionnels du sport s'intéressent à la compréhension des phénomènes biomécaniques qui se produisent lors des mouvements générés par une activité physique. En cyclisme, il est intéressant de comprendre comment les mouvements des membres inférieurs génèrent les forces produites au niveau des manivelles. En connaissant la relation entre ces phénomènes, il est possible de donner les clés pour optimiser la technique de pédalage afin d'obtenir de meilleures performances. Pour garantir cette optimisation, il est nécessaire d'intégrer différents domaines de connaissances et techniques. Un premier domaine à considérer sont les sciences du sport, en tant que domaine pluridisciplinaire visant à comprendre et à améliorer les performances d'un être humain dans le cadre d'une activité sportive [4]. Dans cette optique, les entraîneurs, les intervenants et les cyclistes s'intéressent à l'optimisation de la technique de pédalage, qui influencent les performances [5], les méthodes d'entraînement [6], et la physiologie de l'activité physique [7]. Dans ce domaine se trouve la biomécanique, pour laquelle les structures osseuses, les muscles, et les articulations sont

considérés comme des corps rigides interagissant avec le milieu environnant. De cette manière, grâce aux méthodes biomécaniques appliquées au cyclisme, il est possible d'analyser les mesures cinématiques et cinétiques lors de l'exécution du mouvement de pédalage [8]. Ces mesures cinématiques et cinétiques peuvent être obtenues à l'aide de technologies spécifiquement adaptées au cyclisme. Dans ce contexte, la cinématique est réalisée en capturant le mouvement à l'aide de caméras qui est la technique la plus utilisée pour évaluer la position d'un cycliste [9]. De plus, pour connaître la cinétique du pédalage, il est nécessaire de mesurer les forces produites par l'athlète et, à cette fin, des pédales équipées de capteurs sont utilisées [10]. Toute cette capture de mouvement doit également être post-traitée à l'aide de techniques de traitement du signal, afin d'être correctement visualisés et analysés. Un autre domaine de connaissances intégré dans cette recherche est la bio-instrumentation, définie comme la méthodologie permettant d'obtenir des mesures invasives ou non invasives du corps humain [11]. Ce domaine fait partie de l'ingénierie, et plus particulièrement de l'ingénierie biomédicale. Un exemple est la capture de l'activité musculaire, qui peut être visualisée au moyen d'un traitement mathématique du signal. En cyclisme, la capture des signaux musculaires a été utilisée pour comprendre quels muscles sont activés pendant le pédalage [12]. L'ingénierie fournit également des programmes ou des applications, c'est-à-dire des outils technologiques, qui facilitent l'exécution des tâches et permettent d'économiser du temps et des ressources. Bien que ces approches de l'analyse biomécanique se soient considérablement développées ces dernières années, divers facteurs, notamment économiques et techniques, font que, dans la plupart des cas, l'analyse de la cinétique et du recrutement musculaire n'est pas effectuée. Par conséquent, la principale motivation de ce projet est de proposer des alternatives innovantes basées sur l'apprentissage automatique afin de réaliser une analyse intégrale de la biomécanique du

pédalage. Dans ce contexte, l'objectif global de ce projet, est de développer des outils technologiques capables d'intégrer tous les domaines de connaissance, et toutes les sources d'information susmentionnés, pour analyser la technique de pédalage. Ces outils technologiques pourront guider les entraîneurs, les cyclistes et les intervenants pour optimiser le mouvement ou l'activité musculaire pendant le pédalage, et trouver les points d'amélioration de l'athlète. La figure 1.1 montre le contexte général de ce projet.

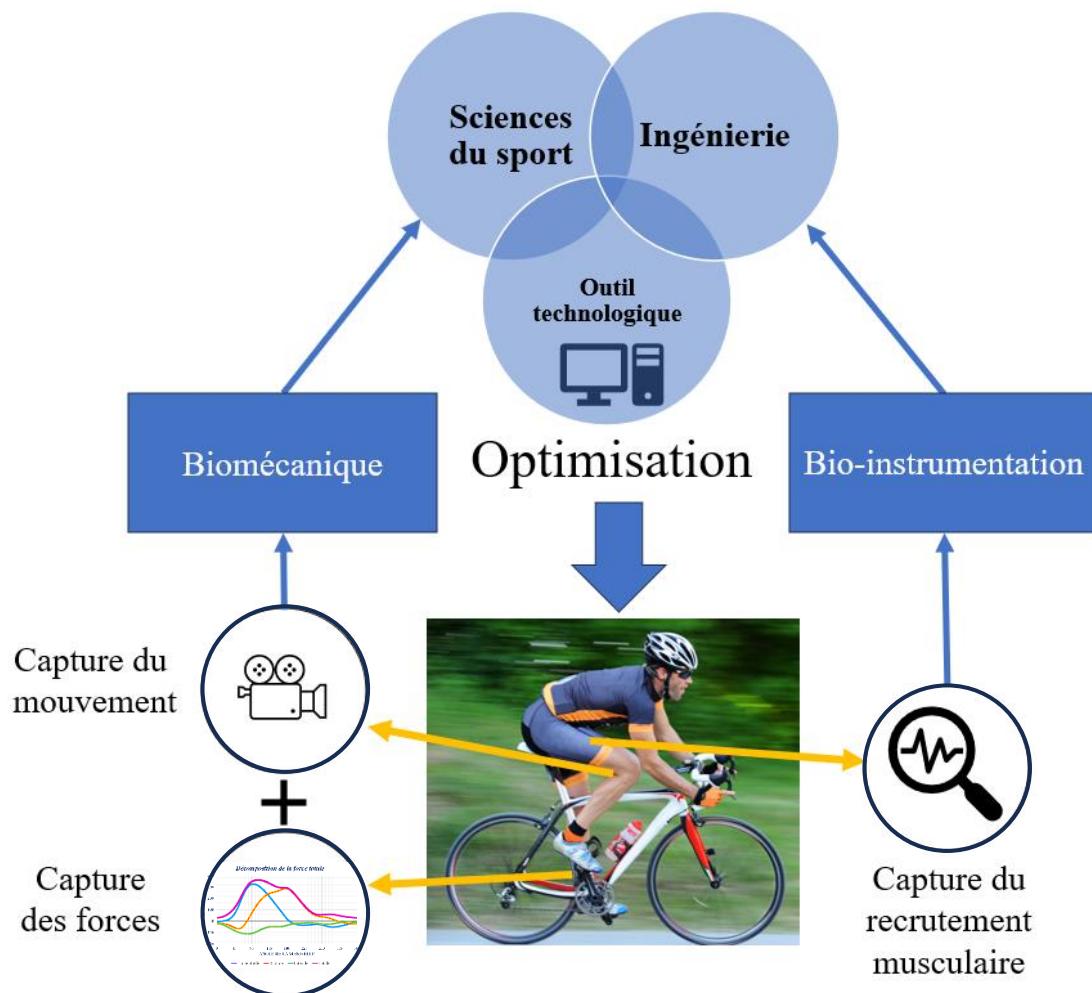


Figure 1.1 : Contexte général de la recherche

Ces outils de modélisation du mouvement pourraient être intégrés dans les systèmes de capture couramment utilisés dans le cyclisme. De cette manière, l'analyse biomécanique sera plus complète, ce qui permettra de comprendre le phénomène de manière plus intégrale. Finalement, compte tenu de la grande quantité de données disponibles, les outils technologiques pourraient s'inspirer de la science des données, pour extraire des informations sur le phénomène étudié et prendre des décisions optimales.

1.2 Problématique de recherche

En général, la capture de mouvement d'un cycliste est utilisée pour effectuer le positionnement correct sur son vélo. À cette fin, la littérature a établi certains angles du membre inférieur ou supérieur qui sont liés au confort, à une technique de pédalage correcte et à la minimisation du risque de blessure [9]. À partir de ces analyses, les éléments mécaniques du vélo sont ajustés (hauteur de la selle, position horizontale de la selle, position du guidon, etc.), ce qui permet théoriquement d'optimiser la biomécanique du pédalage. Dans la majorité des cas, lors d'une séance de positionnement, l'angle du genou dans le plan latéral, le déplacement latéral du genou dans le plan frontal et la position horizontale du genou par rapport à la pédale, sont principalement analysés. Néanmoins, Winter [13] mentionne que le corps humain est un système neuro-musculo-squelettique complexe interconnecté, et qu'il est impossible de comprendre l'ensemble de son fonctionnement avec une seule variable et une seule articulation pendant l'exécution d'un mouvement. En outre, bien que les systèmes de capture puissent fournir des informations sur les vitesses et les accélérations angulaires du mouvement, cette information peu souvent analysée. En conséquence, leur importance pour l'analyse de la biomécanique du pédalage n'est pas évidente.

Concernant la cinétique du pédalage, dans certaines séances de positionnement, des pédales instrumentées sont utilisées pour mesurer les forces de péodalage, mais leur utilisation est très limitée en raison du coût élevé et du manque de fournisseurs de cette technologie. Actuellement, un seul fournisseur est actif, ce qui limite l'acquisition par les intervenants et entraîneurs. Bien qu'il existe des pédales instrumentées économiques telles que l'Assioma Favero (Favero Electronics Srl, Arcade, Veneto, ITALY), elles ne fournissent pas de graphiques des forces tangentialles et radiales exercées sur les pédales pour l'analyse, ni de données brutes pour une analyse plus approfondie. C'est pourquoi ces pédales n'ont probablement pas été utilisées comme outil de travail dans le processus de positionnement des cyclistes. En remplacement, des semelles instrumentées sont parfois utilisés, mais il n'existe pas d'études de validité ou de fiabilité, ni de données normatives pour guider l'utilisation de ces semelles [9]. En conséquence, dans la majorité des cas, lors d'une session de positionnement, seul le point de vue cinématique est pris en compte, mais pas le point de vue cinétique, c'est-à-dire les forces appliquées aux pédales.

Jusqu'à présent, seul le facteur cinématique et cinétique a été discuté, mais un autre facteur à analyser dans la posture du cycliste est le recrutement musculaire. Le mouvement de péodalage est le résultat du recrutement musculaire des membres inférieurs. Le recrutement musculaire est défini comme le processus par lequel le système nerveux active différentes unités motrices au sein d'un muscle pour produire différents niveaux de forces. Il existe des différences de recrutement entre les cyclistes experts et les cyclistes récréatifs [14]. Malgré leur utilité, encore aujourd'hui, les systèmes d'analyse biomécaniques pour connaître les activations musculaires et les systèmes de capture 3D sont considérés trop coûteux pour un usage général [15], par les intervenants et entraîneurs en cyclisme. De plus, les sessions d'analyse nécessitent un temps de

préparation et de traitement prolongé, ce qui entraîne une augmentation des coûts d'intervention. Par conséquent, il reste incertain si la cinématique attendue des membres inférieurs pour un positionnement optimal est directement liée au recrutement musculaire pendant le pédalage. En l'absence de données précises, les analyses se fondent généralement sur une simple observation visuelle de l'activation musculaire lors du pédalage, une méthode qui demeure loin d'être optimale.

Cette situation limite la compréhension de la relation entre les forces exercées sur les pédales et le recrutement musculaire. Dans ce contexte, il n'existe pas des outils technologiques accessibles qui permettent de relier la cinématique articulaire et la cinétique du pédalage à l'activité musculaire, afin d'analyser d'une manière plus holistique la technique du pédalage.

Bien que ces méthodologies soient disponibles pour fournir des informations sur le mouvement de pédalage, elles ne sont pas souvent intégrées et par conséquent, même aujourd'hui, les connaissances sur la manière dont la force de péodalage est générée sont encore limitées. En général, les technologies disponibles pour les analyses biomécaniques, fonctionnent de manière indépendante et il est difficile de guider le cycliste dans l'optimisation de sa technique de pédalage sans un outil capable de les intégrer. Par conséquent, dans la plupart des cas, il reste des inconnues liées à l'impact des changements de posture sur les forces délivrées aux pédales ou sur l'activation des muscles. Dans ce contexte, l'optimisation de la biomécanique du pédalage est limitée et peu explorée lors d'une session de positionnement puisqu'il manque des informations détaillées sur le mouvement. En outre, il existe peu d'outils biomécaniques basés sur la cinématique à disposition des cyclistes et des entraîneurs pour les aider à optimiser leur technique de pédalage. Il existe un outil basé sur des centrales inertielles appelé Leomo (TYPE-R,

Leomo, Boulder, CO, USA), mais sa validité scientifique a été contestée et des erreurs significatives ont été observées dans les mesures [16, 17]. Pour cette raison, en général, les professionnels du positionnement analysent visuellement le geste du pédalage afin de qualifier la technique. Plus précisément, cela constitue une analyse qualitative qui dépendra en grande partie de l'expertise de la personne effectuant l'analyse.

1.3 Proposition méthodologique

Considérant que les technologies de capture du mouvement, la capture des forces de pédalage et les mesures non invasives du corps humain telles que l'électromyographie de surface, fournissent une bonne quantité de données, une nouvelle alternative, avec une tendance croissante en cyclisme et en biomécanique, consiste à utiliser l'apprentissage automatique pour prédire et analyser les variables d'intérêt. Cette méthodologie pourrait constituer le modèle qui relie les variables cinématiques à la force de pédalage et à l'activité musculaire. L'apprentissage automatique offre de nombreuses possibilités, car il est possible d'appliquer différents algorithmes mathématiques qui nous permettent de déterminer la relation entre les variables ou de prédire les variables d'intérêt. À partir de cette relation, il est possible de comprendre les facteurs qui influencent les variables à étudier et de répondre ainsi aux questions de recherche. Plus précisément, l'apprentissage automatique est un domaine de l'intelligence artificielle qui vise à apprendre dans le but de prédire des données sur la base de leurs caractéristiques. [18]. Pour réaliser cet apprentissage, différents algorithmes mathématiques sont utilisés à l'aide d'une quantité de données.

Ces algorithmes d'apprentissage permettent d'intégrer les différentes sources de données. En effet, ces approches constituent la principale innovation de cette recherche,

car il n'existe actuellement aucun outil qui utilise les données provenant du système de capture de mouvement pour estimer les variables liées à la cinétique du pédalage. En d'autres termes, prédire la cinétique du pédalage à partir de la cinématique du membre inférieur. Pour ce faire, un post-traitement correct des données et l'application de protocoles de capture stricts doivent être effectués afin de garantir une base de données appropriée pour la modélisation. De même, la synchronisation et la fréquence de la capture font partie de ce conditionnement nécessaire des données en raison des différents types de capteurs utilisés dans le cadre de la recherche. Dans ce projet, afin d'obtenir les données nécessaires pour élaborer les modèles, deux autorisations éthiques ont été obtenues, avec lesquelles des protocoles de capture de données ont été mis en œuvre sous différentes conditions afin d'alimenter les modèles d'apprentissage automatique. J'applique plusieurs modèles d'apprentissage automatique pour répondre aux questions de recherche proposées dans le projet. Étant donné que chaque modèle utilise des algorithmes différents, une réponse distincte peut être proposée, mais des points communs sont trouvés. Conjointement, ces modèles apportent des réponses à la question fondamentale de la recherche concernant les principaux facteurs influençant la production de force en cyclisme et l'optimisation de la technique de pédalage.

1.4 Objectives de la recherche et contributions

1.4.1 Objectives de la recherche

Pour résoudre le problème d'analyse du cycle de pédalage, il est essentiel de trouver des modèles mathématiques basés sur des méthodes d'apprentissage automatique intégrant les variables cinématiques, cinétiques et électromyographiques. Par conséquent, voici

l'objectif proposé : développer des modèles prédictifs pour identifier les facteurs qui influencent la production de force et le recrutement musculaire en cyclisme.

Objectifs spécifiques :

- Développer des outils informatiques pour le traitement des données cinématiques, cinétiques et électromyographies.
- Développer un modèle prédictif des métriques de pédalage comme l'indice d'efficacité et la proportion d'impulsion positive à partir de la cinématique du pédalage pour identifier les facteurs qui influencent la production de force.
- Développer un modèle basé sur des méthodes d'apprentissage automatique pour prédire les forces de pédalage à partir de la cinématique.
- Développer un modèle basé sur des méthodes d'apprentissage automatique pour prédire le recrutement musculaire à partir de la cinématique et de la cinétique du pédalage.

1.4.2 Principales contributions scientifiques

Au cours de l'élaboration de cette thèse, plusieurs étapes ont été abordées. La première étape a consisté à comprendre les phénomènes de génération de force en cyclisme. Pour ce faire, un modèle de dynamique inverse a été mis en place afin de déterminer les valeurs du moment articulaire et de la puissance du membre inférieur dans le plan sagittal. Cette première étape a permis de comprendre le phénomène du mouvement de pédalage afin de développer les objectifs proposés et de concevoir les contributions scientifiques.

De cette manière, deux études ont été menées dans lesquelles des données cinématiques, cinétiques et électromyographiques ont été collectées avec différents protocoles. La première étude s'est concentrée sur la recherche de modèles basés sur l'apprentissage automatique afin de prédire les paramètres qui évaluent la technique de pédalage. Cela a permis de trouver plusieurs modèles qui permettent de prédire de manière satisfaisante ces mesures, ce qui constitue une contribution scientifique importante dans le domaine du sport. Les résultats obtenus dans cette étude nous ont permis de trouver les facteurs cinématiques les plus importants qui influencent la production de force en cyclisme.

Une deuxième contribution scientifique est obtenue avec l'étude 2. Dans cette étude, un modèle basé sur des réseaux neuronaux a été développé permettant de prédire les forces de pédalage à partir de la cinématique. Cela permettra de remplacer l'utilisation de pédales instrumentées lors d'une séance de pédalage. Un deuxième modèle basé sur des réseaux neuronaux récurrents a également été développé, qui permet de prédire l'activité musculaire de quatre muscles des membres inférieurs, ce qui constitue une autre occasion de fournir des outils permettant d'effectuer ces types d'analyse en une seule séance d'analyse.

Toutes ces contributions scientifiques ouvrent de nouveaux horizons à l'apprentissage automatique appliqué au cyclisme, qui peuvent également être appliqués à d'autres sports et qui pourraient populariser l'utilisation de ces outils aux intervenants et aux entraîneurs. Ci-dessous est présentée en détail les contributions scientifiques acceptées, les articles en préparation et les autres contributions faites au cours de cette recherche.

1.4.2.1 Articles de journaux publiés

- **Torres, A.**, Nougarou, F., & Domingue, F. (2024). Predicting pedalling metrics based on lower limb joint kinematics. Computer Methods in Biomechanics and Biomedical Engineering, 1-15, doi: 10.1080/10255842.2024.2371044.

1.4.2.2 Articles de conférences

- **Torres, A.**, Nougarou, F., & Domingue, F. (2024). Machine learning models to predict kinetic variables in cycling. MLR, 23, 0-88. Science & Cycling Conference, Florence 2024.

1.4.2.3 Chapitre de livre

- **Torres A.**, Yepez M. A., Millour G., Nougarou F., Domingue F. Machine learning for prediction of the index of effectiveness in cycling (2025). Artificial Intelligence, Optimization, and Data Sciences in Sports. doi: 10.1007/978-3-031-76047-1_3

1.4.2.4 Articles en tant que second auteur

- Millour, G., **Velásquez, A. T.**, & Domingue, F. (2023). A literature overview of modern biomechanical-based technologies for bike-fitting professionals and coaches. International Journal of Sports Science & Coaching, 18(1), 292-303, doi: 10.1177/17479541221123960.

1.4.2.5 Publication soumise

- **Torres, A.**, Nougarou, F., & Domingue, F. (2025). Predicting pedalling forces based on lower limb joint kinematics with a feed forward neural network. Computer Methods in Biomechanics and Biomedical Engineering, annexe P.

1.4.2.6 Publication à soumettre

- **Torres, A.**, Coqueret L. Nougarou, F., & Domingue, F. (2025). Predicting muscular recruitment during bicycle pedalling with recurrent neural network. Computer Methods in Biomechanics and Biomedical Engineering, annexe Q.

1.4.2.7 Autres contributions

- Au cours du développement de cette recherche, j'ai participé à différentes activités académiques, conférences et cours. De plus, avec toutes les connaissances acquises sur l'apprentissage automatique, j'ai organisé deux écoles d'été internationales sur la science des données pour le sport et la santé. La première a été organisée conjointement avec l'Universidad Autónoma de Occidente et l'Escuela Nacional del Deporte à Cali, en Colombie, du 29 juillet au 2 août 2024. La deuxième école d'été a été organisée à l'Escuela Colombiana de Ingeniería Julio Garavito à Bogota, en Colombie, du 29 juillet au 2 août 2024. Les deux écoles d'été ont duré de 40 heures. Ces contributions figurent aux annexes A et B.

1.5 Originalité de la recherche

Actuellement, la science des données permet de modéliser n'importe quel phénomène et, par conséquent, de plus en plus d'applications sont trouvées dans le domaine de la biomécanique et dans l'étude du mouvement humain. Un aspect innovant de cette recherche est l'utilisation de la science des données pour modéliser la production de force dans le cyclisme. Cet aspect a été peu développé dans l'environnement du cyclisme, en particulier dans la biomécanique du pédalage. La science des données est une approche multidisciplinaire qui combine la statistique, la programmation et l'apprentissage automatique.

De cette manière, l'apprentissage automatique est un outil qui a été mis en œuvre dans ce projet de manière holistique, en cherchant non seulement à prédire les métriques d'intérêt, mais aussi mettre en œuvre d'autres outils de l'apprentissage automatique qui permettent de sélectionner les meilleurs prédicteurs qui sont utilisées pour prédire ces métriques. Cet aspect constitue l'une des principales originalités de ce travail, car les modèles développés permettront de déterminer clairement les facteurs biomécaniques essentiels pour la production de force en cyclisme, par le développement de modèles explicables ou interprétables tels que la régression et les arbres de régression. De manière originale, plusieurs techniques de sélection de variables, à travers différents algorithmes mathématiques, ont été mises en œuvre pour trouver plusieurs modèles et sélectionner les meilleurs. Cela permettra de mieux comprendre comment la force est produite dans le cyclisme et offrira plusieurs possibilités pour comprendre le phénomène.

Un autre aspect innovant est qu'il sera possible de déterminer les métriques cinétiques liées à la technique de pédalage à partir des valeurs cinématiques qui peut être facilement

capturé dans le cyclisme. À cette fin, ce projet fournira des outils informatiques innovants, basés sur des techniques d'apprentissage automatique, qui peuvent être facilement intégrés dans les systèmes de capture de mouvements disponibles dans le commerce. Ces modèles mathématiques permettront aux intervenants du cyclisme de mettre en œuvre des outils simples à appliquer, rentables et efficaces. Ce projet devrait également contribuer à la réduction des coûts grâce à la possibilité de remplacer des technologies coûteuses.

Autre innovation de ce projet, est la mise en œuvre de l'apprentissage profond, qui est une branche de l'apprentissage automatique où sont utilisés des réseaux neuronaux artificiels. Ces outils visent à prédire les signaux temporels tels que la force de pédalage et la courbe d'enveloppe de l'activité musculaire de certains des muscles les plus importants du membre inférieur du cycliste. Cet aspect permet d'explorer des outils innovants plus complexes comme les réseaux neuronaux récurrents, pour prédire des signaux comme les séries temporelles. De cette manière, les participants et les entraîneurs pourront accéder à plus d'informations sur la biomécanique du cycliste et ainsi compléter les analyses visant à optimiser la technique de pédalage.

1.6 Organisation de la thèse

À cette fin, la présente thèse est organisée en commençant par un premier chapitre dans lequel la présentation du problème est exposée en introduisant principalement les différents domaines de connaissances qui sont impliquées dans le développement du projet de recherche. Postérieurement, les objectifs et les contributions scientifiques du projet jusqu'à le présent sont énumérées, puis complétés par les aspects originaux du projet.

Le deuxième chapitre présente les différentes approches méthodologiques mises en œuvre dans le domaine de la biomécanique. Ensuite, les bases théoriques de la biomécanique du cyclisme sont présentées, exposant les trois visions principales mises en œuvre dans le projet, à savoir la cinématique, la cinétique et le recrutement musculaire lors du pédalage. Par la suite, le chapitre se concentre sur l'apprentissage automatique, présentant les techniques de sélection de variables, et les différentes techniques de prédiction appliquées dans le projet.

Le troisième chapitre présente la technologie utilisée pour la capture des données en exposant les caractéristiques techniques de chacune d'entre elles.

Ensuite, afin d'atteindre les objectifs proposés, la thèse se poursuit sur la présentation des deux études développées. De cette manière, le chapitre 4 présente l'étude 1, dont l'objectif est la prédiction des métriques qui évaluent la technique de pédalage. Dans cette étude, plusieurs techniques d'apprentissage automatique ont été implémentées, montrant les avantages et limitations de chacune d'entre elles. Trois publications découlent de cette étude, citées aux points 1.4.2.1, 1.4.2.2 et 1.4.2.3. Ensuite, dans le chapitre 5, est présentée l'étude 2, dans laquelle est abordée la prédiction de la série temporelle des forces appliquées sur la pédale. Cette étude conduit à une publication qui a été soumise à évaluation, comme indiqué au point 1.4.2.5. En outre, dans cette étude, des travaux de recherche ont été menés sur la prédiction du recrutement musculaire de certains muscles du membre inférieur. De cette étude découle une autre publication qui est en préparation et qui est citée au point 1.4.2.6. Pour ces deux études, deux types de réseaux de neurones ont été implantés. Dans chaque étude, les différentes étapes méthodologiques réalisées sont présentées, en commençant par le protocole de collecte des données, le traitement du

signal, la sélection de variables, l'application des modèles d'apprentissage automatique, la validation des modèles, les résultats et les analyses des résultats.

Un sixième chapitre présente les conclusions finales, complétées par les limitations, perspectives et recommandations.

Enfin, le présent document est complété par les annexes pertinentes pour appuyer et compléter les résultats obtenus dans ce projet de recherche.

De cette manière, cette thèse cherche à présenter une vision holistique de l'apprentissage automatique centrée sur la biomécanique du cyclisme, laissant ouverte la possibilité d'appliquer cette méthodologie à d'autres sports ou mouvements du corps humain.

CHAPITRE 2 - État de l'art

2.1 Biomécanique du cyclisme

La biomécanique en cyclisme analyse la position du cycliste, en se concentrant principalement sur la recherche de la bonne géométrie du vélo qui permet un mouvement de pédalage optimal. Les auteurs [8], affirment que la biomécanique du cyclisme couvre un vaste domaine lié aux forces exercées sur les pédales, aux angles optimaux des membres supérieurs et inférieurs et au recrutement musculaire. De cette manière, le geste de pédalage se caractérise par un mouvement circulaire de la manivelle provoquée par des mouvements de flexion-extension de la hanche, du genou et de la cheville. Ce mouvement est transmis à la manivelle par l'intermédiaire de la pédale sur laquelle se trouve un mécanisme d'ancrage qui permet de transmettre de manière efficace, les moments des forces délivrées par les articulations du membre inférieur. Cependant, avant d'analyser comment ces forces sont transmises à la pédale, il est important de revenir sur le mécanisme de propulsion et la géométrie sur laquelle le cycliste est positionné sur le vélo.

2.1.1 Mécanisme de propulsion du vélo

La propulsion d'un vélo entre dans l'histoire de son invention bien après la première conception d'une structure de deux roues, à la suite de nombreuses tentatives. Bien qu'il existe des croquis de Léonardo de Vinci dans lesquels il est possible d'apercevoir un système de pédales des manivelles et de chaîne, certains auteurs mentionnent l'invention de Eugène Meyer et André Guilmet, brevetée en 1868, en tant que premier système de propulsion de vélo similaire au système utilisé jusqu'à présent [19]. Cependant,

l'authenticité du premier brevet étant mise en doute, et un autre brevet daté de 1879 est déposé, attribué à Henry John Lawson qui présente le même système propulsion [19]. La figure 2.1 illustre ce concept.



Figure 2.1: Vélo d'Henry John Lawson 1879 [19]

Certainement ce système de propulsion a suscité un changement radical au cours de ces années car le mécanisme de propulsion du vélo est principalement basé sur le mouvement du membre inférieur autour des manivelles qui peut être atteint de manière plus sûre. Auparavant les pédales étaient situées sur la roue avant, ce qui demandait beaucoup d'habileté et le cyclisme était plus sujet aux accidents. Bien qu'il s'agisse d'un mouvement relativement simple, de nombreux facteurs permettent d'effectuer ce mouvement de manière appropriée. Cependant, comme le montre la figure 2.1, la différence de diamètre des roues a conduit à une spécification complexe de la conception du cadre et, par conséquent, de la position de la selle. Cela introduit bien sûr l'autre concept important dans l'évolution du vélo, la géométrie du cadre.

2.1.2 Géométrie du cadre

Ce concept relie la position du corps du cycliste par rapport au système de pédales et de manivelles, afin de rendre ce mouvement de propulsion efficace. La géométrie en forme de diamant, qui caractérise aujourd'hui le cyclisme moderne, est probablement une caractéristique remarquable du modèle industrialisé à partir de 1892 [19], après de nombreux concepts. Cette géométrie garanti la mise en place de roues de même diamètre et une position de la selle qui permet au cycliste de positionner confortablement ses bras et ses membres inférieurs, comme le montre la figure 2.2.

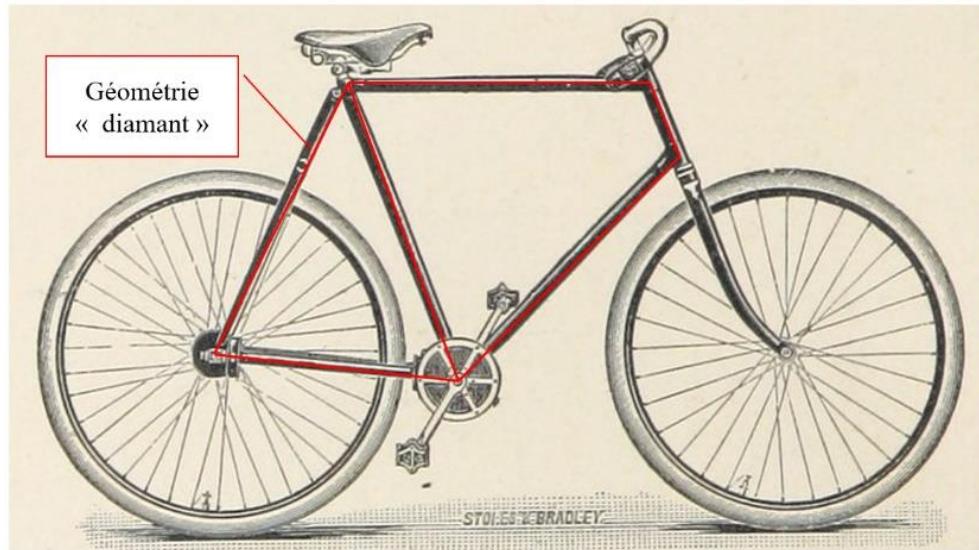


Figure 2.2 : Géométrie du vélo des années 1892, adapté de [21]

À partir de ces deux concepts, la géométrie et le système de propulsion, qui ont évolué jusqu'à aujourd'hui, il est intéressant de comprendre l'interaction du cycliste avec le vélo et de comprendre le phénomène de mouvement. En effet, pour propulser le vélo, le cycliste applique une force sur chaque pédale avec le mouvement du membre inférieur. En d'autres

termes, à travers du mécanisme qui connecte le pied à la pédale, une force est produite. Pour les fins de l'analyse dans le cadre de cette recherche, la force appliquée sur la pédale est divisée en trois composantes F_x, F_y, F_z , selon la référence des axes cartésiens, comme le montre la figure 2.3.

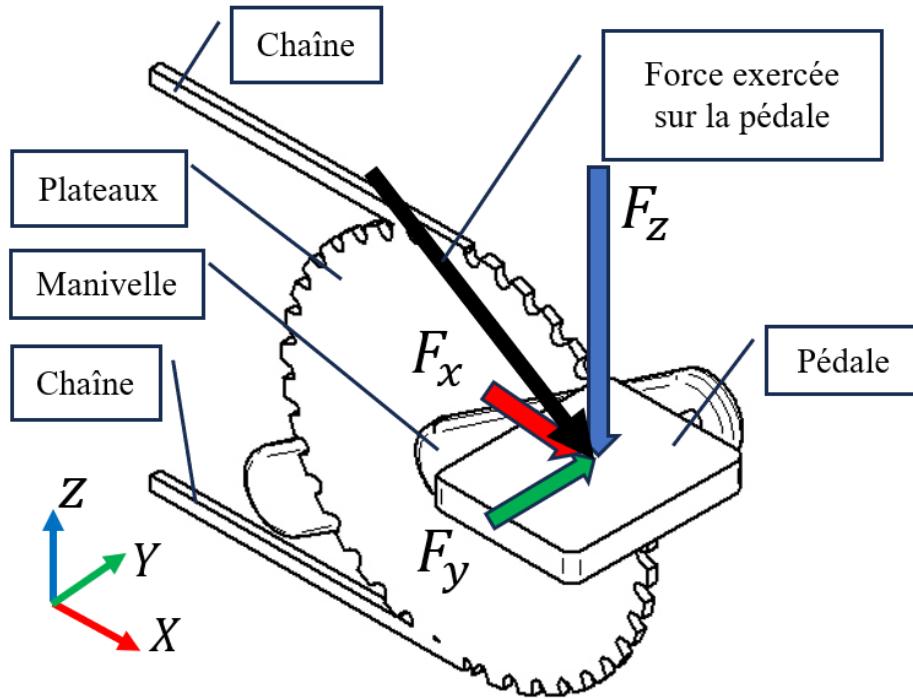


Figure 2.3 : Schéma de la force appliquée sur la pédale

Pour comprendre le comportement de ces forces et leur impact sur la technique de pédalage, il faut les aborder sous le concept utilisé en biomécanique et connu sous le nom de cinétique du pédalage.

2.1.3 Cinétique du pédalage

En général, en biomécanique, il est utilisé des plaques instrumentées pour mesurer les forces, auxquelles sont fixés des capteurs piézorésistifs ou des jauge de contrainte. En cyclisme, les pédales instrumentées intégrant des jauge de contrainte sont conçues pour mesurer les forces appliquées aux pédales dans les trois directions cartésiennes F_x, F_y, F_z de la figure 2.3. À partir de ces forces et en utilisant les formules trigonométriques pour la rotation des axes, il est défini, comme le montre la figure 2.4 (a), la force effective F_{eff} , qui est la force perpendiculaire à la manivelle. De plus, il est défini la force radiale F_R opérant le long du bras de manivelle, en direction ou à l'opposé du pédalier. En outre, la force latérale F_L est définie comme la force perpendiculaire sur le plan colinéaire formé entre F_{eff} et F_R . La figure 2.4 (b) montre la décomposition des angles et des forces F_x et F_z à partir desquels les forces F_{eff} et F_R sont calculées.

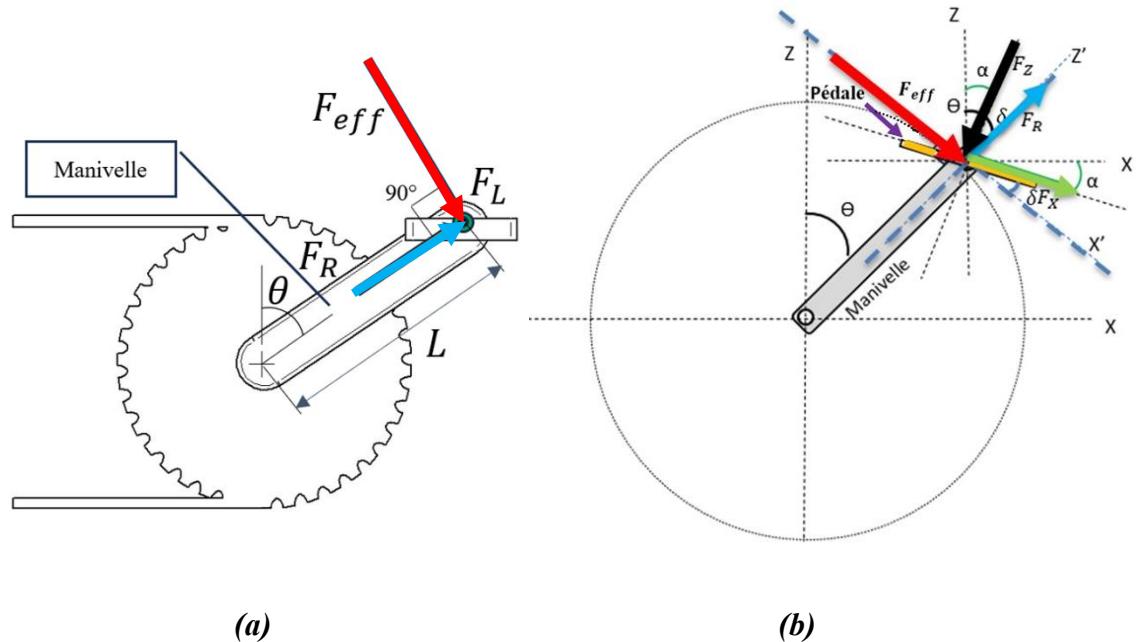


Figure 2.4 : Définition des forces F_{eff} , F_R , F_L (a), et décompositions trigonométriques des forces F_x et F_z (b)

Pour une meilleure clarté, la figure 2.5 montre la décomposition isométrique des forces F_{eff} , F_R , F_L avec les différentes composantes de la manivelle.

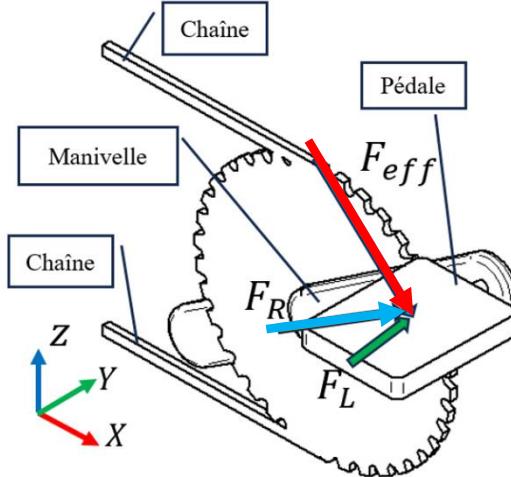


Figure 2.5 : Définition des forces appliquées à la manivelle en isométrique

La figure 2.4 permet de définir l'angle δ (équation 2.1) et les équations trigonométriques de rotation des angles qui permettent de calculer F_{eff} , F_R , (équations 2.2 et 2.3).

$$\delta = \theta - \alpha \quad (2.1)$$

$$F_{eff} = F_x \times \sin(\delta) + F_z \times \cos(\delta) \quad (2.2)$$

$$F_R = F_x \times \cos(\delta) - F_z \times \sin(\delta) \quad (2.3)$$

Où l'angle α est l'angle de la pédale par rapport aux axes cartésiens X-Z.

Cette détermination des forces permet de définir la puissance de pédalage, qui est un paramètre crucial pour mesurer la performance en cyclisme. L'équation de la puissance

est déterminée par le moment de force délivré aux manivelles multiplié par la vitesse angulaire, équation 2.4. En parallèle, le moment de force est défini comme la force effective F_{eff} multipliée par la longueur de la manivelle L .

$$P = F_{eff} \times L \times \omega \quad (2.4)$$

Où P puissance (en watts), F_{eff} est la force effective (en newton N), L la longueur de la manivelle (en mètres m) et ω est la vitesse angulaire (en rad / sec).

Selon la formule, un facteur critique dans la production de la puissance de pédalage est la force effective F_{eff} , d'où l'importance d'étudier la façon dont le cycliste gère cette force tout au long de la rotation de la manivelle. En analysant les forces de pédalage, il est possible de constater que la force effective, la force radiale et la force latérale varient en fonction de la rotation de la manivelle, ce qui se traduit par les profils illustrés à la figure 2.6. Il est illustré dans la figure un cycliste pédalant à 245 W et 80 RPM.

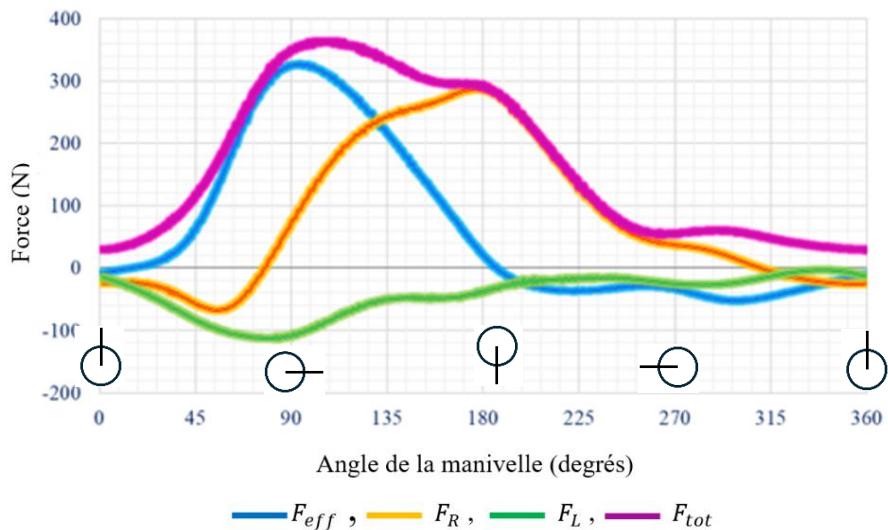


Figure 2.6 : Profils des forces appliquées à la manivelle

À partir de ces forces, il est possible de définir la force totale F_{tot} comme la résultante de toutes les forces appliquées à la manivelle, conformément à l'équation 2.5.

$$F_{tot} = \sqrt{({F_{eff}}^2 + {F_R}^2 + {F_L}^2)} \quad (2.5)$$

Où F_{tot} est la force totale (en newton, N), F_{eff} est la force effective (en newton, N), F_R est la force radiale (en newton, N) et F_L est la force latérale (en newton, N).

Par conséquent, sur la base de l'analyse des forces F_{eff} , il est possible de qualifier ou d'évaluer la technique de pédalage [22] à l'aide des métriques suivantes : l'indice d'efficacité (*IE*) et la proportion d'impulsion positive (*PIP*).

L'indice d'efficacité est le rapport entre la force efficace (tangentielle) et la force totale appliquée aux pédales, équation 2.6. Cette métrique évalue la technique de pédalage et est d'une grande importance dans le cyclisme. Un cycliste amateur a des valeurs comprises entre 30% et 40 %, tandis qu'un cycliste professionnel peut atteindre des valeurs de 40% à 50 % [22].

$$IE = \frac{\int_0^{2\pi} F_{eff}(\theta) d\theta}{\int_0^{2\pi} F_{tot}(\theta) d\theta} \times 100 \quad (2.6)$$

Le *PIP* est la proportion entre l'impulsion positive et l'impulsion totale et est déterminé par l'équation 2.7 [22]. Cette métrique est calculée uniquement à partir de la force effective F_{eff} . En général, cette valeur peut se situer autour de $83.7 \pm 3.9 \%$ pour un cycliste amateur et $86.5 \pm 4.5 \%$ pour un cycliste élite.

$$PIP = \frac{\int_0^{2\pi} P_+(\theta) d\theta}{\int_0^{2\pi} (P_+ + |P_-|)(\theta) d(\theta)} \times 100 \quad (2.7)$$

Où PIP est la proportion d'impulsion positive, P_+ est l'impulsion positive (Nm^*rad), $|P_-|$ est la valeur absolue de l'impulsion négative et θ est l'angle de manivelle (degrés). Le PIP est représenté par un pourcentage sans dimension.

Pour mieux comprendre la façon dont le cycliste produit la force sur les pédales, il est nécessaire de définir en premier lieu les phases du cycle de pédalage. Le cycle de pédalage correspond à une rotation dans le sens horaire de la manivelle autour du pédalier entre le point mort haut PMH (manivelle orientée vers le haut en 0°) et le point mort bas PMB (manivelle orientée vers le bas en 180°). Ce cycle peut être décomposé en 4 phases distinctes qui permettent la propulsion du cycliste, figure 2.7.

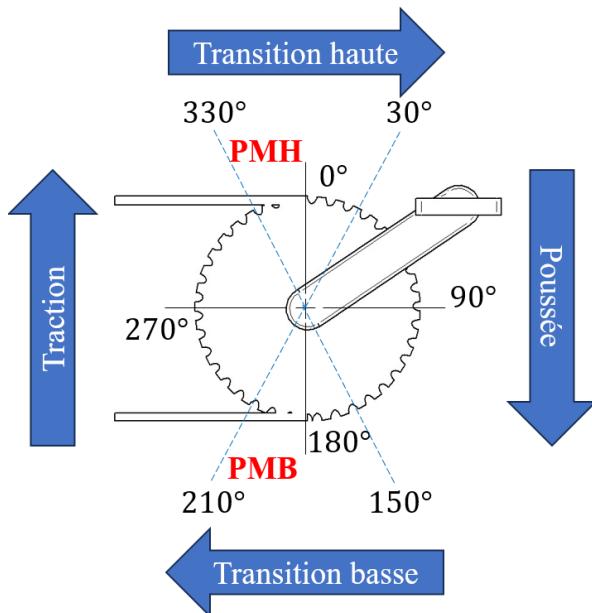


Figure 2.7 : Représentation des différentes phases du cycle de pédalage

De cette manière, les phases suivantes peuvent être distinguées :

- Phase 1 (330° - 30°) : Passage du point mort haut (PMH) où s'effectue la transition entre les phases de flexion et d'extension du membre inférieur.
- Phase 2 (30° - 150°) : Phase de poussée au cours de laquelle le membre inférieur s'étend et produit la majorité de la puissance.
- Phase 3 (150° - 210°) : Passage au point mort bas (PMB) où s'effectue la transition entre les phases d'extension et de flexion du membre inférieur.
- Phase 4 (210° - 330°) : Phase de traction au cours de laquelle le membre inférieur se fléchit.

Aussi, il est important de souligner certains points importants liés aux angles de manivelle qui peuvent être critiques car certains phénomènes biomécaniques se produisent. Il s'agit du point 0° , qui est un point critique de flexion maximale de la hanche et du genou. Le point 90° correspond à l'alignement de la manivelle avec l'horizontale. C'est également à cet endroit que la valeur maximale de la force effective est approximativement générée. Le point 180° est la position près de laquelle l'extension du genou et de la hanche est maximale. Enfin, la position à 270° de la manivelle est la position qui nous intéresse, car c'est le point le plus distal vers l'arrière du membre inférieur par rapport au vertical passant par l'axe du pédalier, et où il est plus difficile d'appliquer des forces de pédalage positives.

2.1.4 Cinématique du pédalage

Le mouvement articulaire du pédalage peut être analysé au moyen de la cinématique articulaire. Cela est réalisé à l'aide de caméras ou par des capteurs attachés au corps de l'athlète, appelés centrales inertielles « Inertial Measurement Unit (IMU) ». Toutes ces méthodologies peuvent être incluses dans le domaine techniquement connu sous le nom de systèmes de capture optique du mouvement « Optical Motion capture – OMC », qui est défini comme le processus d'enregistrement du mouvement d'un sujet vivant sous forme de séquences de coordonnées cartésiennes dans l'espace tridimensionnel [23]. En général, pour chaque segment du corps, la Société internationale de biomécanique « International Society of Biomechanics (ISB) » a proposé des normes pour le système de coordonnées articulaires afin d'analyser le mouvement des articulations humaines [24]. La méthode de capture du mouvement à l'aide de caméras est le système de référence qui a été largement validé pour la marche humaine et a été adapté à la capture de mouvement en cyclisme. Les caméras sont placées autour du corps, qui dans le cas de cette recherche serait un cycliste effectuant le mouvement de pédalage. C'est pourquoi ces systèmes sont appelés tridimensionnels (3D), car il est possible d'obtenir les positions X, Y et Z de marqueurs placés sur le corps humain selon un protocole biomécanique. Ce protocole biomécanique adapté au cyclisme est composé d'un certain nombre de marqueurs et d'une position anatomique sur le corps du cycliste. À l'aide d'un logiciel commercial ou personnalisé, les données brutes sont traitées pour déterminer les angles, les vitesses angulaires et les accélérations angulaires des articulations et des segments selon un modèle biomécanique appliqué. Les modèles biomécaniques, unilatéraux, du corps entier, des

membres inférieurs et autres peuvent être appliqués. La figure 2.8, montre l'environnement de capture adapté au cyclisme.

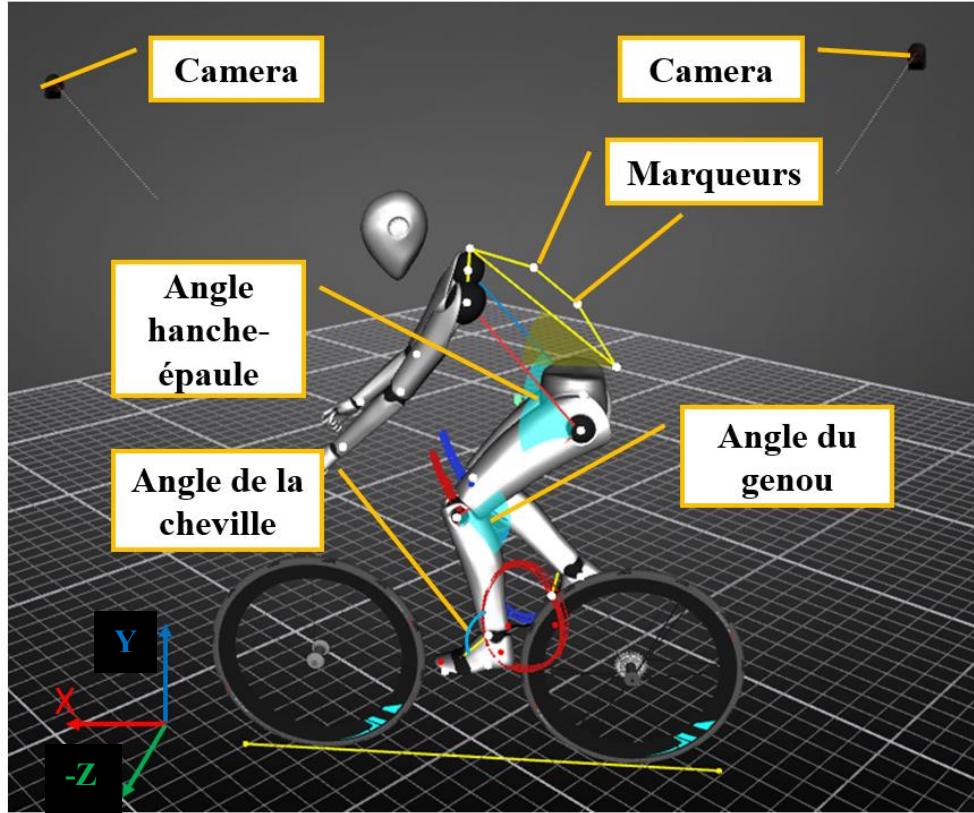


Figure 2.8 : Environnement de capture de mouvement

En général, en cyclisme, cette méthode est utilisée pour trouver la position correcte de la selle et du guidon. Ce processus est connu sous le nom d'ajustement du vélo ou « bike-fitting » en anglais. L'ajustement du vélo, qui consiste à régler les trois positionnements les plus importants entre l'individu et son vélo (hauteur de selle, position de la selle avant/arrière, la hauteur du guidon par rapport à la selle). En tenant compte des mesures anatomiques du cycliste, cette méthode est considérée comme l'un des principaux moyens pour optimiser les performances, le confort et prévenir les pathologies [25, 26]. Les protocoles d'ajustement des vélos s'appuient généralement sur des variables biomécaniques, et en particulier sur la cinématique des articulations dans le plan sagittal.

Le plan sagittal est choisi car l'amplitude des mouvements articulaires est beaucoup plus importante que les plans frontal et horizontal, et le mouvement de flexion/extension joue un rôle primordial dans la propulsion.

Les systèmes de capture de mouvement proposés pour le cyclisme peuvent mesurer les angles et les vitesses angulaires de chaque articulation des membres inférieurs pendant le pédalage. En ce qui concerne les membres supérieurs, la colonne vertébrale et le bassin, les systèmes de capture calculent les angles statiques car le mouvement est minime, les bras étant fixés au guidon par les mains et le bassin étant également considéré comme fixé à la selle. Le système de capture du mouvement permet de créer un modèle simplifié du corps humain, appelé modèle de segment de liaison, qui est largement utilisé pour l'analyse des mouvements en biomécanique, comme le montre la figure 2.9.

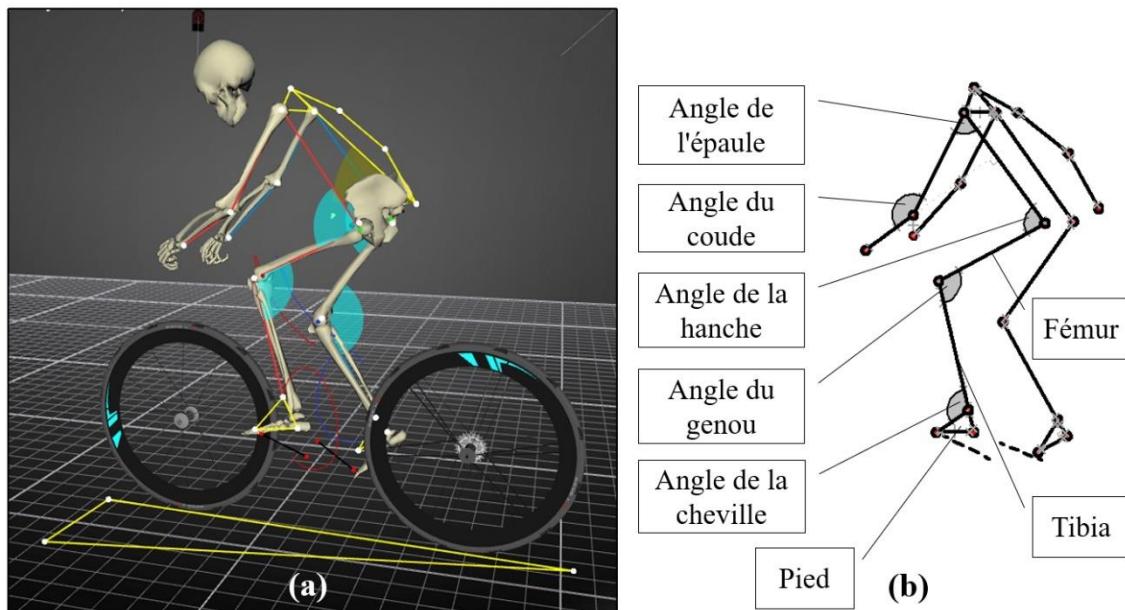


Figure 2.9 : Modèle biomécanique du segment de liaison, (a) représentation informatique de la capture de mouvement, (b) modèle biomécanique du segment de liaison

Pour construire le modèle, des marqueurs sont placés aux points d'articulation à analyser afin que les caméras placées autour d'eux puissent capturer leurs coordonnées dans le temps. Dans le cas de cette recherche, les marqueurs actifs sont utilisés, c'est-à-dire qu'ils émettent une lumière infrarouge, afin d'être capturés par des caméras infrarouges.

Au cours du développement de cette recherche, les informations sur la trajectoire des marqueurs seront principalement axées sur le membre inférieur afin de définir l'angle de la hanche, du genou et de la cheville dans le plan sagittal gauche. Particulièrement, le modèle biomécanique à développer définira l'angle de la hanche par rapport à une ligne horizontale et sera appelé angle du fémur. Un autre paramètre utilisé en biomécanique du cyclisme est la position horizontale du genou par rapport à l'axe de la pédale à un angle de 90° de la manivelle par rapport au point mort haut, également appelée « Knee Over Pedal Spindle (KOPS) ». En définitive, le modèle biomécanique bidimensionnelle simplifié utilisé dans le cadre de cette recherche est présenté à la figure 2.10. Il est illustré dans la figure un cycliste pédalement à 245 W et 80 RPM.

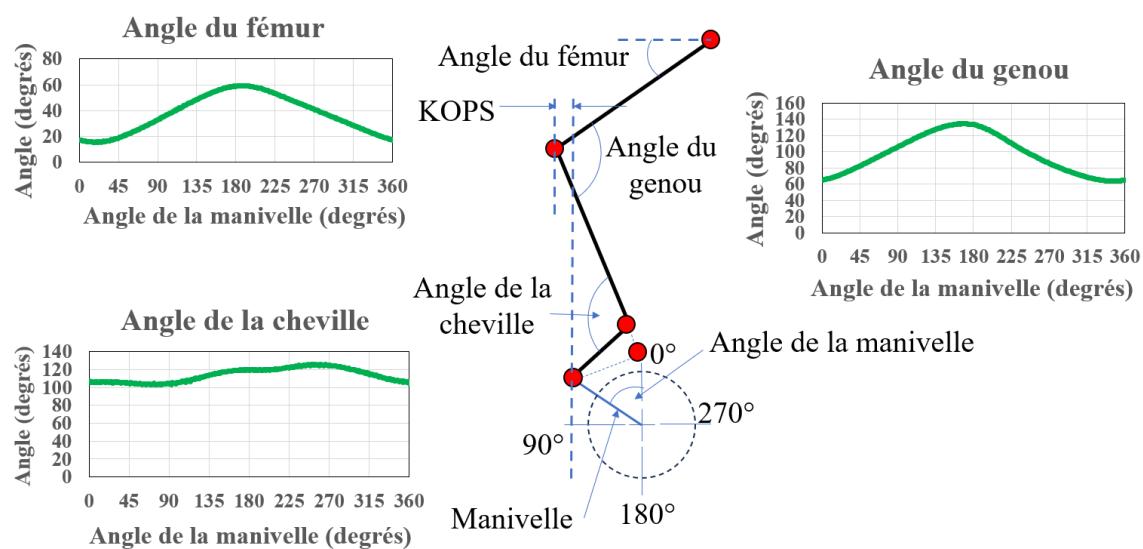


Figure 2.10 : Modèle biomécanique bidimensionnelle

Il est important de noter ici que le modèle biomécanique implémenté au cours de la présente recherche se concentre sur le plan sagittal. En effet, les déplacements des articulations des membres inférieurs dans le plan frontal sont minimes et ne contribuent que marginalement à la production de la puissance de pédalage [27]. La figure 2.10, montre également le profil des courbes angulaires typiques du fémur, du genou et de la cheville, en fonction de l'angle de la manivelle d'un cycliste. Ces profils sont typiques d'un cycliste roulant à une puissance de 150 W à 90 rpm.

2.1.5 Recrutement musculaire pendant le pédalage

L'autre domaine qui permet d'analyser un mouvement en biomécanique est le recrutement musculaire. En cyclisme, le recrutement musculaire est la manière coordonnée dont les muscles sont activés pour fournir de la puissance aux manivelles du vélo. Il y aura des muscles qui seront activés en fonction de la position de la manivelle. En biomécanique, pour déterminer le recrutement musculaire lors d'un mouvement, il est nécessaire d'utiliser des capteurs électromyographiques de surface (sEMG) [13]. Les signaux sEMG correspondent aux potentiels d'action envoyés par le système nerveux central qui sont à l'origine de la contraction musculaire. L'analyse des signaux sEMG permettent de caractériser le niveau d'activité musculaire et le moment de l'activation du muscle [28]. L'institution SENIAM « Surface electromyography for non-invasive muscle assessment » [29] a élaboré des recommandations spécifiques sur l'emplacement des capteurs, le traitement des signaux et la normalisation. Dans le cadre de ces recommandations, la normalisation est un aspect essentiel de l'analyse de l'activité musculaire des muscles évalués. L'objectif de la normalisation de l'amplitude EMG est de faciliter les comparaisons entre les participants, muscles et des sessions de mesure, entre

autres [30]. Il existe six possibilités de normalisation: contraction volontaire maximale dans la même tâche/le même contexte que la tâche d'intérêt, contraction volontaire maximale isométrique (CVM), tâche sous-maximale normalisée, amplitude EMG maximale/moyenne dans la tâche, amplitude maximale de l'onde M normalisée et ne pas normaliser [30]. Il n'y a pas de consensus sur la meilleure méthode de normalisation, chaque méthode a ses avantages et ses inconvénients [31], en particulier pour le cyclisme qui est un mouvement dynamique. Certains auteurs recommandent d'utiliser la méthode (CVM) pour les études de fatigue en cyclisme [32], tandis que d'autres auteurs recommandent la méthode du sprint ou l'amplitude EMG maximale/moyenne dans la tâche [33], pour la plupart des contextes expérimentales. Face à ces controverses, le « Consensus for Experimental Design in Electromyography (CEDE) », affilié à « International Society of Electrophysiology and Kinesiology (ISEK) » (<https://isek.org/cede-project/>), a récemment publié des lignes directrices pour l'expérimentation en électromyographie. Compte tenu de tous ces arguments et du fait que l'objectif de cette recherche est de prédire une activité musculaire au moyen d'une technique d'apprentissage automatique, il est proposé d'utiliser la méthode (CVM). Dans le cadre de cette recherche, il n'est pas prévu d'analyser l'activité musculaire entre les individus, ni de suivre les individus dans le temps ou d'évaluer leur condition musculaire. Le contexte expérimental de cette recherche est la comparaison de l'amplitude entre les muscles au cours de la même tâche (même muscle chez différents participants; ou différents muscles chez le même participant) où le test de contraction volontaire maximale est bien accepté [30, 34]. Dans ce projet, l'objectif n'est pas de réaliser des analyses des synergies musculaires, des comparaisons de force entre les muscles à partir de l'amplitude EMG, ou la détermination de l'activation musculaire à partir de données moyennes, entre

autres expériences, dans lesquelles la méthode CVM n'est pas recommandée [30]. L'évaluation qui sera réalisé, est effectué en un seul instant, dans le but de déterminer le recrutement musculaire de certains muscles des membres inférieurs, afin d'alimenter le modèle d'apprentissage automatique de chaque individu individuellement. Les procédures recommandées par le SENIAM seront strictement respectées afin de garantir la qualité de l'évaluation. Pour effectuer le test CMV du muscle à évaluer, la procédure consiste à demander au participant d'effectuer une contraction isométrique maximale, en bloquant manuellement les articulations afin d'obtenir une contraction maximale de chacun des muscles. Ces contractions sont effectuées trois fois pendant 5 secondes. Avec cette contraction maximale, un post-traitement du signal est ensuite effectué afin de normaliser les signaux de chaque muscle et ainsi présenter le recrutement musculaire en termes de pourcentage 0-100% de contraction [35]. Cette normalisation est essentielle pour comparer le recrutement musculaire entre différents muscles.

Les études sur le recrutement musculaire dans le cyclisme se concentrent sur la mesure de l'activité sEMG des dix principaux muscles recrutés pendant le pédalage. Ces muscles sont mesurables parce qu'ils sont situés à la surface de chaque membre et sont les suivants: le grand fessier (GF), le vaste latéral (VL), le vaste médial (VM), le droit fémoral (DF), le biceps fémoral (BF), le semi-membraneux (SM), le soléaire (SL), le tibial antérieur (TA), le gastrocnémien médial (GM) et le gastrocnémien latéral (GL) [28]. Tous les muscles sont activés pendant la phase de poussée, à l'exception du TA [28]. D'autres muscles ne sont activés que pendant la phase de traction, comme le TA, le DF et, bien qu'il ne soit pas possible de le mesurer par électromyographie de surface, le psoas [36]. La figure 2.11 montre les dix muscles et les lieux d'activation pendant la rotation de la manivelle. Il est illustré dans la figure un cycliste pédalant à 245 W et 80 RPM.

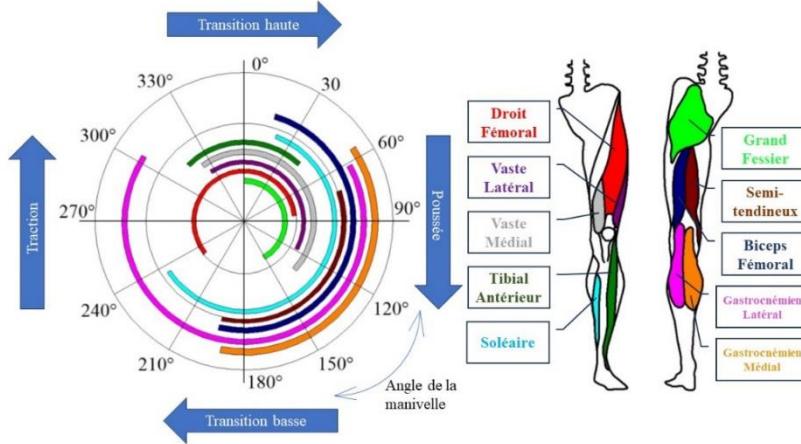


Figure 2.11 : Muscles impliqués pendant le mouvement de pédalage. Extrait du livre:

Optimiser son positionnement pour performer en Cyclisme, L-Tips

Afin de permettre l'analyse du recrutement musculaire, les signaux sEMG capturés doivent être traités. Le traitement du signal comprend trois phases : filtrage du signal, calcul de l'amplitude du signal en calculant la racine carrée moyenne enveloppe « root-mean-square » (RMS) et normalisation. La figure 2.12 présente un schéma de ce processus.

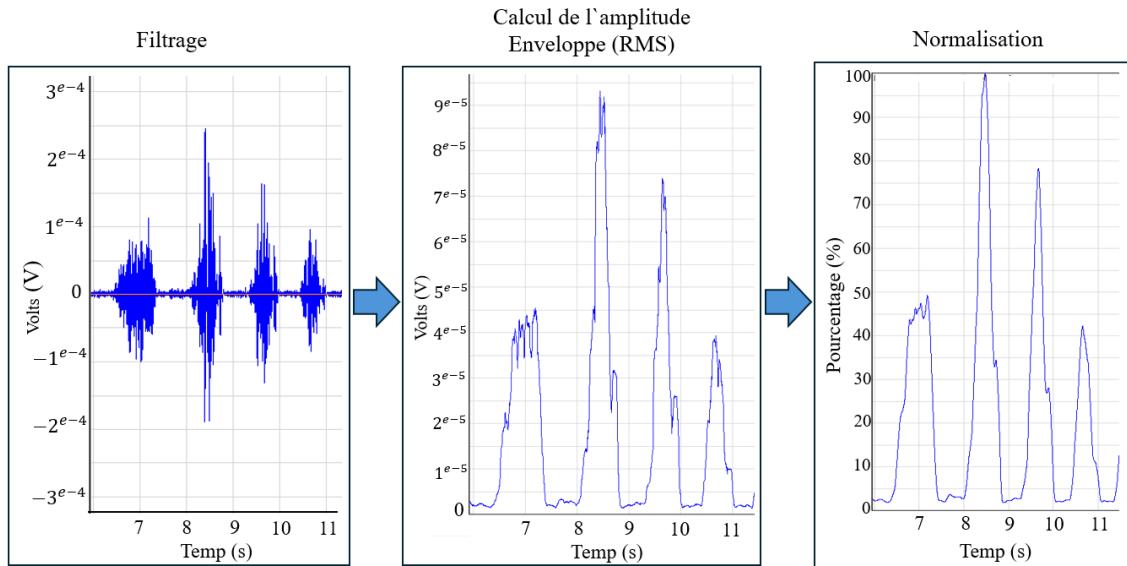


Figure 2.12 : Étapes du traitement du signal sEMG

Le RMS représente l'amplitude et la puissance moyenne du signal (également appelé EMG RMS) et constitue la recommandation préférée pour le lissage [35]. De cette manière, il est possible d'obtenir une courbe d'enveloppe qui permet de connaître l'amplitude maximale d'activation de chaque muscle. L'étude des signaux sEMG permet de comprendre le schéma d'activations musculaires qui a lieu pendant le cycle de pédalage et qui peut être analysé au moyen d'une analyse du cycle. Pour réaliser cette analyse, un accéléromètre est installé sur la manivelle pour indiquer la position de celle-ci pour obtenir le graphique par cycle de pédalage, c'est-à-dire de 0° à 360° de l'activité électromyographique de chaque muscle. Dans le cadre de cette recherche, diverses activités ont été menées pour enregistrer des différents muscles des membres inférieurs pendant le pédalage au moyen d'une analyse du cycle. La figure 2.13 présente un schéma du recrutement musculaire au moyen d'une analyse du cycle pendant le pédalage d'un cycliste effectuant le mouvement de pédalage à 250 W et 80 rpm. Ces résultats sont similaires à ceux obtenus par [28], avec des valeurs de puissance et de cadence similaires. Il peut parfois y avoir des différences entre les individus testés, mais cela est normal pour le signal EMG en raison de la variabilité inter sujet inhérente du signal, un problème qui a été démontré dans de nombreuses études, en particulier dans le domaine du cyclisme [28]. Cette variabilité est due au fait que le système nerveux dispose de plusieurs façons d'exécuter un mouvement donné, car il existe de multiples synergies et diverses combinaisons agoniste/antagoniste qui peuvent contribuer à la même tâche finale et au même schéma de force, entre autres facteurs [37]. Il faut distinguer les muscles mono-articulaires, c'est-à-dire les muscles qui n'influencent qu'une seule articulation, et les muscles bi-articulaires, c'est-à-dire les muscles qui sont impliqués dans deux articulations ou qui les traversent. Par ailleurs, est appelée contraction concentrique celle qui permet le

raccourcissement du muscle et contraction excentrique celle qui consiste en un allongement du muscle. Cette architecture et ces mécanismes de contraction se produisent pendant le pédalage et, par conséquent, certaines actions musculaires du même muscle se produiront à la fois dans les phases d'extension et de flexion de l'articulation.

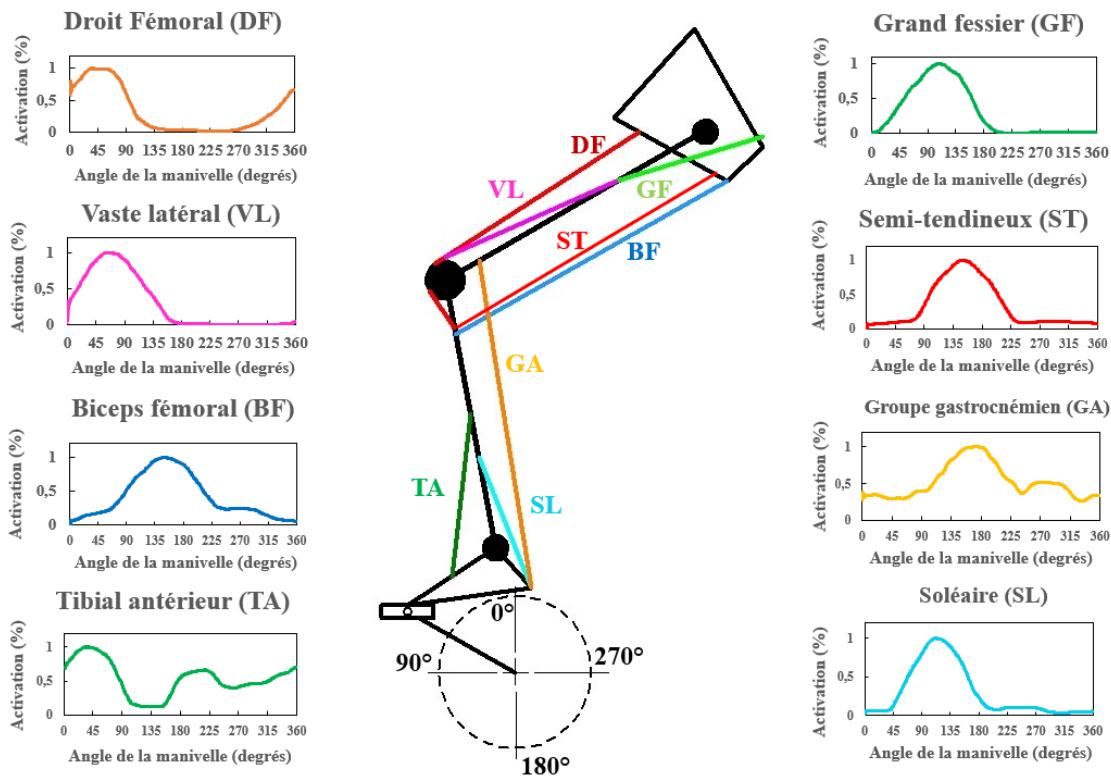


Figure 2.13 : Recrutement musculaire d'un cycliste typique à 250 W et 80 rpm

Le tableau 2.1 résume les caractéristiques des muscles les plus importants impliqués dans le pédalage.

Tableau 2.1 : Caractéristiques des principaux muscles impliqués dans le pédalage

Muscle	Type de contraction	Plage de fonctionnement	Fonctionnalité
Grand fessier	Mono-articulaire	0° - 150°	Extension de la hanche. Principal contributeur à la production de puissance [38].
Droit fémoral	Bi-articulaire	270 ° - 135°	Fournir une force efficace dans la première partie du coup de pédale à 0° ou lors du PMH et flexion de la hanche à 270°. Son action principale en phase de traction est augmenter la force de traction sur les pédales [38].
Biceps fémoral	Bi-articulaire	90° - 315°	Principal fléchisseur du genou, assistance de la phase de traction de la jambe à partir de 180° en améliorant l'efficacité du pédalage [39].
Gastrocnémien latéral	Bi-articulaire	90° - 315°	Principale action en phase de poussée à partir de 90°. Action secondaire à environ 45°.transmission des forces provenant de la hanche et du genou [40].
Vaste latéral	Mono-articulaire	0 ° - 135°	Extenseur du genou et générateur de force effective [12].

Les autres muscles peuvent être considérés comme des accessoires ou des supports de l'action des muscles susmentionnés et, au niveau anatomique, ils sont regroupés au moyen de tissus ou de fascias. Ainsi, le semi-tendineux (ST) travaille avec le biceps fémoral (BF), le soléaire (SL) avec le groupe des gastrocnémiens (GA), le vaste latéral (VL) avec le vaste médial (VM) et le droit fémoral (DF) regroupés sous le nom de quadriceps. Le muscle tibial antérieur (TA) est un muscle mono-articulaire qui intervient essentiellement dans l'orientation de la force dans la pédale plutôt que dans sa production [38]. Il est considéré comme un stabilisateur de la cheville et un transmetteur des forces provenant de la hanche et du genou dans la première phase de pédalage 0°- 90° et la transition de

180 à 225° [8]. En conclusion, il est démontré l'importance des muscles grand fessier, droit fémoral, biceps fémoral, vaste latéral, vaste médial et gastrocnémien latéral dans la production des forces de pédalage, et donc leur importance dans la modélisation du cyclisme.

2.2 Facteurs qui affectent la production de force

Plusieurs auteurs ont étudié les facteurs qui affectent la production de force en cyclisme. Selon la littérature trouvée, il existe des facteurs liés à la géométrie et aux composants du vélo, ainsi que des facteurs biomécaniques. Ces facteurs sont analysés à continuation.

2.2.1 Facteurs liés à la géométrie et aux composants du vélo

Il a été constaté que la géométrie du vélo, joue un rôle crucial dans la gestion des forces de pédalage, parce qu'elle permet de positionner le bassin et les membres pour que le cycliste puisse effectuer confortablement le geste de pédalage. Par conséquent, les mesures suivantes doivent être analysées puisque la modification de ces paramètres change les angles articulaires : la hauteur de la selle par rapport au pédalier (a), le recul de la selle dans la direction horizontale (b), la longueur de la manivelle (c), la position du guidon par rapport à la selle (d, e) et le mécanisme d'engagement mécanique du pied sur la pédale. La figure 2.14 montre les mesures à analyser.

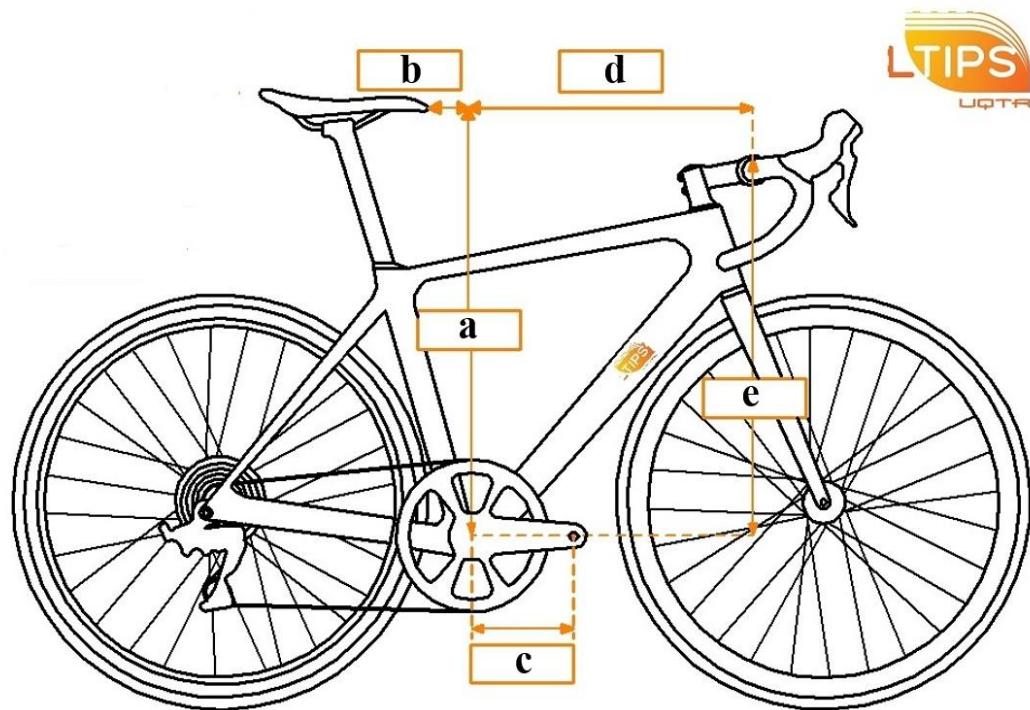


Figure 2.14 : Facteurs géométriques influençant la production de force de pédalage

La position du cycliste par rapport aux manivelles assurera une application correcte des forces sur les pédales [41]. Certains auteurs [42], ont constaté que la hauteur de la selle, mesure (a) dans la figure 2.14, affecte la direction et la magnitude des forces et donc l'efficacité du pédalage. Aussi, dans [43], les auteurs ont constaté qu'en positionnant correctement la selle vers l'arrière, mesure (b) dans la figure 2.14, améliore l'efficacité du pédalage. Au contraire, certains auteurs ont constaté que la position avancée de la selle provoque des activations musculaires différentes de certains muscles des membres inférieurs, qui ne se traduisent pas par une meilleure efficacité du pédalage [44]. Au niveau de la longueur des manivelles, mesure (c) dans la figure 2.14, certains auteurs ont constaté que des manivelles plus longues diminuent le ratio d'impulsion positive et ont donc un impact négatif sur la technique de pédalage [45]. La position du guidon, mesure (d) et (e)

dans la figure 2.14, permet au cycliste de choisir une position confortable pour générer le mouvement des membres inférieurs. Certaines études ont montré une efficacité de pédalage inférieure lors de l'abaissement de l'angle du tronc, c'est-à-dire en position aérodynamique, en raison de l'augmentation de la flexion de la hanche qui induit davantage de dorsiflexion de l'angle de la cheville [46]. Finalement, le mécanisme de fixation du pied sur la pédale exerce une grande influence sur la transmission des forces. Grâce à ce mécanisme, le membre inférieur est en mesure de minimiser les forces négatives dans la phase de traction, c'est-à-dire du point le plus bas de la manivelle jusqu'à le point le plus haut, et de produire une traction sur la pédale, ce qui a un impact direct sur l'efficacité du pédalage [39]. Ces résultats sont liés au fait que l'efficacité du pédalage est plus élevée lorsque cette traction est effectuée dans la phase de récupération [47]. D'autres auteurs [48] ont constaté que l'utilisation du système de cale génère une plus grande force sur la pédale et donc une plus grande puissance. Cela est dû à l'augmentation de l'activité musculaire dans la section de traction de la pédale et à une meilleure coordination musculaire [49]. En conclusion, il est constaté que ces facteurs liés à la géométrie ont un impact sur la production de force, mais il est difficile de savoir si ces facteurs sont uniquement liés aux changements des angles articulaires ou s'il y a d'autres changements au niveau des vitesses angulaires, accélérations angulaires ou même des activations musculaires. Il est également incertain de savoir quelle articulation du membre inférieur est la plus affectée dans la cinématique globale du mouvement de pédalage. Dans la plupart des cas, ces changements sont associés à des modifications de l'articulation du genou, qui est le paramètre d'analyse le plus courant dans le cyclisme.

2.2.2 Facteurs biomécaniques

Lors de l'adaptation du vélo, le plan sagittal ou latéral est généralement analysé, en se concentrant sur la variation des angles du membre inférieur du cycliste. En conséquence, certains auteurs ont observé que les valeurs optimales pour lesquelles des performances élevées peuvent être atteintes sont associées à un angle sagittal du genou compris entre 65° et 70°, pour la flexion maximale au point le plus haut de la pédale, et 135° et 140° pour l'extension maximale du genou au point plus bas de la pédale [50]. Ces performances améliorées sont liées à la puissance maximale, à l'amélioration de la capacité aérobique [9] et à l'amélioration de l'IE [41]. Cette mesure serait liée à l'angle optimal du genou qui réduit le risque de blessure dans cette zone [51]. Il est également constaté que des angles de genoux corrects optimisent la consommation d'oxygène et donc les performances du cycliste [52]. Par conséquent, la détermination de l'angle optimal du genou provoque plusieurs changements biomécaniques, comme l'optimisation de l'activation du gastronemius latéral [53], qui affectent la production de forces [41] et donc l'indice d'efficacité [54]. Il est évident que certains paramètres cinématiques affectent la production de force, mais l'accent est mis uniquement sur la détermination de l'angle du genou. Il est encore difficile de savoir si d'autres paramètres cinématiques des membres inférieurs affectent cette production de force. De même, lors du choix de l'angle du genou, il est problématique de déterminer l'angle optimal du genou dans la plage 130°- 140°, surtout s'il n'y a pas d'informations supplémentaires sur la transmission des forces de pédalage.

D'autres chercheurs ont constaté que la modification de la position sur la selle par rapport à l'axe de la pédale vers l'avant et vers l'arrière affecte aussi la cinématique de la

cheville, mais il n'y a pas de consensus sur la façon dont il affecte l'efficacité du pédalage [44]. La même controverse est générée en ce qui concerne la position du tronc qui affecte l'angle de la hanche, qui est associé à une production de force effective plus faible [55], mais pour d'autres auteurs cette cinématique de la hanche n'affecte pas la transmission des forces [56]. Il est donc conclu qu'encore aujourd'hui les facteurs qui influencent la production de force et donc l'efficacité du pédalage ne sont pas parfaitement clairs. En effet, il n'existe pas de modèles facilement accessibles permettant de relier ces variables cinématiques à la transmission des forces de pédalage.

2.2.3 Facteurs physiologiques

Il existe des facteurs physiologiques qui affectent la production de force, mais ils ne sont pas évalués dans le cadre de la présente recherche car ils sont loin des objectifs réalisables. Par exemple, l'amélioration de l'efficacité brute était associée à des valeurs élevées de l'indice d'efficacité [57]. L'efficacité brute est le rapport entre le travail effectué et la dépense énergétique totale [58]. Bien qu'il n'y ait pas de consensus sur l'existence d'un lien entre l'IE et l'efficacité brute [59], d'autres auteurs mentionnent que il existe une relation entre des valeurs d'indice d'efficacité et de force effective plus élevées avec une consommation d'oxygène diminuée [60, 61]. En revanche, d'autres auteurs établissent un lien entre l'amélioration de l'efficacité du pédalage et l'augmentation de l'activité musculaire des muscles les moins efficaces, ce qui augmente la consommation d'énergie [62].

2.3 La modélisation en biomécanique

Pour comprendre le phénomène de la production de force, la biomécanique utilise généralement différentes approches ou modèles. Ces approches ont la capacité de relier les informations provenant de différentes sources d'information telles que présentées pour la cinématique, la cinétique et le recrutement musculaire [13]. L'idée derrière ces modèles axés sur le cyclisme est de déterminer les paramètres qui affectent la production de la force de pédalage afin de guider l'optimisation de la technique de pédalage. Plusieurs modèles ont été développés pour comprendre les phénomènes de mouvement, mais la plupart d'entre eux ont été développés pour la marche humaine [63]. Ces modèles permettent de mettre en relation la cinématique (i.e. angles, vitesses angulaires, accélérations angulaires, ralentissements linéaires) et la cinétique (i.e. forces internes ou externes, moments articulaires, puissances articulaires) du mouvement évalué. Il est possible de distinguer deux types de méthodologies ou modèles biomécaniques : la dynamique directe, la dynamique inverse [2]. Cependant, ces dernières années, il est possible d'inclure une nouvelle approche, basés sur les techniques d'apprentissage automatique [64].

2.3.1 Modèle de dynamique directe

Dans ce type de modélisation, les activations musculaires autour d'une articulation sont prises en compte en premier lieu. Ces activations sont ensuite converties en forces musculaires et en couples articulaires, et enfin, grâce aux équations du mouvement, elles sont transformées en cinématique du mouvement [63]. Ce type de méthode présente encore aujourd'hui certaines limites, notamment : la transformation de l'activité en force

musculaire présente des limites par les corrélations mathématiques qui doivent être établies, les bras de levier muscle-tendon et les angles d'action sont difficiles à estimer [63]. Pour ces raisons, son application dans le domaine du cyclisme n'est pas très fréquente et, en raison de sa complexité, cette méthodologie n'a pas été adaptée aux intervenants du cyclisme. Certains logiciels ont été développés pour faciliter la sélection des paramètres et le calcul des variables d'intérêt, tels comme OpenSIM et Anybody, entre autres.

2.3.2 Modèle de dynamique inverse

En contraste, le modèle de dynamique inverse est plus traditionnel en biomécanique [13] et dans le contexte du cyclisme, il a été largement appliqué. L'annexe C, présente le développement d'une analyse dynamique inverse réalisée dans le cadre de la présente recherche. Pour réaliser cette modélisation, les segments du corps humain à évaluer sont analysés au moyen de segments connectés avec leurs forces de réaction respectives, c'est-à-dire que chaque segment du corps est analysé séparément et considéré comme un élément rigide [13]. Ce modèle prend d'abord en compte les forces arrivant sur un segment du corps, en particulier, dans le cas du cyclisme, les forces de pédalage arrivant sur le pied au niveau des têtes métatarsiennes, ce qui explique le terme inverse. Ensuite, grâce à la méthode de modélisation du segment de liaison, et en appliquant les équations de Newton-Euler, il est possible de trouver les moments articulaires et les forces présentes sur chaque articulation.

Cette méthode utilise à un modèle bidimensionnel (2D) à 5 barres, largement utilisé en mécanique classique pour comprendre le mouvement à étudier. Grâce à ce modèle, il est possible d'appliquer la méthode de dynamique inverse. Cette méthodologie a été mise en œuvre afin de comprendre comment la force de pédalage est générée, quels pourraient être

les principaux facteurs influençant cette force et comment la cinématique peut être liée à ce phénomène. Plusieurs sessions de collecte de données ont été réalisées pour développer cette méthode, afin d'obtenir des informations sur la cinématique, la cinétique, les moments et la puissance des articulations. À titre d'exemple, les résultats suivants sont typiques d'un cycliste roulant à une puissance de 150 W à 90 rpm. Les équations et la méthodologie utilisées pour appliquer cette méthode sont présentées en Annexe C.

Dans cette approche, le fémur, le tibia, le pied et la manivelle sont représentés par des barres qui ont un déplacement angulaire et qui produisent la rotation de la manivelle qui permettra au cycliste de se déplacer. Les premiers auteurs qui ont effectué cette analyse sont Hull et Jorge [65]. Le modèle bidimensionnel à 5 barres, permet de représenter le mouvement des jambes au moyen de la barre 1, qui représente le sol ou le système fixe en deux points a. (bassin sur la selle) et e. (axe de la manivelle ou boîte de pédalier). Les barres restantes représentent le fémur (barre 2), le tibia (barre 3), le pied (barre 4) et la manivelle (barre 5). Ensuite, à l'aide d'une méthode de modélisation telle que la dynamique inverse, les moments articulaires sont calculés. Dans le mécanisme à 5 barres, il est également courant d'identifier la barre d'entrée, qui serait la barre 2, et la barre de sortie, qui serait la barre 5, qui fournit la puissance totale permettant le mouvement de rotation de cette barre, ce qui, dans le cyclisme, est un paramètre important de la performance. La figure 2.15 illustre les résultats des moments articulaires trouvés en appliquant cette méthodologie.

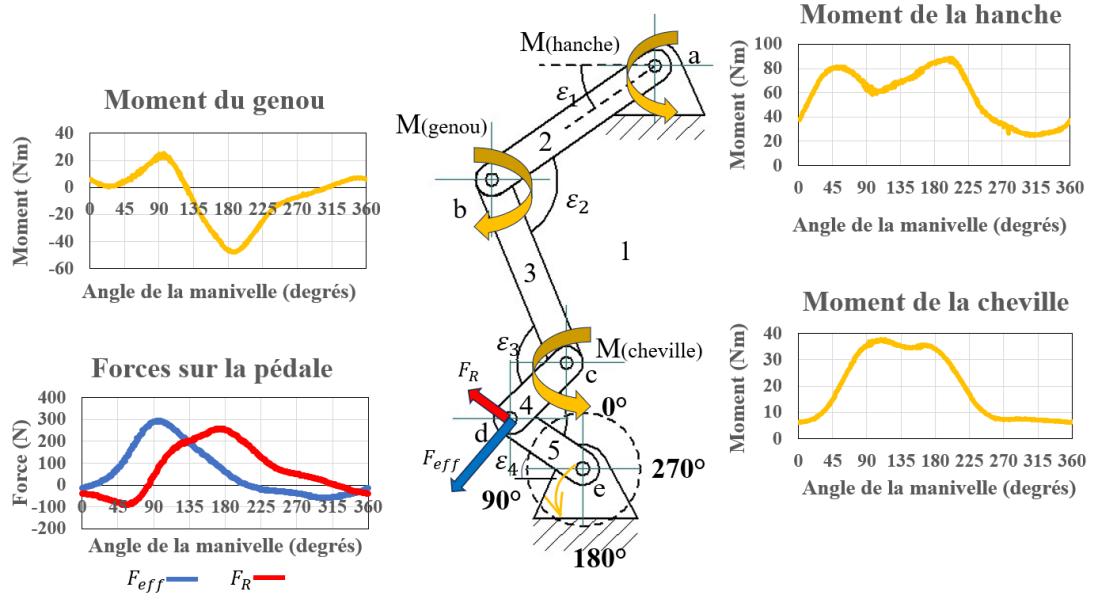


Figure 2.15 : Modèle bidimensionnel à 5 barres du membre inférieur

Les résultats obtenus dans cette analyse sont en accord avec ceux présentés par [40] à des puissances et des cadences similaires. Comme le montre la figure 2.15, les valeurs maximales de la force effective à 90° coïncident avec les moments d'extension maximaux au niveau du genou et de la cheville. La hanche contribue également à atteindre la force effective maximale car les couples sont maintenus à des valeurs maximales pendant la phase de poussée, avec une valeur supérieure à celle du genou et de la cheville. L'analyse de ces résultats permet d'identifier la relation entre les moments maximaux de la hanche, du genou et de la cheville et les forces maximales effectives F_{eff} et radiales F_R exercées sur la pédale (point d. sur la figure 2.15). Cela explique son importance dans la biomécanique du pédalage et de la transmission de la puissance au niveau des manivelles du vélo. Ces moments articulaires représentent les actions musculaires au niveau de chaque articulation des membres inférieurs. De plus, lorsque les muscles se contractent

autour d'une articulation, ils provoquent un mouvement, c'est-à-dire une vitesse angulaire du segment. De cette façon la puissance articulaire du membre inférieur peut être déterminée par la vitesse angulaire du segment multipliée par le moment musculaire dans l'articulation (Annexe C, point 6). Grâce à ces analyses, il est également possible de connaître la puissance délivrée par chaque articulation, ce qui permet de comprendre la contribution de chaque segment musculaire à la puissance totale délivrée aux manivelles. Il est constaté qu'environ 50 % de la puissance totale provient de l'articulation de la hanche, environ 30 % du genou et 20 % de la cheville. Les profils de puissance des articulations du membre inférieur sont présentés dans la figure 2.16.

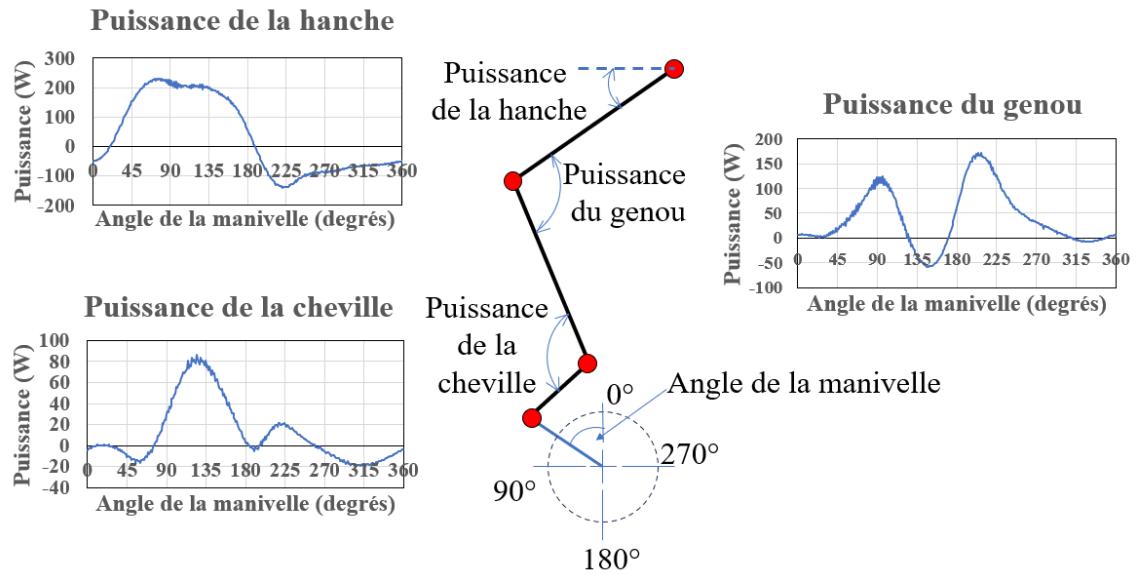


Figure 2.16 : Profils de puissance des articulations du membre inférieur

Comme dans l'analyse présentée par [40] et [66], se produit un pic maximal de production de puissance de la hanche à environ 50° dû à l'action des muscles extenseurs de la hanche. Également pour le genou, sont établies deux pics maximaux à 90°, causés par les muscles extenseurs du genou et un autre à environ 200° causé par les muscles fléchisseurs du genou. Au niveau de la cheville, on observe un pic maximal à environ 130°

causé par les fléchisseurs de la cheville. Un autre, plus petit, près de 225°, causé par les muscles dorsi-flexeurs de la cheville, permettant de réduire les forces négatives qui se produisent dans la phase de traction de la pédale.

De plus, certains auteurs ont constaté que l'angle et la vitesse angulaire de l'articulation influencent la production de la force et des moments articulaires [67]. En particulier, en [67], les auteurs Hahn, Herzog, et Schwirtz ont étudié le genou et ont présenté la relation entre le couple maximal atteint dans une articulation en fonction de l'angle de l'articulation et de la vitesse angulaire de l'articulation. Notamment, le couple maximum est atteint à des vitesses angulaires faibles 0,5 à 1,1 rad/sec et proches des angles de flexion du genou entre 110° à 120 °. Ces angles de genou sont atteints lorsque la manivelle est positionnée près de 90°, où la force effective maximale atteint sa valeur la plus élevée. Cette analyse montre l'importance de la cinématique articulaire dans la génération des forces de pédalage et de la puissance, et fournit ainsi un lien direct qui permettra la construction de modèles biomécaniques de pédalage. Herzog [68] a trouvé que la capacité maximale de production de force et de puissance musculaire est obtenue lorsque les muscles atteignent une longueur appropriée qui correspond approximativement à la position de la manivelle de 90°.

Bien que la dynamique inverse offre la possibilité de résoudre des problèmes complexes en biomécanique, cette approche nécessite l'emploi des méthodes mathématiques supplémentaires en raison du problème des variables redondantes, c'est-à-dire les muscles entourant chaque articulation. Il est également nécessaire de formuler des hypothèses relatives à diverses constantes géométriques et physiques (masse du segment, moments d'inertie, distances au centre de masse, etc.), ce qui limite la généralisation à toute anatomie.

2.3.3 Modèles biomécaniques basés sur techniques d'apprentissage automatique

L'apprentissage automatique est une branche de l'intelligence artificielle (IA) et de l'informatique qui vise à réaliser des modèles de classification ou de prédiction qui apprennent directement des données. Cette approche est particulièrement pertinente quand les systèmes à modéliser sont trop complexes ou nécessitent des interactions avec de nombreuses sources d'information. De la même manière que les modèles mentionnés précédemment ont initialement été utilisés pour étudier la biomécanique de la marche, l'application des approches d'apprentissage automatique s'est progressivement étendue au domaine du sport ces dernières années [69]. Dans le large spectre des possibilités de l'apprentissage automatique, cette méthodologie est considérée comme étant orientée vers un objectif, en tenant compte principalement des variables d'entrée et des variables de sortie [70]. Au moyen de différentes méthodologies, un modèle mathématique est établi entre les variables. En d'autres termes, il n'est pas nécessaire d'analyser chaque composant et de faire des hypothèses du système analysé, mais plutôt d'apprendre ou de découvrir comment les données d'entrée sont liées aux données de sortie ou à la variable à déterminer. En conséquence, les limitations rencontrées avec les approches analytiques sont réduites. Cependant l'approche par apprentissage automatique présente également des limitations, notamment le fait qu'il faut disposer d'une grande quantité de données pour permettre aux algorithmes d'apprendre à partir des caractéristiques des données. Une autre limitation réside dans le fait qu'une majorité des techniques agissent comme des boîtes noires où la relation interne des données n'est pas explicite. Toutefois, certaines approches d'apprentissage automatique conçoivent des modèles qui permettent une

interprétation claire des relations entre les variables. On parle ici d'IA explicable « Explainable AI XAI » [71].

Actuellement, les techniques d'apprentissage automatique sont implémentées dans le domaine du sport afin de créer des modèles mathématiques pour prédire des variables [72]. Un exemple de ces applications est la possibilité d'entraîner un algorithme d'apprentissage automatique pour prédire la réponse de la fréquence cardiaque d'un cycliste au cours d'une séance d'entraînement [73]. Un autre exemple est la prédiction de la cadence de pédalage à l'aide de centrales inertielles et d'un réseau neuronal [74]. Il est donc intéressant de présenter dans ce projet de recherche une alternative pour guider l'athlète dans l'optimisation biomécanique au moyen de modèles qui peuvent déterminer les métriques d'intérêt en cyclisme. De plus, il est important de proposer des modèles d'apprentissage automatique interprétables afin de déterminer les prédicteurs qui sont fortement liés à ces métriques.

Il existe quelques exemples d'application d'apprentissage automatique dans le contexte du cyclisme. Un exemple de ces applications est la sélection de la hauteur de la selle en fonction de l'anthropométrie, de la géométrie du vélo et de la cinématique des articulations [75]. Les auteurs ont développé une équation basée sur une régression LASSO (« least absolute shrinkage and selection operator ») dont les variables indépendantes sont la hauteur entre les jambes, l'angle maximal du genou et l'angle du tube de selle d'un cadre du vélo. Dans d'autres travaux [76], ils ont utilisé plusieurs sélections de prédicteurs pour associer, au moyen de régressions linéaires, les paramètres cinématiques qui affectent les charges sur le genou pendant le pédalage. Ils ont constaté que le prédicteur le plus important était l'angle minimal du genou. Certains auteurs [77], ont récemment présenté une étude qui permet de prédire les forces exercées sur la pédale au moyen d'une technique

basée sur des réseaux neuronaux récurrents utilisant comme variables d'entrée la cinématique des articulations du membre inférieur. D'autres auteurs [78], ont également étudié deux types de réseaux neuronaux permettant la capture de mouvements sans marqueurs pendant le pédalage. Ils ont constaté qu'en dépit des progrès réalisés dans cette méthodologie de capture, il est problématique de capturer le mouvement des pieds. Dans une autre étude [79], les auteurs ont évalué deux autres méthodes basées sur l'apprentissage profond pour l'estimation des angles dans le contexte de l'ajustement des vélos. Ils ont constaté que le système Openpose [80] est plus performant et a un meilleur avenir dans cette application. De même, il est possible de trouver dans la littérature des modèles qui utilisent les phénomènes cinématiques du mouvement de pédalage pour prédire des variables physiologiques telles que la performance d'un cycliste au moyen d'une régression linéaire multiple [81].

Comme il est possible de constater, l'application de modèles basés sur l'apprentissage automatique dans différents domaines du cyclisme est prometteuse, d'où l'intérêt de la présente recherche pour le développement de modèles basés sur l'apprentissage automatique pour la prédiction de variables dans le cyclisme.

2.4 Concepts de base de l'apprentissage automatique

L'apprentissage automatique, communément appelé «machine-learning», est utilisé pour enseigner aux machines comment interpréter les données et modéliser leurs interactions. Comme les sections précédentes l'ont montré, lors de l'analyse biomécanique en cyclisme, une grande quantité de données est obtenue à partir de la capture cinématique du mouvement, des forces de pédalage et des activations musculaires impliqués dans le

geste de pédalage. En conséquence, l'apprentissage automatique permet de modéliser la relation entre ces données à l'aide d'un algorithme. Pour effectuer cette modélisation, l'apprentissage automatique peut se baser sur deux approches distinctes en fonction des besoins : l'apprentissage supervisé et l'apprentissage non supervisé [82]. L'apprentissage supervisé prédit une variable dépendante avec des données étiquetées. Au contraire, l'apprentissage non supervisé identifie des caractéristiques en utilisant les données disponibles. Dans ce projet de recherche, l'apprentissage supervisé est utilisé car les données cinématiques, cinétiques ou électromyographiques sont étiquetées. Aujourd'hui, dans le domaine du sport, de nombreuses approches peuvent être utilisées pour créer des modèles prédictifs dans le contexte de l'apprentissage supervisé, à commencer par la régression linéaire, la régression linéaire multiple, la régression rigide, la régression logistique, la régression LASSO, la régression polynomiale, la régression linéaire bayésienne, entre autres [83]. Il existe également des techniques plus complexes qui offrent une plus grande robustesse et une plus grande précision pour faire des prédictions, telles que : les réseaux neuronaux artificiels « Artificial Neural Networks (ANN) », les machines vectorielles de soutien « Support Vector Machines (SVM) », les arbres de décision, les arbres randomisés « random forest », et les K-voisins les plus proches « K-Nearest Neighborhoods » [83].

Les réseaux neuronaux artificiels simples ainsi que d'autres algorithmes, tels que les arbres de décision, peuvent être inclus dans le terme d'apprentissage automatique superficiel [84]. Quand il existe plus d'une couche avec des architectures profondément intégrées, ils sont appelés réseaux neuronaux profonds et font partie de l'apprentissage profond ou « Deep learning » en anglais [85]. Les principaux architectures de réseaux neuronaux sont les suivantes : « Feedforward Neural Network (FFNN) », « Recurrent

Neural Network (RNN) », « Convolutional Neural Networks (CNN), « Radial Basis Function Neural Network », « Kohonen Self Organizing Neural Network », « Modular Neural Network » [85]. Mahesh [86], signale que pour les scientifiques des données, il n'existe pas de meilleur algorithme pour résoudre un problème. C'est pourquoi, dans le cadre de cette recherche, plusieurs modèles seront présentés, chacun d'entre eux présentant des avantages et des limitations. Le choix d'un modèle est donc influencé par l'une des caractéristiques de ce type de modélisation par apprentissage automatique, à savoir la propriété d'interprétabilité ou d'explicabilité. Le concept d'interprétabilité, selon Miller [87], est le degré auquel un observateur comprend comment le modèle réalise la prédiction. Par conséquent, conformément aux objectifs définis dans la présente recherche, il est intéressant de rechercher des modèles dans lesquels il est possible de savoir comment le modèle effectue la prédiction, c'est-à-dire de savoir quelles variables sont utilisées et dans quelle proportion elles sont utilisées. Dans cet ordre d'idées, l'une des limites des méthodes d'apprentissage profond est que, compte tenu de la complexité de leur architecture, il n'est pas possible de comprendre comment les variables d'entrée sont en relation avec la variable de sortie [88]. Ce phénomène est communément appelé « black-box » ou boîte noire, cependant ces algorithmes ont la capacité de résoudre des problèmes complexes [88].

Afin de développer les modèles d'apprentissage automatique, un flux de travail est proposé, qui distingue les différentes étapes à développer afin d'atteindre les objectifs proposés dans cette recherche. La figure 2.17 présente un schéma de cette méthodologie.

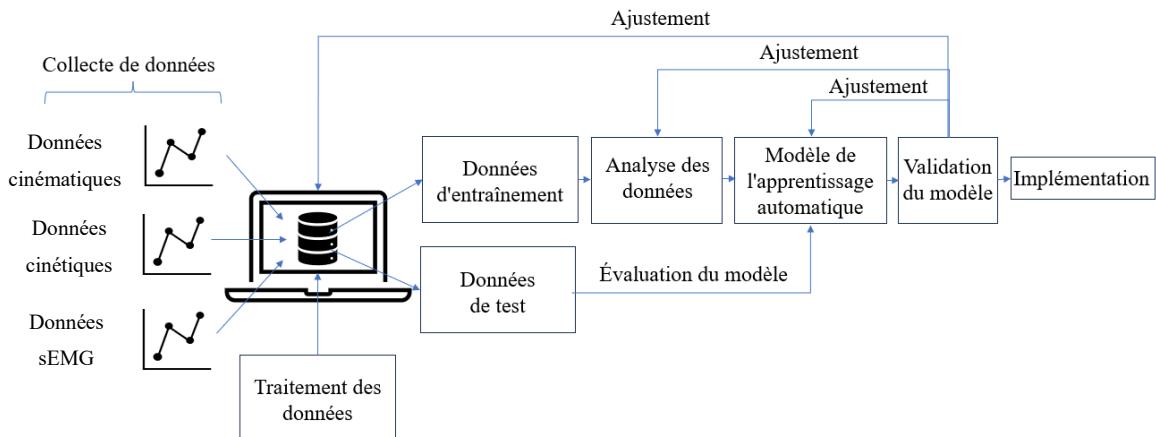


Figure 2.17 Flux de travail apprentissage automatique

Dans la figure 2.17, les étapes du développement des modèles d'apprentissage automatique sont présentées, en commençant par la capture des données au moyen d'un protocole développé à cet effet, puis en organisant et en traitant ces données. Les données sont ensuite divisées en données d'apprentissage et données de test, généralement réparties à 80 % et 20 % respectivement. Par la suite, les données sont analysées afin de trouver des relations entre les variables d'entrée et de sortie et d'effectuer la sélection la plus appropriée de ces variables. Le modèle souhaité est ensuite appliqué en sélectionnant initialement les paramètres du modèle. La validation du modèle est subséquemment effectuée à l'aide de techniques telles que la validation croisée et l'analyse de l'erreur du modèle. Si les résultats souhaités ne sont pas satisfaisants, les paramètres du modèle sont ajustés et le modèle est réévalué. En outre, si nécessaire, d'autres variables d'entrée sont recherchées afin de trouver un meilleur modèle. Enfin, une fois le modèle validé, la dernière phase est appelée implémentation, au cours de laquelle le modèle sera intégré dans un système de capture de mouvement. Dans ce qui suit, certaines des phases les plus

importantes du processus de travail proposé seront présentées en détails. Les modèles d'apprentissage mis en œuvre dans le cadre de cette recherche sont présentés ci-dessous.

2.4.1 Régression linéaire multiple

Cette technique permet d'estimer ou de prédire une sortie ou une variable dépendante à partir d'une série de variables indépendantes. La régression linéaire multiple est une méthode d'apprentissage automatique qui appartient à la catégorie des modèles interprétables, car il est connu comme le modèle effectue la prédiction. En d'autres termes, les poids de chaque prédicteur est connu. La théorie de base d'une régression linéaire multiple consiste à prédire une variable d'intérêt au moyen de quelques prédicteurs ou variables indépendantes. L'équation générale d'une régression linéaire multiple est l'expression suivante, équation (2.8) :

$$Y = a + b_1 \times X_1 + b_2 \times X_2 \dots b_k \times X_k \quad (2.8)$$

Où Y est la variable dépendante, X_1, X_2, \dots, X_k sont les variables indépendantes, b_1, b_2, \dots, b_k , sont les coefficients de régression. Ces coefficients donnent un poids aux variables prédictives en fonction de leur contribution relative à la prédiction de Y .

Le rapport des poids de régression est le rapport de l'importance prédictive des prédicteurs. En d'autres termes, une régression linéaire multiple montre les relations mathématiques entre chaque prédicteur et la variable dépendante. Cette méthode est considérée comme une technique permettant de modéliser le monde réel complexe d'une manière simple et pratique [89] et constitue l'une des méthodes les plus transparentes et

les plus facilement compréhensibles pour les humains [90]. Pour construire un modèle de régression linéaire multiple valide, il est indispensable de respecter quatre hypothèses : linéarité, absence de multi colinéarité, homoscédasticité et variables normalement distribuées [91]. Ces hypothèses sont vérifiées au moyen de graphiques ou de fonctions à l'aide du logiciel utilisé. Ces analyses ne sont pas concluantes, et doivent être complétées par d'autres mesures telles que la valeur du facteur d'inflation de la variance (VIF) une fois que la régression est construite. Des valeurs VIF comprises entre 5 et 10 sont considérées comme acceptables pour la non colinéarité [92]. Un modèle de régression doit également respecter le principe de parcimonie, c'est-à-dire qu'un modèle parcimonieux se caractérise par sa simplicité et son pouvoir explicatif et prédictif élevé. Ils expliquent les données avec un nombre minimal de paramètres ou de variables prédictives [93].

2.4.2 Arbre de décision et forêt aléatoires

La méthode de l'arbre de décision, par analogie avec un arbre, est une structure de calcul composée de nœuds et de feuilles par lesquels passe l'information, et qui permet d'estimer la probabilité d'une valeur à partir d'une gamme de valeurs. Cette gamme est ajustée à chaque nœud au moyen du signe mathématique d'inégalité $Y \leq X_n$, où Y est la variable dépendante et X est une valeur numérique que les prédicteurs n peuvent prendre dans le modèle et les feuilles représentent l'estimation de la variable indépendante [94]. Le principal avantage d'un arbre de décision est qu'il est hautement interprétable puisqu'il est possible de déterminer exactement quelles prédicteurs sont plus importants. L'intérêt d'utiliser un arbre de régression dans la présente recherche, réside dans le fait qu'il permet de connaître les variables les plus critiques dans la prédiction de l'indice d'efficacité et de

la proportion d'impulsions positives. La figure 2.18 montre la structure composée d'un nœud racine, qui identifie le prédicteur le plus important. Ensuite, les prédicteurs sont énumérés par ordre d'importance de haut en bas, chacun d'entre eux prenant la forme de décisions sur la base d'inégalité évaluée. Cette inégalité donne quelques pistes sur la valeur des prédicteurs critiques dans les prévisions. Enfin, les feuilles sont les prédictions accompagnées d'un pourcentage qui indique dans quel pourcentage se trouve la valeur prédictive. Particulièrement, deux arbres seront obtenus pour chaque variable dépendante qui sera utilisée à des fins d'interprétabilité des facteurs affectant la production de force dans le cyclisme. Ils sont appelés arbres de régression lorsque la variable dépendante est continue ou discrète et arbres de classification lorsque la variable dépendante est qualitative.

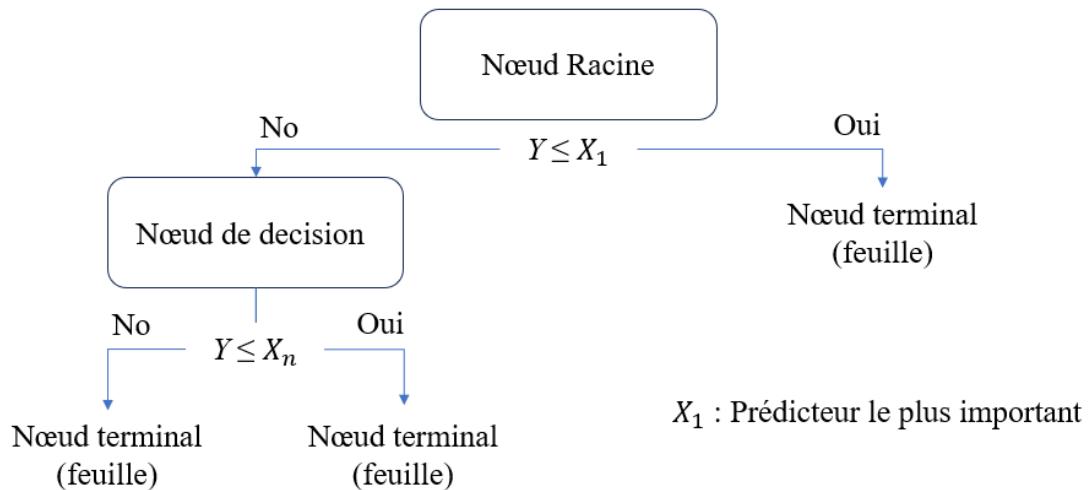


Figure 2.18 : Architecture d'un arbre de décision

Par la suite, l'arbre doit être optimisé, c'est-à-dire coupé, au moyen d'une fonction qui calcule un coefficient d'élagage et qui permet de trouver moins de feuilles prédites.

Cependant, un arbre de décision se caractérise par une très faible précision [95], et pour cela il existe dans la littérature une technique appelé forêts aléatoires, qui combine plusieurs arbres de décision dans son architecture pour calculer la variable dépendante. Cette architecture améliore la précision et la réduction de la variance, mais perdant l'interprétabilité en raison de la combinaison des modèles [96]. Toutefois, il est très utile pour comprendre le problème à résoudre puisqu'à l'aide de l'algorithme l'ordre d'importance des variables sélectionnées est déterminé.

2.4.3 Régression LASSO

Le nom est dérivé de l'acronyme en anglais : « Least absolute shrinkage and selection operators (LASSO) ». Son utilité réside dans le phénomène qui se produit lorsque, dans le processus de modélisation, il existe plusieurs prédicteurs lesquelles peuvent être redondantes et provoquer le phénomène de multi colinéarité. Un algorithme appelé régularisation s'applique, qui réduit la taille (contraction) des coefficients des prédicteurs qui n'expliquent pas la variable dépendante, afin qu'ils tendent vers zéro [97]. Cela signifie que la méthode limite la complexité du modèle puisque les variables dont les coefficients sont égaux à zéro sont éliminées du modèle [98].

2.4.4 Régression rigide

Comme indiqué précédemment, l'élimination de la multi colinéarité est l'un des principaux objectifs de l'élaboration d'un modèle de régression linéaire multiple. La méthode de régression rigide est l'une des méthodologies les plus utilisées. Cette méthode a été introduite dans les années 1970 par Hoerl et Kennard [99], pour résoudre des

problèmes dans le domaine de l'ingénierie. Cette technique de régression modifie une régression par les moindres carrés ordinaires ou « Ordinary Least Squares regression OLS » et détermine les paramètres du modèle de régression linéaire au moyen d'une solution linéaire standard des moindres carrés. Cette solution est la meilleure estimation linéaire sans biais qui minimise la variance en introduisant un léger biais en faveur de la variable dépendante [100].

2.4.5 Régression en composantes principales (PCR)

La technique appelée régression en composantes principales ou « Principal Component Analysis (PCA) » offre également la possibilité d'optimiser le nombre de variables et donc de trouver un modèle optimisé. Cette technique crée de nouvelles variables par des combinaisons linéaires des variables originales. À partir de cette analyse, il est possible d'effectuer une régression linéaire multiple en combinant les composantes obtenues lors de la méthode PCA. Cette régression est donc appelée régression en composantes principales (PCR) [101].

2.4.6 Régression partielle du dernier carré (PLSR)

Une autre méthode explorée dans le cadre de cette recherche est la régression partielle du dernier carré (PLSR), en particulier lorsque les variables et les données sont peu nombreuses. Il s'agit d'un problème dérivé de l'expérimentation humaine, lorsqu'il est nécessaire que les critères d'inclusion souhaités soient remplis. Cette méthode n'essaie pas de choisir des variables significatives. La méthode cherche à trouver un sous-espace des variables explicatives, en d'autres termes, en utilisant simultanément la méthode PCA et effectue ensuite une régression avec les nouvelles variables explicatives [102].

2.4.7 Réseaux neuronaux artificiels

Un réseau neuronal artificiel ou « Artificial Neural network (ANN) », est un modèle informatique inspiré des neurones naturels. L'architecture se compose d'entrées, qui sont multipliées par des poids représentant le flux d'informations. Les poids sont calculés à l'aide d'une fonction mathématique qui détermine l'activation du neurone, puis une autre fonction calcule la sortie du neurone artificiel. À partir de ce concept, des architectures plus complexes ont été créées, telles que les réseaux neuronaux « Feed Forward Neural Network (FFNN) », et les réseaux neuronaux récurrents « Recurrent Neural Network (RNN) », entre autres. Les réseaux neuronaux se caractérisent par des structures plus interconnectées, avec des couches cachées et des flux d'informations dans différentes directions, ce qui rend impossible pour les humains de suivre la correspondance exacte entre l'entrée des données et la prédiction. Par conséquent, ce sont des modèles qui sont classés comme des modèles moins interprétables mais dans les réseaux à une seule couche, il est possible de savoir au moins quelle variable donne le plus de poids à la prédiction. Pour cette raison, bien qu'il existe certaines méthodes mathématiques pour comprendre le fonctionnement d'un réseau neuronal, en général, les réseaux neuronaux sont considérés comme un système de boîte noire. Cependant, ils peuvent modéliser des phénomènes complexes [103], et sont plus perméables aux variables indépendantes colinéaires (multi colinéarité) entre autres avantages. En général, l'architecture d'un réseau neuronal est illustrée à la figure 2.19.

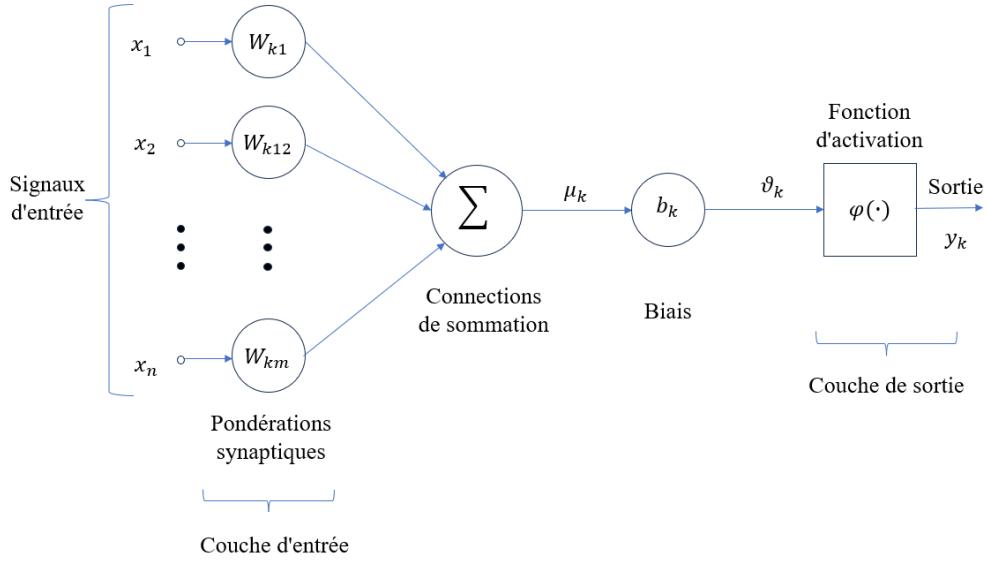


Figure 2.19 : Architecture générale d'un réseau de neurones

Où:

- x_1, x_2, \dots, x_n sont les signaux d'entrée
- $W_{k1}, W_{k2}, \dots, W_{kn}$ sont les poids synaptiques du neurone k .
- ϑ_k est la sortie du combinateur linéaire due aux signaux d'entrée.
- b_k est le biais.
- $\varphi(\cdot)$ est la fonction d'activation.
- y_k est le signal de sortie du neurone.
- μ_k est la fonction de transfert, qui combine plusieurs entrées en une valeur de sortie afin que la fonction d'activation puisse être appliquée.

En termes mathématiques, un neurone k est décrit par les équations 2.9, 2.10 et 2.11.

$$\mu_k = \sum_{j=1}^m w_{kj} x_j \quad (2.9)$$

$$\vartheta_k = \mu_k + b_k \quad (2.10)$$

$$y_k = \varphi(\vartheta_k) \quad (2.11)$$

Les fonctions d'activation définissent le neurone de sortie et doivent être choisies parmi les possibilités suivantes : fonction de seuil « threshold function », fonction linéaire par segments « piecewise-linear function », fonction sigmoïde « sigmoid function », rectified linear unit ReLU, tangente hyperbolique Tanh, entre autres fonctions.

Pour développer un modèle d'apprentissage automatique basée sur des réseaux des neurones, les signaux d'entrée ou variables d'entrée, le nombre des neurones, le nombre de couches entre la couche d'entrée et la couche de sortie sont ajustés. Par conséquent, les architectures de type réseau neuronal multicouche sont définies comme dans la figure 2.20.

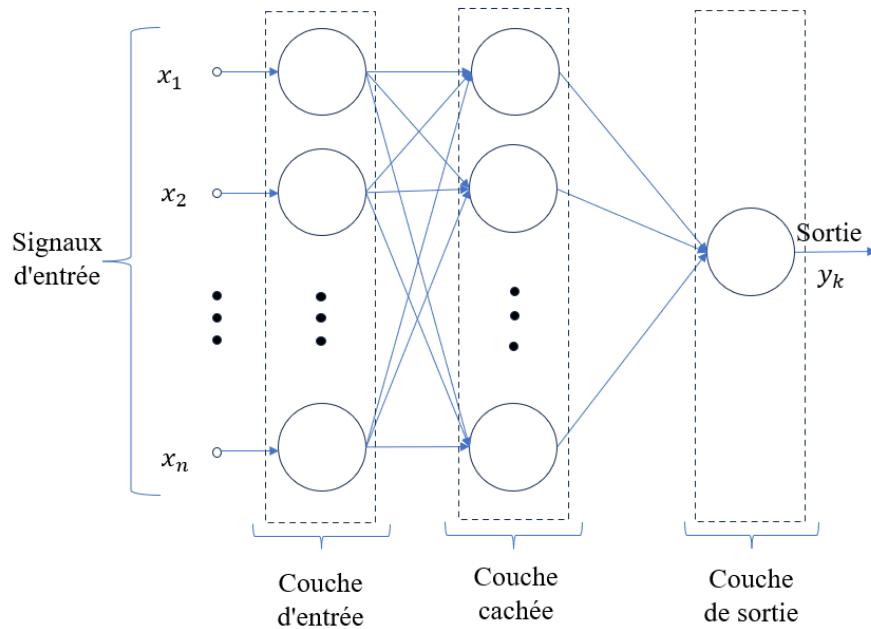


Figure 2.20 : Architecture des réseaux neuronaux multicouches

Dans la famille des réseaux neuronaux récurrents (RNN), il est possible de distinguer deux types de réseaux neuronaux récurrents, le « Nonlinear autoregressive models with exogenous input (NARX) » et le « long short-term memory (LSTM) ».

2.4.7.1 Réseaux neurones vers l'avant « Feed Forward Neural Network (FFNN) »

Ce type de réseau neuronal est composé de plusieurs couches et de plusieurs neurones interconnectés entièrement et se caractérisent en particulier par la mise en œuvre d'un algorithme d'apprentissage par rétro-propagation [104]. Dans un réseau neuronal de type feedforward, chaque nœud se connecte au nœud de la couche suivante. Les données circulent constamment, d'une couche à l'autre, sans boucle ni cycle. Les données ne font qu'avancer, d'où le nom du réseau neuronal.

2.4.7.2 Réseau neuronal recurrent « long short-term memory LSTM »

Ce type de réseau neuronal récurrent est également très utile pour les séries temporelles et dispose d'une mémoire explicite. À cette fin, il incorpore des unités cachées spécifiques qui ont naturellement tendance à retenir les entrées pendant une période prolongée. Une unité LSTM est composée d'une cellule, d'une porte d'entrée, d'une porte de sortie et d'une porte d'oubli. La cellule mémorise des valeurs sur des intervalles de temps arbitraires et les trois portes régulent le flux d'informations associé à la cellule. L'architecture LSTM consiste en un ensemble de sous-réseaux connectés de manière récurrente, appelés blocs de mémoire. L'idée du bloc de mémoire est de maintenir son état dans le temps et de

réguler le flux d'informations en pensant à des unités de contrôle non linéaires. Dans l'architecture LSTM, il est possible de distinguer un bloc d'entrée, cette étape est consacrée à la mise à jour de la composante d'entrée du bloc, qui combine l'entrée actuelle $x^{(t)}$ la sortie de cette unité LSTM $y^{(t-1)}$ lors de la dernière itération. Porte d'entrée, dans cette étape, la porte principale est mise à jour pour combiner l'entrée actuelle, la sortie de cette unité LSTM $y^{(t-1)}$ et la valeur de la cellule $c^{(t-1)}$ à la dernière itération. Porte de l'oubli, dans cette étape, l'unité LSTM détermine quelles informations doivent être supprimées de ses états de cellule précédents $c^{(t-1)}$. Cellule, cette étape calcule la valeur de la cellule, qui combine l'entrée du bloc $z^{(t)}$, la porte d'entrée $i^{(t)}$ et les valeurs de la porte d'oubli $f^{(t)}$, avec la valeur de la cellule précédente. Porte de sortie, cette étape calcule la porte de sortie, qui combine l'entrée actuelle $x^{(t)}$, la sortie de cette unité LSTM $y^{(t-1)}$ et la valeur de la cellule $c^{(t-1)}$ dans la dernière itération. Bloc de sortie, calcule finalement la sortie du bloc, qui combine la valeur actuelle de la cellule $c^{(t)}$ avec la valeur actuelle de la porte de sortie [105]. La figure 2.21 montre l'architecture d'un réseau LSTM.

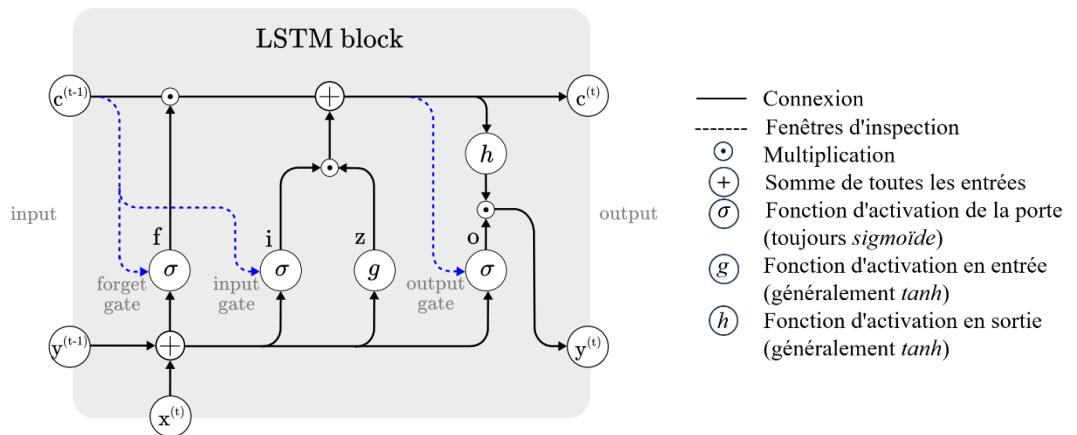


Figure 2.21 : Architecture d'un réseau LSTM, adapté de [105]

2.4.7.3 Hyperparamètres du réseau neuronal

Lors du développement d'un réseau neuronal, les hyperparamètres suivants doivent être sélectionnés :

- Algorithme d'optimisation

Un algorithme d'optimisation est la méthode par laquelle un modèle ajuste ses paramètres internes afin d'obtenir des prévisions beaucoup plus précises et de réduire les erreurs. Ce paramètre doit être sélectionné lors du développement d'un réseau neuronal.

Les algorithmes suivants peuvent être utilisés : « Gradient descent », « Newton method », « Root Mean Squared Propagation (RMSprop) », « Adaptive Moment Estimation (ADAM) », « Levenberg-Marquardt (LM) » sont couramment utilisés. RMSprop est un algorithme qui modifie le taux d'apprentissage en fonction de la situation afin d'accumuler le gradient [106]. ADAM combine les algorithmes RMSprop et momentum, dans lesquels la mise à jour ne prend en compte que la version lisse du gradient et inclut également un mécanisme de correction des biais [106]. LM est une technique itérative qui localise le minimum d'une fonction multivariée exprimée comme la somme des carrés de fonctions non linéaires à valeurs réelles [107].

- Le taux d'apprentissage

Le taux d'apprentissage est un hyperparamètre qui contrôle le degré de modification du modèle en réponse à l'erreur estimée à chaque fois que les poids du modèle sont mis à jour.

- Le mini-lots « mini batch »

La descente de gradient par mini-lots est une variante de l'algorithme de descente de gradient qui divise l'ensemble de données d'apprentissage en petits lots utilisés pour calculer l'erreur du modèle et mettre à jour les coefficients du modèle.

- Nombre d'époques

Une époque signifie entraîner le réseau neuronal avec toutes les données d'entraînement pour un cycle. Une passe avant et une passe arrière sont comptées ensemble comme une seule passe. Une époque se compose d'un ou de plusieurs lots, dans lesquels il faut utiliser une partie de l'ensemble de données pour former le réseau neuronal.

- Probabilité d'abandon « drop probability »

La probabilité d'abandon définit si tous les paquets ou seulement une fraction d'entre eux sont abandonnés lorsque le seuil maximal est atteint. Un taux d'abandon de 0,5 signifie que chaque neurone a 50 % de chances d'être abandonné à chaque étape d'entraînement.

- Période d'abandon du taux d'apprentissage

Nombre d'époques pour abandonner le taux d'apprentissage spécifié. Ce facteur multiplie le taux d'apprentissage global par le facteur d'abandon à chaque fois que le nombre d'époques spécifié arrive à son terme.

2.4.8 Méthodes de sélection des variables

Un point critique pour la mise en œuvre d'une méthode d'apprentissage automatique est la sélection des prédicteurs ou des variables indépendantes, également appelée sélection des caractéristiques. Il s'agit de trouver les meilleures variables indépendantes qui permettent d'estimer la variable dépendante avec précision. Cela signifie aussi trouver

les meilleures variables indépendantes ou prédicteurs qui fonctionnent le mieux ensemble dans le modèle. Les conséquences d'une mauvaise sélection des variables peuvent provoquer des phénomènes tels que le surajustement « overfitting », la faible précision des poids des prédicteurs et l'instabilité computationnelle en matière d'estimation de modèles [108]. Le processus de sélection est un algorithme mathématique qui permet de sélectionner les variables indépendantes qui sont fortement liées à la variable dépendante. En fonction des algorithmes, les méthodes de sélection des caractéristiques peuvent être classées dans quatre catégories : méthodes de filtrage, méthodes d'enveloppe, méthodes intégrées, méthodes hybrides [109]. Dans les méthodes enveloppantes, il est possible de trouver une sous-catégorie appelée emboîtée [110] où il est possible de trouver : la sélection du meilleur sous-ensemble « best subset selection », la sélection pas à pas à rétroaction « backward », la sélection pas à pas à l'avance « forward », la sélection hybride de sous-ensembles « hybrid » [111]. Dans les méthodes d'enveloppement, il est également possible d'identifier une autre sous-catégorie basée sur le poids des prédicteurs, appelée élimination récursive des caractéristiques « recursive feature elimination (RFE) ». Une autre façon de réduire la multi colinéarité est d'utiliser des méthodes de régularisation qui ont été expliquées précédemment comme la régression LASSO, la régression rigide et la régression élastique nette « Elastic Net regression » [97]. La sélection de la méthode est liée au degré de multi colinéarité détecté, au nombre de variables indépendantes et à la quantité de données disponibles. Voici une brève explication de chaque méthode de sélection des caractéristiques et de la méthode d'évaluation d'un modèle d'apprentissage automatique.

2.4.8.1 Sélection du meilleur sous-ensemble

La meilleure sélection de sous-ensembles ou « best subset selection » est une méthode mathématique développée par Gupta et Yang [112]. La procédure consiste à calculer un critère d'information statistique pour chaque variable indépendante pour tous les modèles possibles $2^k - 1$, où k est le nombre de variables indépendantes [113]. Si la variable indépendante ne remplit pas les conditions données, le coefficient estimé est égal à 0 [114]. La sélection du meilleur sous-ensemble est considérée comme l'une des procédures les plus exigeantes en termes de calcul si le nombre de variables indépendantes est considérable [113], mais elle permet de trouver le meilleur modèle parmi toutes les combinaisons possibles compte tenu d'un critère. En outre, la meilleure sélection de sous-ensembles parvient à trouver un ensemble de variables indépendantes avec un niveau élevé de signification, c'est-à-dire une valeur une valeur inférieure à $p < 0.05$, et par conséquent, la méthode de la meilleure sélection de sous-ensembles est considérée comme l'une des plus performantes dans les méthodologies de sélection de variables [115].

2.4.8.2 Sélection pas à pas à rétroaction

La sélection pas à pas à rétroaction est un algorithme qui part d'un modèle complet prenant en compte toutes les variables indépendantes disponibles et les élimine ensuite une à une. La procédure élimine la variable indépendante qui contribue le moins à réduire l'erreur de prédiction. Le processus s'achève lorsque toutes les variables sont significatives [116]. Cependant, certains auteurs suggèrent que cette méthode est très inefficace dans les modèles de grandes bases de données [117].

2.4.8.3 Sélection pas à pas à l'avance

Contrairement à la méthode précédente, la procédure de sélection pas à pas consiste à partir d'un modèle vide et à ajouter progressivement des variables ayant une forte corrélation avec la variable dépendante. L'algorithme ajuste les poids des variables indépendantes à chaque fois qu'une nouvelle variable est ajoutée et se termine lorsque la dernière variable a un coefficient non significatif [116]. Cette méthode est considérée comme une version de la méthode de sélection du meilleur sous-ensemble. Bien qu'une discussion approfondie sur sa validité soit en cours dans la littérature [111], certains auteurs s'interrogent sur le fait que parfois, au cours du processus d'élimination des variables, celles-ci ne peuvent pas être réintroduites comme le fait la méthode de sélection des sous-ensembles.

2.4.8.4 Sélection hybride de sous-ensembles

Les méthodes hybrides de sélection des caractéristiques sont un type de méthodes basées sur l'ensemble qui visent à construire un groupe de sous-ensembles de caractéristiques à partir de différents algorithmes et à produire un résultat agrégé à partir du groupe. Les méthodes hybrides de sélection des caractéristiques se composent de deux étapes : (1) construire un ensemble de résultats différents de sélection de caractéristiques et (2) agréger les différents résultats en un résultat consensuel. Les méthodes hybrides peuvent être considérées comme une combinaison d'algorithmes de sélection de caractéristiques multiples (par exemple méthodes de filtrage, méthodes d'enveloppe, méthodes intégrées) [118].

2.4.8.5 Élimination récursive des caractéristiques (RFE)

L'élimination récursive des caractéristiques ou « recursive feature elimination (RFE) » est une adaptation de l'élimination à rétroaction comme algorithme intégré. L'approche RFE implique un processus itératif d'entraînement d'un modèle de réseaux des neurones en utilisant initialement toutes les variables d'entrée candidates, puis en supprimant une ou plusieurs variables d'entrée à chaque fois sur la base de l'importance des poids correspondant à chaque variable d'entrée. La technique a été développée et appliquée avec succès à l'aide de machines à vecteurs de support (SVM), mais elle est extensible à d'autres architectures ANN en fournissant une expression appropriée pour le poids d'une variable d'entrée donnée [110].

2.4.9 Validation du modèle

La dernière étape de la sélection du modèle est l'évaluation de la performance des modèles. Cette étape consiste à estimer l'erreur de prédiction du modèle et à ajuster les paramètres de prédiction si nécessaire [119]. Les données disponibles doivent être divisées en données d'entraînement, de validation et de test. Il existe plusieurs stratégies de sous-échantillonnage des données, telles que le sous-échantillonnage aléatoire à maintien unique, le sous-échantillonnage aléatoire, la validation croisée k-fold, la validation croisée « leave-one-out » et le « Jackknife » [119]. En général, la validation croisée avec $k = 10$ est la stratégie la plus recommandée pour la validation d'un modèle en l'apprentissage automatique. Cette stratégie consiste à évaluer le modèle avec différents paquets de données de test formés aléatoirement et à trouver l'erreur et la performance du modèle sur chaque sélection de données. Les paquets de données de test peuvent être formés avec une

proportion de 10 %, 20 %, 30 %, 40 % par rapport au total des données disponibles. De cette manière l'évaluation du modèle avec ces données de test est effectuée 10 fois en observant comment se comporte l'erreur et la performance du modèle. Si la variation est significative, il est dit que le modèle n'est pas généralisable. Également la variation indique que le modèle ne s'ajuste pas correctement aux données. Aussi, cette problématique peut être liée à la nécessité de disposer de plus de données ou à une sélection inadéquate de prédicteurs.

2.4.10 Critères d'information, métriques de validation et indicateurs d'erreur

Pour la sélection des prédicteurs, il existe des critères statistiques de sélection des prédicteurs qui mesurent la performance des modèles et sélectionnent le meilleur modèle parmi un ensemble de modèles potentiels [120]. À cet effet, les critères d'information, également appelés critères de pénalité, généralement utilisés à cette fin sont les suivants: le critère d'information d'Akaike (AIC), le Cp de Mallows, le R^2 , le R^2 ajusté, le critère d'information bayésien (BIC), le « mean squared error (MSE) » et le « root mean squared error (RMSE) » [121]. L'AIC et le BIC sont des critères largement discutés dans la littérature statistique. Les deux offrent un bon critère de pénalité. Cependant, certains auteurs suggèrent le critère BIC car il pénalise davantage que le critère AIC [122].

De plus, pour évaluer les performances d'un modèle, certaines métriques doivent être sélectionnées comme le coefficient de détermination R^2 et indicateurs d'erreur de prédiction du modèle [123]. Les indicateurs d'erreur de prédiction suivantes seront utilisés dans le présent projet pour évaluer le modèle : MSE, RMSE, MAE, et RSE.

Le coefficient de corrélation de Pearson r est également utilisé pour analyser la mesure de la relation linéaire entre deux variables.

2.4.10.1 Coefficient de détermination R^2

Le coefficient de détermination R^2 peut être interprété comme la proportion de la variance de la variable dépendante qui est prévisible à partir des variables indépendantes. L'équation 2.12 définit cet indicateur.

$$R^2 = 1 - \frac{\sum_{i=1}^m (X_i - Y_i)^2}{\sum_{i=1}^m (\bar{y} - Y_i)^2} \quad (2.12)$$

Ou, m est la quantité de données disponibles, X_i est la valeur prédite, Y_i est la valeur actuelle ou mesuré et \bar{y} est la moyenne de la valeur mesurée. Une valeur mauvaise est égale à $-\infty$, une bonne valeur est proche de +1. Il est considéré un R^2 comme mauvais en dessous de 0.3, bon entre 0.4 et 0.7, très bon entre 0.7 et 0.99 [124].

2.4.10.2 L'erreur quadratique moyenne « Mean square error » MSE

L'erreur quadratique moyenne peut être utilisée s'il existe des valeurs aberrantes à détecter. En fait, l'erreur quadratique moyenne est très utile pour attribuer des poids plus importants à ces valeurs. Si le modèle arrive à une seule prédiction très mauvaise, la partie au carré de la fonction amplifie l'erreur. L'équation 2.13 définit cet indicateur.

$$MSE = \frac{1}{m} \sum_{i=1}^m (X_i - Y_i)^2 \quad (2.13)$$

Une valeur mauvaise est égale à $+\infty$, une bonne valeur est proche de 0.

2.4.10.3 La racine de l'erreur quadratique moyenne « Root mean square error » RMSE

Mesure la différence moyenne entre les valeurs prédites par rapport aux valeurs mesurées du modèle, et est utilisé s'il est souhaité que les données aberrantes n'influencent pas l'erreur. L'équation 2.14 définit cet indicateur.

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (X_i - Y_i)^2} \quad (2.14)$$

Une valeur mauvaise est égale à $+\infty$, une bonne valeur est proche de 0.

2.4.10.4 L'erreur absolue moyenne « Mean absolute error » MAE

L'erreur absolue moyenne MAE peut être utilisée si les valeurs aberrantes représentent des parties corrompues des données. MAE ne pénalise pas trop les valeurs aberrantes de l'apprentissage et fournit donc une mesure de performance générique et limitée pour le modèle. L'équation 2.15 définit cet indicateur.

$$MAE = \frac{1}{m} \sum_{i=1}^m |X_i - Y_i| \quad (2.15)$$

Une valeur mauvaise est égale à $+\infty$, une bonne valeur est proche de 0.

2.4.10.5 L'erreur type résiduelle « Residual standard error » RSE

L'erreur type résiduelle (RSE) est une mesure utilisée pour évaluer la qualité de l'ajustement d'un modèle de régression. Elle mesure la distance moyenne entre les valeurs observées et les valeurs prédictives. L'équation 2.16 définit cet indicateur.

$$RSE = \sqrt{\frac{1}{m-2} \sum_{i=1}^m (X_i - Y_i)^2} \quad (2.16)$$

Une valeur mauvaise est égale à $+\infty$, une bonne valeur est proche de 0.

2.4.10.6 Coefficient de corrélation de Pearson

Le coefficient de corrélation de Pearson r est défini en statistique comme la mesure de la force de la relation entre deux variables et leur association. La lettre r indique la valeur, qui se situe entre -1 et +1. Le coefficient de corrélation peut être calculé à l'aide de la formule 2.17.

$$r = \frac{\sum(x_i - \bar{x}) \times (y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \times (y_i - \bar{y})^2}} \quad (2.17)$$

Si r est compris entre 0,8 et 1, il existe une forte corrélation entre les variables évaluées. Cette métrique permet également d'analyser un groupe de variables indépendantes dont il est intéressant de savoir lesquelles sont liées ou colinéaires afin d'éviter la multicolinéarité. À cette fin, une matrice de corrélation est réalisée, où les coefficients r entre les variables sont calculés. Ces matrices peuvent être illustrées par des cercles et des couleurs indiquant le degré de corrélation avec la taille et l'intensité du couleur.

CHAPITRE 3 - Infrastructure de capture des données

Afin de construire les modèles d'apprentissage automatique, l'infrastructure disponible dans le laboratoire L-TIPS a été configuré pour la capture des données. Comme indiqué, deux études ont été conçues. En particulier, pour l'étude 1, a été utilisé un système de capture de mouvement, une paire de pédales instrumentées et un système de capture et de contrôle de la puissance et de la cadence. Pour l'étude 2, ces mêmes équipements ont été utilisés, accompagnés d'un dispositif de capture de l'activité électromyographique. Les descriptions de ces technologies sont présentées ci-dessous.

3.1 Système de capture de mouvement

Afin d'obtenir les données cinématiques, le système de capture de mouvement Cycling 3DMA de la compagnie Stt (Stt Systems, San Sebastian, Espagne) est utilisé. La fréquence de capture était réglée sur 100 Hz. Ce système était composé de huit caméras Optitrack (NaturalPoint, Inc., Corvallis, États-Unis) positionnées autour du cycliste. Dans la présente recherche le protocole « Bike-fitting left side & floor » a été utilisée. Ce protocole est composé de huit marqueurs actifs positionnés sur le côté gauche du cycliste au niveau du pied (alignement de l'axe de la pédale), de la malléole latérale, du calcanéum (90° entre les marqueurs du pied et de la malléole), de l'épicondyle latéral du fémur, du grand trochanter, du grand tubercule de l'humérus, de l'épicondyle latéral de l'humérus et de la radio-ulnaire distale, figure 3.1. Trois marqueurs supplémentaires ont été placés sur le sol pour définir le plan de référence.

- ① Marqueur sur radio - ulnaire distale
- ② Marqueur sur épicondyle latéral de l'humérus
- ③ Marqueur sur grand tubercule de l'humérus
- ④ Marqueur sur grand trochanter
- ⑤ Marqueur sur l'épicondyle latéral du fémur
- ⑥ Marqueur sur malléole latérale
- ⑦ Marqueur sur le pied
- ⑧ Marqueur sur calcanéum et en ligne avec le 7

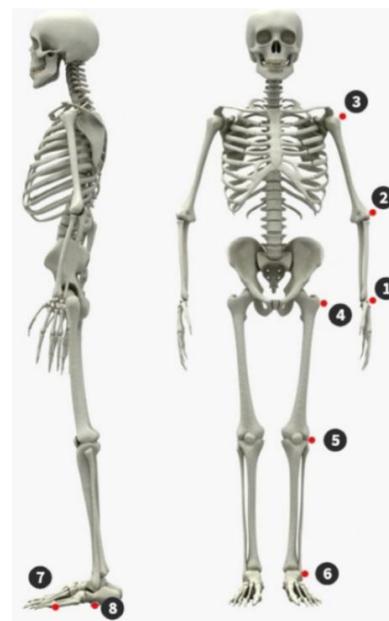


Figure 3.1 : Protocole biomécanique « Bike-fitting left side & floor »

Les données brutes du système de capture 3D peuvent fournir des informations sur les trajectoires des marqueurs placés sur des points anatomiques. Comme mentionné au point 2.14 ci-dessus, la présente étude se concentre sur le plan sagittal, c'est-à-dire sur les trajectoires dans les directions X et Y, ce qui correspond à un modèle bidimensionnel. Aussi, seul le membre inférieur a été pris en compte, car il est fortement lié à la production de force en cyclisme. Ce système a été utilisé pour l'ensemble des travaux, car il permet de déterminer les coordonnées des marqueurs avec lesquels toute la cinématique du membre inférieur est calculée. Comme mentionné ci-dessus, en raison du mouvement minimal du membre inférieur dans le plan frontal, les forces dans ce plan sont également minimes. En particulier, au cours de la présente étude, il a été constaté que les forces medio-latérales représentaient entre 10 % et 15 % des forces de pédalage effectives. Leur utilité dans le cyclisme réside principalement dans les pathologies liées au genou [125,

126]. Un code Matlab (MATLAB 2021a, MathWorks Inc, Natick, Massachusetts, États-Unis), développé par l'auteur de cette thèse, a été développé pour déterminer les angles (α, β, θ), les vitesses angulaires ($\dot{\alpha}, \dot{\beta}, \dot{\theta}$) et les accélérations angulaires ($\ddot{\alpha}, \ddot{\beta}, \ddot{\theta}$) de la hanche, du genou et de la cheville respectivement, comme le montre la figure 3.2. Dans la présente étude, l'angle α sera défini comme l'angle du fémur par rapport à l'horizontale et que nous appellerons l'angle du fémur. Les équations utilisées pour créer le modèle se trouvent en Annexe D.

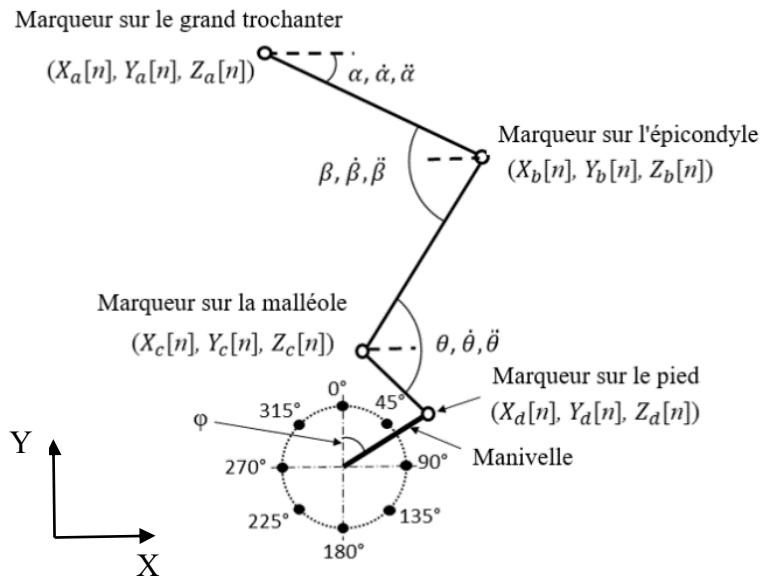


Figure 3.2 : Schème du protocole cinématique

Toutes les variables cinématiques ont été calculées par rapport à l'angle de la manivelle, de 0° à 360° . Ensuite, les valeurs d'angle, de vitesse angulaire et d'accélération angulaire de la hanche, du genou et du pied ont été sélectionnés à chaque 45° , soit 72 points de données cinématiques par participant ($8 \text{ points} \times 3 \text{ articulations} \times 3 \text{ valeurs cinématiques}$). Le code Matlab développé calcule la moyenne des valeurs cinématiques à chaque degré, mais tous les 45° ont été choisis car ils couvrent les points normalement analysés dans la littérature ($0^\circ, 90^\circ, 180^\circ$ et 270°) et le point intermédiaire entre ces valeurs, ce qui permet

de disposer d'une bonne variété de points pour faciliter l'analyse. La valeur discrète a été choisie car les variables dépendantes (IE, PIP) sont calculées de cette manière en tant que moyenne sur une période donnée.

3.2 Système de mesure des forces de pédalage

Pour la détermination des forces de pédalage, le système Forped (Forped, Cycling Science, Ljubljana, Slovenia) a été utilisé. La fréquence de capture des données est définie à 450 Hz. Ce système fournit les données brutes dans les directions X, Y et Z par rapport à la pédale. Ce système a été utilisé pour toutes les études, car il permet de déterminer les forces exercées sur les pédales. Ces pédales instrumentées ont enregistré des données cinétiques pour chaque sélection de puissance et cadence afin de déterminer les métriques de pédalage pour l'étude 1, et des séries temporelles de forces du pédalage dans les directions F_x, F_y, F_z pour l'étude 2. Le code Matlab mentionné pour calculer la cinématique des membres inférieurs, permet également de calculer les métriques utilisées dans le cyclisme. La figure 3.3 montre le système de pédales utilisé.

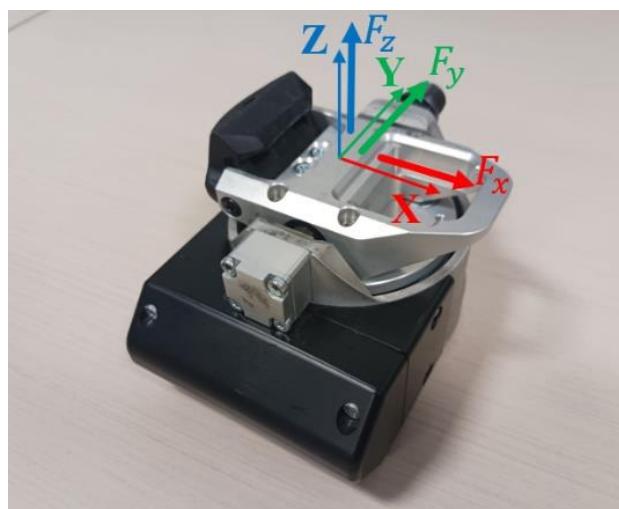


Figure 3.3 : Système de pédales Forped

3.3 Système de capture de l'activité musculaire

L'électromyographie de surface a été mesuré avec le système Delsys Trigno Wireless System (Delsys Inc, Boston, MA, USA). Le système est composé de 16 capteurs bipolaires de dimensions 27 mm × 37 mm × 13 mm, 14 g/ch. et plage d'entrée du signal EMG 11/22 mV. Le capteur est placé sur le ventre de chaque muscle, conformément aux recommandations du SENIAM. La fréquence d'acquisition des données est réglée à 2160 Hz. Afin d'obtenir un signal sans interférence et à faible impédance, la peau doit être préparée : rasage, abrasion avec un papier abrasif approprié et nettoyage avec de l'alcool pour enlever les impuretés. Les capteurs ont été fixés à l'aide d'un ruban adhésif double face Delsys renforcé par une bande « kinesio tape ». Ce système est utilisé dans l'étude 2, afin d'obtenir l'activité myoélectrique des muscles utilisés dans le mouvement de pédalage. Dans le but de normaliser les données électromyographiques, le test appelé contraction maximale volontaire communément abrégée MCV sera exécuté, selon les recommandations du SENIAM. La figure 3.4 montre le système d'électromyographie de surface utilisé.



Figure 3.4 : Système d'électromyographie Delsys

3.4 Système de contrôle et mesure de cadence et puissance

La cadence et la puissance sont contrôlées par la base d'entraînement Saris H3 (C+A Global, Madison, Wi, USA), sur laquelle le vélo est monté. Pour contrôler la puissance, l'application Saris a été utilisée en mode « ergo ». De plus, les données de puissance et de cadence sont enregistrées au moyen d'un compteur de vélo Garmin (Garmin International, Kansas City, MO, USA). La fréquence de capture était réglée sur 1 Hz. La figure 3.5 montre l'environnement de mesure installé.

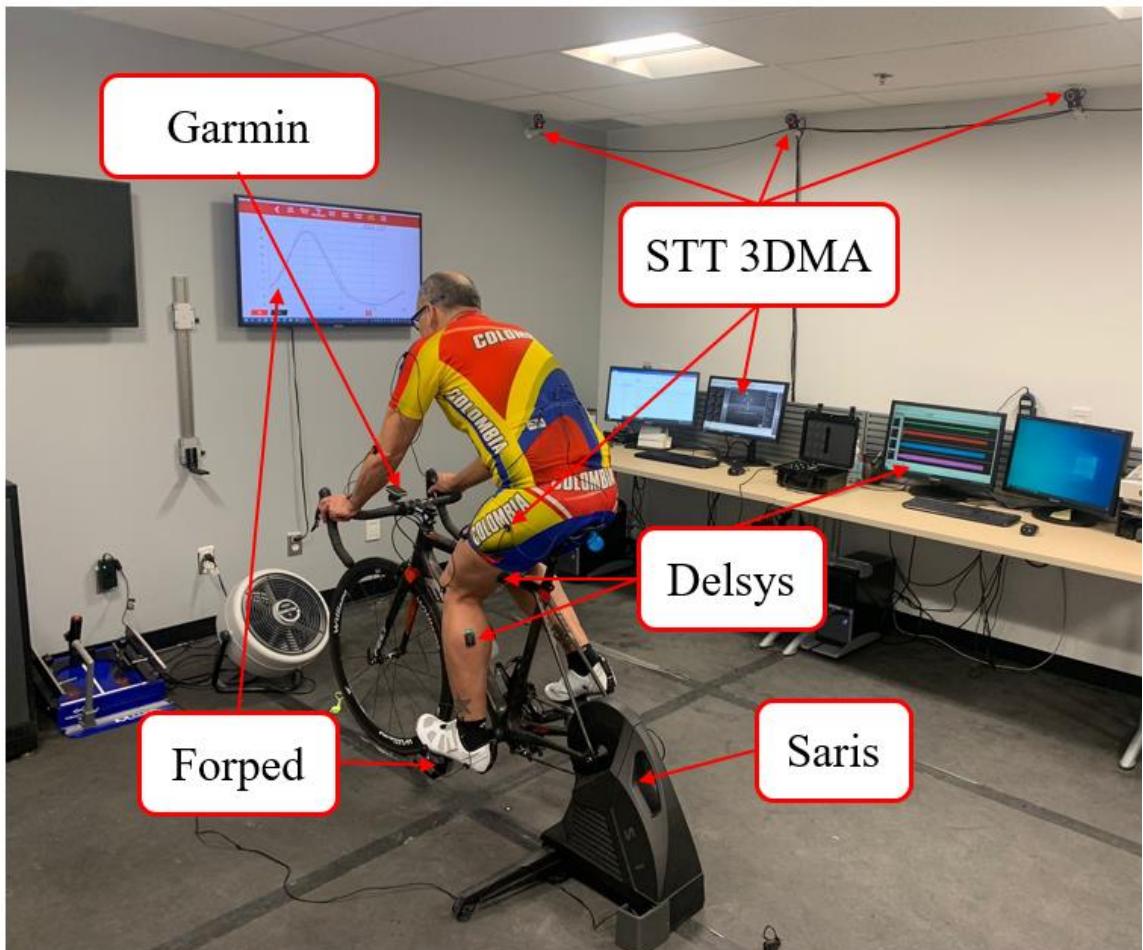


Figure 3.5 : Environnement de mesure

CHAPITRE 4 - Étude 1, prédition de l'indice d'efficacité et proportion d'impulsions positives

L'objectif de l'étude 1 consiste à prédire les paramètres qui évaluent la technique de pédalage à partir de la cinématique du pédalage. Les métriques sélectionnées sont l'indice d'efficacité et la proportion d'impulsions positives. À cette fin, la méthodologie suivante a été appliquée.

4.1 Méthodologie

Pour le développement de l'étude 1, une méthodologie a été élaborée à partir de l'infrastructure sélectionnée, puis la conception d'un protocole d'acquisition des données. Ensuite, un traitement postérieur des données afin d'implémenter les modèles et enfin une évaluation et une optimisation des modèles. Un schéma général de la méthodologie appliquée est présenté dans la figure 4.1



Figure 4.1 : Méthodologie appliquée

Ci-après est présentée la description de chacune de ces étapes.

4.2 Protocole d'acquisition des données

Pour sélectionner les différents ensembles de données dans l'étude 1, une échelle de Borg [127], sur l'échelle CR10, a été présentée aux participants, qui devaient adopter une puissance correspondant à 1/10 sur l'échelle de Borg pour l'échauffement pendant 5 min. Les participants ont effectué trois tests d'une minute à un effort perçu facile (2/10 sur l'échelle de Borg), modéré (3/10 sur l'échelle de Borg) et difficile (5/10 sur l'échelle de Borg) avec leur cadence de pédalage préférée (basée sur leur perception de confort), entrecoupés d'une minute de récupération à une intensité très facile (1/10 sur l'échelle de Borg) et à une cadence de pédalage libre. Ensuite, la session de test a été reproduite deux fois avec leur cadence de pédalage préférée de +10 rpm et -10 rpm (les trois puissances choisies lors de la première session ont été maintenues pour les deuxièmes et troisième sessions). L'application Saris en mode « ERG » a été utilisée pour la sélection de la puissance. Un schéma du protocole de l'étude 1 est présenté dans la figure 4.2.

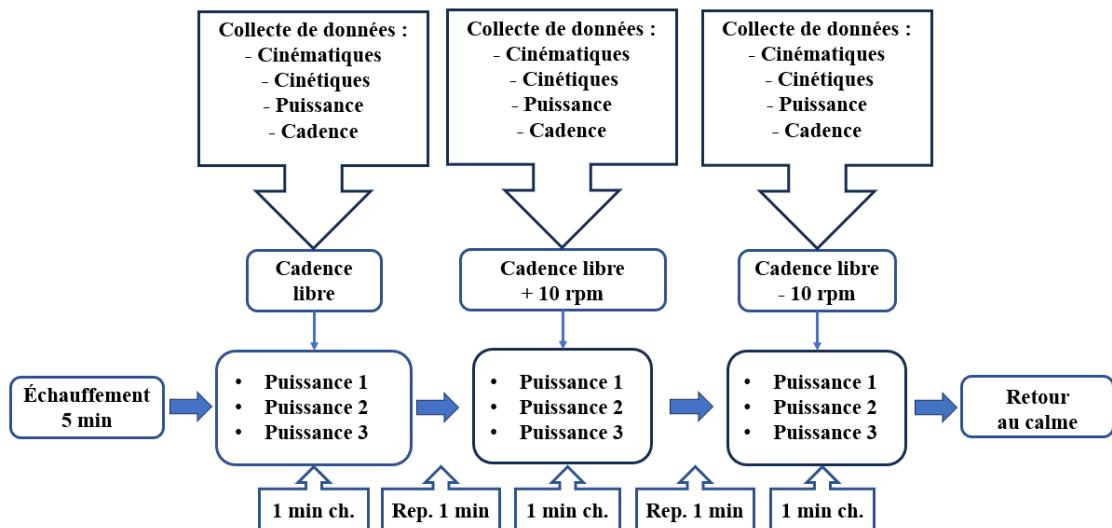


Figure 4.2 : Schéma du protocole d'étude 1

4.3 Participants

Dix-sept cyclistes en bonne santé se sont portés volontaires pour participer à cette étude (36 ± 12 ans, $1,74 \pm 0,06$ m, $72,4 \pm 8,6$ kg). Tous les participants étaient des cyclistes récréatifs sur route, selon la classification de Priego Quesada et al. [128]. Les cyclistes ont déclaré parcourir en moyenne 5 000 km par an. Sur la base de cette classification, et selon Ansley et Cangley [129], un cycliste amateur peut développer une puissance moyenne de 200 W à 275 W. Le tableau 4.1 présente les caractéristiques moyennes des cyclistes évalués.

Tableau 4.1 : Caractéristiques moyennes des cyclistes évalués

	Cadence	Puissance	IE	PIP
Moyen	85,42	203,2	34,23	81,43
Écart-type	10,06	52,01	7,69	7,06
Val. max	102	330	51,3	96,65
Val. min	60	100	15	60,26

Comme le montre le tableau 4.1, la valeur moyenne de la puissance ($203,2 \pm 52.01$ W) est proche de celle mentionnée par Asley et Cangley. De même, après avoir analysé les puissances sélectionnées selon l'échelle de Borg montrée aux cyclistes, il a été constaté que la puissance moyenne facile était de 160 ± 31.8 W (2/10), la puissance moyenne modérée de 203.2 ± 52.01 W (3/10) et la puissance moyenne difficile de 243.52 ± 47.88 W (5/10). Seules trois personnes ont obtenu des niveaux de puissance supérieurs à 300 W comme niveau de puissance difficile. Cela confirme la catégorie récréative de la plupart des cyclistes analysés.

Les participants n'ont pas rapporté de douleur ou de pathologie pouvant affecter leur technique de pédalage. Avant les expériences, ils ont été informés des risques et des bénéfices de l'étude et ont donné leur consentement éclairé par écrit. De plus, ils ont rempli un questionnaire sur l'aptitude à l'activité physique afin de s'assurer de leur capacité à réaliser le protocole (Questionnaire sur l'aptitude à l'activité physique - Q-AAP 2002). Cette étude a été approuvée par le comité d'éthique de l'UQTR (CER-21-277-07.27) conformément à la Déclaration d'Helsinki (Annexe E).

4.4 Post-traitement des données

Tous les signaux obtenus dans les deux études ont été traités afin d'obtenir le signal le plus propre possible, en éliminant le bruit causé par les mouvements vigoureux des membres inférieurs lors de l'acquisition des données du signal. Les données de capture du mouvement et les données provenant des pédales de force ont été filtrées avec un filtre type Butterworth passe-bas d'ordre 2 avec une fréquence de 10 Hz.

Enfin, si applicable, les données cinématiques et cinétiques collectées dans les deux études seront normalisées à l'aide de l'équation 3.1 [130]. Cette technique permet de mettre toutes les données à l'échelle dans la plage (0, 1).

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (3.1)$$

Ou X' est la valeur normalisée, X est la valeur originale, X_{min} est la valeur minimale de X , est la valeur maximale de X .

En raison des différentes fréquences de capture des trois technologies utilisées dans cette recherche, 100 Hz pour le système de capture du mouvement, 450 Hz pour les

pédales instrumentées et 1 Hz pour le système de capture de la puissance et de la cadence, un rééchantillonnage à 100 Hz a été nécessaire.

4.5 Implémentation des modèles

Un modèle basé sur des méthodes d'apprentissage automatique consiste en la prédiction d'une variable d'intérêt ou d'un signal temporel continu, à partir de l'entraînement d'un algorithme avec des données du mouvement étudié. Par conséquent, à partir de données d'entrée, également appelées variables indépendantes ou prédicteurs, le modèle prédit une variable de sortie, appelée variable dépendante. Comme stratégie, l'étude 1 cherche à appliquer le plus grand nombre possible de techniques d'apprentissage automatique disponibles dans la littérature afin de sélectionner les meilleures. De cette manière, pour le développement des modèles, l'indice d'efficacité IE et la proportion d'impulsion positive PIP ont été définis comme des variables dépendantes, en calculant la moyenne de ces métriques sur la période d'évaluation (1 min). Comme mentionné précédemment, cette recherche se concentre sur le plan sagittal, en tenant compte que les mouvements dans le plan frontal sont minimes. En conséquence, au niveau des prédicteurs candidats, les 72 variables cinématiques décrites à la section 3.1.1 sont définis comme une partie des prédicteurs. Ceci est justifié par le fait que d'un point de vue biomécanique, la production de force en cyclisme combinerait différents phénomènes cinématiques : les angles des articulations, les vitesses angulaires, et les accélérations angulaires. Les angles entre les segments du corps sont importants car ils affectent la production de force dans le cyclisme, comme expliqué au point 2.2.2 ci-dessus. Aussi, les vitesses angulaires des segments sont les indicateurs d'une contraction musculaire [131]. En plus, les accélérations angulaires

des segments du corps sont liés aux couples produits au niveau des articulations [132]. En outre, la puissance, la cadence de mouvement, la masse de l'individu, et KOPS influencent aussi la production de force en cyclisme [133]. Le paramètre KOPS est extrait du logiciel STT 3DMA et est calculé comme défini dans le point 2.1.4. Par conséquent, 76 prédicteurs au total sont définis.

4.5.1 Modélisation par régression linéaire

Les différents modèles de régression de l'étude 1, ont été développés à l'aide du logiciel R Studio (R Core Team, Vienne, Autriche, 2021 ; <https://www.R-project.org/>). La figure 4.3 montre un schéma de la méthodologie employée pour le développement des modèles de régression.

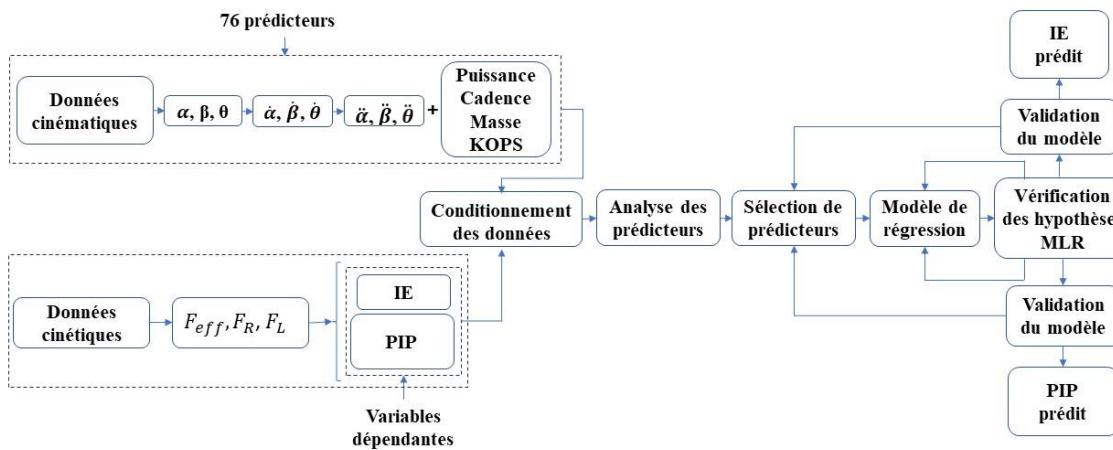


Figure 4.3 : Diagramme de la méthodologie modèles de régression

Comme le montre la figure 4.3, une fois que les données sont disponibles, la première étape de la méthodologie est le conditionnement des données. Ce processus consiste à normaliser les données entre 0 et 1. Ensuite, une analyse des prédicteurs est effectuée,

pour laquelle une matrice de corrélation est réalisée afin de calculer le coefficient de corrélation de Pearson r, qui doit être proche de 0 pour l'absence de relation entre les variables et proche de 1 pour la corrélation entre les variables. En raison du nombre considérable de variables, le résultat est une matrice de 76 x 76, ce qui n'est pas facile à présenter en forme graphique dans la présente thèse, en raison de sa taille. Néanmoins, en l'analysant cette matrice, la conclusion qui apparaît est qu'il existe de multiples colinéarités, ce qui permet de conclure qu'un modèle de régression composé de 76 prédicteurs serait un modèle non fonctionnel en raison de la présence de ce phénomène. Ce modèle poserait le problème du surapprentissage ou « overfitting », et il est donc nécessaire de passer à l'étape suivante, la sélection des prédicteurs. Toutefois, la régression linéaire multiple composée de 76 prédicteurs sera présentée dans les résultats à des fins de comparaison et de discussion. La figure 4.4 présente un diagramme de flux de travail des méthodes de sélection des prédicteurs.

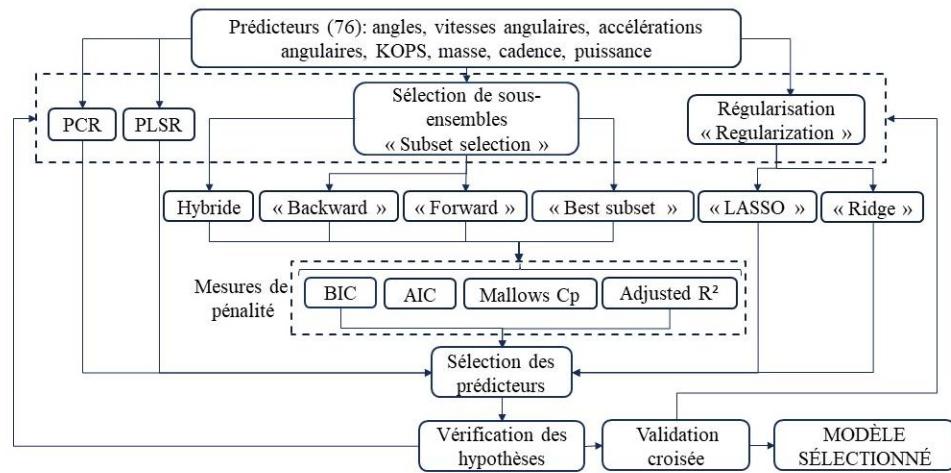


Figure 4.4 : Flux de travail sélection des prédicteurs

Comme le montre la figure 4.4, à l'aide du logiciel R Studio, 8 méthodologies de sélection de prédicteurs ont été exécutées. De cette manière, 8 modèles de régression

multiple sont obtenus qui doivent ensuite être évalués via la vérification d'hypothèses de régression linéaire multiple (linéarité, absence de multi colinéarité, homoscédasticité et variables normalement distribuées). La dernière étape est la validation du modèle via une validation croisée. Dans ce projet, la validation croisée a été appliquée avec des packages de 10 %, 20 %, 30 % et 40 % de données de test et 90 %, 80 %, 70 % et 60 % de données d'entraînement respectivement. Pour évaluer les performances du modèle, le R^2 a été utilisé et pour évaluer l'erreur du modèle, l'erreur standard résiduelle RSE a été utilisée.

4.5.2 Modélisation par arbre de régression

La figure 4.5 montre un schéma de la méthodologie pour le développement des modèles basé sur les arbres de régression.

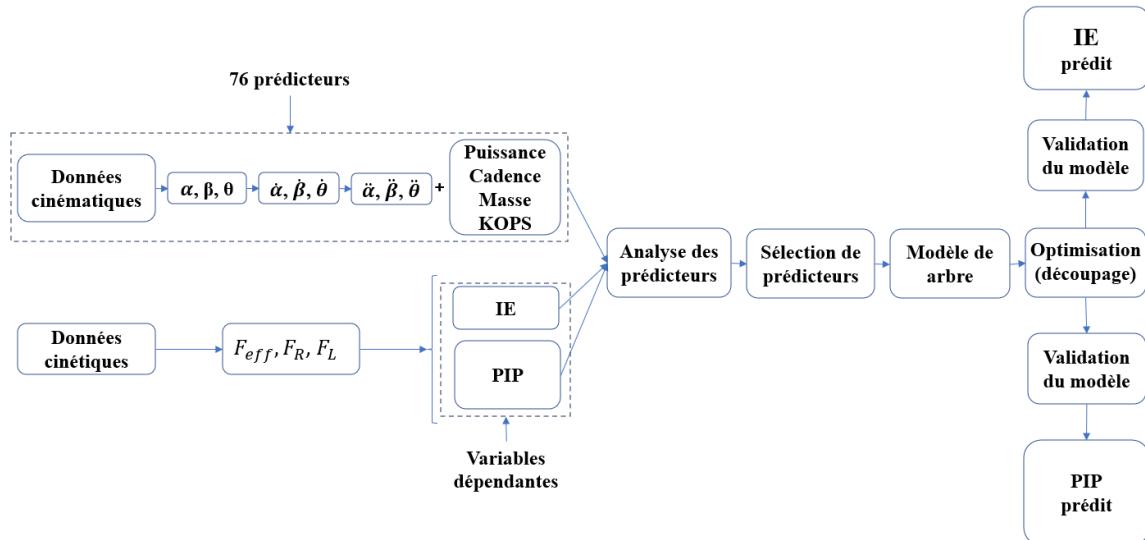


Figure 4.5 : Diagramme méthodologie des modèles basé sur arbres de régression

Pour le développement de modèles basés sur des arbres de régression, la méthodologie est similaire à celle de la régression linéaire multiple. La différence réside dans le fait qu'il n'est pas nécessaire de normaliser les données et dans l'apparition d'une étape supplémentaire appelée optimisation. L'optimisation consiste à améliorer la précision de

l'arbre par l'élagage des feuilles. Cet élagage est en fait une fonction mathématique qui permet d'optimiser l'arbre en réduisant sa complexité initiale.

4.5.3 Modélisation par réseaux de neurones

La figure 4.6 montre un schéma de la méthodologie pour le développement des modèles basés sur réseaux neuronaux.

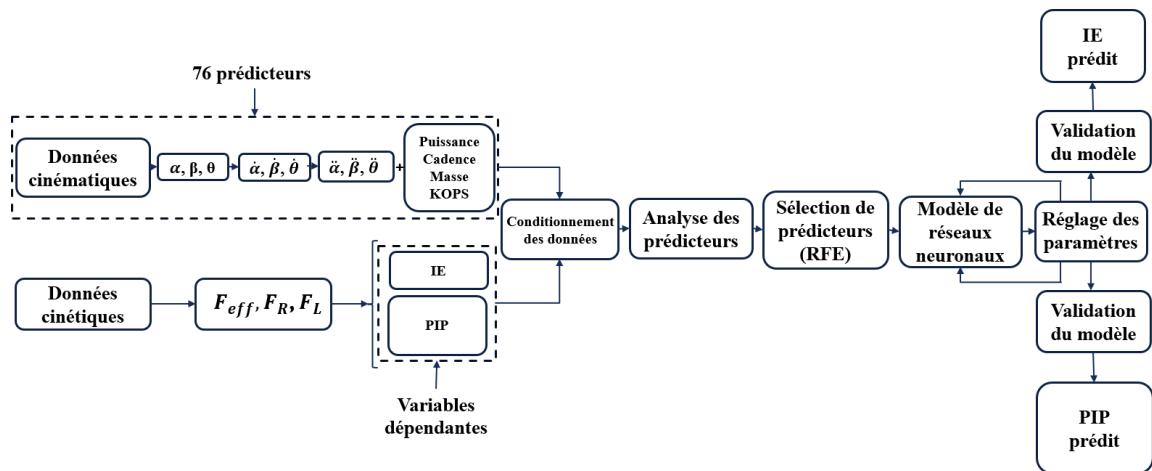


Figure 4.6 : Diagramme méthodologie modèles basés sur réseaux neuronaux

La méthodologie est similaire à celle suivie pour un modèle de régression, mais un réglage doit être effectué pour la sélection des paramètres du réseau neuronal. Le réglage consiste à déterminer le nombre de couches, le nombre de neurones, la fonction d'activation et le nombre d'itérations c'est-à-dire, le nombre de fois que le réseau est entraîné. Pour cette recherche de paramètres, il n'existe pas de méthode spécifique, la recommandation étant de procéder par essais et erreurs jusqu'à ce que le réseau le plus optimal soit trouvé, avec le coefficient de détermination R^2 le plus élevé et l'erreur RSE le plus basse.

Il convient de noter ici qu'une autre différence par rapport à la méthode de régression linéaire multiple réside dans l'utilisation d'une méthode différente de sélection des prédicteurs, la méthode d'élimination récursive des caractéristiques (RFE). Notamment, dans le cadre du présent projet, des réseaux neuronaux ont également été développés avec la méthode des meilleurs sous-ensembles « subset selection », mais aucun meilleur résultat n'a été obtenu pour les mêmes prédicteurs. Par conséquent, une autre méthode a été recherchée pour produire plus de prédicteurs, puisque les réseaux neuronaux sont mieux adaptés à la présence de multi-colinéarités. Cela s'explique par le fait qu'ils disposent de plus d'informations pour prédire la variable de sortie [134]. De cette manière, 21 prédicteurs sont déterminés pour prédire l'IE et 28 prédicteurs pour prédire le PIP.

- Paramètres du réseau de neurones

Deux réseaux neuronaux ont été développés pour chaque métrique. Chaque réseau a un nombre différent de prédicteurs obtenus par la méthode d'élimination récursive des caractéristiques (RFE). Cependant, ils ont été développés avec les mêmes hyperparamètres. De cette manière, la fonction d'activation est la fonction logistique, le taux d'apprentissage atteint 0,01, l'algorithme d'apprentissage utilisé est l'algorithme de rétropropagation résiliente « resilient backpropagation algorithm (RMSprop+) ». Enfin, le réseau neuronal est composé de trois couches (une couche d'entrée, une couche cachée et une couche de sortie) de 35 neurones pour l'IE et pour le PIP. Le nombre d'époques a été fixé à 50, mais l'erreur minimum pour les deux réseaux neuronaux est atteinte à l'époque 15. Les réseaux neuronaux ont été entraînés avec un ratio de 90 % de données d'entraînement et 10 % de données de test. La figure 4.7 montre l'architecture des deux réseaux neuronaux développés.

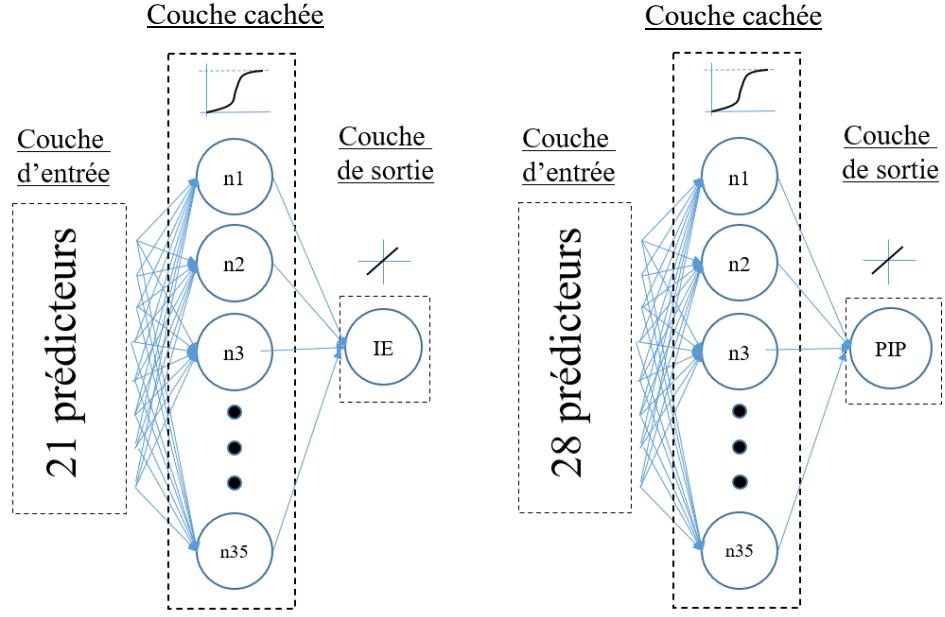


Figure 4.7 : Architecture des deux réseaux neuronaux étude 1

4.6 Résultats

4.6.1 Résultats sélection de prédicteurs

Plusieurs procédures de sélection de variables avec différents critères de pénalisation ont été testées pour trouver les meilleurs modèles identifiant les variables indépendants correspondantes. Avec chaque sélection de prédicteurs, différentes régressions linéaires multiples sont construites. De cette manière, les résultats de chaque sélection de prédicteurs (voir section 2.4.8.) sont présentés dans le tableau 4.2, 4.3 et 4.4 pour prédire l'indice d'efficacité IE et le tableau 4.5, 4.6 et 4.7 pour prédire la proportion d'impulsions positives. Cependant, bien que tous les critères d'information aient été utilisés pour évaluer le pouvoir explicatif de chaque prédicteur sélectionné dans chaque méthode, seules les variables sélectionnées avec le critère BIC sont présentées. Ce critère a été retenu car c'est

celui qui permet d'obtenir le nombre réduit de prédicteurs en tenant compte sa valeur de pénalité élevée, afin d'obtenir le modèle le plus parcimonieux possible. Une fois les prédicteurs sélectionnés, une régression linéaire multiple est effectuée, indiquant le RSE et le R^2 ajusté, ainsi que le niveau de signification, c'est-à-dire le *p-value* de chaque prédicteur. À des fins explicatives, le premier modèle développé est une régression linéaire multiple prenant en compte toutes les prédicteurs, c'est-à-dire 76, pour prédire chaque variable dépendante. À partir des paramètres définis dans la figure 3.2, dans les tableaux suivants pour identifier le prédicteur, la lettre grecque indique l'angle de l'articulation, c'est-à-dire α = hanche, β = genou, θ = cheville, et le sous-indice l'angle de la manivelle entre 0° et 360° à laquelle la variable cinématique est mesurée. De plus, KOPS est « Knee Over Pedal Spindle », MASS est la masse du cycliste, CAD est la cadence de pédalage et PUISS est la puissance délivrée aux manivelles.

Tableau 4.2 : Sélection de prédicteurs, indice d'efficacité « best subset selection »

Méthode	Aucune sélection	Best subset selection
Critère de pénalisation	N/A	BIC
# Prédicteurs sélectionnées	76	11
RSE	2.39	2.81
R ² Ajusté	0.90	0.86
Variables les plus significatives (***) p-value = 0	PUISS	$\alpha_{45^\circ} - \dot{\alpha}_{90^\circ} - \dot{\alpha}_{270^\circ} - \beta_{225^\circ} - \dot{\beta}_{180^\circ} - \dot{\beta}_{270^\circ} - \theta_{45^\circ} - \theta_{270^\circ} - \dot{\theta}_{270^\circ}$ MASS - PUISS
Variables les plus significatives (**) p-value < 0.001		$\ddot{\alpha}_{135^\circ} - \ddot{\beta}_{135^\circ} - \ddot{\theta}_{135^\circ}$ /
Variables les plus significatives (*) p-value < 0.01		$\dot{\alpha}_{135^\circ} - \ddot{\theta}_{45^\circ} - \ddot{\beta}_{45^\circ} - \ddot{\alpha}_{45^\circ} - \dot{\theta}_{135^\circ} - \dot{\beta}_{135^\circ} - \dot{\beta}_{90^\circ} - \dot{\beta}_{45^\circ}$ /
Variables moins significatives (.) p-value < 0.05		$\alpha_{270^\circ} - \dot{\alpha}_{90^\circ} - \dot{\theta}_{90^\circ} - \ddot{\alpha}_{225^\circ} - \ddot{\beta}_{0^\circ} - \ddot{\beta}_{225^\circ} - \ddot{\theta}_{225^\circ}$ /
Variables non significatives	$\alpha_0^\circ - \alpha_{45^\circ} - \alpha_{90^\circ} - \alpha_{135^\circ} - \alpha_{180^\circ} - \alpha_{225^\circ} - \alpha_{315^\circ} - \beta_0^\circ - \beta_{45^\circ} - \beta_{90^\circ} - \beta_{135^\circ} - \beta_{225^\circ} - \beta_{270^\circ} - \beta_{315^\circ} - \theta_0^\circ - \theta_{45^\circ} - \theta_{90^\circ} - \theta_{135^\circ} - \theta_{180^\circ} - \theta_{225^\circ} - \theta_{270^\circ} - \dot{\alpha}_0^\circ - \dot{\alpha}_{45^\circ} - \dot{\alpha}_{90^\circ} - \dot{\alpha}_{135^\circ} - \dot{\alpha}_{180^\circ} - \dot{\alpha}_{225^\circ} - \dot{\alpha}_{315^\circ} - \dot{\beta}_0^\circ - \dot{\beta}_{180^\circ} - \dot{\beta}_{225^\circ} - \dot{\beta}_{270^\circ} - \dot{\beta}_{315^\circ} - \dot{\theta}_0^\circ - \dot{\theta}_{45^\circ} - \dot{\theta}_{90^\circ} - \dot{\theta}_{135^\circ} - \dot{\theta}_{180^\circ} - \dot{\theta}_{225^\circ} - \dot{\theta}_{270^\circ} - \ddot{\alpha}_0^\circ - \ddot{\alpha}_{270^\circ} - \ddot{\alpha}_{315^\circ} - \ddot{\beta}_0^\circ - \ddot{\beta}_{180^\circ} - \ddot{\beta}_{270^\circ} - \ddot{\beta}_{315^\circ} - \ddot{\theta}_0^\circ - \ddot{\theta}_{180^\circ} - \ddot{\theta}_{270^\circ} - \ddot{\theta}_{315^\circ} - \text{KOPS CAD}$	/

**Tableau 4.3 : Sélection de prédicteurs, indice d'efficacité « forward selection »,
« backward selection »**

Méthode	Forward selection	Backward selection
Critère de pénalisation	BIC	BIC
# Prédicteurs sélectionnées	23	21
RSE	2.63	2.51
R² Ajusté	0.88	0.89
Variables les plus significatives (***)	$\dot{\theta}_{0^\circ} - \dot{\beta}_{45^\circ} - \dot{\alpha}_{90^\circ} - \dot{\alpha}_{0^\circ} - \beta_{315^\circ}$ $\beta_{45^\circ} - \text{PUISS} - \text{CAD}$	$\dot{\beta}_{180^\circ} - \dot{\theta}_{0^\circ} - \dot{\theta}_{45^\circ} - \dot{\theta}_{90^\circ} - \ddot{\alpha}_{135^\circ}$ $\ddot{\alpha}_{180^\circ} - \ddot{\alpha}_{225^\circ} - \ddot{\beta}_{90^\circ} - \ddot{\alpha}_{225^\circ} - \ddot{\theta}_{45^\circ}$ $\ddot{\theta}_{135^\circ} - \ddot{\theta}_{180^\circ} - \ddot{\theta}_{225^\circ} - \text{PUISS}$ MASS
p-value = 0		
Variables significatives (**) p-value < 0.001	$\ddot{\beta}_{90^\circ} - \dot{\theta}_{90^\circ} - \dot{\alpha}_{45^\circ}$	/
Variables significatives (*) p-value < 0.01	$\ddot{\theta}_{90^\circ} - \ddot{\alpha}_{180^\circ} - \ddot{\alpha}_{90^\circ} - \dot{\alpha}_{315^\circ}$	$\ddot{\alpha}_{270^\circ}$
Variables moins significatives (.) p-value < 0.05	$\dot{\alpha}_{180^\circ} - \dot{\alpha}_{0^\circ} - \dot{\alpha}_{45^\circ} - \ddot{\beta}_{0^\circ} - \ddot{\beta}_{135^\circ}$ $\ddot{\theta}_{180^\circ} - \ddot{\alpha}_{270^\circ} - \text{MASS}$	/
Variables non significatives	/	/

**Tableau 4.4 : Sélection de prédicteurs, indice d'efficacité, « hybrid selection »,
LASSO, Rigide**

Méthode	Hybrid selection	LASSO	Rigide
Critère de pénalisation	BIC	N/A	N/A
# Prédicteurs sélectionnées	49	45	73
RSE	2.20	2.57	3.87
R ² Ajusté	0.91	0.88	0.74
$\alpha_{0^\circ} - \alpha_{45^\circ} - \alpha_{90^\circ} - \alpha_{270^\circ}$			
$\beta_{0^\circ} - \beta_{180^\circ} - \beta_{225^\circ} - \beta_{270^\circ}$			
$\alpha_{45^\circ} - \alpha_{270^\circ} - \dot{\alpha}_{90^\circ} - \dot{\alpha}_{135^\circ}$			
$\dot{\beta}_{45^\circ} - \dot{\beta}_{90^\circ} - \dot{\beta}_{135^\circ} - \dot{\beta}_{180^\circ}$			
Variables les plus significatives (***) p-value = 0	$\dot{\theta}_{45^\circ} - \dot{\theta}_{90^\circ} - \ddot{\alpha}_{0^\circ} - \ddot{\alpha}_{45^\circ}$	$\dot{\alpha}_{135^\circ} - \dot{\alpha}_{180^\circ} - \dot{\alpha}_{225^\circ} - \dot{\alpha}_{270^\circ}$	
	$\ddot{\alpha}_{135^\circ} - \ddot{\alpha}_{255^\circ} - \ddot{\alpha}_{270^\circ} - \ddot{\beta}_{0^\circ}$	$\dot{\alpha}_{315^\circ} - \dot{\beta}_{45^\circ} - \dot{\beta}_{135^\circ} - \dot{\beta}_{180^\circ}$	Toutes les variables
	$\ddot{\beta}_{45^\circ} - \ddot{\alpha}_{90^\circ} - \ddot{\beta}_{135^\circ} - \ddot{\beta}_{225^\circ}$	$\dot{\beta}_{225^\circ} - \dot{\beta}_{270^\circ} - \dot{\beta}_{315^\circ} - \dot{\theta}_{45^\circ}$	sauf : $\ddot{\alpha}_{315^\circ} - \ddot{\beta}_{315^\circ} - \ddot{\theta}_{315^\circ}$
	$\ddot{\beta}_{270^\circ} - \ddot{\theta}_{0^\circ} - \ddot{\theta}_{45^\circ} - \ddot{\theta}_{135^\circ}$	$\dot{\theta}_{135^\circ} - \dot{\theta}_{180^\circ} - \dot{\theta}_{270^\circ} - \dot{\theta}_{315^\circ}$	
	$\ddot{\alpha}_{225^\circ} - \ddot{\theta}_{270^\circ} - \text{PUISS-KOPS}$	$\ddot{\alpha}_{90^\circ} - \ddot{\alpha}_{135^\circ} - \ddot{\beta}_{0^\circ} - \ddot{\beta}_{45^\circ}$	
		$\ddot{\beta}_{90^\circ} - \ddot{\beta}_{225^\circ} - \ddot{\theta}_{0^\circ} - \ddot{\theta}_{90^\circ}$	
		$\ddot{\theta}_{135^\circ} - \ddot{\theta}_{225^\circ} - \text{PUISS}$	
		CAD - MASS	
Variables significatives (**) p-value < 0.001	$\beta_{270^\circ} - \theta_{0^\circ} - \theta_{135^\circ} - \dot{\alpha}_{45^\circ}$	/	/
Variables significatives (*) p-value < 0.01	$\dot{\alpha}_{315^\circ} - \dot{\theta}_{135^\circ} - \ddot{\theta}_{180^\circ}$	/	/
Variables non significatives	MASS		

Tableau 4.5 : Sélection de prédicteurs, proportion d'impulsions positives
« best subset selection »

Méthode	Aucune sélection	Best subset selection
Critère de pénalisation	N/A	BIC
# Prédicteurs sélectionnées	76	11
RSE	2.85	3.23
R ² Ajusté	0.83	0.81
Variables les plus significatives (***)	PUISS	$\alpha_{90^\circ} - \ddot{\alpha}_{0^\circ} - \ddot{\alpha}_{135^\circ} - \beta_{90^\circ} - \ddot{\beta}_{135^\circ} - \dot{\theta}_{180^\circ} - \ddot{\theta}_{0^\circ} - \ddot{\theta}_{180^\circ} - \ddot{\theta}_{225^\circ} - \text{MASS} - \text{PUISS}$
p-value = 0		
Variables significatives (**)	$\alpha_{225^\circ} - \alpha_{270^\circ} - \beta_{225^\circ} - \beta_{270^\circ} - \theta_{225^\circ} - \ddot{\alpha}_{135^\circ}$	/
p-value < 0.001		
Variables significatives (*)	$\theta_{135^\circ} - \theta_{270^\circ} - \dot{\alpha}_{135^\circ} - \dot{\beta}_{45^\circ} - \dot{\beta}_{90^\circ}$	/
p-value < 0.01		
Variables moins significatives (.)	$\alpha_{315^\circ} - \dot{\alpha}_{90^\circ} - \dot{\beta}_{135^\circ} - \dot{\theta}_{135^\circ}$	/
p-value < 0.05	KOPS	
Variables non significatives	$\alpha_{0^\circ} - \alpha_{45^\circ} - \alpha_{90^\circ} - \alpha_{135^\circ} - \alpha_{180^\circ} - \beta_{0^\circ}$ $\beta_{45^\circ} - \beta_{90^\circ} - \beta_{135^\circ} - \beta_{180^\circ} - \beta_{315^\circ} - \theta_{0^\circ}$ $\theta_{45^\circ} - \theta_{90^\circ} - \theta_{180^\circ} - \theta_{315^\circ} - \dot{\alpha}_{0^\circ} - \dot{\alpha}_{45^\circ}$ $\dot{\alpha}_{180^\circ} - \dot{\alpha}_{225^\circ} - \dot{\alpha}_{270^\circ} - \dot{\alpha}_{315^\circ} - \dot{\beta}_{0^\circ} - \dot{\beta}_{180^\circ}$ $\dot{\beta}_{225^\circ} - \dot{\beta}_{270^\circ} - \dot{\beta}_{315^\circ} - \dot{\theta}_{0^\circ} - \dot{\theta}_{45^\circ} - \dot{\theta}_{90^\circ} - \dot{\theta}_{180^\circ}$ $\dot{\theta}_{225^\circ} - \dot{\theta}_{270^\circ} - \dot{\theta}_{315^\circ} - \ddot{\alpha}_{0^\circ} - \ddot{\alpha}_{90^\circ} - \ddot{\alpha}_{180^\circ}$ $\ddot{\alpha}_{225^\circ} - \ddot{\alpha}_{270^\circ} - \ddot{\alpha}_{315^\circ} - \ddot{\beta}_{0^\circ} - \ddot{\beta}_{45^\circ} - \ddot{\beta}_{90^\circ}$ $\ddot{\beta}_{180^\circ} - \ddot{\beta}_{225^\circ} - \ddot{\beta}_{270^\circ} - \ddot{\beta}_{315^\circ} - \ddot{\theta}_{0^\circ} - \ddot{\theta}_{90^\circ}$ $\ddot{\theta}_{180^\circ} - \ddot{\theta}_{225^\circ} - \ddot{\theta}_{270^\circ} - \ddot{\theta}_{315^\circ} - \text{CAD} - \text{MASS}$	/

Tableau 4.6 : Sélection de prédicteurs, proportion d'impulsions positives « forward selection », « backward selection »

Méthode	Forward selection	Backward selection
Critère de pénalisation	BIC	BIC
# Prédicteurs sélectionnées	17	16
RSE	3.29	3.16
R² Ajusté	0.78	0.8
Variables les plus significatives (***) p-value = 0	$\alpha_{45^\circ} - \beta_{315^\circ} - \theta_{0^\circ} - \theta_{270^\circ} - \dot{\alpha}_{45^\circ}$ $\dot{\alpha}_{90^\circ} - \dot{\beta}_{45^\circ} - \text{PUISS} - \text{MASS}$	$\alpha_{135^\circ} - \alpha_{225^\circ} - \alpha_{270^\circ} - \beta_{45^\circ} - \beta_{135^\circ}$ $\beta_{225^\circ} - \beta_{270^\circ} - \theta_{135^\circ} - \theta_{225^\circ} - \dot{\alpha}_{270^\circ}$ $\dot{\beta}_{45^\circ} - \dot{\beta}_{270^\circ} - \dot{\beta}_{315^\circ} - \dot{\beta}_{270^\circ} - \text{PUISS}$
Variables significatives (**) p-value < 0.001	$\alpha_{270^\circ} - \theta_{90^\circ} - \text{CAD}$	$\dot{\theta}_{90^\circ}$
Variables significatives (*) p-value < 0.01	$\dot{\alpha}_{225^\circ} - \dot{\beta}_{180^\circ}$	/
Variables moins significatives (.) p-value < 0.05	/	/
Variables non significatives	$\alpha_{90^\circ} - \theta_{45^\circ} - \dot{\alpha}_{0^\circ}$	/

Tableau 4.7 : Sélection de prédicteurs, proportion d'impulsions positives « hybrid selection », LASSO, Rigide

Méthode	Hybrid selection	LASSO	Rigide
Critère de pénalisation	BIC	N/A	N/A
# Prédicteurs sélectionnées	49	38	73
RSE	2.6	3.24	3.87
R ² Ajusté	0.86	0.79	0.74
<i>Variables les plus significatives (***) p-value = 0</i>			
$\alpha_{0^\circ} - \alpha_{45^\circ} - \alpha_{90^\circ} - \alpha_{225^\circ}$ $\alpha_{45^\circ} - \alpha_{225^\circ} - \alpha_{270^\circ} - \beta_{225^\circ}$ $\beta_{270^\circ} - \theta_{45^\circ} - \theta_{135^\circ} - \theta_{225^\circ}$ $\dot{\alpha}_{45^\circ} - \dot{\alpha}_{45^\circ} - \dot{\alpha}_{90^\circ} - \dot{\alpha}_{135^\circ}$ $\dot{\beta}_{45^\circ} - \dot{\beta}_{90^\circ} - \dot{\beta}_{135^\circ} - \dot{\beta}_{180^\circ}$ $\dot{\beta}_{270^\circ} - \dot{\theta}_{0^\circ} - \dot{\theta}_{45^\circ} - \dot{\theta}_{90^\circ}$ $\dot{\theta}_{135^\circ} - \ddot{\alpha}_{135^\circ} - \ddot{\alpha}_{255^\circ} - \ddot{\beta}_{90^\circ}$ $\ddot{\beta}_{135^\circ} - \ddot{\beta}_{180^\circ} - \ddot{\theta}_{45^\circ} - \ddot{\theta}_{135^\circ}$			
$\beta_{90^\circ} - \beta_{180^\circ} - \beta_{270^\circ} - \beta_{315^\circ}$ $\theta_{0^\circ} - \theta_{45^\circ} - \theta_{90^\circ} - \theta_{135^\circ}$ $\theta_{180^\circ} - \theta_{225^\circ} - \theta_{270^\circ} - \theta_{315^\circ}$ $\dot{\alpha}_0 - \dot{\alpha}_{45^\circ} - \dot{\alpha}_{90^\circ} - \dot{\alpha}_{135^\circ}$ $\dot{\alpha}_{180^\circ} - \dot{\alpha}_{225^\circ} - \dot{\alpha}_{270^\circ} - \dot{\beta}_0$ $\dot{\beta}_{45^\circ} - \dot{\beta}_{135^\circ} - \dot{\beta}_{180^\circ} - \dot{\beta}_{225^\circ}$ $\dot{\beta}_{315^\circ} - \dot{\theta}_0 - \dot{\theta}_{45^\circ} - \dot{\theta}_{90^\circ}$ $\dot{\theta}_{225^\circ} - \dot{\theta}_{270^\circ} - \dot{\theta}_{315^\circ}$ -CAD			
PUISS PUISS-KOPS-MASS			
<i>Variables significatives (**) p-value < 0.001</i>			
$\alpha_{315^\circ} - \theta_{0^\circ} - \theta_{270^\circ} - \dot{\alpha}_{180^\circ}$ $\ddot{\alpha}_{45^\circ} - \ddot{\alpha}_{270^\circ} - \ddot{\beta}_{270^\circ} - \ddot{\theta}_{225^\circ}$			
<i>Variables significatives (*) p-value < 0.01</i>			
$\dot{\alpha}_0 - \ddot{\alpha}_0 - \ddot{\beta}_{45^\circ}$ - KOPS			
<i>Variables moins significatives (.) p-value < 0.05</i>			
$\theta_{180^\circ} - \theta_{315^\circ} - \dot{\beta}_0$ - MASS			
<i>Variables non significatives</i>			
$\dot{\alpha}_{315^\circ} - \dot{\beta}_{225^\circ} - \dot{\theta}_{270^\circ} - \ddot{\beta}_0$			

Comme le montrent les résultats des différentes sélections de prédicteurs, c'est la méthode « Best subset selection » qui permet d'obtenir le plus petit nombre de prédicteurs

pour les deux variables dépendantes. Cela en fait une méthode privilégiée pour construire les régressions linéaires multiples puisque le principe de parcimonie est respecté. Cette sélection de prédicteurs est en accord avec [115], qui mentionne qu'il s'agit de l'une des méthodes les plus performantes en matière de sélection de prédicteurs. Cette méthode permet également de trouver les prédicteurs ayant la valeur de signification la plus élevée ($p < 0.05$), un aspect qui sera démontré plus tard. En outre, les valeurs de performance, évalué au moyen de R^2 sont bonnes et la valeur d'erreur est faible. Il est important de rappeler ici qu'une valeur R^2 est considérée comme mauvaise en dessous de 0.3, bonne entre 0.4 et 0.7, et très bonne entre 0.7 et 0.99 [124, 135]. En poursuivant les analyses, cette méthode est suivie par la méthode « Backward selection », qui obtient plus de prédicteurs mais avec une performance légèrement supérieur et une erreur légèrement inférieur pour les deux variables dépendantes. Ensuite apparaît la méthode « Forward selection », avec une sélection de prédicteurs légèrement plus élevée et des valeurs de performance légèrement inférieures pour les deux variables dépendantes. Enfin, les méthodes « Hybrid selection » , régression LASSO et régression « Ridge » donnent plus de prédicteurs et de bonnes valeurs de performance, pour les deux variables dépendantes. Cependant, la méthode LASSO sera analysée afin de vérifier si elle peut être utilisée pour construire un modèle de régression, même si la valeur des prédicteurs est élevée.

4.6.2 Résultats des régressions linéaires multiples

Sur la base des résultats obtenus avec le processus de sélection des prédicteurs, il est recherché un modèle de régression linéaire multiple qui réponde à la variable dépendante et qui respecte le principe de parcimonie. En outre, il est indispensable que les quatre hypothèses de la régression linéaire multiple soient respectées (linéarité, absence de multi

colinéarité, homoscédasticité et normalité). Dans l'étude 1, pour l'entraînement des modèles, l'analyse et la comparaison des modèles, a été utilisé 90 % des données d'entraînement et 10 % des données de test. Par conséquent, les deux modèles composés de 11 prédicteurs obtenus avec la méthode de sélection du meilleur sous-ensemble « best subset selection », sont sélectionnés pour chaque variable dépendante évaluée. Les coefficients des prédicteurs des régressions linéaires multiples sont présentés dans le tableau 4.8.

Tableau 4.8 : Coefficients régressions linéaires multiples 11 prédicteurs. A. Indice d'efficacité, B. Proportion d'impulsion positive

A			B		
Nom du prédicteur	p-value	Coefficient estimé	Nom du prédicteur	p-value	Coefficient estimé
Intercept	2.32e-07	21.66	Intercept	< 2e-16	97.21
$\dot{\alpha}_{270^\circ}$	8.42e-16	70.15	α_{90°	< 2e-16	-29.59
$\dot{\beta}_{270^\circ}$	1.18e-13	-53.93	$\ddot{\beta}_{135^\circ}$	3.39e-05	-28.6
θ_{270°	< 2e-16	41.51	$\ddot{\alpha}_0^\circ$	3.52e-10	-28.51
θ_{45°	< 2e-16	-34.24	$\dot{\theta}_{180^\circ}$	< 2e-16	27.43
α_{45°	< 2e-16	-23.74	$\ddot{\alpha}_{135^\circ}$	1.78e-09	23.11
$\dot{\theta}_{270^\circ}$	9.31e-10	19.92	$\ddot{\theta}_0^\circ$	2.45e-15	-21.13
Power	< 2e-16	19.09	Power	< 2e-16	18.89
$\dot{\beta}_{180^\circ}$	< 2e-16	-16.31	β_{90°	< 2e-16	18.04
$\dot{\alpha}_{90^\circ}$	0.000326	-10.41	$\ddot{\theta}_{180^\circ}$	7.99e-13	-17.38
β_{225°	2.21e-08	8.93	$\ddot{\theta}_{225^\circ}$	1.76e-10	12.59
Mass	1.45e-07	-5.67	Mass	2.8e-08	-5.92

Comme le montre le tableau 4.8, ces modèles se caractérisent également par le fait que tous les prédicteurs sont significatifs $p\text{-value} < 0.05$, c'est-à-dire qu'ils sont fortement liés à la variable dépendante. De même, l'ajustement des modèles par le calcul du R^2 ajusté et de l'erreur estimée RSE se situe à de bonnes valeurs. Comme le RSE mesure la distance

typique entre les points de données observés et la ligne de régression, et qu'il est exprimé dans les mêmes unités que la variable dépendante, on constate que le RSE est inférieur ou proche de 3 ou moins, ce que je considère comme très bon.

La figure 4.8 montre les graphiques de régression et Bland-Altman du modèle composé de 11 prédicteurs pour l'indice d'efficacité et la proportion d'impulsion positive.

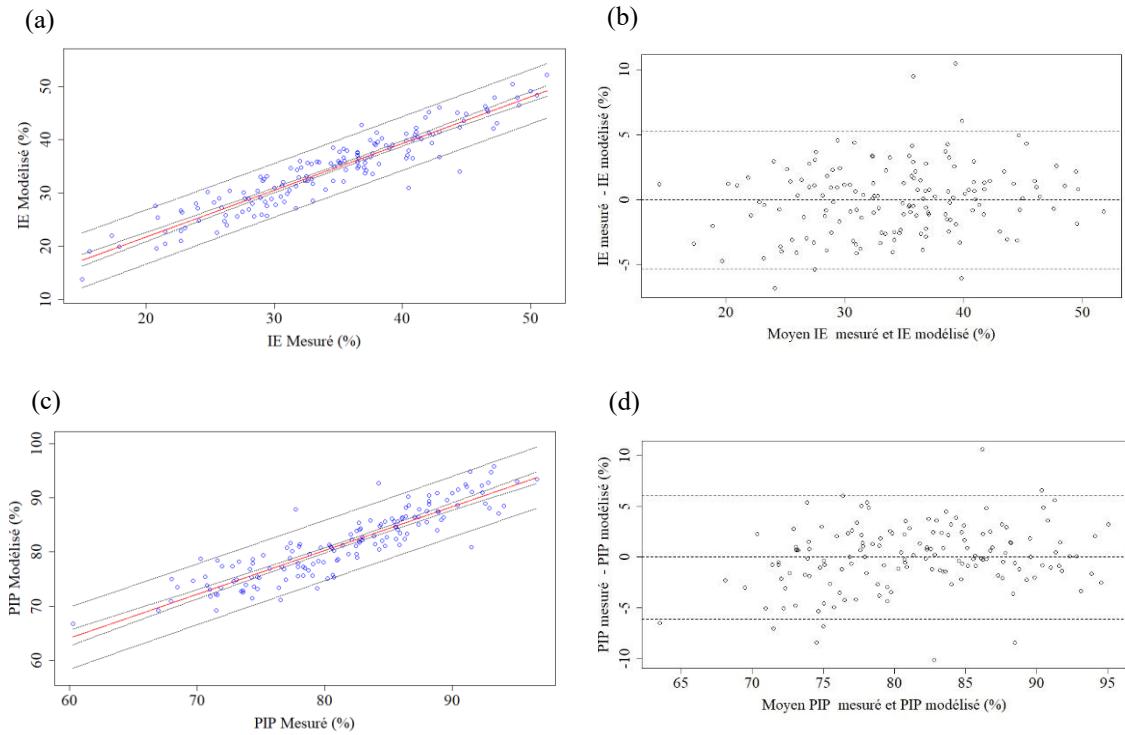


Figure 4.8 : Graphiques de régression, (a) IE, (c) PIP et Bland-Altman (b) IE, (d) PIP régressions linéaires multiples 11 prédicteurs

En ce qui concerne la régression LASSO et la régression rigide présentées dans les tableaux 4.3 et 4.6, comme l'algorithme sélectionne les prédicteurs de manière différente, cela permet d'obtenir un autre type de régression qu'il peut être utile d'analyser. Il est donc proposé de sélectionner les deux régressions LASSO qui éliminent les variables les moins significatives. Il est également possible de constater que l'ajustement des modèles par le calcul du R^2 ajusté et RSE se situent dans de bonnes valeurs, similaires à celles d'une

régression avec 11 prédicteurs. Les graphiques de régression et Bland-Altman sont présentés dans la figure 4.9.

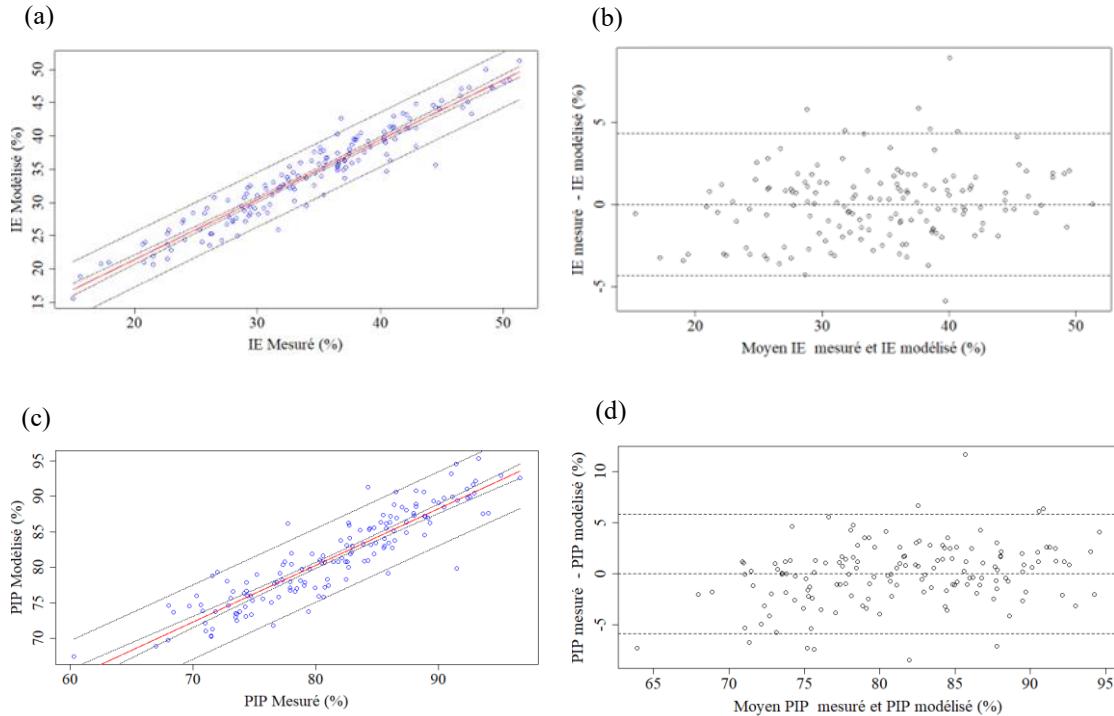


Figure 4.9 : Graphiques de régression, (a) IE, (c) PIP et Bland-Altman, (b) IE, (d) PIP, régression LASSO

Les modèles régression de composantes principales PCR et la régression partielle du dernier carré PLSR, sont présentés dans le tableau 4.9 ci-dessous. Ce type de modèle ne donne pas les variables prédictives originales mais définit les nouveaux prédicteurs à partir des variables indépendantes originales.

Tableau 4.9 : Résultats méthodes régression de composantes PCR principales et régression partielle du dernier carré PLSR

Méthode	PCR		PLSR	
Variable dependent	IE	PIP	IE	PIP
# Composants	34	35	16	15
RSE	3.34	3.98	3.06	4
R ² Ajusté	0.9	0.84	0.92	0.84

Sur la base des résultats du tableau 4.9 et de l'analyse de la faisabilité de la mise en œuvre des modèles basés sur le nombre réduit de prédicteurs, il est proposé d'explorer la méthode de régression par moindres carrés partiels PLSR. Il est également constaté qu'il s'agit de modèles plus précis, avec une erreur de prédiction plus faible. Les graphiques de de régression et Bland-Altman ce type de modèle sont présentés à la figure 4.10.

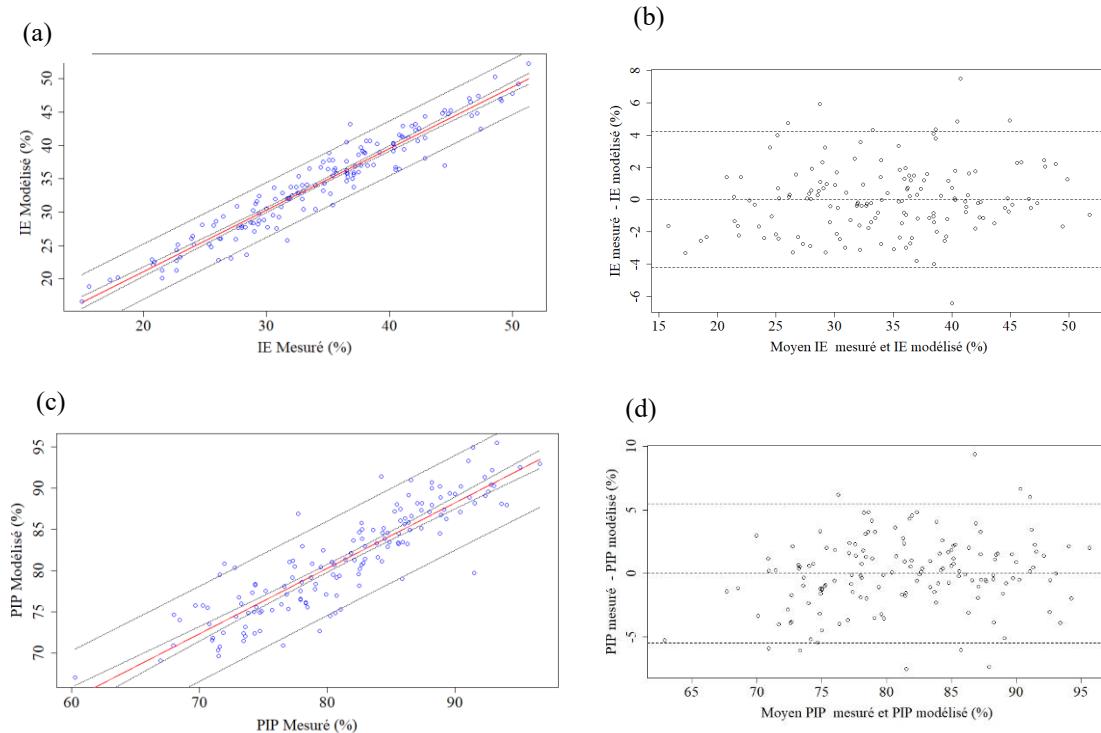


Figure 4.10 : Graphiques de régression, (a) IE, (c) PIP et Bland-Altman, (b) IE, (d) PIP, régression par moindres carrés partiels PLSR

Pour les réseaux neuronaux développés dans l'étude 1, la méthode élimination récursive des caractéristiques RFE a été utilisée. Le tableau 4.10 présente les prédicteurs sélectionnés pour chaque variable dépendante.

Tableau 4.10 : Sélection de prédicteurs, méthode élimination récursive des caractéristiques pour prédire l'indice d'efficacité et la proportion d'impulsions positives

Méthode	RFE	
Variable dependent	IE	PIP
# Prédicteurs sélectionnées	21	28
	$\alpha_{0^\circ} - \alpha_{45^\circ} - \alpha_{135^\circ}$	
	$\alpha_{0^\circ} - \alpha_{225^\circ} - \alpha_{315^\circ}$	$\alpha_{180^\circ} - \alpha_{225^\circ} - \alpha_{270^\circ}$
	$\dot{\alpha}_{180^\circ} - \ddot{\alpha}_{135^\circ} - \dot{\beta}_{180^\circ}$	$\alpha_{315^\circ} - \dot{\alpha}_{180^\circ} - \ddot{\alpha}_{135^\circ}$
	$\theta_{270^\circ} - \dot{\theta}_{0^\circ} - \dot{\theta}_{90^\circ}$	$\beta_{0^\circ} - \beta_{45^\circ} - \dot{\beta}_{0^\circ}$
Prédicteurs sélectionnées	$\dot{\theta}_{180^\circ} - \dot{\theta}_{225^\circ} - \dot{\theta}_{315^\circ}$	$\dot{\beta}_{180^\circ} - \theta_{180^\circ} - \theta_{225^\circ}$
	$\ddot{\theta}_{90^\circ} - \ddot{\theta}_{135^\circ} - \ddot{\theta}_{180^\circ}$	$\theta_{270^\circ} - \theta_{315^\circ} - \dot{\theta}_{0^\circ}$
	$\ddot{\theta}_{225^\circ} - \ddot{\theta}_{270^\circ} - \ddot{\theta}_{315^\circ}$	$\dot{\theta}_{45^\circ} - \dot{\theta}_{180^\circ} - \dot{\theta}_{225^\circ}$
	KOPS - MASS - PUISS	$\ddot{\theta}_{90^\circ} - \ddot{\theta}_{180^\circ} - \ddot{\theta}_{225^\circ}$
		$\ddot{\theta}_{270^\circ} - \text{KOPS}$
		MASS-PUISS

4.6.3 Résultats des arbres de régression

Le tableau 4.11 présente les caractéristiques des arbres de régression optimisé.

Tableau 4.11 : Caractéristiques des modèles basés sur des arbres de régression

Variable dépendant	IE	PIP
RSE	8.78	6.08
R ²	0.57	0.69
# Prédicteurs sélectionnées	9	9

La figure 4.11 montre les caractéristiques de l'arbre de régression optimisé pour prédire l'indice d'efficacité. Dans les graphiques suivants qui représentent les arbres, chaque rectangle en couleur représente le nœud de l'arbre où figure la valeur moyenne de la variable dépendante évaluée, ainsi que le pourcentage des données qui se concentrent sur cette valeur.

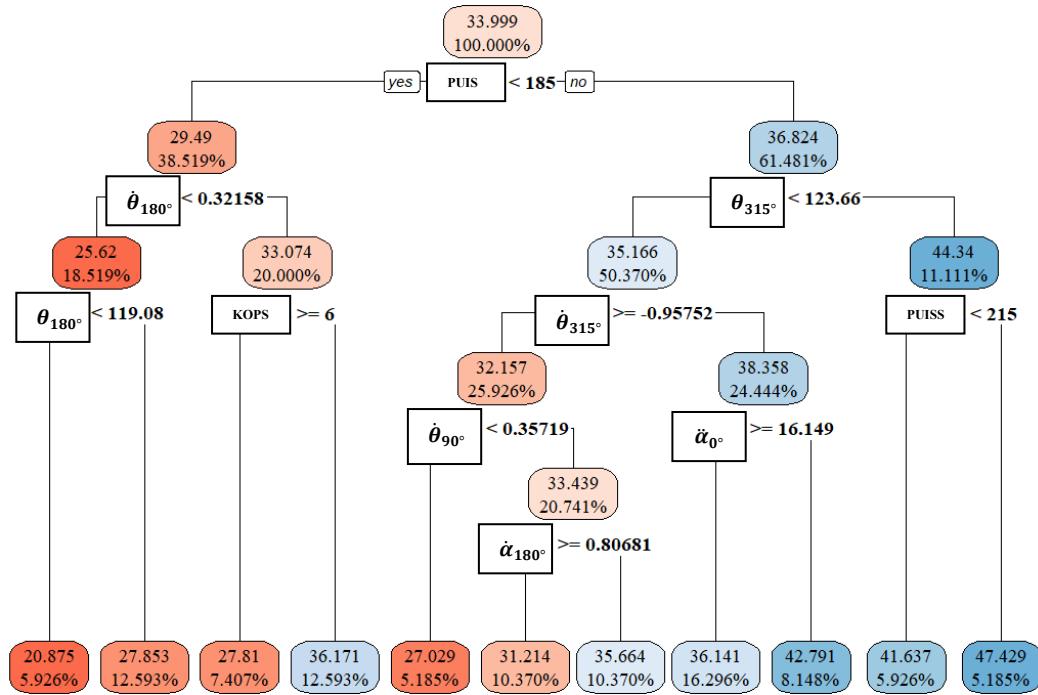


Figure 4.11 : Arbre de régression optimisé pour prédire l'indice d'efficacité

La figure 4.12 montre l'ordre d'importance des prédicteurs pour prédire l'indice d'efficacité. Un algorithme est utilisé pour déterminer l'ordre d'importance des prédicteurs

à l'aide d'une valeur UA ou d'Unités Arbitraires. Plus cette valeur est élevée, plus le prédicteur est important pour le modèle.

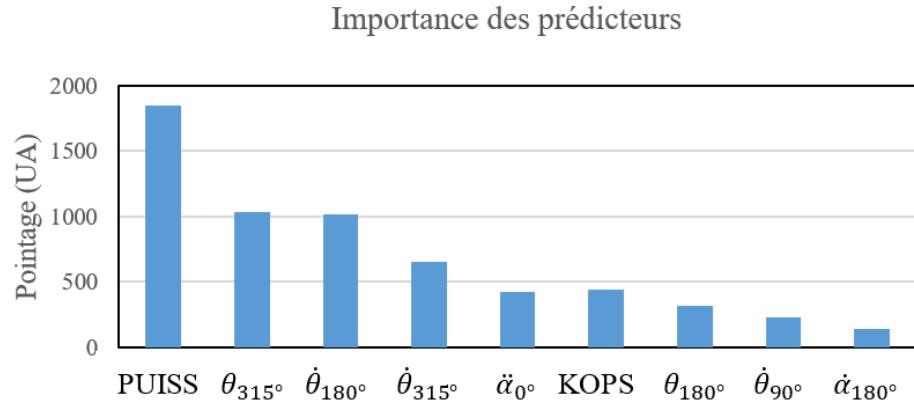


Figure 4.12 : Importance des prédicteurs arbre de régression indice d'efficacité

La figure 4.13 montre les caractéristiques de l'arbre de régression optimisé pour prédire la proportion d'impulsions positives.

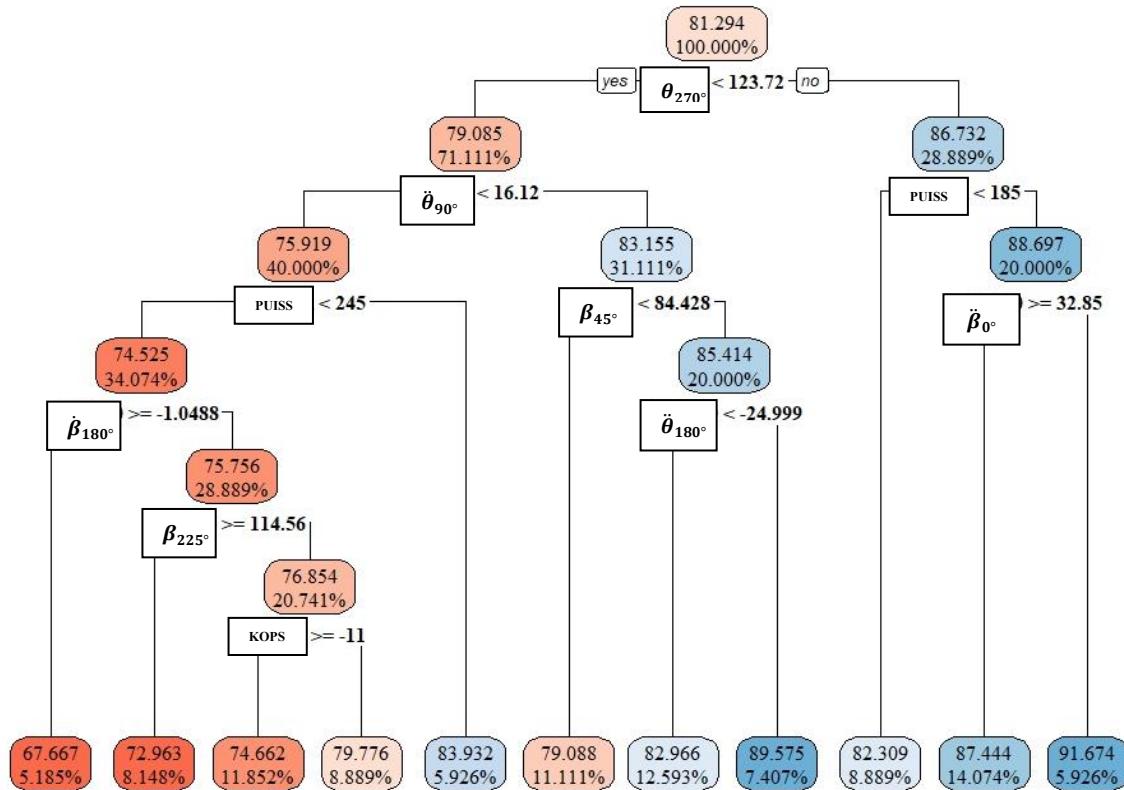


Figure 4.13 : Arbre de régression pour prédire la proportion d'impulsions positives

La figure 4.14 montre l'ordre d'importance des prédicteurs pour prédire la proportion d'impulsions positives.

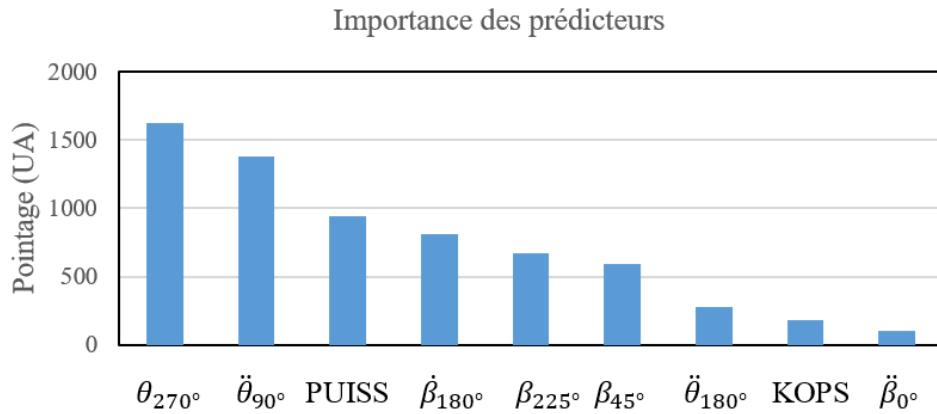


Figure 4.14: Importance des prédicteurs arbre de régression proportion d'impulsions positives

4.6.4 Résultats des modèles de forêts aléatoires

En partant d'un modèle de forêt aléatoire avec tous les prédicteurs disponibles, une procédure d'optimisation par validation croisée est appliquée pour chaque variable dépendante. Cette procédure sélectionne le nombre optimal d'arbres et le nombre optimal de prédicteurs, ce qui donne le modèle final. Les caractéristiques des forêts aléatoires optimisées sont présentées dans le tableau 4.12.

Tableau 4.12 : Caractéristiques des forêts aléatoires optimisés

Variable dépendant	IE	PIP
RSE	1.75	1.6
MAE	1.36	1.2
R ²	0.95	0.96
# Arbres	500	400
# Prédicteurs sélectionnées	37	37

La figure 4.15 ci-dessous présente les graphiques de régression et Bland-Altman du modèle de forêt aléatoire.

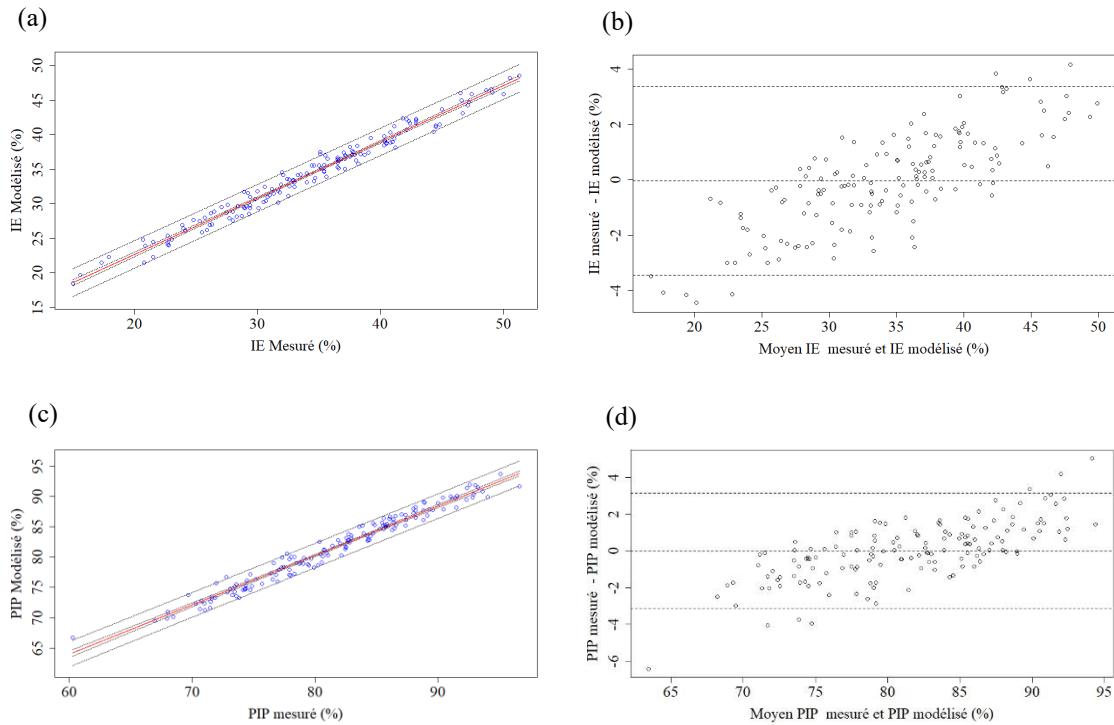


Figure 4.15 : Graphiques de régression, (a) IE, (c) PIP et Bland-Altman, (b) IE, (d) PIP forêt aléatoire

Le modèle de forêt aléatoire pour prédire l'indice d'efficacité utilise 37 prédicteurs. Les 10 prédicteurs les plus importants sont présentés dans la figure 4.16 ci-dessous. Pour déterminer l'ordre d'importance des prédicteurs, un algorithme est utilisé pour déterminer la valeur « Increase in Node Purity » (IncNodePurity) de chaque prédicteur [136]. Comme la méthode de la forêt aléatoire utilise plusieurs arbres, la valeur « IncNodePurity » est utilisée pour mesurer la qualité de la division de chaque nœud des arbres. Dans les arbres, chaque division représente un prédicteur. Le « IncNodePurity » est basé sur l'indice d'impureté de Gini [137], qui indique si le prédicteur apporte un bénéfice en termes

d'information. Plus cette valeur est élevée, plus le prédicteur évalué est important pour le modèle, c'est-à-dire que les nœuds sont beaucoup plus « purs ».

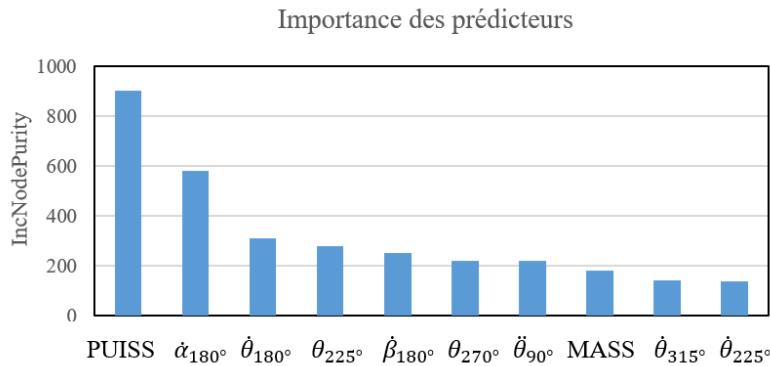


Figure 4.16 : Importance des prédicteurs modèle forêts aléatoires pour prédire l'indice d'efficacité

Le modèle de forêt aléatoire pour prédire la proportion d'impulsions positives utilise 37 prédicteurs. Les 10 prédicteurs les plus importants sont présentés dans la figure 4.17 ci-dessous.

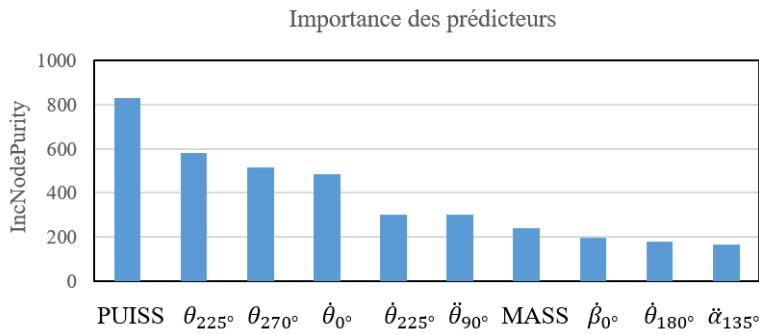


Figure 4.17 : Importance des prédicteurs modèle forêts aléatoires pour prédire la proportion d'impulsions positives

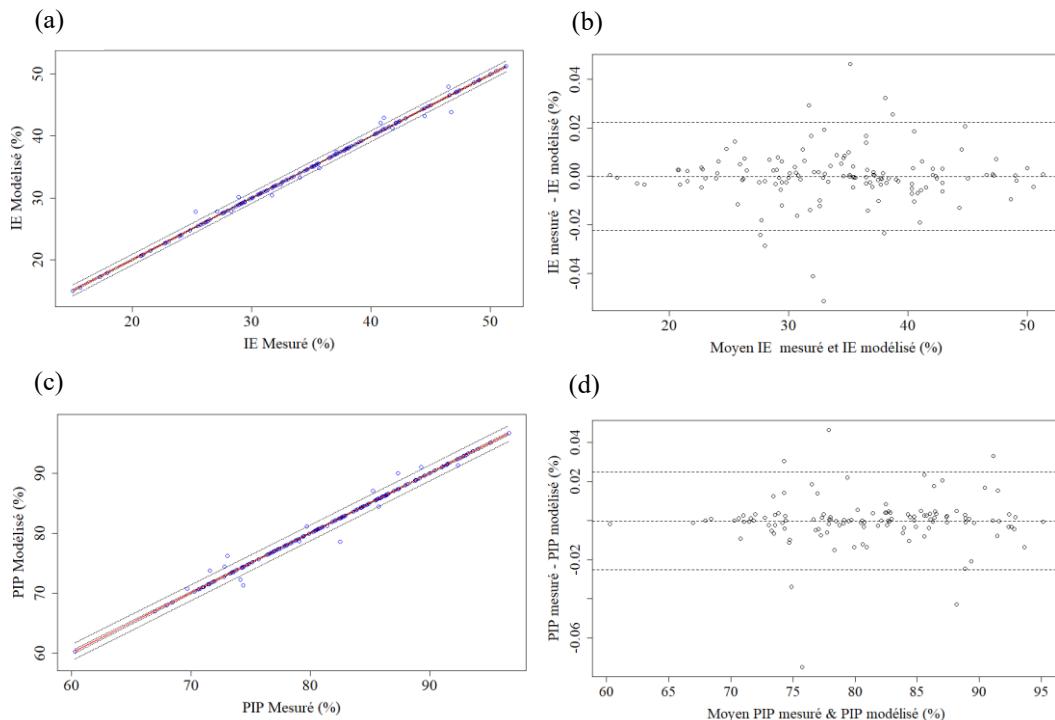
4.6.5 Résultats réseaux neuronaux

Deux réseaux neuronaux artificielles ont été développés pour chaque variable dépendante, indice d'efficacité et proportion d'impulsions positives. Les caractéristiques de ces réseaux neuronaux sont présentées dans le tableau 4.13.

Tableau 4.13 : Caractéristiques des réseaux neuronaux étude 1

Variable dépendante	Nombre de prédicteurs	Nombre de couches cachées	Nombre de neurones	R ²	MAE	RSE
Indice d'efficacité	21	1	35	0.96	1.16	0.03
Proportion d'impulsions positives	28	1	35	0.91	1.82	0.09

La figure 4.18 ci-dessous présente les graphiques de régression et Bland-Altman des réseaux neuronaux artificiels développés.

**Figure 4.18 : Graphiques de régression, (a) IE, (c) PIP et Bland-Altman, (b) IE, (d) PIP, des réseaux neuronaux**

Au moyen d'un algorithme appelé Garson [138], il est possible de déterminer l'ordre d'importance des prédicteurs d'un réseau neuronal à couche cachée. La figure 4.19, montre le résultat de cette analyse pour la variable dépendante indice d'efficacité. Seuls 10

prédicteurs sont présentés dans cette figure, car les prédicteurs restants se situent dans la marge de 0.02 à 0.03 de l'échelle de Garson,

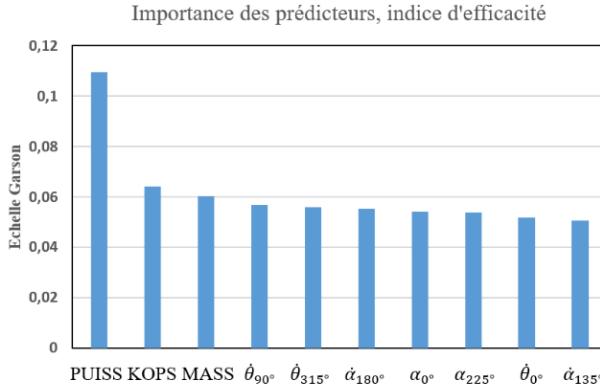


Figure 4.19 : Importance des prédicteurs algorithme Garson, réseau de neurones indice d'efficacité

La figure 4.20, montre le résultat de l'analyse de Garson pour la variable dépendante proportion d'impulsions positives. Seuls 10 prédicteurs sont présentés dans cette figure, car les prédicteurs restants se situent dans la marge de 0.02 à 0.03 de l'échelle de Garson.

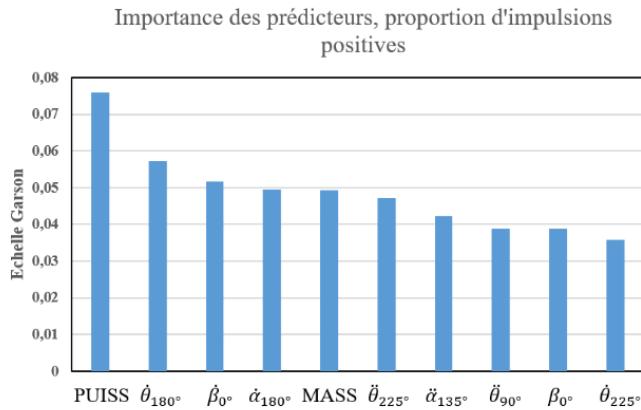


Figure 4.20 : Importance des prédicteurs algorithme Garson, réseau de neurones proportion d'impulsions positives

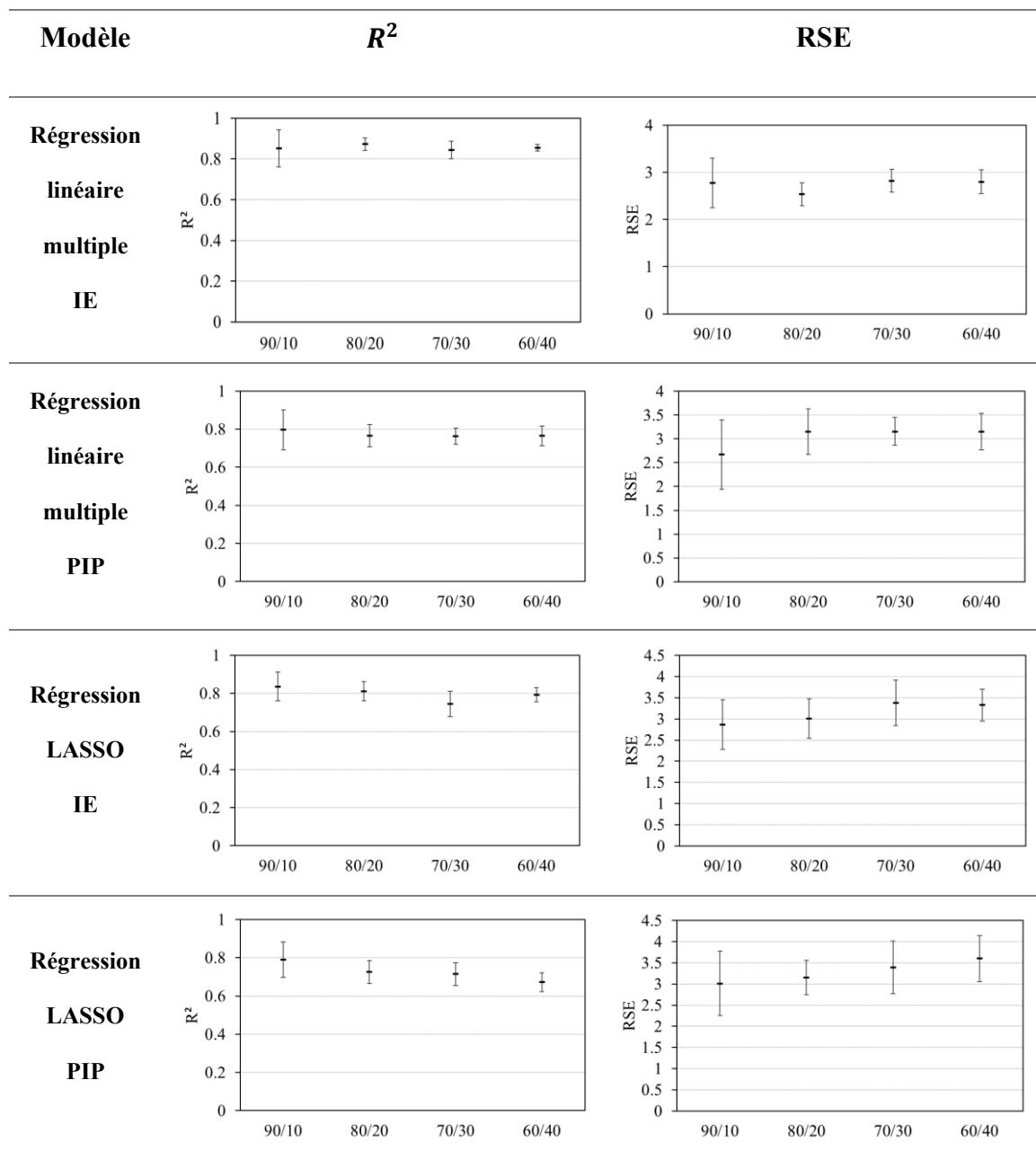
4.6.6 Résultats validation croisée

Le tableau 4.14, présente les graphiques de la validation croisée des modèles de régression linéaire multiple et régression LASSO dans l'étude 1. À cette fin, la méthode

des 10-fold a été mise en œuvre, en divisant les données dans les proportions 90/10, 80/20, 70/30 et 60/40, dont le numérateur indique les données d'apprentissage et le dénominateur les données de test.

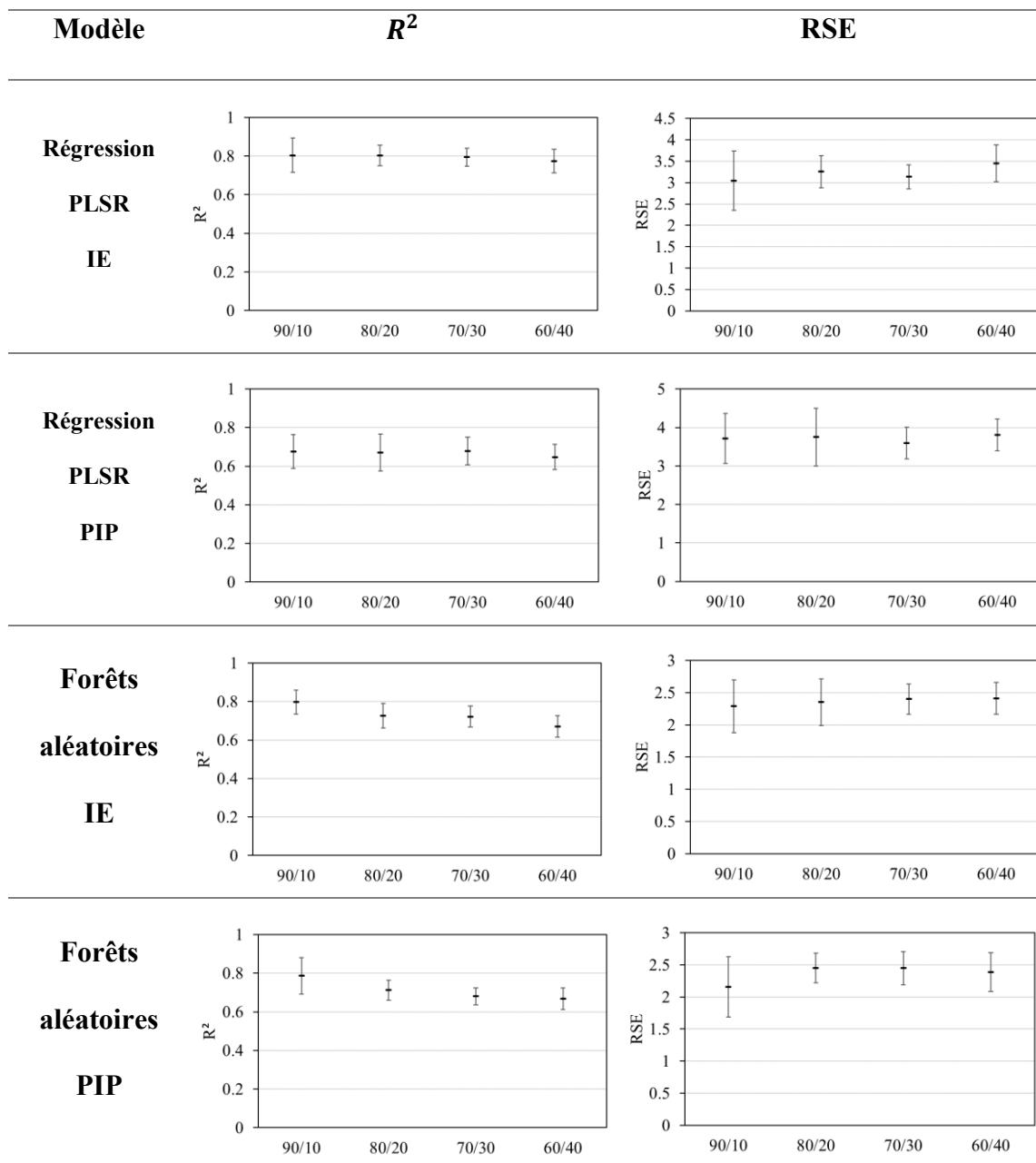
Tableau 4.14 : Validation croisée étude 1, régression linéaire multiple, régression

LASSO



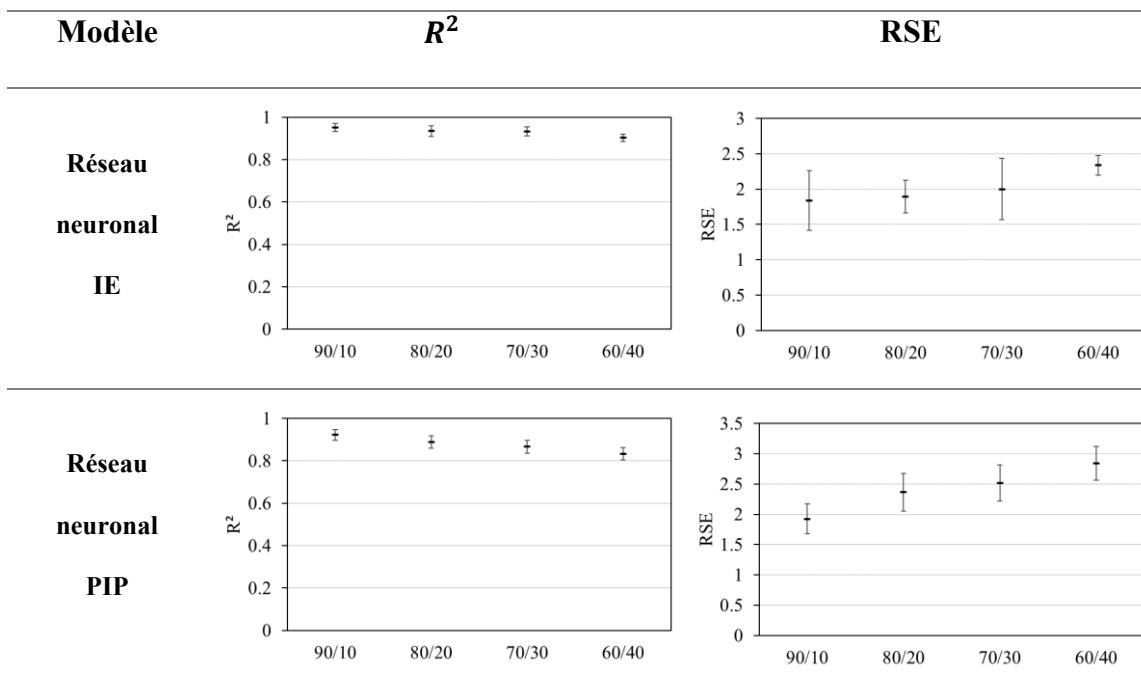
Le tableau 4.15, présente les graphiques de la validation croisée des modèles de régression PLSR et forêts aléatoires.

Tableau 4.15 : Validation croisée l'étude 1, régression PLSR et forêts aléatoires



Le tableau 4.16, présente les graphiques de validation croisée pour les réseaux de neurones implémentés dans l'étude.

Tableau 4.16 : Validation croisée étude 1, réseau neuronal



4.7 Analyse des résultats

4.7.1 Analyse des méthodes de sélection des prédicteurs

Pour l'élaboration de la régression linéaire multiple, plusieurs méthodes de sélection des prédicteurs ont été appliquées afin de trouver le meilleur ensemble possible de prédicteurs. Lors de l'analyse des ensembles de prédicteurs trouvés par les différents algorithmes, l'objectif est de sélectionner l'ensemble comportant un nombre minimal de prédicteurs afin de respecter le principe de parcimonie. Par conséquent, la meilleure

méthode de sous-ensemble « best subset selection » est celle qui fournit un minimum de prédicteurs pour les deux variables dépendantes IE et PIP, c'est-à-dire 11 prédicteurs, avec l'avantage qu'ils ont tous une valeur significative élevée (valeur $p < 0,05$), ce qui garantit qu'ils sont fortement liés aux variables dépendantes. Cette méthode permet d'obtenir de bonnes mesures de la qualité des régressions, $R^2 = 0.86$ et $RSE = 2.81$ pour l'IE et $R^2 = 0.81$ et $RSE = 3.81$ pour le PIP. Cependant, les autres méthodes sélectionnent également les prédicteurs de manière efficace, mais l'interprétation des résultats est difficile en raison du nombre considérable de variables. Il est également important d'analyser les résultats obtenus avec les arbres de régression qui sélectionnent 9 prédicteurs pour les deux variables dépendantes. Cependant, malgré leur faible précision et erreur élevée, $R^2 = 0.57$ et $RSE = 8.78$ pour l'IE et $R^2 = 0.69$ et $RSE = 6.08$ pour le PIP, les arbres de régression ont l'avantage de dévoiler certains prédicteurs susceptibles d'avoir une relation importante avec les deux variables dépendantes. Dans ce contexte, il est observé que la puissance a une relation remarquable avec l'IE et également avec le PIP. Il est aussi observé que la cinématique de la cheville (5 prédicteurs) joue un rôle très important dans la prédiction de l'IE et, dans le cas de le PIP, la cinématique du genou (4 prédicteurs) est significativement importante. En suivant les méthodes décrites ci-dessus, les algorithmes de sélection descendante « backward selection » donnent un bas nombre de prédicteurs, 21 pour l'IE et 16 pour le PIP. Avec cette méthode, il est possible d'obtenir des prévisions avec une bonne marge, toutefois, elles ne sont pas exceptionnelles en termes de précision et d'erreur et par conséquent et ne seront donc pas pris en compte. En analysant la sélection des prédicteurs effectuée par chaque algorithme, les prédicteurs trouvés dans la plupart des méthodes sont la masse du sujet et la puissance délivrée. Ceci serait en accord avec les

études précédentes montrant son influence sur l'IE [47] et le PIP [5]. Les deux autres modèles de régression qui respectent le principe de parcimonie sont ceux obtenus par régression des moindres carrés partiels « partial least squares régression (PLSR) », parce qu'ils sont constitués un nombre réduit de prédicteurs, 16 pour IE et 15 pour PIP. Aussi, ces modèles ont des valeurs de qualité de régression acceptables pour les deux variables dépendantes, $R^2 = 0.92$ et $RSE = 3.06$ pour IE, $R = 0.84$ et $RSE = 4$ pour PIP. Cependant, comme cette méthode génère des composantes dérivées des prédicteurs originaux, le principe d'interprétabilité n'est pas applicable. Une autre analyse a consisté à évaluer la multicollinéarité des régressions linéaires multiples obtenues avec chaque algorithme. Cette analyse a été réalisée au moyen de la matrice de corrélation et du calcul de l'indice d'inflation de la variance (VIF), qui doit être inférieur à 10. La présence de multi colinéarités a été constatée dans tous les sous-ensembles de prédicteurs, à l'exception des sous-ensembles obtenus avec 11 prédicteurs à l'aide de la méthode du meilleur sous-ensemble. Il est conclu que ce sous-ensemble de 11 variables est le nombre idéal de prédicteurs pour les régressions linéaires multiples pour les deux variables dépendantes. Il est également possible de vérifier que les sous-ensembles comportant un grand nombre de variables prédictives n'atteignent pas une valeur élevée de signification $p\text{-value} < 0.05$, telle que celle obtenue avec l'ensemble de 11 variables dans lequel toutes les valeurs de signification sont exceptionnelles $p\text{-value} < 0.0001$. Cela indique une forte relation entre tous les prédicteurs et la variable dépendante.

Conclusions : la méthode du meilleur sous-ensemble « best subset selection », permet de trouver un modèle de régression linéaire multiple qui est caractérisé par un minimum de prédicteurs (11), une bonne qualité d'ajustement et une faible erreur de régression. En

outre, ce modèle se caractérise par le fait que tous les prédicteurs ont une forte relation avec les variables dépendantes ($p\text{-value} < 0.05$). Les prédicteurs qui sont récurrents dans la plupart des méthodes de sélection sont la masse et la puissance délivrée, ce qui indique leur impact sur la prédiction.

4.7.2 Évaluation des hypothèses de régressions linéaires multiples

Sur la base de l'analyse ci-dessus, des régressions linéaires multiples composées de 11 prédicteurs correspondants ont été sélectionnées pour prédire chaque variable dépendante analysée. Ces régressions doivent répondre aux 4 hypothèses suivantes : linéarité, normalité, homoscédasticité et non-colinéarité. L'hypothèse de linéarité a été évaluée en traçant les données prédites par rapport aux résidus, ce qui a permis de trouver une ligne de tendance horizontale sur l'axe 0 des résidus. L'hypothèse de normalité a été évaluée à l'aide du graphique Q-Q, qui présente une ligne allant de la partie inférieure gauche à la partie supérieure droite, ce qui indique que l'hypothèse de normalité est respectée. L'hypothèse d'homoscédasticité a été évaluée à l'aide du diagramme échelle-emplACEMENT, qui a montré que les résidus sont distribués approximativement selon une ligne horizontale. La figure 4.21 présente les graphiques de ces analyses. Enfin, l'hypothèse de non-colinéarité a été évaluée à l'aide de la matrice de corrélation et le calcul de l'indice d'inflation de la variance VIF.

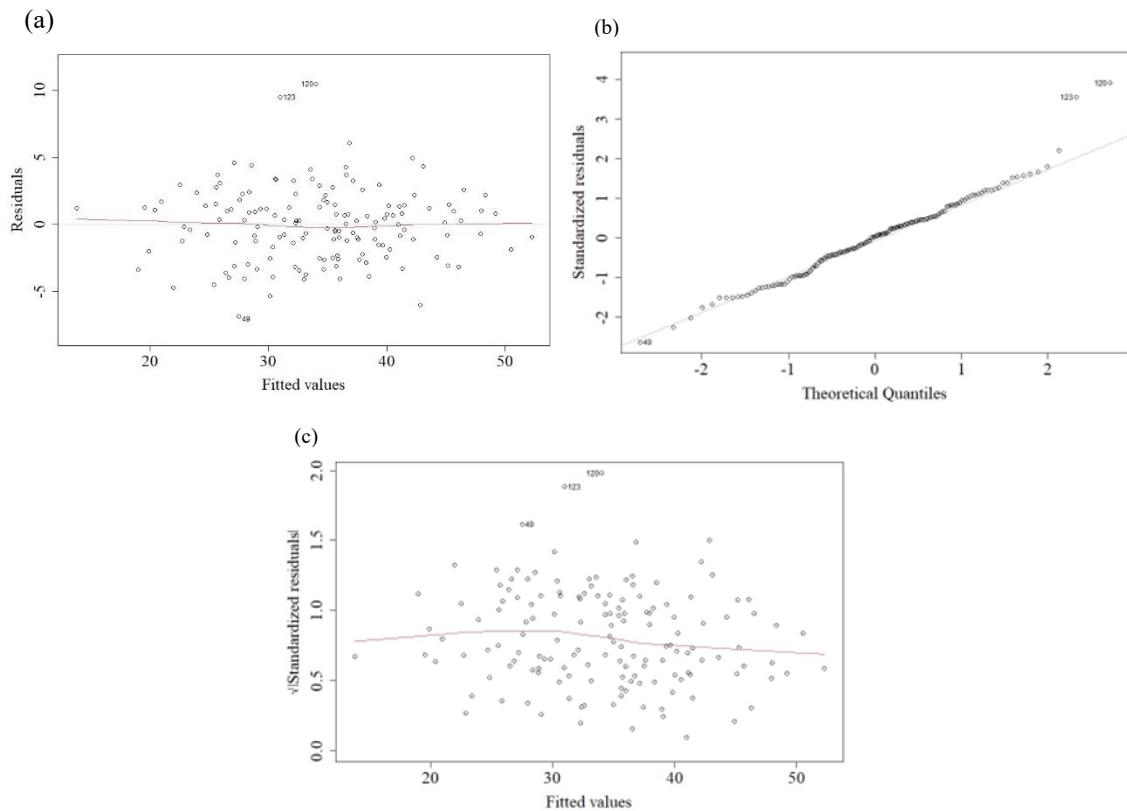


Figure 4.21 : Graphiques des évaluations des hypothèses modèle de régression linéaire multiple pour prédire l'IE. (a) linéarité, (b) normalité, (c) homoscédasticité

La figure 4.22 présente la matrice de corrélation et le calcul du VIF pour le modèle de régression linéaire multiple qui permet de prédire l'IE.

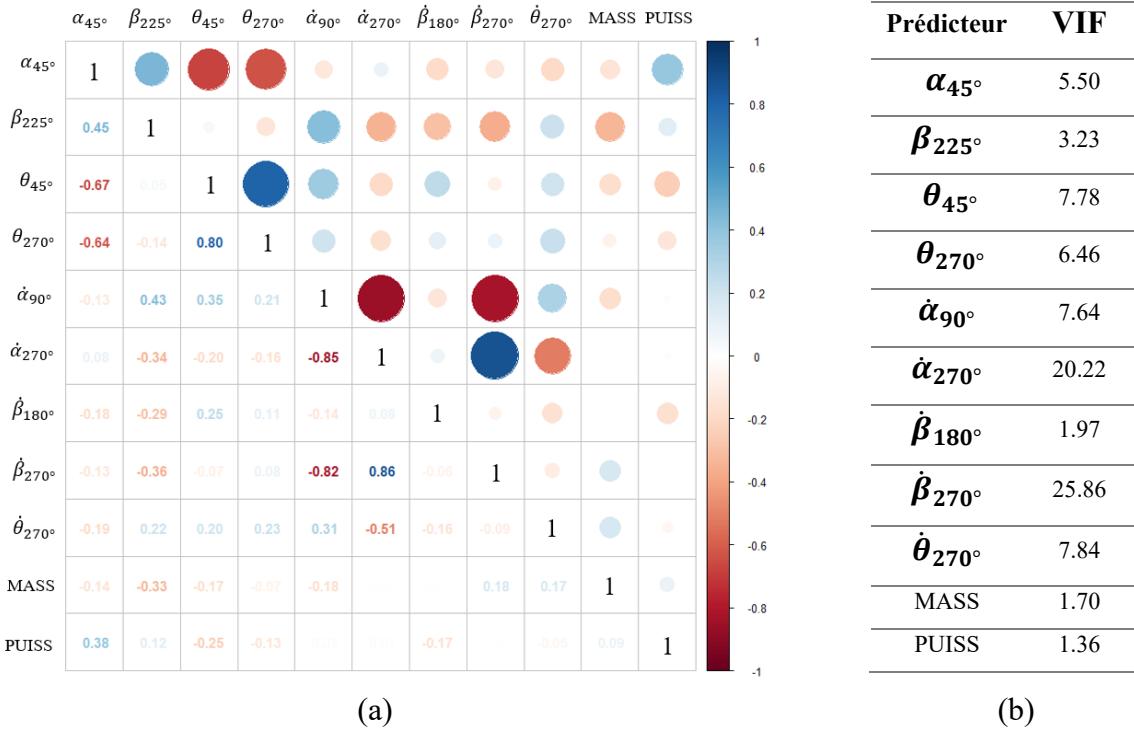


Figure 4.22 : Matrice de corrélation (a) et calcul du VIF (b) modèle de régression linéaire multiple pour la prédiction de l'IE

L'analyse des graphiques présentés dans la figure 4.21 conclut que la régression linéaire qui prédit l'IE respecte les hypothèses de linéarité, de normalité et d'homoscédasticité. En ce qui concerne la non-colinéarité, analysée dans la figure 4.22, seule une valeur élevée du coefficient de corrélation $r = 0.86$ est observée entre les prédicteurs $\dot{\alpha}_{270^\circ}$ et $\dot{\beta}_{270^\circ}$ en analysant la matrice de corrélation. Également un VIF élevé de 20.22 et 25.86 est également observé pour ces deux prédicteurs respectivement. La première solution consisterait à éliminer l'une de ces deux variables. Après avoir analysé cette possibilité, il a été constaté qu'un modèle à 10 prédicteurs perdait en précision et performance. Il a été décidé de laisser ces deux variables car le VIF n'est pas très élevé. D'un point de vue biomécanique, il est supposé que les deux articulations travaillent

ensemble car elles sont reliées à des muscles bi-articulaires connectés par les mêmes muscles et leur action se fait conjointement. En outre, une corrélation entre $\dot{\alpha}_{90^\circ}$ et $\dot{\beta}_{270^\circ}$ de $r = 0.85$ et entre $\dot{\beta}_{270^\circ}$ et $\dot{\alpha}_{90^\circ}$ de $r = 0.82$ a été trouvée, mais le VIF du prédicteur $\dot{\alpha}_{90^\circ}$ est bas ($VIF = 7.64$), et par conséquent il a été décidé de conserver ces variables.

De la même manière, l'analyse est effectuée pour la régression linéaire multiple composé de 11 prédicteurs qui prédit le PIP. La figure 4.23 présente les graphiques de ces analyses.

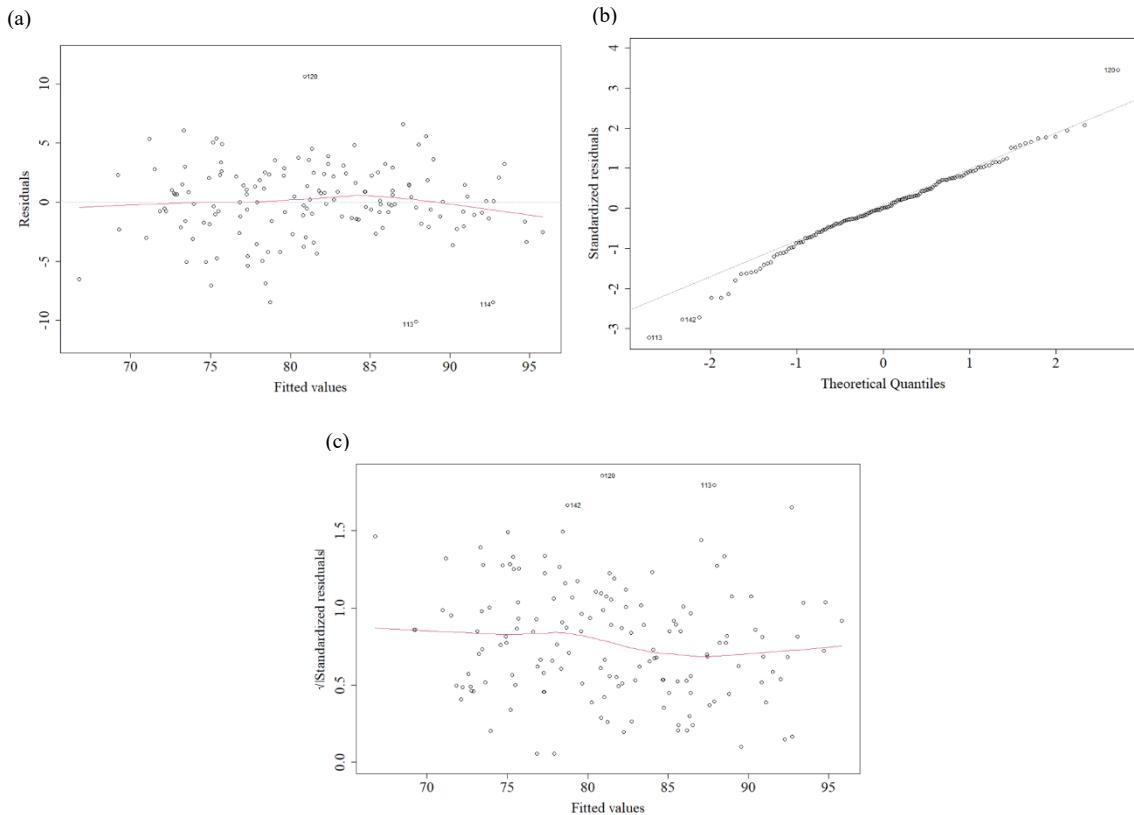


Figure 4.23 : Évaluations des hypothèses modèle de régression linéaire multiple pour prédire le PIP. (a) linéarité, (b) normalité, (c) homoscédasticité

La figure 4.24 présente la matrice de corrélation et le calcul du VIF pour le modèle de régression linéaire multiple qui permet de prédire le PIP.

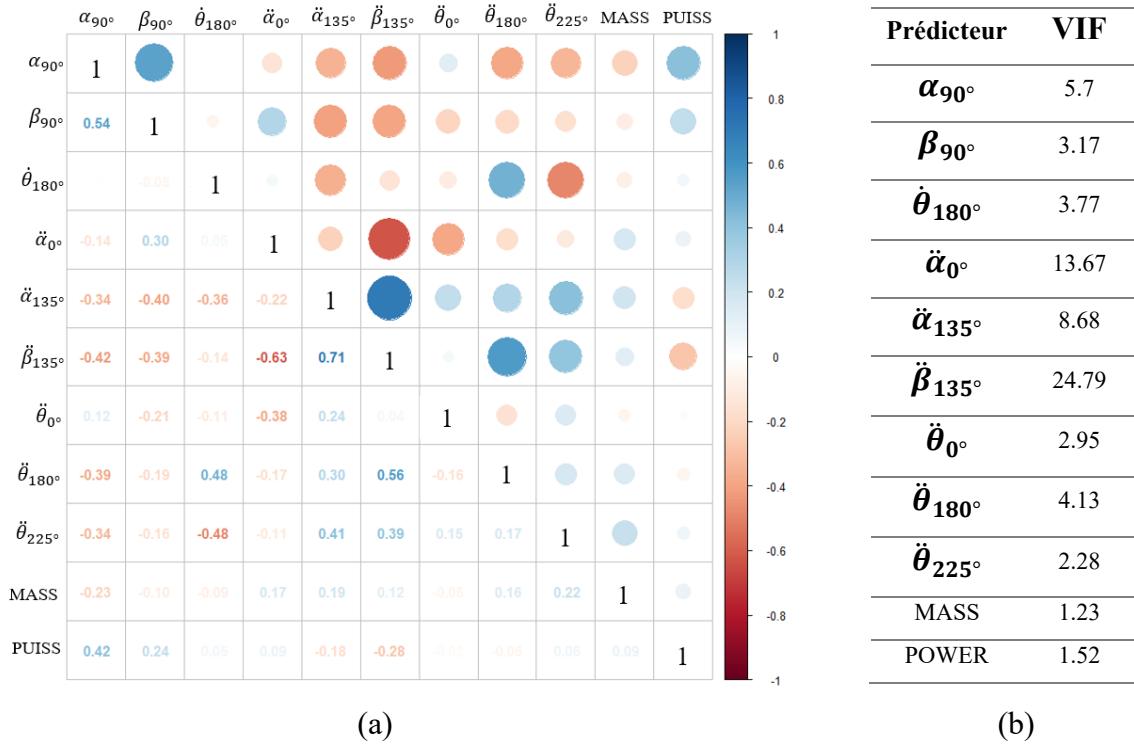


Figure 4.24 : Matrice de corrélation (a) et calcul du VIF (b) modèle de régression linéaire multiple pour la prédiction de PIP

L'analyse des graphiques présentés dans la figure 4.23 conclut que la régression linéaire qui prédit le PIP respecte les hypothèses de linéarité, de normalité et d'homoscédasticité. En ce qui concerne l'analyse de la non-colinéarité, analysée dans la figure 4.24, un phénomène similaire a été constaté avec la régression linéaire qui prédit l'IE. Cependant, la matrice de corrélation montre une valeur $r = 0.71$ entre les prédicteurs $\ddot{\alpha}_0^\circ$ et $\ddot{\beta}_{135^\circ}$. Cependant, il se situe dans une marge faible et le VIF de chaque variable est également à une faible valeur, et par conséquent, il a été décidé de conserver les deux prédicteurs.

Conclusion : Les deux régressions linéaires multiples composées de 11 prédicteurs satisfont aux quatre hypothèses de ce type de modèle, à savoir la linéarité, normalité, homoscédasticité et non-colinéarité.

4.7.3 Analyse de validation croisée des régressions linéaires multiples

Un autre aspect à évaluer dans une régression linéaire multiple est la validation croisée, qui vise à vérifier si le modèle est généralisable. À cette fin, différentes fractions des données totales ont été évaluées afin de vérifier si le R^2 et le RSE sont stables et si l'écart-type de chaque mesure et de chaque fraction n'est pas élevé. L'analyse du tableau 4.14 montre que les deux modèles de régression linéaire multiple qui prédisent l'IE et le PIP, conservent le $\bar{R^2}_{IE} = 0.84 \pm 0.05$ et $\bar{RSE}_{IE} = 2.28 \pm 0.29$ pour le IE et $\bar{R^2}_{PIP} = 0.77 \pm 0.06$ et $\bar{RSE}_{PIP} = 3.03 \pm 0.46$ pour le PIP, dans une marge moyenne optimale en analysant toutes les fractions de données. Il est également évident que les deux modèles peuvent prédire avec une bonne précision avec un nombre réduit de données (60 % entraînement/40 % test), ce qui montre leur capacité de généralisation.

Conclusion : Les deux modèles de régression linéaire multiple qui prédisent l'IE et le PIP sont généralisables.

4.7.4 Analyse des prédicteurs sélectionnés pour l'IE et le PIP

Les deux modèles de régression linéaire multiple qui prédisent l'IE et le PIP ont l'avantage de connaître la relation mathématique entre tous les prédicteurs ainsi que le

signe positif ou négatif qui accompagne chaque prédicteur. De plus, comme le poids de chacun des prédicteurs est connu, c'est-à-dire le coefficient qui accompagne chaque prédicteur, le modèle est explicable ou interprétable. En conséquence, les principaux facteurs influençant la production de force en cyclisme sont connus.

En examinant les caractéristiques du modèle de régression linéaire multiple pour prédire l'indice d'efficacité (tableau 4.8, A), il est constaté que le prédicteur vitesse angulaire du fémur à 270° ($\dot{\alpha}_{270^\circ}$) a le coefficient (70.15) qui présente la relation mathématique la plus forte avec la variable dépendante, suivi de la vitesse angulaire du genou à 270° ($\dot{\beta}_{270^\circ}$, -53.93). Cela peut s'expliquer par le recrutement musculaire des fléchisseurs du genou et de la hanche pendant la course ascendante afin d'optimiser l'action de traction du pédalage [5, 139]. Ce fait corrobore également l'impact sur l'indice d'efficacité de la phase ascendante, qui est largement influencé par le groupe musculaire des fléchisseurs de la hanche [140]. La zone autour de 270° est un moment critique pour limiter la force négative qui ralentit la propulsion [141], ce qui fait que cette phase du cycle est cruciale pour l'optimisation de l'indice d'efficacité. Dans la même position, il a été trouvé comme prédicteur la vitesse angulaire de la cheville ($\dot{\theta}_{270}$), qui est accompagnée d'un coefficient de relation moyenne (19.92) avec la variable dépendante, ce qui pourrait indiquer la contribution de la cheville dans l'action de traction puisqu'il s'agit de la valeur la plus élevée de la puissance négative [40]. La puissance apparaît également comme un prédicteur dont le coefficient a une relation médiane avec la variable dépendante (19.09), ce qui serait en accord avec les études précédentes montrant leur influence sur l'indice d'efficacité [62]. Enfin, la masse de l'individu est le prédicteur ayant le moins de relation (-5.67). Néanmoins, leur importance dans le modèle peut être liée au

fait que si les cyclistes plus lourds ont une masse plus importante au niveau des membres inférieurs, une augmentation de la masse du cycliste peut augmenter la force négative pendant le mouvement ascendant, ce qui affecte l'indice d'efficacité [5].

En relation à la régression linéaire multiple pour prédire la proportion d'impulsion positive, il est bien connu que cette métrique est liée à la transmission du couple et de la puissance aux manivelles [45]. Lors de l'analyse de l'importance du prédicteur du modèle (tableau 4.8, B), le prédicteur angle du fémur à 90° (α_{90°) a le coefficient (-29,59) qui présente la relation mathématique la plus forte avec la variable dépendante. Cela pourrait s'expliquer par le fait que c'est à ce moment-là que la valeur maximale du couple et de la puissance de la hanche est atteinte [40], qui pourraient influencer la contribution des impulsions positives dans les manivelles. Par conséquent, l'angle du fémur serait une valeur critique pour atteindre ces valeurs élevées, d'où son importance. Après ce prédicteur, les coefficients des prédicteurs accélération angulaire du genou à 135° ($\ddot{\beta}_{135^\circ}$) et accélération angulaire du fémur à 0° ($\ddot{\alpha}_{0^\circ}$) apparaissent dans un niveau de relation proche 28.6 et 28.5 respectivement, mais à ma connaissance, il y a peu d'études qui mettent en relation l'accélération angulaire des membres inférieurs et la production de couple. Je formule l'hypothèse que l'importance de ces deux prédicteurs réside dans le fait que le pic négatif maximal de l'accélération angulaire du fémur genou et de l'accélération angulaire trouvée par [65]. Ce qui précède pourrait également être lié à la génération de moments articulaires totaux positifs du genou et du fémur respectivement analysés par ces auteurs et également constaté dans l'analyse de la dynamique inverse présentée dans le point 2.3.2. En outre, plusieurs prédicteurs liés à la cinématique de l'articulation de la cheville ont été trouvés dont les coefficients ont une relation moyenne avec la proportion d'impulsions

positives. Cela pourrait être lié au rôle important de la cheville dans la transmission du couple positif, en particulier dans la phase de récupération [40], qui influence positivement la proportion d'impulsion positive. Ceci est également lié au fait que les cyclistes d'élite ayant une bonne technique de pédalage présentent dans cette phase moins de couple négatif et moins de puissance négative. Il est intéressant de noter que deux prédicteurs sont dans le même ordre d'influence par rapport à la variable dépendante et également avec un poids similaire. Ces deux prédicteurs sont la puissance et la masse de l'individu, ce qui renforce l'argument précédemment avancé selon lequel ces deux prédicteurs influencent la technique de pédalage.

Conclusions : Les prédicteurs qui influencent le plus l'indice d'efficacité sont liés à la cinématique articulaire du membre inférieur dans la zone de 270°. En revanche, les prédicteurs qui influencent le plus la proportion d'impulsion positive sont distribués dans le premier (0° - 90°), le deuxième quadrant (90° - 180°) et combinent principalement diverses valeurs de la cinématique de la hanche et du genou. Ces deux modèles partagent la puissance et la masse à une valeur de poids mathématique similaire.

4.7.4 Analyse des modèles de régression

Dans cette section, sont analysés les modèles de régression de type linéaire multiple, LASSO et PLSR, en mettant l'accent sur la précision et le faible taux d'erreur, indépendamment de la possibilité de les interpréter. Les graphiques de régression, figures 4.8 (a, c), 4.9 (a, c) et 4.10 (a, c), montrent que les données se situent dans les intervalles de prédiction et de confiance et qu'il y a très peu de valeurs aberrantes. Les tracés sont très

similaires, mais une plage plus large est identifiée dans les intervalles de prédiction PIP et dans le modèle de type PLSR. En observant les intervalles de prédiction il est évident que les modèles sont capables de prédire des valeurs dans une bonne fourchette et donc que les trois modèles de régression sont valides. L'analyse des diagrammes de Bland-Altman, figures 4.8 (b, d), 4.9 (b, d) et 4.10 (b, d) des deux variables dépendantes montre qu'il n'y a pas de biais de prédiction, c'est-à-dire que la ligne centrale est à 0. Les trois modèles présentent des différentiels positifs et négatifs dans une fourchette très similaire autour de ± 5 unités, mais légèrement meilleure pour les deux modèles PLSR.

L'analyse des coefficients de détermination R^2 ajustés et du RSE montre des valeurs légèrement meilleures pour les deux variables dépendantes selon le modèle PLSR, $R^2_{Aj,IE} = 0.92$, $RSE_{IE} = 3.06$ et $R^2_{Aj,PIP} = 0.84$, $RSE_{PIP} = 4$. Ensuite, les modèles de régression LASSO, $R^2_{Aj,IE} = 0.88$, $RSE_{IE} = 2.57$ et $R^2_{Aj,PIP} = 0.79$, $RSE_{PIP} = 3.24$. Postérieurement, les modèles de régression linéaire multiple $R^2_{Aj,IE} = 0.86$, $RSE_{IE} = 2.81$ et $R^2_{Aj,PIP} = 0.81$, $RSE_{PIP} = 3.23$. Toutefois, les différences sont marginales et non concluantes. L'analyse de tous ces modèles, la variable dépendante IE est la variable avec laquelle de meilleures valeurs sont atteintes. Cependant, lorsque l'on analyse la valeur moyenne de la validation croisée de ces modèles, le résultat est meilleur pour les modèles de régression linéaire multiple concernant les deux variables dépendantes $\overline{R^2}_{IE} = 0.87$, $\overline{RSE}_{IE} = 2.41$ et $\overline{R^2}_{PIP} = 0.77$, $\overline{RSE}_{PIP} = 3.03$. Ensuite, le modèle LASSO $\overline{R^2}_{IE} = 0.79$, $\overline{RSE}_{IE} = 3.14$ et $\overline{R^2}_{PIP} = 0.72$, $\overline{RSE}_{PIP} = 3.28$ et finalement le modèle PLSR $\overline{R^2}_{IE} = 0.79$, $\overline{RSE}_{IE} = 3.22$ et $\overline{R^2}_{PIP} = 0.66$, $\overline{RSE}_{PIP} = 3.71$. Bien que les performances soient légèrement inférieures, cette diminution des performances de la régression LASSO peut être due au fait que ce type de modèle nécessite un plus grand nombre de données

[98]. En outre, en ce qui concerne le PLSR, certains auteurs soulignent que les méthodes traditionnelles de validation croisée, telles que « 10-fold cross validation », ne permettent pas toujours d'obtenir une bonne mesure de la validité du modèle [142]. En outre, sa faible performance peut être due au fait que ce type de régression est plus indiqué lorsqu'il y a plus de variables que de données [143].

Conclusions : La qualité du modèle en termes de précision, faible erreur et validation, est supérieure dans la régression linéaire multiple par rapport aux autres méthodes de régression comme la régression LASSO et la régression PLSR. Néanmoins, ces méthodes peuvent être améliorées avec l'augmentation du nombre de données et la sélection d'une méthode spécifique de validation du modèle. Par conséquent, leur application est possible en tenant compte des différences minimales de performance.

4.7.5 Analyse et comparaison de tous les modèles

Dans ce qui suit, il sera présenté une analyse de toutes les méthodes implémentées pour prédire les variables dépendantes, indice d'efficacité et la proportion d'implosions positives. L'analyse de tous les modèles développés dans l'étude 1 montre que le modèle de réseau neuronal est le meilleur modèle de tous pour les deux variables dépendantes, ces modèles se caractérisant par des mesures de performance $R^2_{IE} = 0.96$, $RSE_{IE} = 0.03$ et $R^2_{PIP} = 0.91$, $RSE_{PIP} = 0.09$. De même, lorsque l'on analyse les diagrammes de régression de la figure 4.18 (a,c), les intervalles de confiance et de prédition sont les plus proches de tous les modèles pour les deux variables dépendantes. De même, il y a peu de valeurs aberrantes dans les deux graphiques. L'analyse du diagramme de Bland-Altman de la figure 4.18 (b,d), illustre également des différences exceptionnelles entre les données

mesurées et modélisées. Ces marges sont proches de ± 0.02 par les deux variables dépendantes, qui est l'intervalle le plus bas de tous les modèles et sans biais entre les mesures. La procédure de validation croisée donne également des résultats exceptionnels pour les deux variables dépendantes. Les valeurs moyennes d'ajustement du modèle et de l'erreur étant inférieures à celles des autres modèles $\overline{R^2}_{IE} = 0.93$, $\overline{RSE}_{IE} = 2$ et $\overline{R^2}_{PIP} = 0.87$, $\overline{RSE}_{PIP} = 2.4$. Par conséquent, ce modèle est considéré comme généralisable, stable ou robuste. La robustesse se réfère à la capacité du modèle à maintenir stable la performance prédictive face aux variations et aux changements des données d'entrée, qui est corroboré par la validation croisée [144]. La limite de ce type de modèle réside dans sa faible interprétabilité. Cependant, l'algorithme développé par Garson [138], permet d'identifier l'importance relative des prédicteurs pour une variable dépendante. La figure 4.19 montre que le prédicteur qui influence principalement la variable dépendante IE est la puissance, suivie du KOPS et de la masse de l'individu. Lors de l'analyse des prédicteurs restants, il est important de souligner l'importance de la cinématique de la cheville, qu'il pouvait être justifié comme un bon transmetteur de forces. Bien que ces résultats soient différents de ceux d'un modèle de régression linéaire multiple, cette procédure renforce l'importance de la puissance et de la masse dans la prédiction de l'indice d'efficacité. Ces deux prédicteurs sont communs aux deux modèles. Il est remarquable que la puissance soit près du double des autres prédicteurs. La figure 4.20, montre que le prédicteur qui influence principalement la variable dépendante PIP est également la puissance, suivie de la cinématique de la cheville et du genou. Comme dans le cas précédent, le KOPS et la masse exercent une influence sur la prédiction, bien que dans un rôle moyen. Mais il est remarquable que le réseau neuronal utilise également ces

deux variables pour prédire la PIP. Ensuite, il y a une combinaison de divers prédicteurs cinématiques du genou, de la cheville et d'une partie de la hanche, qui est compliquée à analyser car elle combine différents phénomènes biomécaniques qui ont été peu étudiés.

Après ce modèle, le modèle de forêt aléatoire apparaît, avec lequel de bons résultats sont obtenus pour les deux variables dépendantes, $R^2_{IE} = 0.95$, $RSE_{IE} = 1.75$ et $R^2_{PIP} = 0.96$, $RSE_{PIP} = 1.6$. Bien que de meilleurs résultats soient obtenus avec le coefficient de corrélation R^2 par rapport aux réseaux neuronaux, l'erreur RSE augmente dans les deux modèles. Par conséquent, lorsque les graphiques de régression de la figure 4.15 (a,c) sont examinés, les intervalles de prédiction et de confiance sont légèrement plus larges. De même, dans le diagramme de Bland-Altman, figure 4.15 (b,d), les écarts entre les données sont plus importants, atteignant environ ± 3 . De même, il n'y a pas de biais et peu de données aberrantes. En ce qui concerne la validation croisée, les valeurs de performance R^2 , sont inférieures à celles des réseaux neuronaux même à celles des régressions linéaires multiples, $\overline{R^2}_{IE} = 0.72$, $\overline{RSE}_{IE} = 2,36$ et $\overline{R^2}_{PIP} = 0.71$, $\overline{RSE}_{PIP} = 2.36$. Cependant, des erreurs RSE mineures sont obtenues. Ce phénomène peut être dû au fait que cet algorithme nécessite une grande quantité de données et que, par conséquent, lorsque les données d'entraînement sont sélectionnées dans une proportion inférieure, les performances diminuent et pénalisent le R^2 . Enfin, cet algorithme permet également de trouver l'ordre d'importance des prédicteurs. Dans les graphiques 4.16 et 4.17, il apparaît qu'une fois de plus la puissance est le prédicteur le plus important pour les deux prédicteurs. La masse apparaît également, mais dans une moindre influence. Il est également observé que dans les deux modèles, la cinématique de la cheville est l'un des prédicteurs les plus importants.

Après ces deux modèles, qui donnent de très bons résultats en termes de R^2 et RSE, suivent dans l'ordre le modèle de régression linéaire multiple, puis la régression PLSR et enfin le modèle LASSO. Cependant, la capacité d'explication des modèles de réseaux neuronaux et des arbres aléatoires n'est pas comparable à celle de la régression linéaire multiple. Il est possible de distinguer deux groupes de modèles ayant deux objectifs différents, l'un se concentrant sur l'interprétabilité et l'autre sur la performance.

Enfin, le modèle d'arbre fournit des informations importantes concernant la sélection des prédicteurs et de leurs valeurs, car il n'est pas nécessaire de normaliser les données. Comme cela peut être constaté, ces modèles ne donnent pas de meilleures performances en termes de R^2 et de RSE. Cependant, ces valeurs sont un peu meilleures avec la variable dépendante PIP, $R^2_{IE} = 0.57$, $RSE_{IE} = 8.78$ et $R^2_{PIP} = 0.69$, $RSE_{PIP} = 6.08$. Il est visible dans la figure 4.11 que la puissance est le principal facteur explicatif de l'IE. C'est surtout une puissance supérieure à 215 W, qui permet d'obtenir les valeurs d'IE les plus élevées. Cela signifie que la puissance est directement liée à l'IE. Le prédicteur suivant est la cheville qui, lorsqu'elle atteint une valeur supérieure à 123.6° , c'est-à-dire en flexion planaire, donne les valeurs d'IE les plus élevées. Il est possible de conclure que la cheville joue un rôle important dans la phase de traction de la jambe, en particulier dans le quatrième quadrant ($270^\circ - 0^\circ$). En ce qui concerne la variable dépendant PIP en analysant la figure 4.13, le prédicteur angle de cheville à 270° avec une valeur de $123,7^\circ$ est le plus important. Cette situation est similaire à celle observée pour l'IE et l'importance de l'action de la cheville dans le quatrième quadrant. Ensuite, la puissance, avec des valeurs supérieures à 185 W, permet d'obtenir une PIP plus élevée. Le prédicteur KOPS est également commun aux deux modèles, ce qui corrobore son influence sur la technique de pédalage. Il est donc confirmé que le prédicteur puissance avec des valeurs supérieures à

185 W exerce une influence positive importante sur les indices qui évaluent la technique de pédalage. Le tableau 4.17 suivant, présente une synthèse des modèles développés dans l'étude 1.

Tableau 4.17 : Synthèse des modèles développés, étude 1

Modèle	Variable dépendante	R ²	RSE	Caractéristiques
Réseau neuronal	IE	0.96	0.03	Performance élevée, généralisable, robuste,
	PIP	0.91	0.09	non interprétable
Forêt aléatoire	IE	0.95	1.75	Performance élevée, généralisable, robuste,
	PIP	0.96	1.6	non interprétable
Régression linéaire multiple	IE	0.86	2.81	Bonne performance, généralisable, robuste,
	PIP	0.81	3.23	hautement interprétable
Régression PLSR	IE	0.92	3.06	Bonne performance, généralisable, robuste,
	PIP	0.84	4	non interprétable
Régression LASSO	IE	0.88	2.57	Bonne performance, généralisable, robuste,
	PIP	0.79	3.24	non interprétable
Arbre de régression	IE	0.57	8.78	Faibles performances, généralisable,
	PIP	0.69	6.08	robuste, hautement interprétable

À ma connaissance, il existe très peu d'études portant sur l'utilisation d'algorithmes d'apprentissage automatique pour prédire les paramètres cinétiques en cyclisme. Cette étude est donc pionnière puisque dans la littérature publiée jusqu'à présent, pour évaluer

les paramètres de pédalage (IE, PIP), il faut utiliser des pédales instrumentées. Il est donc très restrictif de comparer les résultats avec la littérature et aussi parce que les conditions de puissance et de cadence sont très variées dans cette étude. Par conséquent, il est possible d'affirmer qu'il s'agit d'une innovation importante dans le domaine du cyclisme qui permettra une avancée importante dans l'étude de la biomécanique du pédalage.

Conclusion : Le modèle de réseau neuronal permet de prédire avec une précision très grande et une faible erreur les variables dépendantes indice d'efficacité et les proportions d'impulsion positive. Ces modèles sont généralisables, robustes et utilisent la puissance de pédalage comme l'un des principaux prédicteurs. La limite de ce modèle est son explicabilité limitée, mais grâce à un algorithme, il est possible de savoir quelles sont les variables les plus importantes. Le prédicteur puissance est le plus important dans presque tous les modèles étudiés.

4.7.6 Conclusion générale chapitre 4

Ce chapitre présente une gamme de modèles d'apprentissage automatique qui permettent de prédire l'IE et le PIP avec différents niveaux de précision en utilisant la cinématique obtenue à partir de systèmes de capture du mouvement. De même, il est possible de comprendre les facteurs qui influencent ces mesures et de guider les cyclistes dans le processus d'amélioration de leur technique de pédalage. À partir de ces résultats et corroborant le fait que les informations cinématiques sont liées à la cinétique, une étude est proposée pour estimer les forces verticales et antéropostérieures sur la pédale, qui sera présentée dans le chapitre 5 ci-dessous. Aussi, sur la base de la même philosophie d'utilisation des données cinématiques et cinétiques, un modèle qui prédit l'activité musculaire de 4 muscles des membres inférieurs sera également proposé.

CHAPITRE 5 – Étude 2, prédition des forces de pédalage et du recrutement musculaire

Sur la base des résultats obtenus dans l'étude 1, concernant la possibilité de prédire des variables cinétiques à partir de données cinématiques, cette étude 2 explore la possibilité de prédire d'autres variables telles que les forces appliquées sur les pédales et le recrutement musculaire. Comme dans l'étude 1, il serait possible que les applications basées sur l'apprentissage automatique soient intégrées dans les systèmes d'analyse biomécanique en cyclisme. Par conséquent, l'étude 2 consiste à développer des modèles basés sur des réseaux neuronaux pour prédire le signal temporel des forces sur la pédale, F_x , F_z et le recrutement musculaire de 4 muscles du membre inférieur : le grand fessier (GF), le biceps fémoral (BF), le droit fémoral (DF) et le gastrocnémien latéral (GL). Ces muscles ont été sélectionnés de manière à retrouver des moments d'activation dans chaque quadrant de la rotation de la manivelle. Par conséquent, le grand fessier est sélectionné parce qu'il est principalement actif dans le quadrant 1, le biceps fémoral qui est essentiellement actif dans le quadrant 2, le gastrocnémien latéral qui est majoritairement actif dans le quadrant 3 et le droit fémoral qui est en majorité actif dans le quadrant 4. La méthodologie implémentée est la même que celle expliquée dans la section 4.1, commençant par la sélection de la technologie nécessaire, un protocole ultérieur d'acquisition des données, puis un post-traitement des données, l'implémentation des réseaux neuronaux et enfin une validation du modèle. Chacune de ces étapes est décrite ci-dessous.

5.1 Protocole d'acquisition des données

Afin d'obtenir les différents ensembles de données, les participants ont été invités à choisir 3 réglages de puissance fixes à 100 W, 150 W, 200 W et une sélection de puissance variable pendant 5 minutes avec des périodes de repos entre chaque collecte de données. Pour l'étude 2, des cyclistes considérés comme récréatifs ont été recrutés, selon Priego Quesada et al. [128], de manière similaire à ceux recrutés dans l'étude 1. De cette manière, dans l'étude 1, il a été constaté que les cyclistes effectuaient en moyenne des puissances faciles à partir de 100 W, des puissances modérées autour de 150 W et des puissances difficiles autour de 200 W. Compte tenu de ces paramètres, il a été décidé de sélectionner trois gammes de puissance : facile (100 W), modérée (150 W) et difficile (200 W). Pour l'entraînement du réseau neuronal, les puissances sont, comme cela sera discuté plus tard, présentes des conditions cinématiques et cinétiques différentes, une stratégie qui a été appliquée par des auteurs tels qu'Ahmadi et al. [145] dans des objectifs similaires à ceux de la présente étude. Dans la sélection de puissance variable, les individus choisissent au hasard la puissance de leur choix et la font varier sans limite de niveau pendant toute la durée de l'évaluation. De cette manière, la puissance variable moyenne sélectionnée par les cyclistes était 238 ± 93 W (valeur maximale = 755.77 W, valeur minimale = 5.82 W). Un schéma du protocole de l'étude 2 est présenté dans la figure 5.1. Pour la synchronisation des signaux, il a été demandé aux cyclistes de ne pas faire de mouvement pendant 3 secondes et de démarrer avec la manivelle en position 0°. Au début du mouvement, les cyclistes ont réalisé un pic de puissance et par conséquent les 500 données après le début du mouvement n'ont pas été prises en compte. Il en résulte un total de 30000 données (5

$\text{min} \times 60 \times 100$), dont 28000 seront utilisées, en sélectionnant 25000 données (4,16 min) pour l'entraînement et 3000 données pour le test de chaque sujet.

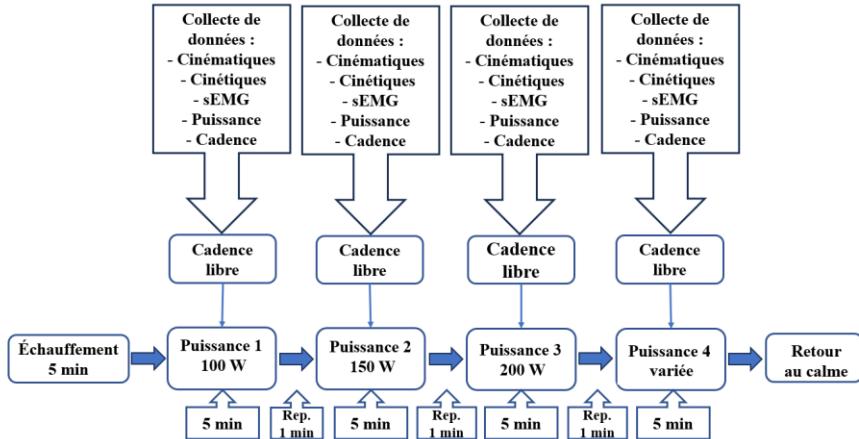


Figure 5.1 : Schéma du protocole d'étude 2

5.2 Participants

Douze cyclistes en bonne santé se sont portés volontaires pour participer à cette étude (36 ± 12 ans, $1,74 \pm 0,06$ m, $72,4 \pm 8,6$ kg). Tous les participants étaient des cyclistes récréatifs sur route comme pour le cas précédent. Les participants ont déclaré le même nombre moyen de kilomètres par an, n'ont déclaré aucune blessure, ont rempli les mêmes formulaires que ceux de l'étude 1 et ont été informés des mêmes aspects. Cette étude a été approuvée par le comité d'éthique de l'UQTR (CER-22-291-07.03) conformément à la Déclaration d'Helsinki (Annexe F).

5.3 Post-traitement des données

Dans cette deuxième étude, au contraire de l'étude 1, les séries temporelles ont été analysées pour l'ensemble des variables à évaluer. Les données cinématiques et cinétiques ont été filtrées de la même manière que dans l'étude 1. De plus, les données d'électromyographie ont été filtrées avec un filtre passe-bande de type Butterworth d'ordre

4 avec des fréquences de coupure comprises entre 20 et 450 Hz. Par la suite, la valeur « root mean square » (RMS) des signaux électromyographiques a été calculé pour obtenir l'amplitude du signal. La fenêtre de glissant a été définie comme étant de 200 ms et un recouvrement de 10 %. Chaque signal est normalisé par rapport à la contraction volontaire maximale obtenue pour chaque muscle au moyen d'une manœuvre spécifique pour le GF, BG, DF et GL. Cela permet d'obtenir l'activité musculaire de chaque muscle calculé en pourcentage (0 - 100%). Les données cinématiques et cinétiques ont été normalisées comme dans l'étude 1. De même, en raison des différentes fréquences de capture des quatre technologies utilisées dans cette recherche, 100 Hz pour le système de capture du mouvement, 450 Hz pour les pédales instrumentées, 1 Hz pour le système de capture de la puissance et de la cadence et 2148 Hz pour le système de capture sEMG, un rééchantillonnage à 100 Hz a été nécessaire. Pour ce signal électromyographique, le rééchantillonnage a été effectué après avoir obtenu le RMS. Afin de synchroniser les signaux, tous les individus évalués ont commencé le mouvement de la manivelle à la position 0°, c'est-à-dire au point mort haut, en commençant le mouvement 5 secondes après le démarrage de tous les systèmes de capture. En outre, à l'aide du logiciel Matlab, les signaux ont été synchronisés en fonction du début de l'activation du muscle grand fessier.

5.4 Implémentation des modèles

Pour cette étude 2, les variables dépendantes suivantes ont été définies : les forces de pédalage F_x, F_z et la courbe enveloppe d'activation EMG RMS de 4 muscles des membres inférieurs GF, BF, DF, GL. La force F_y n'a pas été analysée conformément au modèle bidimensionnel défini. En outre, selon le modèle cinématique bidimensionnelle

dans le plan sagittal, il n'y a pas de données cinématiques dans ce plan frontal. Les prédicteurs des forces de pédalage ont été définis comme la cinématique des membres inférieurs, c'est-à-dire les angles et les vitesses angulaires de la hanche, du genou et de la cheville (six valeurs), conjointement avec la puissance et la cadence, pour un total de 8 prédicteurs. Enfin, les prédicteurs de la courbe enveloppe d'activation EMG RMS des quatre muscles des membres inférieurs ont été définis comme les forces de pédalage à partir des pédales instrumentées (F_x, F_y, F_z) et la cinématique des membres inférieurs, c'est-à-dire, comme ci-dessus, les angles et les vitesses angulaires des membres inférieurs en liaison avec la puissance et la cadence, pour un total de 11 prédicteurs.

5.4.1 Réseau neuronal pour prédire les forces de pédalage

Les signaux temporels cinématiques et cinétiques d'une durée totale de 4.16 minutes, soit 25000 données pour chaque participant, ont été regroupés pour chaque valeur de puissance c'est-à-dire à 100 W, 150 W et 200 W. Il en résulte trois bases de données avec un total de 275000 données de séries temporelles d'entraînement pour chaque puissance évaluée afin d'entraîner le réseau neuronal. Le coefficient de détermination R^2 a été utilisée pour évaluer les performances des résultats par rapport au modèle de régression. Pour déterminer les différences entre les valeurs prédites et mesurées, le RMSE a été calculée. Pour les données de test, ont été sélectionnées les 30 secondes suivantes de séries temporelles cinématiques et cinétiques, soit 3000 points de données, suivants à chaque sélection de données d'entraînement pour chaque individu. Pendant la période d'entraînement, 10 % des données d'apprentissage ont été usées pour disposer de données de validation. À la fin de chaque « epoch » d'entraînement, le modèle est évalué à l'aide de ces données de validation. Ces données sont utilisées pour éviter le sur-apprentissage

du modèle [146]. Pour valider le modèle, la méthode de validation croisée 5-fold a été utilisée. Chaque portion (ou « fold ») a été composée de 20 %, 40 %, 60 %, 80 % et 100 % de chaque individu. Les données de chaque portion de données ont été rassemblées et le réseau neuronal a été entraîné. Il a ensuite été déterminé le moyenne et l'écart-type du RMSE et du coefficient de détermination R^2 . Le processus général est illustré dans la figure 5.2. Un réseau neuronal a été développé pour toutes les conditions d'essai, c'est-à-dire 100 W, 150 W et 200 W. D'autres possibilités ont également été évaluées au niveau de la puissance variable et au niveau inter-sujets. Pour implémenter les modèles intra-sujet, 25000 données de séries temporelles provenant de chacune des trois puissances de chaque sujet sont regroupées en un seul ensemble de données de manière à obtenir 275000 données (25000×11) de séries temporelles pour l'entraînement du réseau neuronal. De ces données, 10 % ont été consacrées aux données de test, 10 % aux données de validation et 80 % aux données d'entraînement. Pour le modèle à puissance variable, un ensemble unique de 275000 données de séries temporelles a été regroupé en sélectionnant la même proportion de données d'entraînement, validation et test.

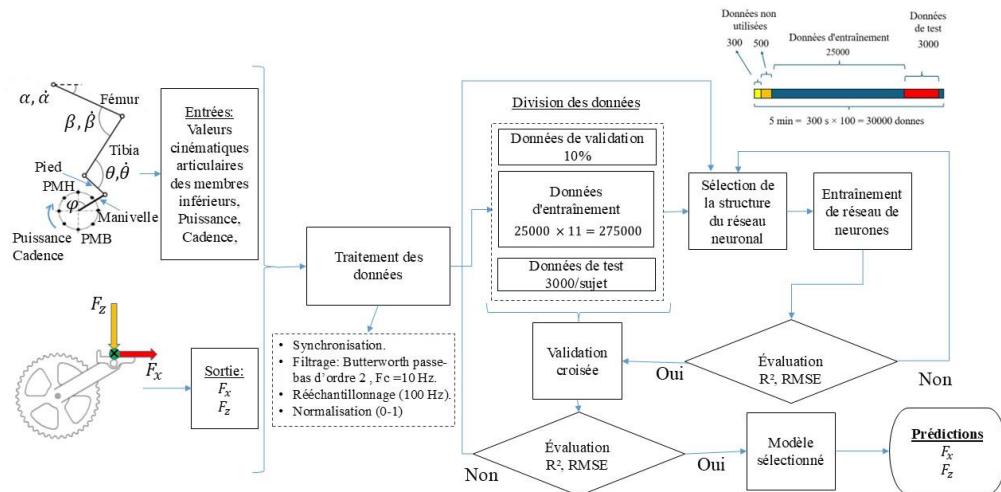


Figure 5.2 : Diagramme méthodologie réseau neuronal pour prédire forces de pédalage

pédalage

Comme le montre la figure 5.2, le réseau de neurones est composé de huit entrées. Les données de sortie du réseau des neurones consistaient en des séries temporelles de force dans les directions F_x et F_z , en utilisant une seule série temporelle comme sortie ou variable dépendante. Le réseau neuronal a été développé à l'aide de RStudio en utilisant les librairies du logiciel Python.

- Paramètres du réseau neuronal

La fonction ReLU a été sélectionnée comme fonction d'activation. Cette fonction a été choisie parce qu'elle converge beaucoup plus rapidement que les autres fonctions, ce qui facilite l'entraînement, et aussi parce qu'elle est plus facile à implémenter d'un point de vue computationnel, entre autres avantages [147]. Au cours du processus de sélection des paramètres, d'autres fonctions d'activation telles que Tanh et sigmoïdale ont été testées et des problèmes de non-convergence du réseau neuronal ont été rencontrés. Le MSE a été utilisée comme fonction de perte pour comparer la forme d'onde force-temps prédite et mesurée. Pour minimiser la fonction de perte pendant l'entraînement des réseaux neuronaux, il est utilisé l'algorithme ADAM « Adaptive Moment Estimation », qui est un algorithme pour la technique d'optimisation de la descente de gradient. Le taux d'apprentissage a été défini à une valeur de 0.01. Enfin, le réseau de neurones est composé de 5 couches cachées de 2000 neurones chacune. Une valeur de 10 époques a été mise en œuvre avec une fonction d'arrêt précoce du modèle avant qu'il ne s'adapte trop aux données d'apprentissage. La figure 5.3 montre l'architecture du réseau neuronal « Feed Forward » développé. Tous les neurones entre chaque couche sont connectés et il s'agit donc d'un type dense.

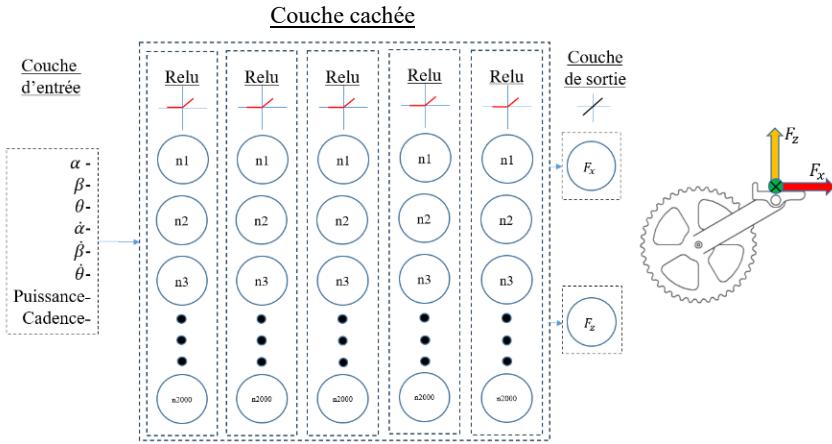


Figure 5.3 : Architecture réseau neuronal pour prédire forces de pédalage étude 2

5.4.2 Réseau neuronal pour prédire la courbe enveloppe RMS EMG

Le réseau neuronal implémenté est le LSTM (voir section 2.4.7.2) à l'aide du logiciel Matlab. Le processus général est illustré à la figure 5.4. Ce réseau neuronal a été sélectionné, car il a donné les meilleurs résultats parmi les autres réseaux évalués, tels que le réseau FFNN et un autre de la famille des réseaux récurrents, tel que le réseau NARX. Les données d'entrée des réseaux neuronaux étaient constituées de 11 prédicteurs. Ils sont composés des angles et les vitesses angulaires des articulations de la hanche, du genou et de la cheville, les données de force, c'est-à-dire F_x, F_y, F_z , la puissance et la cadence. Les données de sortie des réseaux neuronaux comprenaient chaque courbe enveloppée RMS EMG de quatre muscles des membres inférieurs, à savoir grand fessier (GF), droit fémoral (DF), biceps fémoral (BF), gastrocnémien latéral GL. Une seule donnée de sortie a été utilisée pour chaque réseau, en fonction du muscle testé. Une LSTM a été étudiée pour chaque condition de test, c'est-à-dire 100 W, 150 W et 200 W.

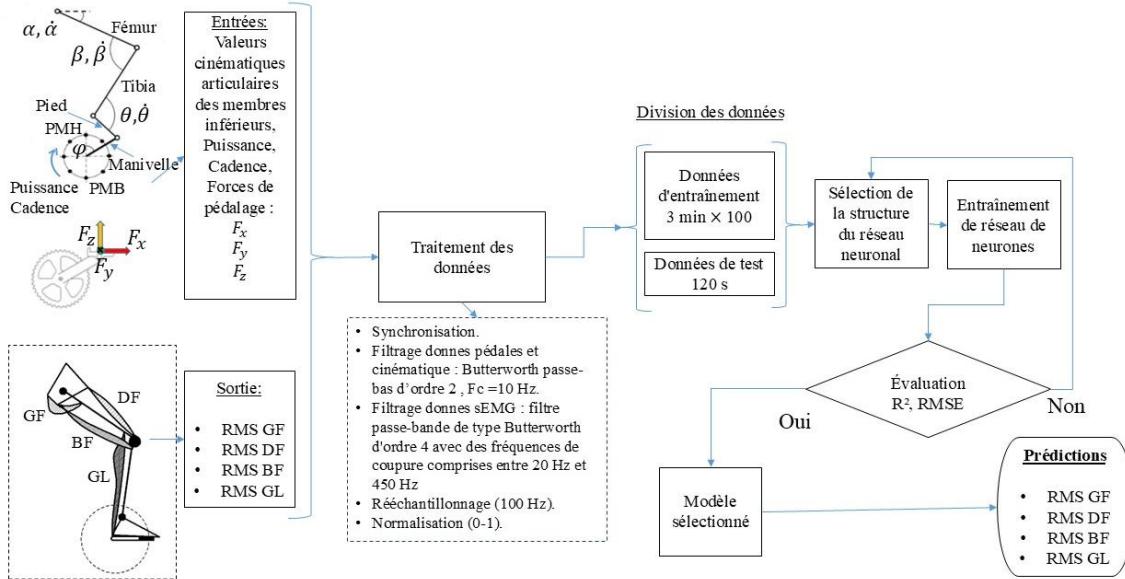


Figure 5.4 : Diagramme méthodologique réseau neuronal LSTM pour prédire la courbe enveloppe EMG RMS

Pour l'entraînement du LSTM, les données de force, de cinématique articulaire, c'est-à-dire les angles et les vitesses angulaires du fémur, du genou, de la cheville, la puissance et la cadence, des 20 premières secondes des tests n'ont pas été utilisées, dans le but de ne pas prendre en compte la phase d'accélération où ces variables partent de zéro pour atteindre leur état stable. Les données entre la 20ème seconde et la 200ème seconde de chaque test ont été sélectionnées pour arriver à un total de 3 min. Afin d'améliorer l'entraînement du réseau, ces données ont été multipliées 100 fois. Cette valeur a été trouvée par essai et erreur jusqu'à ce qu'une valeur permettant de trouver de bonnes valeurs de performance pour le modèle soit obtenue, en commençant par 10 fois, puis 20 fois jusqu'à ce que la valeur trouvée soit atteinte. Un code Matlab concatène les paquets de données d'entrée du début à la fin de la série temporelle de chaque prédicteur pour effectuer l'apprentissage du réseau neuronal. Cette valeur de multiplication a été choisie par essai/erreur pour optimiser les performances des réseaux et le temps de calcul. Ces

données ont été utilisées comme données d'entrée pour les réseaux. Ensuite, les données de force, la cinématique des articulations, la puissance et la cadence à partir de 200e à la 320e seconde ont été utilisées comme données de test (120 sec). Les réseaux prédisent en sortie la courbe d'enveloppe EMG des quatre muscles étudiés pendant cette période, pour chaque valeur de puissance, c'est-à-dire à 100 W, 150 W et 200 W. Le coefficient de détermination R^2 , basé sur la corrélation entre les signaux EMG réels et prédits pour chaque muscle pendant le test, a été utilisé pour quantifier la qualité des prédictions faites par chaque réseau.

L'erreur quadratique moyenne arrondie (RMSE) est utilisée pour déterminer l'erreur entre les données mesurées et les données prédites. Le LSTM a été créé pour chaque muscle, pendant chaque condition de puissance et pour chaque sujet. Les hyperparamètres du réseau ont été fixés pour tout l'entraînement des réseaux. Ils ont été définis en tenant compte des caractéristiques des données de sortie et par essai/erreur afin d'optimiser les performances des réseaux et le temps de calcul.

- Paramètres du réseau de neurones LSTM

Le réseau de neurones LSTM est composé d'une couche d'entrée, une couche LSTM entièrement connectée, une couche d'unité linéaire rectifiée (ReLU), une couche d'exclusion entièrement connectée avec un facteur d'exclusion de 0,5 et une période de 125, et finalement une couche de sortie de régression. Le nombre de neurones a été fixé à 200 unités. De cette manière, le nombre total de couches est de 5. Le réseau a été entraîné avec une taille de « mini batch » de 32. Le taux d'apprentissage initial était de 0.001 et le gradient de seuil de 1. Le solveur utilisé était ADAM « Adaptive Moment Estimation ». L'entraînement a été arrêté après 10 époques, parce que la performance du réseau n'était

pas meilleure avec plus d'époques et pour réduire le temps de calcul. L'architecture du LSTM est présentée à la figure 5.5.

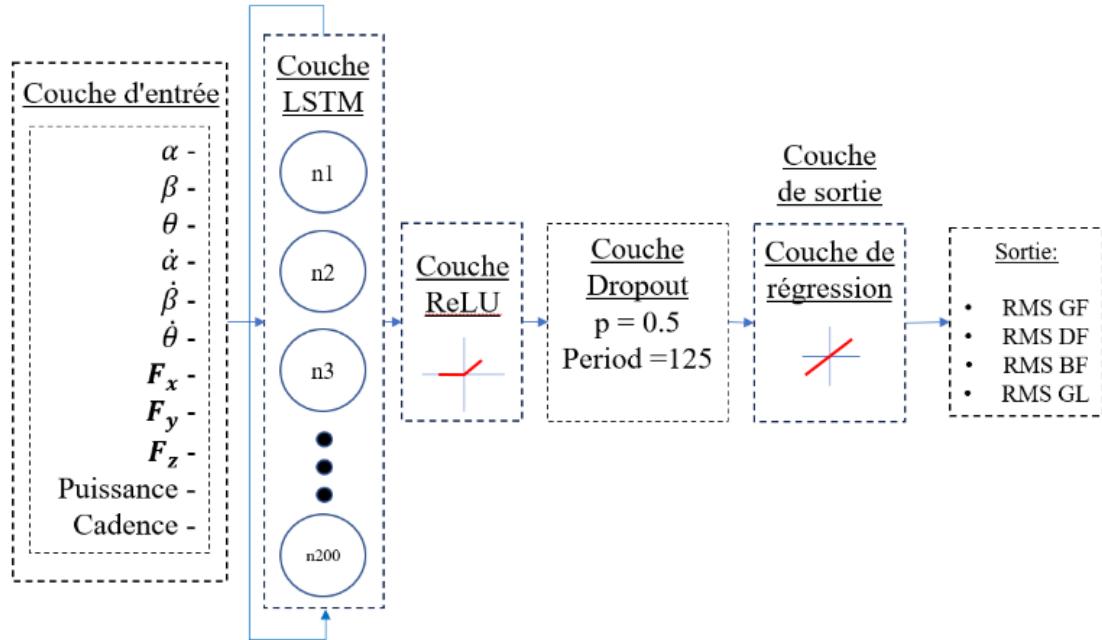


Figure 5.5 : Architecture du réseau de neurones LSTM pour prédire la courbe enveloppe EMG RMS

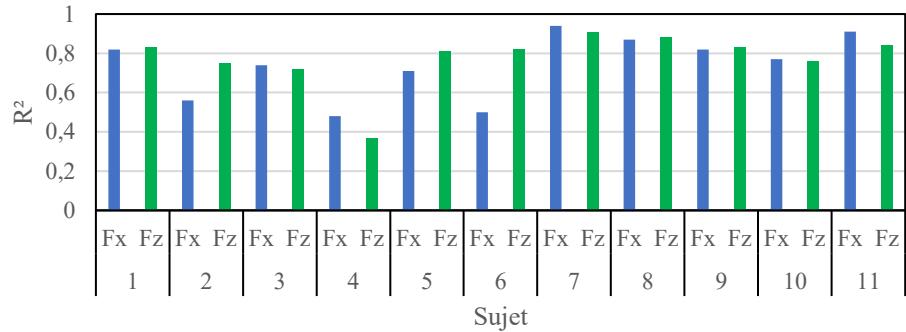
5.5 Résultats

5.5.1 Résultats réseau neuronal prédiction forces de pédalage

Quatre réseaux neuronaux ont été développés pour chaque puissance évaluée à 100 W, 150 W, 200 W et a puissance variable, c'est-à-dire que quatre réseaux neuronaux différents ont été entraînés. Les 500 premières données ont été supprimées, afin d'obtenir des données cinématiques et cinétiques stables. Pour entraîner les réseaux de neurones, un total de 25000 données pour chaque individu ont été sélectionnées. Les données cinématiques corrompues d'un participant ont été supprimées. De cette façon, les données de 11 participants ont ensuite été regroupées pour entraîner chaque réseau neuronal, avec

un total de 275000 données d'entraînement. Pour le test du réseau neuronal, les 30 secondes restantes, c'est-à-dire que 3000 données de séries temporelles ont été utilisées pour chaque participant, c'est-à-dire les données que le réseau n'a pas utilisées pour l'apprentissage. Cette sélection de données donne un ratio d'environ 90 % de données d'entraînement et 10 % de données de test. Cela permet de prédire les forces F_x et F_z de chaque individu. À partir des données d'apprentissage, 10 % des données ont été sélectionnées pour validation afin de surveiller la courbe d'apprentissage. De cette manière, la figure 5.6 montre le R^2 et le RMSE de la prédiction des forces F_x et F_z appliquées aux pédales pour une puissance de 100 W pour chacun des individus évalués. Aussi, en annexe F et annexe G, un segment de 5 secondes et un graphique de corrélation de prédiction des forces F_x et F_z à une puissance de 100 W sont présentés pour chaque sujet évalué. Le RMSE est exprimé dans les mêmes unités que la variable cible. Par conséquent, considérant que dans la présente étude, la variable F_z a atteint des valeurs maximales d'environ de 500 N et que la variable F_x a atteint des valeurs maximales d'environ - 20 N et 80 N, les valeurs de l'erreur RMSE atteignent des niveaux différents. Par conséquent, des différences dans les valeurs RMSE peuvent être observées entre les deux variables cibles.

(a)

 R^2 par sujet à 100 W

(b)

RMSE par sujet à 100 W

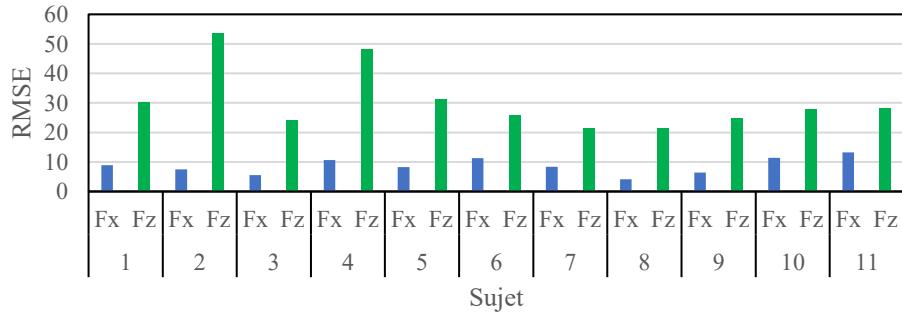


Figure 5.6 : Prédiction des forces Fx et Fz à 100 W, R^2 (a), RMSE (b)

Pour valider le modèle, une validation croisée a été effectuée en regroupant 5 fractions différentes de l'ensemble des données disponibles pour chaque sujet, comme suit : 20 % (Fraction 1), 40 % (Fraction 2), 60 % (Fraction 3), 80% (Fraction 4) et 100 % (Fraction 5) des données. Cela donne les fractions de données suivantes : fraction 1 = 55000, fraction 2 = 110000, fraction 3 = 165000, fraction 4 = 220000 et fraction 5 = 275000. Chaque fraction de données a été divisée en 80 % pour l'entraînement du réseau neuronal et 20 % comme données de test afin de déterminer le R^2 et le RMSE. Sur les 80 % de données d'entraînement, 10 % ont été utilisés comme données de validation pendant

l'entraînement du réseau neuronal. La figure 5.7 montre le résultat de la validation croisée de chaque portion de données à 100 W.

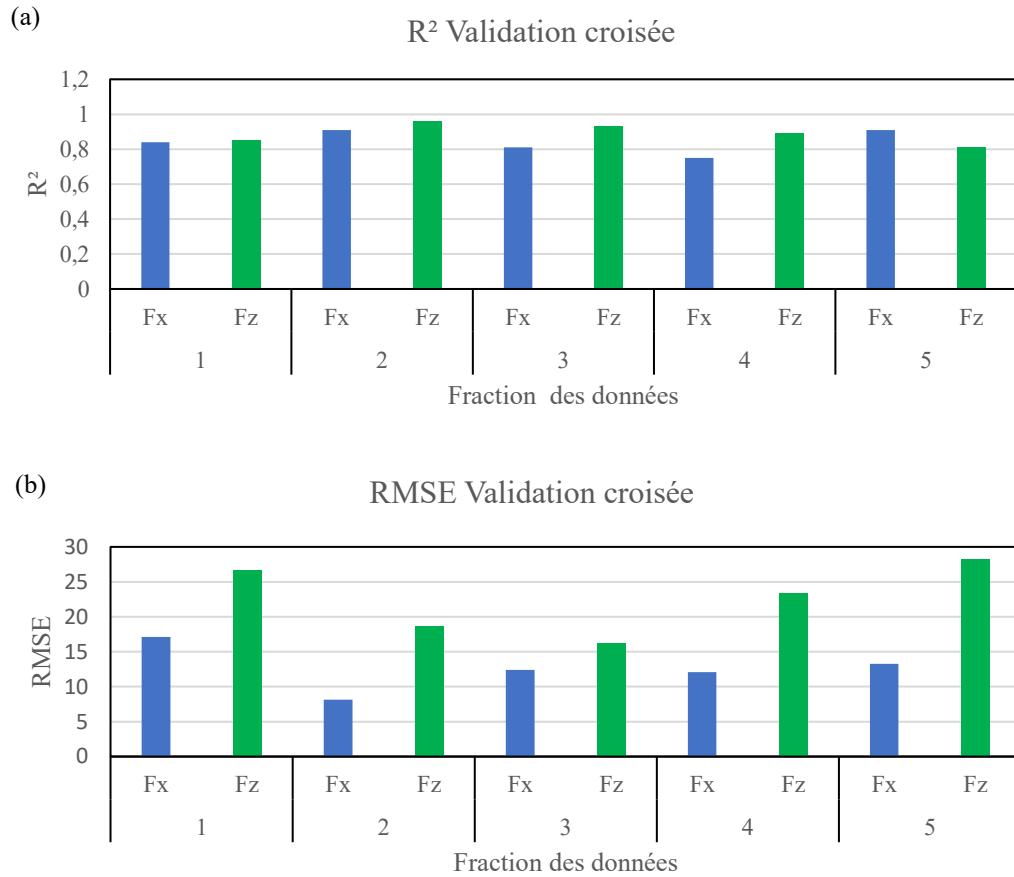


Figure 5.7 : Résultats validation croisée à 100 W, R^2 moyenne (a), RMSE moyenne (b)

La figure 5.8 montre le R^2 et le RMSE de la prédiction des forces F_x et F_z appliquées aux pédales pour une puissance de 150 W pour chacun des individus évalués.

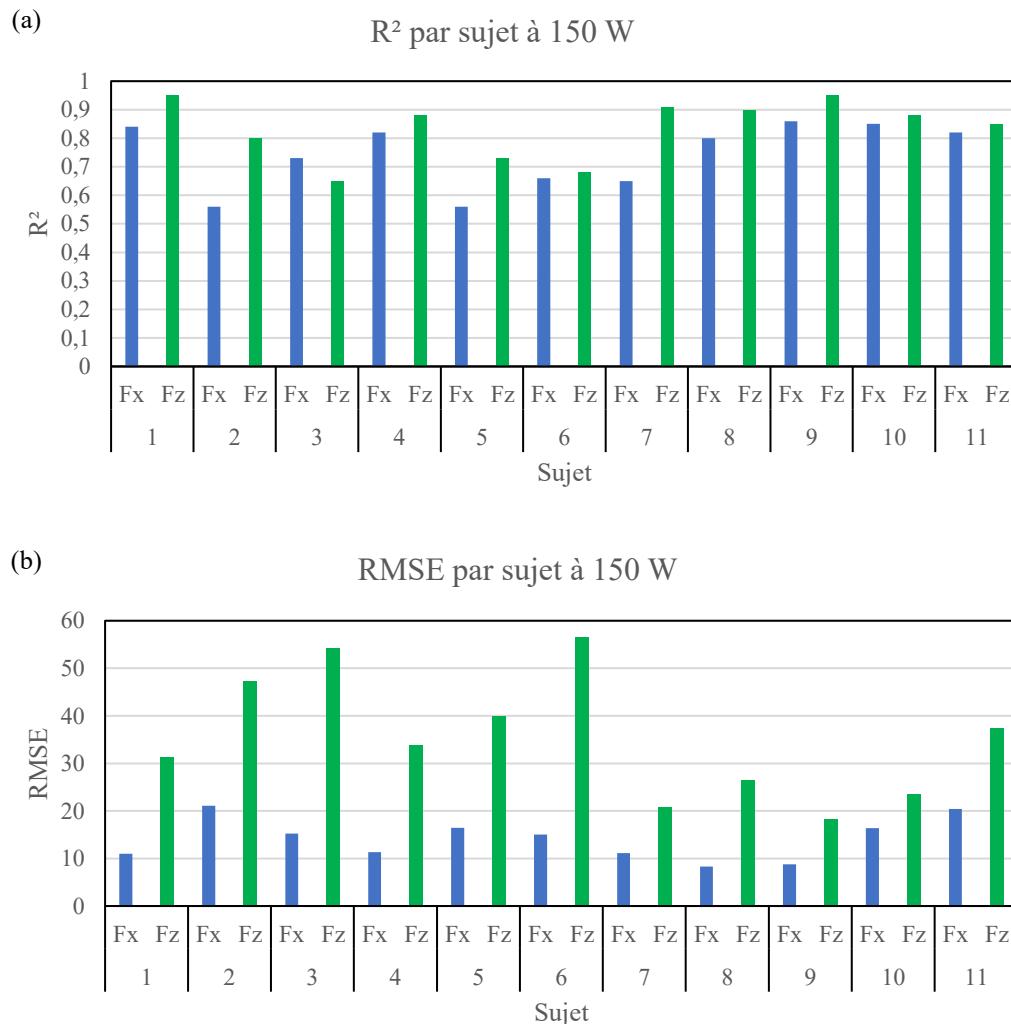
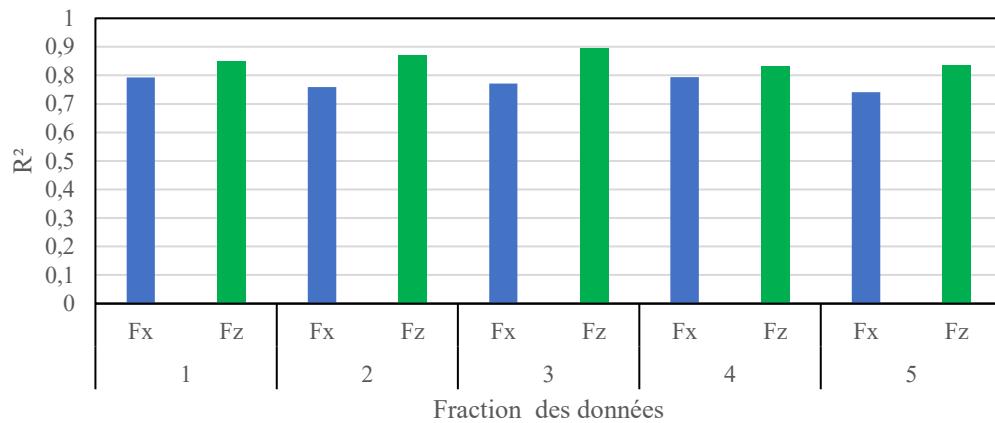


Figure 5.8 : Résultats prédition des forces Fx et Fz à 150 W, R^2 (a), RMSE (b)

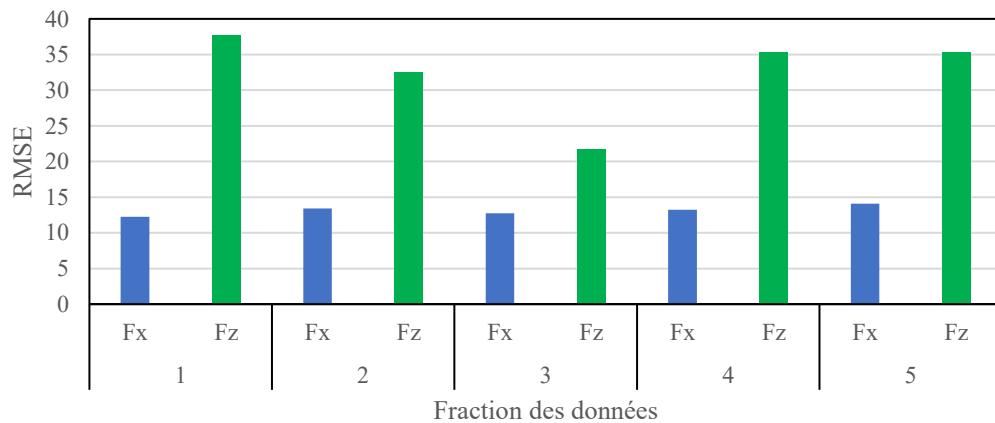
De la même manière, les résultats de la validation croisée du réseau neuronal à 150 W sont présentés dans la figure 5.9.

(a)

 R^2 Validation croisée à 150 W

(b)

RMSE Validation croisée à 150 W

**Figure 5.9 : Résultats validation croisée à 150 W, R^2 moyenne (a), RMSE moyenne (b)**

La figure 5.10 montre le R^2 et le RMSE de la prédiction des forces Fx et Fz appliquées aux pédales pour une puissance de 200 W pour chacun des individus évalués.

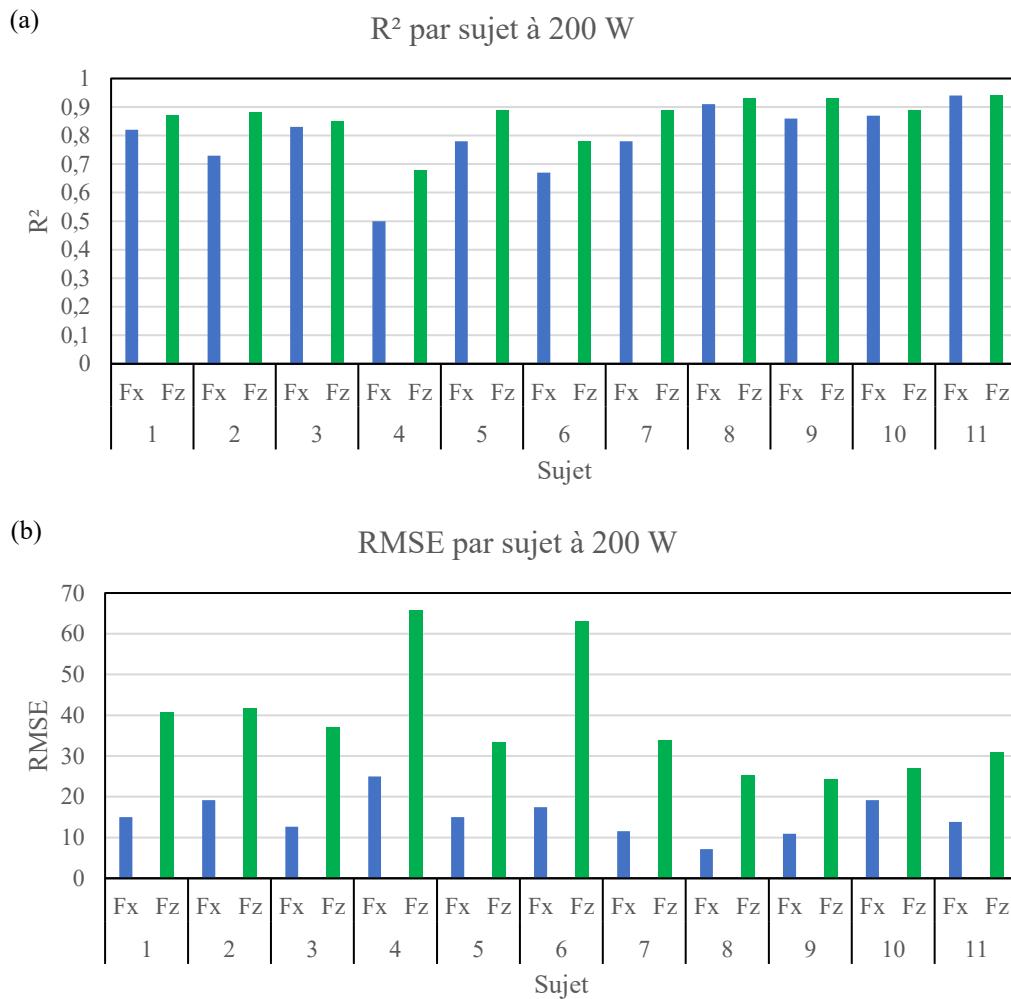


Figure 5.10 : Résultats prédition des forces F_x et F_z à 200 W, R^2 (a), RMSE (b)

De la même façon, les résultats de la validation croisée du réseau neuronal à 200 W sont présentés dans la figure 5.11.

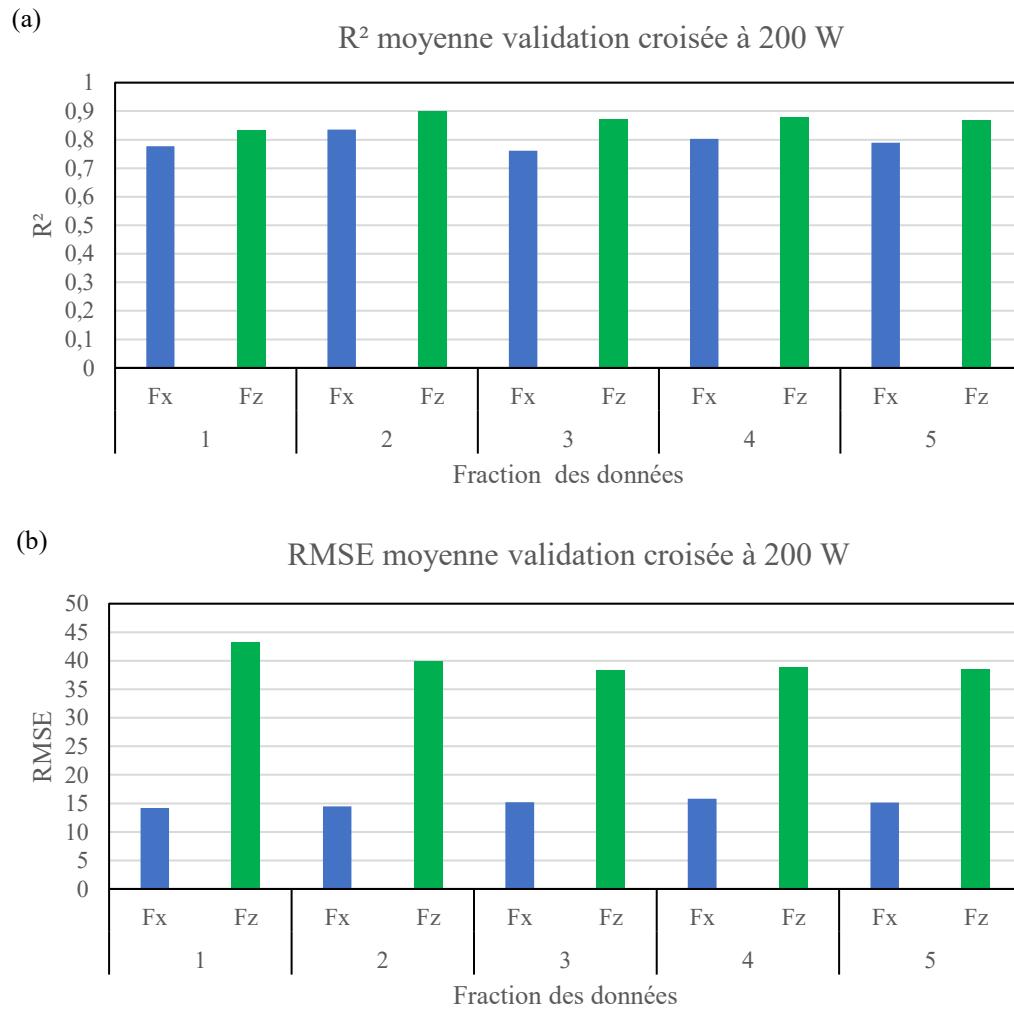


Figure 5.11 : Résultats validation croisée à 200 W, R^2 moyenne (a),

RMSE moyenne (b)

La figure 5.12 (a,b), montre le R^2 et le RMSE de la prédiction des forces F_x et F_z appliquées aux pédales pour une puissance variable pour chacun des individus évalués. De plus, sont également montrés dans la figure 5.12 (c,d), le R^2 et le RMSE intra-sujet en regroupant toutes les puissances évaluées. Pour obtenir ces résultats, l'entraînement du réseau neuronal a été effectué séparément pour chaque sujet en regroupant les puissances

évaluées (100 W, 150 W, 200 W) avec les proportions 80 % d'entraînement, 10 % de test et 10 % de validation.

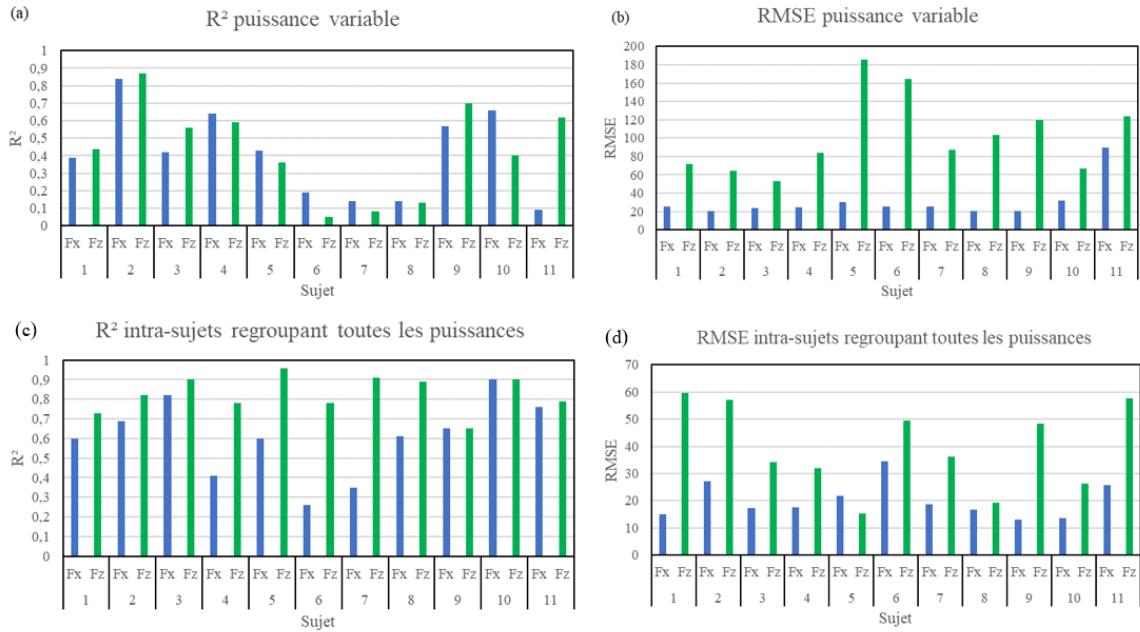


Figure 5.12 : R^2 (a), RMSE (b) de la prédiction des forces Fx et Fz à puissance variable, et R^2 (c), RMSE (d) de la prédiction des forces Fx et Fz intra-sujets regroupant toutes les puissances

Comme synthèse, la figure 5.13 montre les moyennes et l'écart-type des valeurs F_x et F_z prédites pour chaque puissance évaluée.

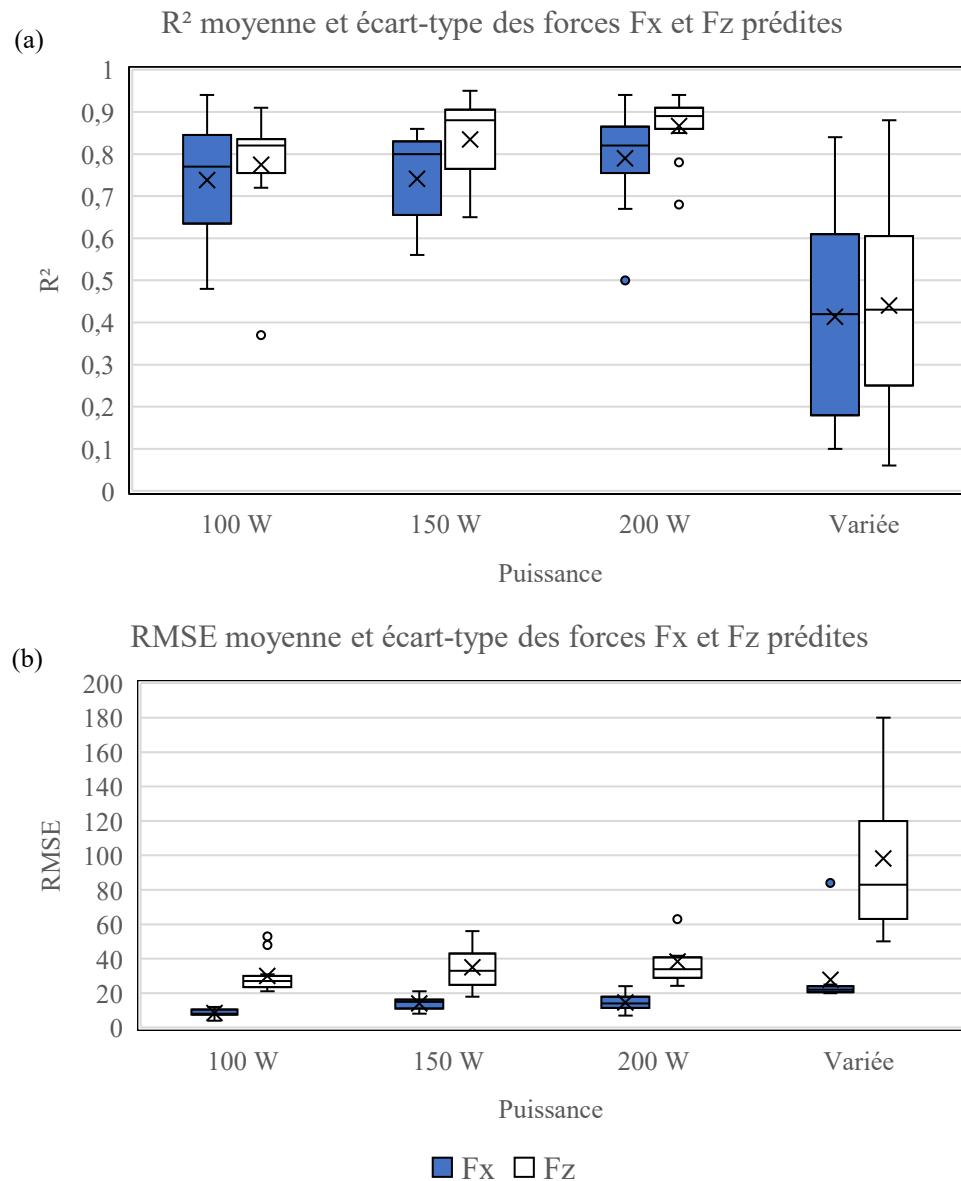
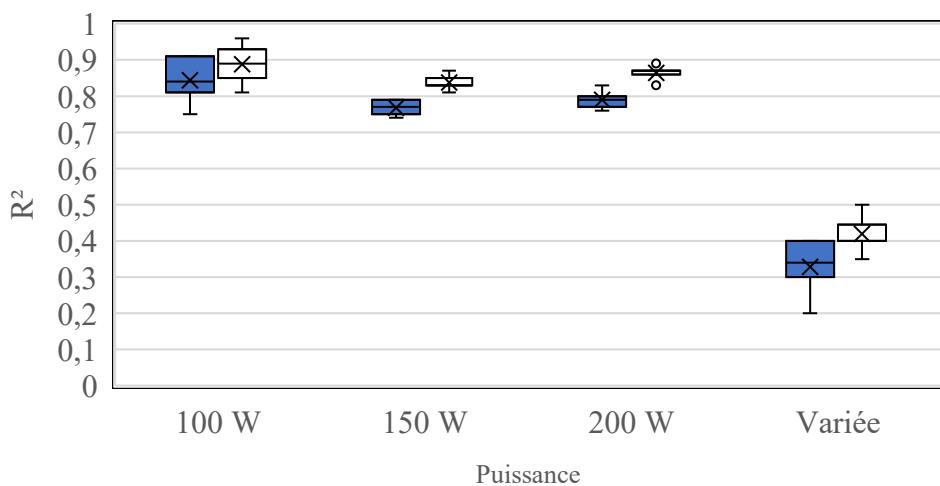


Figure 5.13 : R² (a) et RMSE (b) moyenne et écart-type des forces prédites F_x et F_z dans les gammes de puissance évaluées

De même, la figure 5.14 ci-dessous résume la moyenne obtenue avec la validation croisée en tenant compte de tous les « folds » construits.

(a) R^2 moyenne et écart-type validation croisée

(b) RMSE moyenne et écart-type validation croisée

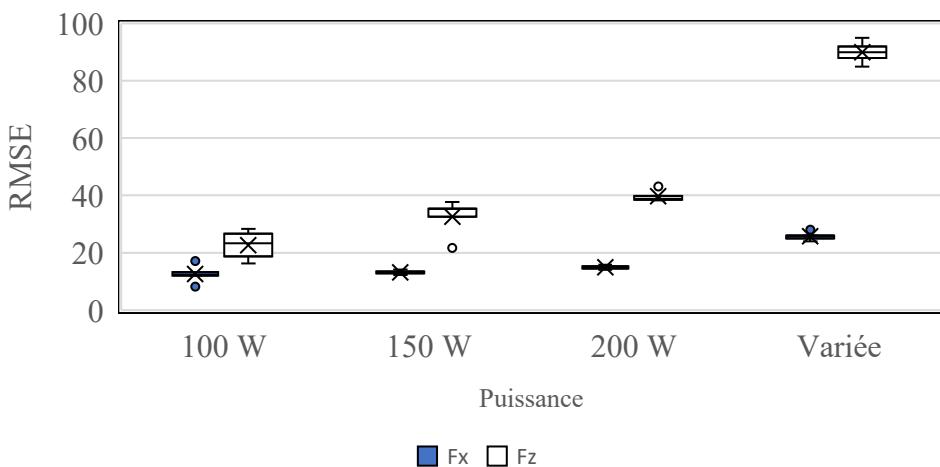


Figure 5.14 : R^2 (a) et RMSE (b) moyenne et écart-type de la validation croisée dans les gammes de puissance évaluées

5.5.2 Résultats réseau neuronal prédiction courbe enveloppe EMG RMS

Six sujets ont été sélectionnés pour modéliser la courbe enveloppe linéaire EMG RMS au moyen d'un réseau neuronal de type LSTM, de quatre muscles des membres inférieurs, le grand fessier (GF), le biceps fémoral (BF), le droit fémoral (DF) et le gastrocnémien

latéral (GL), pour chaque valeur de puissance évaluée. L'Annexe M présente cinq secondes de chaque prédition pour chaque muscle testé à une puissance de 100 W, l'Annexe N à la puissance de 150 W et l'Annexe O pour la puissance de 200 W. De plus, des graphiques de corrélation sont présentés pour l'ensemble des données.

La figure 5.15, montre les valeurs obtenues de R^2 et l'erreur RMSE des prédictions de la courbe linéaire enveloppe EMG RMS avec le réseau neuronal LSTM à 100 W.

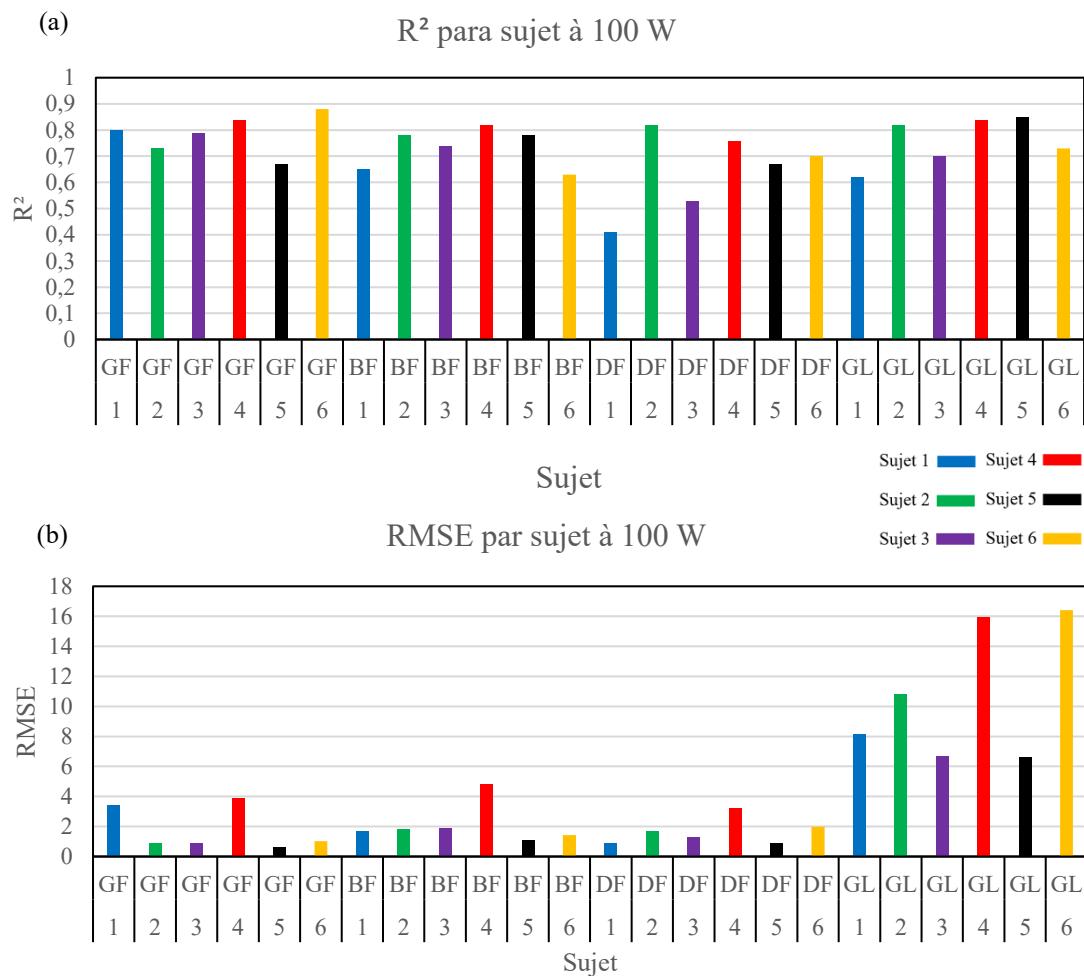


Figure 5.15 : R^2 (a) et RMSE (b) prédictions de la courbe linéaire enveloppe EMG RMS avec le réseau neuronal LSTM à 100 W

RMS avec le réseau neuronal LSTM à 100 W

La figure 5.16, montre les valeurs obtenues de R^2 et l'erreur RMSE des prédictions de la courbe linéaire enveloppe EMG RMS avec le réseau neuronal LSTM à 150 W.

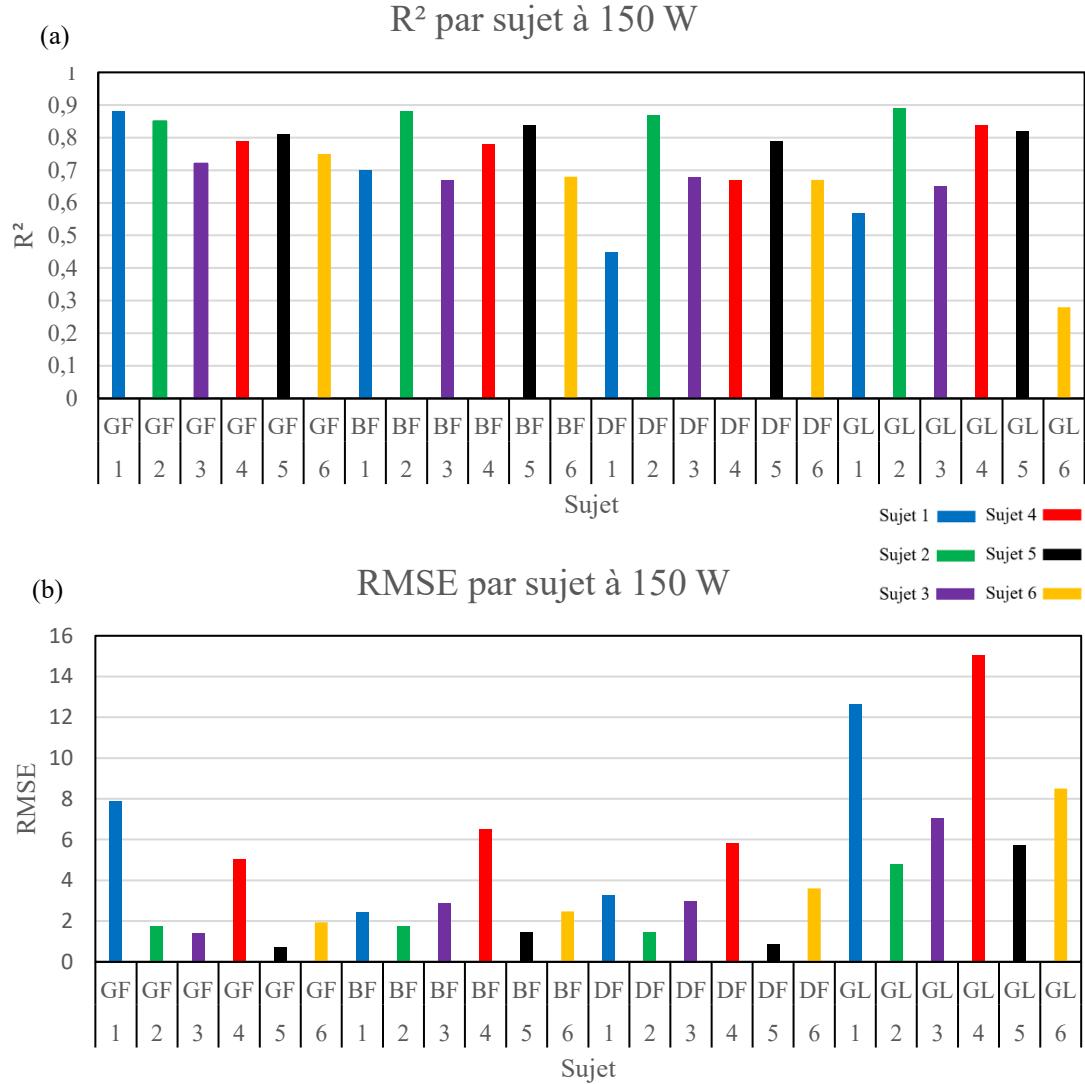


Figure 5.16 : R^2 (a) et RMSE (b) des prédictions de la courbe linéaire enveloppe EMG RMS avec le réseau neuronal LSTM à 150 W

La figure 5.17, montre les valeurs obtenues de R^2 et l'erreur RMSE des prédictions de la courbe linéaire enveloppe EMG RMS avec le réseau neuronal LSTM à 200 W.

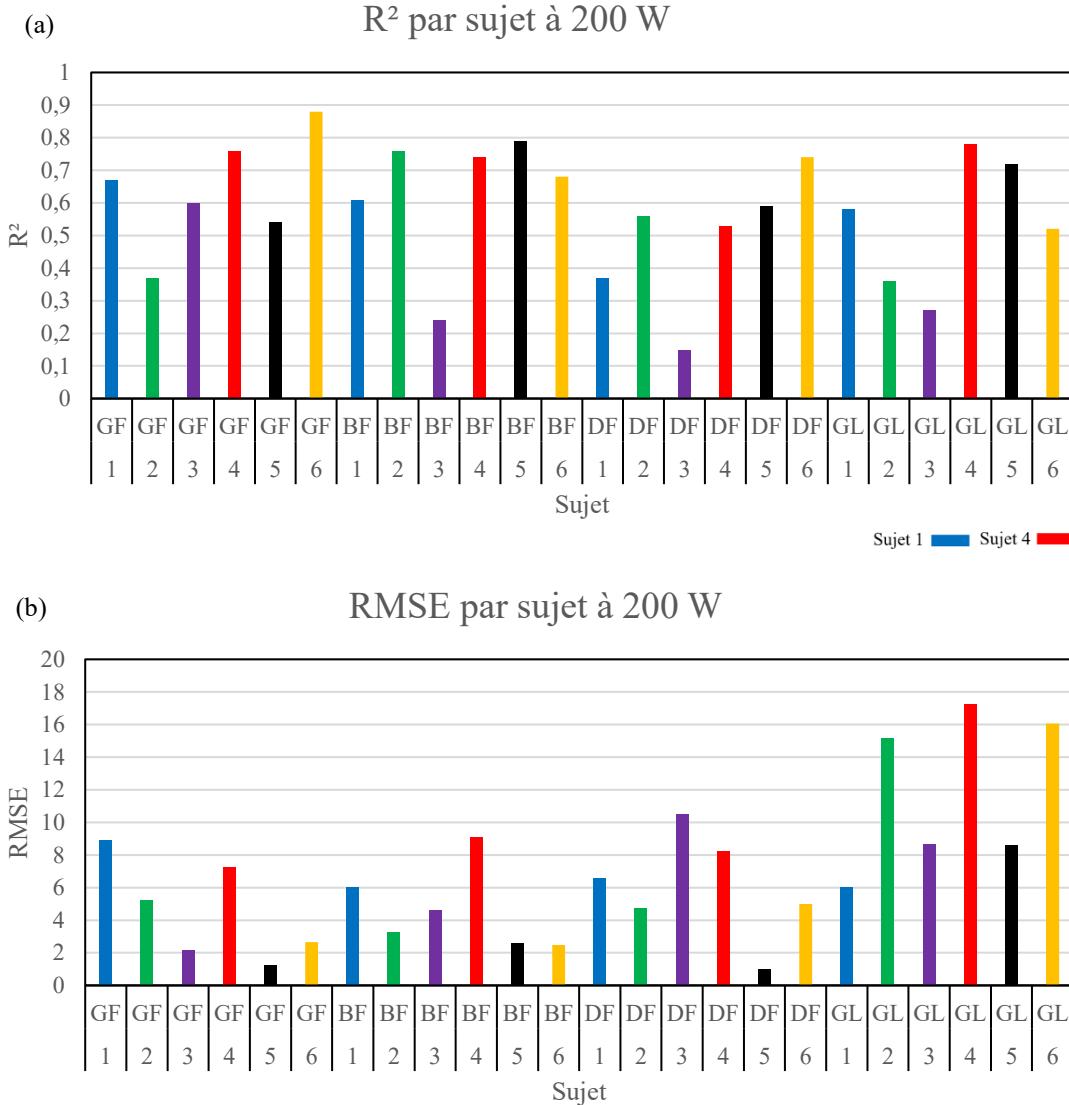


Figure 5.17 : R^2 (a) et RMSE (b) prédictions de la courbe linéaire enveloppe EMG

RMS avec le réseau neuronal LSTM à 200 W

La figure 5.18, montre les valeurs obtenues de R^2 et l'erreur RMSE des prédictions de la courbe linéaire enveloppe EMG RMS avec le réseau neuronal LSTM à puissance variable.

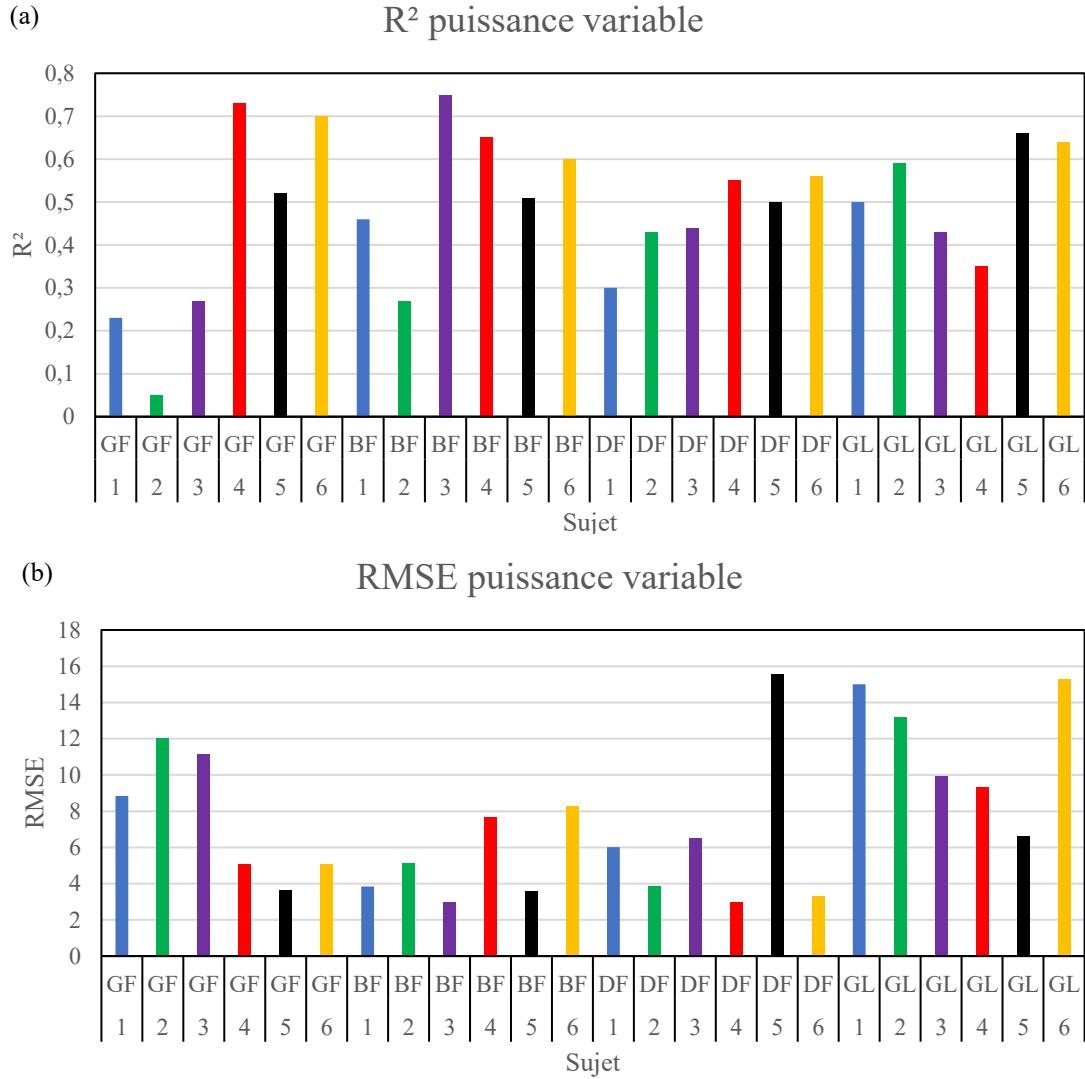


Figure 5.18: R^2 (a) et RMSE (b) prédictions de la courbe linéaire enveloppe EMG

RMS avec le réseau neuronal LSTM à puissance variable

La figure 5.19 montre le R^2 par puissance pour chaque muscle analysé pour tous les sujets testés.

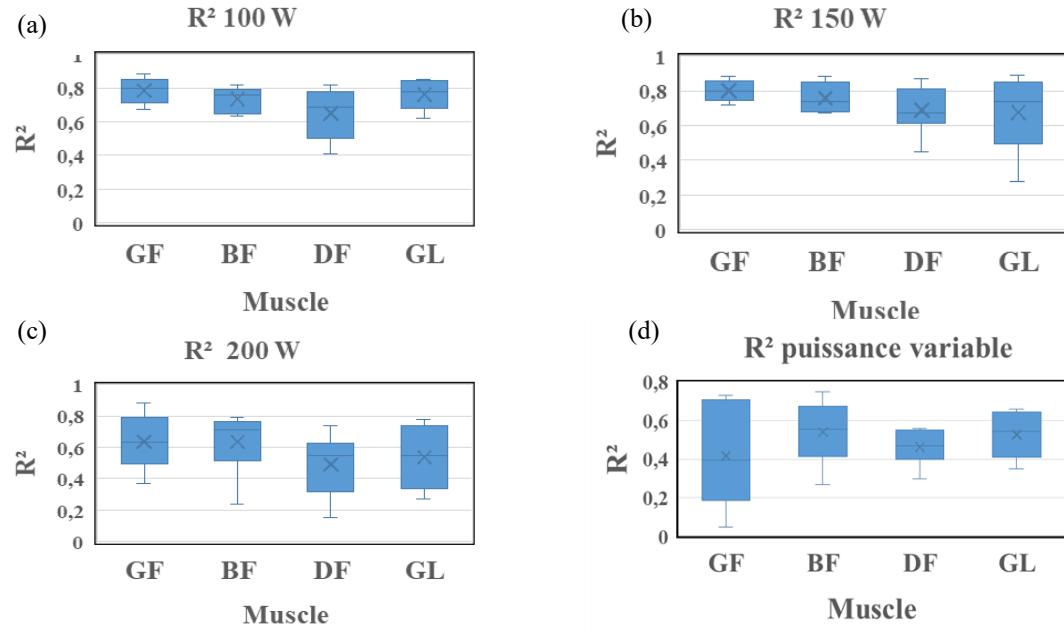


Figure 5.19 : R^2 par puissance (a) 100 W, (b) 150 W, (c) 200 W et (d) à puissance variable pour chaque muscle analysé

La figure 5.20 montre le RMSE par puissance pour chaque muscle analysé pour tous les sujets testés.

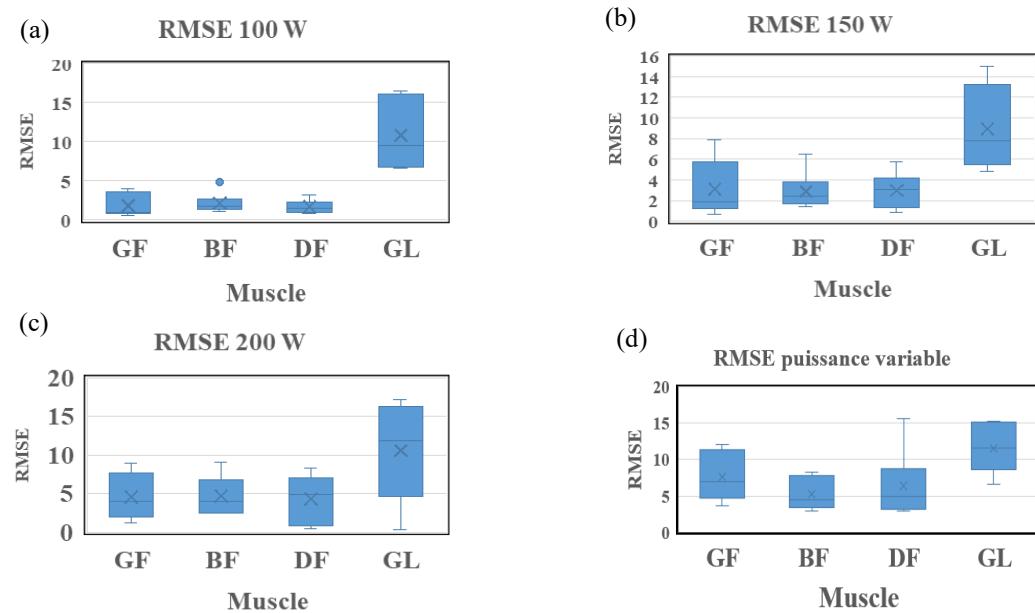


Figure 5.20 : RMSE par puissance (a) 100 W, (b) 150 W, (c) 200 W et (d) à puissance variable pour chaque muscle analysé

5.6 Analyse des résultats

5.6.1 Analyse des résultats prédition forces du pédalage

Dans cette section, les résultats obtenus avec le réseau neuronal développé à partir des données de l'étude 2, concernant la prédition de la force de pédalage, seront analysés. Tout d'abord, il a été constaté qu'avec les données prises dans le plan sagittal, il est possible de prédire avec une bonne précision la force F_x dans la direction de l'axe X et la force F_z dans la direction de l'axe Z selon les axes cartésiens définis dans la figure 3.3. Cependant, la méthodologie appliquée et les procédures d'apprentissage automatique peuvent être extrapolées à une étude future afin d'obtenir des données sur le plan frontal et de déterminer ainsi les profils de force F_y . Une autre particularité de ces résultats est que le réseau neuronal dépend de la puissance, c'est-à-dire que quatre réseaux neuronaux différents ont été développés pour les puissances 100W, 150W et 200W. L'hypothèse de ce phénomène est qu'il peut être dû au fait que, comme le montrent les annexes G, H, I, J, K et L, les différences dans la façon de générer la force dans chaque condition de puissance diffèrent d'une condition à l'autre chez les individus qui ont été évalués. D'abord l'amplitude des forces est différente dans chaque condition de puissance. Il en résulte une grande variabilité des courbes avec des détails spécifiques à chaque sujet et à chaque condition de puissance. En particulier pour la force F_x , certains individus présentent des valeurs positives et négatives, tandis que d'autres n'en présentent pas. Par conséquent, pour le développement d'un réseau neuronal généralisable à toutes les puissances, d'autres sources d'informations ou des prédicteurs peuvent être nécessaires. Comme expliqué, les prédicteurs sélectionnés sont basés principalement sur la cinématique du membre

inférieur, en cherchant un modèle avec le moins de prédicteurs possibles, c'est-à-dire un modèle qui soit conforme au principe de parcimonie (modèle le plus simple possible). Je suggère parmi les prédicteurs potentiels, il serait intéressant de considérer, la masse de l'individu, la position horizontale du corps à travers le KOPS, les accélérations angulaires et la cinématique du membre inférieur dans le plan frontal. Cependant, la production de puissance de pédalage est un phénomène complexe qui est influencé par différents facteurs. Pour illustrer ce phénomène, certains profils sont extraits des annexes G, H, I, J, K et L, comme le montre la figure 5.21.

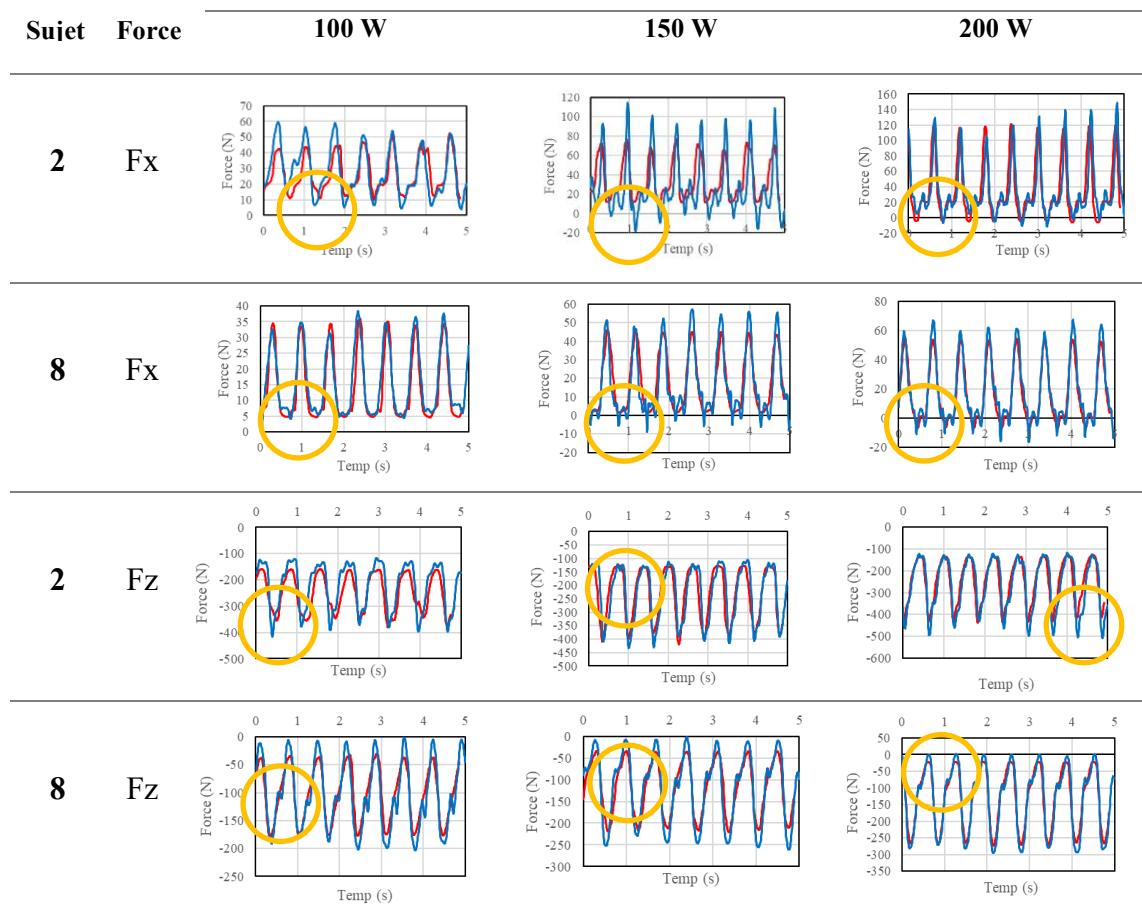


Figure 5.21 : Caractéristiques des profils de force F_x et F_z à différentes puissances chez les individus 2 et 8 (ligne bleue mesurée, ligne rouge prédictée)

Il est remarquable que la production de force Fx et Fy soit différente inter-sujets et intra-sujets entre les puissances évaluées. Comme cela peut être constaté, des changements d'amplitude et certaines formes différentes (dans le cercle orange) sont observés. Toutes ces différences ont été constatées par certains auteurs en raison de la variabilité inhérente à la cinématique du pédalage, influencée par des facteurs tels que le temps de pratique, le sexe, l'âge, les caractéristiques corporelles et la puissance de sortie [148]. En outre, la cinétique de pédalage peut varier en fonction de la technique, comme le style de pédalage préféré, en mettant l'accent sur la propulsion pendant la phase de traction ou pendant la phase de poussée [47]. Il a également été discuté dans la littérature qu'à des niveaux de puissance élevés tels que 200 W, le cerveau a moins de temps pour coordonner les muscles afin de générer des forces productives [145]. Ceci est cohérent avec les études montrant une asymétrie réduite dans les fonctions des membres inférieurs à des niveaux de puissance élevés, où l'efficacité est plus élevée [149], par rapport à une faible puissance (100 W) [150]. Ce fait est corroboré par le tableau 5.1, qui indique les indices d'efficacité calculés à partir du système de pédales instrumentées.

Tableau 5.1 : Indice d'efficacité des individus évalués

Sujet	Puissance		
	100 W	150 W	200 W
1	25.1	31.8	35.5
2	12.8	14.5	21.5
3	28.1	31.3	40
4	23.6	28.7	32
5	28.9	40.2	41.3
6	35.6	42.6	44.9
7	23.9	30	33.6
8	26.8	34.9	44.1
9	24.7	33	39.8
10	15.7	21.7	24.1
11	32.3	37	43.4
Moyenne	25.22 ± 6.26	31.42 ± 7.61	36.38 ± 7.57

Comme il est évident, l'indice d'efficacité est faible (25.22 ± 6.26 , 31.42 ± 7.61) pour les faibles puissances (100 W, 150 W) et élevé (36.39 ± 7.57) à haut puissance (200 W), pour tous les individus évalués. Cela indique que la production de force à faible puissance est très inefficace, peut-être parce que les commandes du cerveau aux muscles ne sont pas organisées efficacement en raison des faibles exigences de l'exercice. Il est donc laborieux d'obtenir un modèle unique capable de reproduire chaque caractéristique à chaque puissance. À puissance élevée, des changements importants sont observés chez la plupart des sujets, favorisant probablement une force verticale plus importante sur les pédales dans la phase de poussée et une force dirigée vers le bas plus faible dans la phase de traction afin d'améliorer l'efficacité. En résumé, toutes ces différences signifient qu'il est nécessaire de disposer des prédicteurs supplémentaires pour que le réseau neuronal puisse reconnaître toutes les différences inter et intra-sujets à différentes puissances.

Au niveau de la prédiction intra-sujet, en regroupant toutes les trois puissances évaluées pour l'entraînement du réseau neuronal, il est obtenu de bons résultats, mais moins satisfaisants que dans le cas de la prédiction des forces séparément pour chaque puissance. Comme le montre la figure 5.12 (c, d), pour la force Fx , la moyenne R^2 est de 0.6 ± 0.18 et la moyenne de RMSE est de 20.06 ± 6.28 . Pour la force Fz , la valeur moyenne de R^2 est de 0.82 ± 0.08 et la valeur moyenne de RMSE est de 39.56 ± 15.05 . Ces résultats sont en accord avec ce qui a été discuté ci-dessus concernant la variabilité de la production de force due aux différents facteurs mentionnés. De plus, il est constaté que certains sujets obtiennent de moins bons résultats (sujets 4, 6 et 7), notamment en ce qui concerne la prédiction de la force Fx . Lorsque les courbes des séries temporelles de ces sujets sont

examinées, ils présentent des formes plus complexes avec des pics positifs et négatifs, ce qui peut être la cause de leurs faibles performances.

En ce qui concerne à la prédiction des forces à puissance variable, comme le montre la figure 5.12 (a,b), la performance est très faible et l'erreur très élevée. Seuls deux individus (2, 9) ont obtenu des valeurs modérément bonnes, similaires à celles obtenues avec les puissances fixes. Il est possible que ce mauvais résultat soit dû au fait que, lors de l'évaluation, les individus ont produit différentes puissances sur de courtes périodes, de sorte que, avec les données disponibles, il n'y a pas suffisamment d'informations pour prédire les forces variables exercées sur la pédale. Il est possible que l'expérience n'ait pas été correctement contrôlée et que les individus ont à leur convenance, ce qui n'a pas permis d'obtenir un ensemble de données adéquat pour l'entraînement du réseau neuronal. En outre, certains individus ont sélectionné les puissances maximales, ce qui peut perturber le schéma de pédalage normal des individus. Aussi, certains individus ont même choisi de sélectionner des pignons différents au cours de l'expérimentation à puissance variable afin d'obtenir de meilleures performances à des puissances plus élevées. La conclusion est que l'expérience de la puissance variable doit être mieux paramétrée et contrôlée et inclure d'autres prédicteurs.

Malgré ces limites, les résultats obtenus sont bons et nous permettent de comprendre comment la force de péodalage est générée dans le cyclisme et d'obtenir des réponses intéressantes au problème de recherche. De plus, en général, lorsqu'un cycliste est évalué, la puissance et la cadence sont choisies pendant le test et maintenues constantes pendant l'évaluation, et ces valeurs se situent dans les gammes évaluées dans l'étude 2. L'analyse des résultats présentés dans les figures 5.6, 5.8, 5.10 montre que le réseau neuronal peut prédire les forces F_x et F_z pour les trois valeurs de puissance en présentant de bonnes

valeurs de précision avec des faibles erreurs. La figure 5.22 présente les diagrammes de Bland-Altman qui comparent la qualité des prédictions à chaque puissance évaluée.

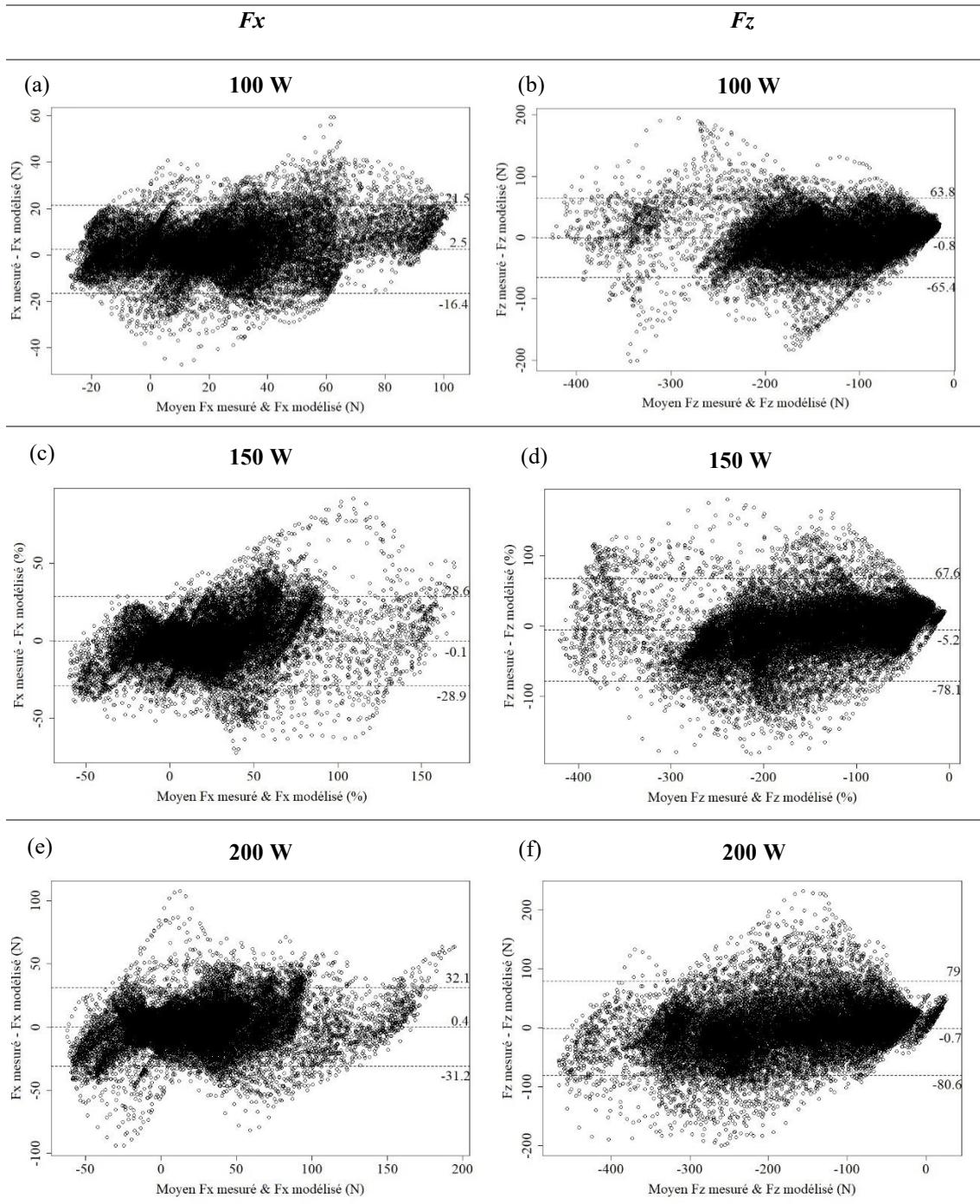


Figure 5.22 : Graphiques Bland-Altman des forces prédictes (a) 100 W F_x , (b) 100 W F_z , (c) 150 W F_x , (d) 150 W F_z , (e) 200 W F_x , (f) 200 W F_z

Premièrement, il est observé qu'il n'y a pas de biais significatif dans la prédiction pour toutes les conditions, c'est-à-dire que la ligne moyenne se situe autour du point 0. Malgré la variété des valeurs atteintes par les forces dans les deux sens, en raison de la dissemblance des individus évalués, aucune valeur aberrante fréquente n'est observée dans la prédiction et dans la plupart des cas les points sont situés sur la ligne centrale. En observant les valeurs limites supérieure et inférieure de concordance pour toutes les prédictions faites, il est possible de conclure que la prédiction faite avec le réseau neuronal développé est concordante.

Sur la base de ces résultats, une moyenne des coefficients de détermination R^2 et des RMSE est présentée dans le tableau 5.2.

Tableau 5.2 : Moyenne des coefficients, R^2 et RMSE des forces prédictives

Puissance									
	100 W		150 W		200 W		Variée		
Force	\bar{R}^2	\overline{RMSE}	\bar{R}^2	\overline{RMSE}	\bar{R}^2	\overline{RMSE}	\bar{R}^2	\overline{RMSE}	
F_x	0.73 ± 0.15	8.7 ± 2.6	0.74 ± 0.1	14.09 ± 4.14	0.79 ± 0.11	15.15 ± 4.64	0.41 ± 0.23	$30,78 \pm 18.99$	
F_z	0.73 ± 0.14	30.7 ± 10.1	0.83 ± 0.1	35.37 ± 12.48	0.86 ± 0.07	38.46 ± 13.4	$0,43 \pm 0.25$	$102,17 \pm 4,57$	

En analysant les résultats du tableau 5.2, ces moyennes sont pénalisées par certains des sujets testés, qui présentent des valeurs de performance plus faibles et des valeurs d'erreur plus élevées par rapport au reste des individus testés dans certaines valeurs des puissances évaluées. En général, les valeurs de performance des prédictions de force Fx et Fz sont similaires pour les puissances de 150 W et 200 W. Les performances sont plus faibles

pour une puissance de 100 W. De même, à puissance variable, les valeurs de performance obtenues sont faibles et l'erreur est de deux à trois fois supérieure à l'erreur correspondant aux trois conditions à puissance fixe. Ces valeurs de moyenne et d'écart-type sont également illustrées dans la figure 5.13. De légères différences dans l'écart-type et une valeur moyenne supérieure à 0,75 pour le coefficient de détermination R^2 pour les deux forces sont également corroborées. De même, en termes d'erreur RMSE, très peu de différences sont constatées, y compris en dessous de 20 pour F_x et de 40 pour F_z . Il est constaté que les meilleurs résultats sont obtenus pour les puissances de 150 W et 200 W et que les performances se détériorent pour les puissances variables.

En analysant les figures 5.6 (a,b), 5.8 (a,b) et 5.10 (a,b), un groupe d'individus (1,7,8,9,10,11) est identifié avec les valeurs les plus élevées de performance du réseau neuronal pour les deux forces et les trois puissances. Aucune caractéristique particulière ne les distingue des autres individus testés. Cela peut être dû à des caractéristiques particulières de l'interface pied-pédale, à des aspects techniques du matériel ou même à des facteurs physiologiques. Il n'est donc pas possible de formuler une hypothèse à partir de ce fait particulier. L'indice d'efficacité a également été évalué et s'est avéré être dans de bonnes valeurs pour la majorité des cyclistes évalués ($IE \geq 35$), c'est-à-dire qu'il se situe dans la fourchette normale des cyclistes récréatifs $39.2 \pm 6.2\%$ [3], ce qui suggère une bonne technique de pédalage. Cependant, les valeurs faibles d' IE ($IE \leq 30$) pour les individus 2, 3 et 10, où, dans certains cas, les valeurs de performance du réseau neuronal sont faibles, mais cela n'est pas cohérent pour toutes les sélections de puissance et de force.

En analysant les courbes de prédiction de la force F_x et F_z pour chaque sujet dans les annexes G, H, I, J, K et L, il est observé différents profils des forces évaluées. En général, le réseau neuronal est capable de reproduire les profils de chaque individu évalué. Il y a

quelques différences dans certaines formes qui peuvent conduire à une augmentation du RMSE. Il faut considérer que ces courbes de force n'ont pas été lissées, de sorte que le réseau neuronal tente de reproduire ces formes complexes, en général de manière adéquate. Il est aussi observé que le réseau neuronal tend à reproduire plus facilement le profil F_z , ce qui peut s'expliquer par le fait qu'il tend à être plus simple par rapport au profil F_x . Ceci est corroboré par les valeurs plus élevées du coefficient de détermination R^2 et l'erreur RMSE proportionnellement moins élevée, en tenant compte des valeurs de résistance atteintes. Le profil F_x présente chez certains sujets des zones négatives et positives ainsi que des formes plus complexes, en particulier chez les sujets qui n'obtiennent pas d'excellente F_x . Le réseau neuronal est également capable de reproduire la période de tous les profils de force, aucun déplacement n'étant observé tout au long de la prédiction de la série temporelle continue. Les graphiques de corrélation montrent qu'en général les données prédites et mesurées sont corrélées et qu'en moyenne le coefficient de corrélation r est égal à 0.85. Cela signifie que les données mesurées et prédites au moyen du réseau neuronal, sont fortement corrélées ($r > 0.85$) pour toutes les puissances évaluées et pour les deux forces exercées sur les pédales. Les résultats montrent que la plupart des données sont proches de la ligne de corrélation. Les valeurs de corrélation les plus élevées sont obtenues à 150 W et 200 W, ce qui peut s'expliquer par le fait que les valeurs d'efficacité du pédalage sont atteintes à mesure que la puissance augmente.

À partir de l'observation des données évaluées, le développement des forces de pédalage présente des caractéristiques spécifiques pour chaque sujet, c'est-à-dire que chaque sujet a une façon unique de générer ces forces. Un échantillon de 11 sujets est

obtenu, ce qui est considéré comme relativement modeste mais donne une idée de la grande variété des profils des forces générées pendant le pédalage.

Finalement, en ce qui concerne la validation croisée, les graphiques 5.7, 5.9 et 5.11 montrent que les modèles développés sont généralisables et robustes parce qu'ils peuvent maintenir le coefficient de détermination R^2 et le RMSE dans les différentes sélections de données appliquées. Le tableau 5.3 présente les valeurs moyennes de R^2 et RMSE obtenues lors de la validation croisée. Comme cela est évident, les valeurs moyennes présentent des différences et des écarts-types minimes. Ces résultats sont également représentés graphiquement dans la figure 5.14. Il est remarquable que le modèle demeure robuste puisqu'il maintient les R^2 et les erreurs RMSE dans de bonnes valeurs moyennes et avec des différences minimes dans l'écart-type, ce qui indique que le modèle est généralisable. Les meilleurs résultats sont également observés pour les puissances de 150 W et 200 W.

Tableau 5.3 : Moyenne R^2 et RMSE validation croisée des forces prédites

	100 W		150 W		200 W	
Force	R^2	RMSE	R^2	RMSE	R^2	RMSE
F_x	0.84 ± 0.06	12.59 ± 2.85	0.77 ± 0.02	$13.14 \pm 0,7$	0.79 ± 0.02	$14.97 \pm 0,65$
F_z	0.88 ± 0.05	22.66 ± 4.57	0.85 ± 0.02	33.5 ± 4.24	0.86 ± 0.02	39.71 ± 2

À ma connaissance, il existe jusqu'à présent deux articles qui prédisent de manière similaire la cinétique des membres inférieurs. Cependant, l'article proposé par Ahmadi et al. [145] prédit la force radiale et la force latérale, de sorte qu'il n'est pas possible de faire une comparaison objective. Cependant, ils utilisent également un réseau neuronal « feed

forward » avec 5 prédicteurs, partageant seulement deux prédicteurs - la cadence et la puissance - avec celui proposé dans ce chapitre, ce qui démontre la faisabilité de ce type de réseaux neuronaux. En revanche, dans l'article de Cordero-Sánchez, Bini et Serrancoli [151], les auteurs implémentent un protocole avec des plages de cadence variées, ce qui rend difficile la comparaison des données obtenues. Les auteurs mettent en œuvre un réseau neuronal récurrent avec des prédicteurs cinématiques 2D, quelques prédictions en relation avec le plan frontal et la masse du sujet. Cependant, en comparant les erreurs RMSE obtenues pour la force verticale, il est constaté que les valeurs obtenues dans le modèle développé ($38,46 \pm 13,4$ N) sont inférieures à celles rapportées dans la publication susmentionnée ($58,9 \pm 21,8$ N). En ce qui concerne la force antéropostérieure, de meilleures valeurs RMSE sont également obtenues pour le modèle proposé ($15,15 \pm 4,64$ N) par rapport à la valeur de l'article mentionné ($25,5 \pm 14,8$ N). Ces auteurs ne rapportent pas le coefficient de détermination R^2 mais le coefficient de Pearson (r). Ils rapportent un r pour la force verticale de $0,89 \pm 0,13$ et pour la force antéropostérieure de $0,77 \pm 0,12$. En analysant ceux obtenus au cours de cette thèse, sont obtenus $0,92 \pm 0,1$ pour la force verticale et $0,88 \pm 0,21$ pour la force antéropostérieure. Il est conclu que ceux obtenus sont légèrement supérieurs, ce qui confirme la bonne performance du modèle obtenu. En conclusion, les résultats sont comparables à ceux de la littérature, ce qui rend possible la mise en œuvre d'une technique d'apprentissage automatique différente avec de bonnes performances.

Conclusions : Le réseau neuronal « feed forward » composé de 5 couches cachées et de 2000 neurones pour chaque couche, peut prédire avec une bonne performance et une erreur acceptable les séries temporelles des forces de pédalage Fx et Fy . Le réseau de neurones utilise comme données d'entrée la puissance, la cadence, les angles et les vitesses

angulaires de la hanche, du genou et de la cheville dans le plan sagittal, ce qui permet de conclure que la cinématique du membre inférieur est fortement liée à la production des forces de pédalage. Ce réseau neuronal est dépendant de la puissance, c'est pourquoi trois réseaux neuronaux différents ont été développés pour chaque valeur de puissance évaluée.

5.6.2 Analyse des résultats prédition courbe enveloppe EMG RMS

Conformément aux résultats présentés à la figure 5.15, 5.16, 5.17, il est observé qu'en général, le réseau neuronal LSTM peut prédire avec une bonne précision la courbe linéaire de l'enveloppe EMG RMS des muscles grand fessier, biceps fémoral et droit fémoral pour les trois valeurs de puissance évaluées. Néanmoins, dans le cas du muscle gastrocnémien latéral, les résultats ne sont pas satisfaisants. Dans ce muscle, le RMSE est presque deux fois plus élevé avec les trois puissances testées et pour tous les sujets évalués. Bien que pour deux sujets (3,5) ce phénomène ne soit pas très négatif, il semble que la dynamique complexe de l'articulation de la cheville nécessite d'autres sources de données pour alimenter le réseau neuronal. Comme mentionné précédemment (voir section 2.1.5), la cheville fournit une grande variété de mouvements et contribue à la fois à la cinématique et à l'orientation du pied, car il s'agit d'un muscle bi-articulaire qui présente deux activations à deux moments différents. Les données supplémentaires pourraient être les couples et la puissance fournie par la cheville. Ce phénomène se produit à tous les niveaux de puissance évalués. En général, le réseau neuronal est capable d'atteindre des valeurs acceptables et similaires du coefficient de détermination R^2 pour tous les sujets testés. Le RMSE reste également à des niveaux acceptables. Il est observé que les meilleurs résultats sont obtenus dans les gammes de puissance de 100 W et 150 W. Dans la gamme de

puissance de 200 W, les résultats sont plus modestes pour R^2 et RMSE des muscles examinés. Pour certains sujets, comme le sujet 3 et pour le droit fémoral et le biceps femoris, les résultats ne sont pas satisfaisants. Ce fait est aussi accentué à une puissance de 200 W. En analysant la figure 5.18, qui montre les résultats obtenus pour une puissance variable, il est possible d'observer une faible performance et une erreur plus élevée par rapport aux prédictions pour des puissances fixes. Ce résultat est similaire à ce qui s'est produit dans le réseau neuronal qui prédit les forces de pédalage. Ces faibles performances peuvent être dues à la grande variabilité du signal EMG, un phénomène caractéristique du signal EMG, en plus de ce qui a été expliqué ci-dessus, sur la façon dont l'expérience de la puissance variable a été réalisée. Par conséquent, cette baisse de performance peut être due au manque de données provenant d'autres prédicteurs pour obtenir les caractéristiques du signal EMG des muscles lorsque la puissance varie. Comme suggéré précédemment, un meilleur contrôle de l'expérimentation doit être réalisé et aussi une plus grande quantité de données en condition variable du signal étudié.

L'analyse des courbes EMG RMS qui sont présentés dans les Annexes M, N et O, présente plusieurs particularités pour analyser. En général, il est observé que le réseau neuronal LSTM peut prédire la période d'activation de la plupart des muscles, puissances évaluées, et sujets évalués. En général, le profil de courbe qui reproduit le mieux le réseau neuronal LSTM est le muscle grand fessier. Ce muscle a une caractéristique d'être mono-articulaire, c'est-à-dire qu'il n'a qu'une seule période d'activation (0° - 90°), et sa fonction principale est l'extension de la hanche. Le réseau neuronal prédit donc mieux un profil de signal plus simple. Cependant, cette fonction n'est pas très cohérente pour les sujets évalués. Le profil d'activité électromyographique global du muscle biceps fémoral peut

également être reproduit satisfaisante pour la majorité de tous les individus et pour les trois puissances testées. En ce qui concerne le profil d'activité électromyographique du droit fémoral, une activité plus cohérente est observée au niveau de la double fonction du muscle. Dans ce cas, le réseau neuronal reproduit approximativement cette double fonction. Bien que le gastrocnémien latéral ne reproduise pas les profils de recrutement musculaire de manière cohérente pour tous les sujets et toutes les puissances testées, les profils prédits sont proches des profils mesurés. Un aspect intéressant à évaluer est la variabilité de l'activité musculaire au cours de la période d'évaluation. Bien que ce document ne montre qu'un segment de 5 s, cette variabilité était évidente pendant toute la durée de la période d'enregistrement. Cette variabilité a été étudiée par plusieurs auteurs et est également liée à la charge imposée au cycliste [28, 37]. Cette variabilité est l'un des aspects du réseau neuronal les plus difficiles à reproduire. Dans certains cas, le réseau neuronal s'adapte aux changements et dans d'autres cas, il n'y arrive pas de manière satisfaisante. En raison de cette variabilité, il peut donc s'avérer complexe d'entraîner le réseau neuronal avec l'ensemble des données. Des tentatives ont été faites dans ce sens, mais aucun résultat satisfaisant n'a été obtenu. Il est possible qu'afin d'entraîner un réseau neuronal avec plus d'informations pour reconnaître ces caractéristiques de variabilité du signal électromyographique. Les variables d'entrée utilisées pour entraîner le réseau neuronal LSTM étaient les variables cinématiques du membre inférieur, telles que les angles, les vitesses angulaires, les forces de pédalage (F_x , F_y et F_z), la puissance et la cadence pour un total de 11. En plus de ces variables, les valeurs de couples des articulations des membres inférieurs peuvent être des valeurs telles que les couples de chaque articulation peuvent être nécessaires pour fournir davantage d'informations au réseau neuronal. Les graphiques de corrélation montrent une corrélation généralement

bonne entre les données mesurées et les données prédictes, les données se regroupant autour d'une ligne centrale.

Enfin, en analysant les valeurs moyennes de R^2 présentées dans les figures 5.19 (a, b, c) pour les différentes gammes de puissance évaluées, celles-ci montrent en général une valeur similaire pour les muscles GF, BF, DF et différent pour le muscle GL. De même, selon la figure 5.20 (a, b, c) les RMSE moyennes présentent un comportement similaire en ce qui concerne les GL. Concernant la puissance variable, figure 5.19 (d), le R^2 montre des valeurs plus faibles et avec une augmentation de l'écart type, ce qui indique une grande variabilité. Une dégradation des performances est même observée au niveau du grand fessier et du biceps fémoral, qui sont les muscles les plus performants dans les autres conditions à puissance fixe. De plus, dans la figure 5.20 (d), des erreurs plus importantes sont observées dans tous les muscles par rapport aux puissances fixes. Le tableau 5.4 présente les valeurs de R^2 et RMSE.

Tableau 5.4 : Moyenne des coefficients R^2 et RMSE des muscles évalués

Muscle	100 W		150 W		200 W		Puissance variable	
	R ²	RMSE	R ²	RMSE	R ²	RMSE	R ²	RMSE
GF	0.78 ± 0,06	1.79 ± 1.32	0.79 ± 0.05	2.26 ± 2.51	0.61 ± 0.16	3.67 ± 2.81	0.41 ± 0.25	7.62 ± 3.22
BF	0.73 ± 0,07	2.12 ± 1.24	0.75 ± 0.08	2.56 ± 1.66	0.59 ± 0.18	4.18 ± 2.33	0.54 ± 0.15	5.23 ± 2.03
DF	0.64 ± 0.13	1.65 ± 0.79	0.67 ± 0.12	2.52 ± 1.58	0.43 ± 0.18	2.89 ± 2.81	0.46 ± 0.08	6.45 ± 4.33
GL	0.76 ± 0.08	10.7 ± 4.05	0.63 ± 0.02	8.24 ± 3.69	0.50 ± 0.18	6.51 ± 6.12	0.52 ± 0.11	11.54 ± 3.17

En analysant le tableau 5.4, certaines particularités sont corroborées, comme le fait qu'en moyenne les meilleures performances sont observées à 100 W avec tous les muscles

en valeurs très similaires et aussi avec des écarts-types semblables. De même, à 150 W, les performances moyennes sont similaires mais moins dans les muscles DF et GL. La performance moyenne la plus faible est atteinte à 200 W pour tous les muscles. Les erreurs de prédiction moyennes les plus élevées sont observées avec le muscle GL et à la puissance de 200 W, où les écarts types sont également les plus élevés. Les erreurs de prédiction moyennes et les écarts types mineurs sont atteints à 100 W. Les erreurs moyennes légèrement plus élevées sont à 150 W avec des écarts types similaires. Ces particularités permettent de corroborer le fait qu'à 100 W et 150 W, le réseau neuronal LSTM atteint de bonnes valeurs de performance et de faibles erreurs moyennes. Enfin, il est constaté que dans la condition de puissance variable, l'écart type du R^2 et RMSE augmente dans tous les muscles, indiquant que les données sont dispersées sur une large gamme de valeurs. Il est observé également, qu'en moyenne les valeurs atteintes dans les muscles GF, BF et DF sont similaires à la puissance fixe de 200 W, où le réseau neuronal ne peut pas prédire les activations musculaires avec une haute performance. Ceci est probablement dû à la grande variabilité du signal musculaire à une puissance légèrement supérieure à la puissance sous-maximale.

La prédiction du signal EMG RMS au moyen d'une technique d'apprentissage automatique utilisant un réseau neuronal récurrent tel que LSTM est réalisable. Bien que certains des résultats obtenus ne soient pas totalement satisfaisants, ils peuvent être améliorés en ajustant les paramètres, en augmentant les variables d'entrée ou en augmentant le nombre d'échantillons. Les résultats obtenus sont prometteurs et ouvrent la possibilité de prédire le signal électromyographique en cyclisme, un sujet sur lequel peu de recherches ont été publiées.

Certains aspects techniques sont essentiels à la précision de la prédiction et peuvent être étudiés en profondeur afin d'être améliorés. Dans cette optique, le système de capture du mouvement, les pédales instrumentées, le système d'électromyographie de surface et le cyclo-compteur sont des systèmes qui fonctionnent de manière indépendante. Une manœuvre a été implémentée pour minimiser l'erreur de synchronisation en demandant au sujet de rester immobile pendant 3 s. Cela a permis de synchroniser tous les signaux via Matlab à partir du signal du grand fessier. Cependant, ceci pourrait être amélioré au moyen d'un plug-in pour synchroniser tous les signaux.

L'analyse électromyographique est influencée par des phénomènes physiologiques tels que le retard neuro-mécanique. Le délai neuro-mécanique est la différence de temps entre la commande neuronale au muscle et la force générée. Ce retard pourrait être significatif sur la cinématique du membre inférieur. Les estimations de la littérature varient entre 20 ms, 40 ms, 50 ms et 100 ms [152], de sorte qu'il n'y a pas de consensus sur cette valeur. Cependant, ces délais n'ont pas été pris en compte dans la recherche, ce qui pourrait induire des erreurs dans le signal prédit. Certains auteurs suggèrent également qu'il est différent pour chaque muscle et avec l'augmentation de l'intensité [153]. Ces analyses n'ont pas été prises en compte afin de simplifier le projet.

Comme indiqué précédemment dans cette étude, la contraction volontaire maximale a été utilisée pour normaliser l'activité musculaire. La mise en œuvre de cette méthode a été justifiée par les recommandations du projet CEDE. Considérant que ce projet est basé sur l'estimation de l'activité de quatre muscles des membres inférieurs, et qu'il ne repose pas sur une comparaison interindividuelle, un suivi dans le temps de l'activité musculaire, et d'autres objectifs dans le cadre de l'analyse de l'activité physique, je considère que la méthode utilisée est appropriée. L'utilisation de l'une ou l'autre méthode n'a pas d'impact

négatif sur la technique d'apprentissage automatique mise en œuvre, puisque l'objectif initial est de trouver une solution au problème cible, de sorte que d'autres méthodes de normalisation peuvent être explorées, ce qui peut contribuer à améliorer les résultats trouvés.

Il a été démontré que les angles et la vitesse angulaire de la hanche, du genou et de la cheville dans le plan sagittal ainsi que les forces de pédalage F_x , F_y et F_z , la puissance et la cadence peuvent être liés à l'activité musculaire des muscles des membres inférieurs au moyen d'un modèle d'apprentissage automatique de type réseau neuronal récurrent LSTM.

Conclusion : La courbe EMG RMS de l'enveloppe du grand fessier, droit fémoral, biceps fémoral et du gastrocnémien latéral peut être prédite avec une bonne précision et une faible erreur à l'aide d'un réseau neuronal récurrent de type LSTM. Ce réseau neuronal donne de meilleures performances dans les gammes de puissance 100 W et 150 W.

5.6.3 Conclusion générale chapitre 5

Ce chapitre présente un modèle de réseau neuronal pour prédire la force verticale et la force antéro-postérieure appliquée sur la pédale. Il est aussi présenté un modèle de réseau neuronal récurrent qui prédit l'activité de 4 muscles des membres inférieurs pendant le pédalage. Les deux modèles utilisent la cinématique du membre inférieur dans le plan sagittal comme prédicteur, ce qui permet de valoriser les informations fournies par les systèmes d'analyse du mouvement actuellement utilisés. Cela permettra également d'obtenir des informations supplémentaires dans une analyse biomécanique puisque ces modèles peuvent être intégrés aux systèmes actuellement utilisés dans le cyclisme et contribuera indirectement à réduire les coûts de l'infrastructure nécessaire.

CHAPITRE 6 – Conclusions, limitations et recommandations

6.1 Conclusions générales

Comme le montrent les résultats et les analyses effectuées, l'apprentissage automatique offre un large éventail de possibilités pour prédire les variables qui évaluent la performance biomécanique en cyclisme avec un haut niveau de précision. Il est également possible de développer des modèles qui permettent de connaître les variables les plus importantes dans le phénomène de la production de force. Cela permet aux cyclistes et aux entraîneurs de contribuer à l'amélioration de la technique de pédalage.

La sélection des prédicteurs est le point le plus critique de cette recherche, non seulement du point de vue biomécanique, mais aussi du point de vue de l'apprentissage automatique, car l'algorithme est capable de prédire correctement la variable étudiée si les prédicteurs correctes alimentent le modèle.

Il a également été constaté que, bien qu'il existe des modèles complexes permettant d'obtenir une grande précision dans les variables prédites, une technique simple telle que la régression linéaire multiple peut répondre de manière adéquate à ce besoin.

Bien que les modèles développés ne disposent pas d'une grande quantité de données, les techniques de validation garantissent que les modèles sont généralisables.

La plupart des modèles développés ont été mis en œuvre avec le logiciel RStudio, qui est un logiciel libre, offrant une grande variété de fonctions et de librairies qui permettent la mise en œuvre de toutes sortes d'algorithmes d'apprentissage automatique. Cela permet une diffusion pratique des modèles développés, ce qui facilite l'accès aux personnes intéressées par la recherche sur le cyclisme. Aussi, bien que ce projet soit axé sur le

cyclisme, ces techniques peuvent être appliquées à d'autres sports dans le même but de comprendre la production de force au cours de n'importe quel mouvement humain.

6.2 Limitations

Il est bien connu que toutes ces méthodologies d'apprentissage automatique nécessitent une bonne quantité de données. Par conséquent, afin d'améliorer les performances des modèles, il est suggéré d'augmenter le nombre de données et à des niveaux plus variés.

Dans l'étude 1, plusieurs méthodes d'apprentissage automatique ont été mises en œuvre. Cependant, il existe plusieurs possibilités d'apprentissage automatique qui peuvent être explorées pour déterminer si il est possible d'améliorer les résultats obtenus.

Dans l'étude 2, des limites ont été trouvées dans la prédiction de la force dans la direction Fy . En d'autres termes, il est orienté vers l'analyse 2D ou bidimensionnelle. Cela s'explique par le fait que toutes les analyses ont été effectuées dans le plan sagittal et qu'il n'y a pas d'informations cinématiques dans le plan frontal. Lors de l'analyse du mouvement dans le plan frontal, ce sont des données cinématiques du membre inférieur qui présentent des caractéristiques très différentes de celles du plan X-Z sagittal ou latéral. Il est considéré que l'information cinématique pourrait être dérivée à d'autres mesures plus appropriées telles que les déplacements linéaires ou les accélérations par l'intermédiaire des centrales inertielles. De cette manière, les analyses développées avec des méthodologies 3D pourraient être intégrées.

Comme pour l'étude 1, l'étude 2 porte sur un petit échantillon de 11 personnes dans le contexte de la prédiction de la force de pédalage, ce qui permet d'obtenir une première approximation et des connaissances afin d'analyser les différents profils de force. Dans l'étude 2, seuls 4 muscles ont été testés, ce qui permet de trouver quelques conclusions sur

la possibilité de prédire l'activité musculaire des muscles des membres inférieurs du cycliste.

Comme autres limites de la présente étude, il y a les aspects qui n'ont pas été pris en compte dans la modélisation musculaire concernant les structures élastiques que le système musculaire combine. Cela signifie que ces effets pourraient être considérés comme non linéaires. C'est peut-être l'un des facteurs qui rendent difficile la reproduction de la forme exacte de la courbe d'enveloppe EMG RMS au moyen du réseau neuronal. Cet aspect devrait être analysé dans d'autres évaluations de ce type.

Bien que lors de la capture de l'activité musculaire, la même personne ait installé les capteurs d'électromyographie, les différences anatomiques entre un individu et un autre fait que le capteur n'est pas été placé au même endroit, d'où la variabilité du signal. Cette limitation peut être minimisée par des exécutants plus proches du domaine de la physiologie musculaire tels que des kinésiologues ou des physiothérapeutes.

Une autre limite de la présente recherche est qu'il n'y a pas d'évaluation en temps réel des prédicteurs, car toutes les informations sont extraites de différentes sources d'information qui doivent suivre des processus internes de calcul et de traitement des données pour présenter les données brutes. Ces processus sont cryptés par les entreprises qui produisent ces outils de collecte de données. Pour réaliser une étude en temps réel, il faudrait développer différentes technologies de capture cinématique, cinétique et physiologique. Une autre solution pourrait consister à négocier avec le fournisseur de technologie pour qu'il décode les informations capturées et que toutes les sources d'information soient disponibles en temps réel.

6.3 Perspectives et recommandations

- Parmi les perspectives, il y a l'intégration de ces modèles d'apprentissage automatique dans les systèmes de capture du mouvement et les logiciels associés à ces systèmes. Le but est essentiellement de permettre aux experts en positionnement cycliste d'obtenir des informations supplémentaires telles que des métriques qui évaluent la technique de pédalage. Actuellement, ces informations ne sont pas disponibles dans une analyse de la posture du cycliste et seraient d'une grande aide pour permettre une analyse plus complète de la position des cyclistes.
- Grâce aux modèles interprétables qui ont été développés au cours de cette recherche, les facteurs qui influencent la technique de pédalage sont mieux connus, ce qui permettra de proposer de nouveaux paramètres ou métriques liés à la cinématique permettant de qualifier la technique de pédalage d'un cycliste.
- L'apprentissage automatique et la science des données offrent une nouvelle perspective en matière d'analyse du cyclisme, car des techniques telles que l'apprentissage profond peuvent être utilisées pour déterminer les variables cinétiques et physiologiques, éliminant ainsi la nécessité de recourir à des technologies coûteuses.
- Il est recommandé d'augmenter le nombre de données pour l'étude des paramètres qui évaluent la technique de pédalage.
- Il est recommandé d'appliquer ces techniques d'apprentissage automatique à d'autres mesures utilisées pour évaluer la technique de pédalage, telles que la fluidité de la pédale « pedal smoothness (PS) », « leg smoothnes (LS) », taille du

point mort « dead center size (DC) », « dead spot score (DSS) », entre autres mesures.

- Il est recommandé appliquer d'autres techniques pour prédire les paramètres de pédalage tels que machines à vecteurs de support et le « gradient boosting » en raison de la grande précision qu'il est possible d'atteindre et de sa robustesse.
- Il est également recommandé d'analyser la cinématique du membre inférieur dans le plan frontal afin de fournir des données dans cette direction pour prédire la force de pédalage Fy. L'analyse de cette force permettrait aux cyclistes d'améliorer leur technique de pédalage en observant son impact sur la production de puissance.
- Il est suggéré d'inclure la masse de l'individu, la position horizontale du corps à travers le KOPS, les accélérations angulaires et la cinématique du membre inférieur dans le plan frontal, afin d'améliorer les performances du réseau neuronal FFNN qui prédit les forces de pédalage.
- Il est également souhaitable d'implémenter un réseau neuronal récurrent tel que LSTM pour prédire les forces de pédalage afin d'évaluer s'il est possible d'améliorer les résultats obtenus.
- Il est suggéré d'étudier les autres muscles du membre inférieur avec le réseau neuronal récurrent LSTM, tels que le vaste latéral, le vaste médial, le semi-tendineux, le soléaire et le gastrocnémien médial.
- Il est suggéré d'augmenter le nombre de sujets pour améliorer les résultats du réseau neuronal LSTM qui prédit le signal d'activité musculaire EMG RMS.
- Au cours de l'étude, le processus de recrutement des individus n'a pas permis d'inclure des cyclistes professionnels. Il est proposé de collecter des données

auprès de cette population, car cela permettrait de connaître les paramètres de haute performance. Cela aussi, permettrait également d'établir des comparaisons plus conformes avec celles publiées dans la littérature.

Bibliographie

- [1] N. Özkaya, M. Nordin, D. Goldsheyder, and D. Leger, *Fundamentals of biomechanics*. Springer, 2012.
- [2] D. A. Winter, *Biomechanics and motor control of human movement*. John wiley & sons, 2009.
- [3] G. Millour, A. T. Velásquez, and F. Domingue, "A literature overview of modern biomechanical-based technologies for bike-fitting professionals and coaches," *International Journal of Sports Science & Coaching*, vol. 18, no. 1, pp. 292-303, 2023.
- [4] D. Bishop, "An applied research model for the sport sciences," *Sports medicine*, vol. 38, pp. 253-263, 2008.
- [5] J. García-López, S. Díez-Leal, A. Ogueta-Alday, J. Larrazabal, and J. A. Rodríguez-Marroyo, "Differences in pedalling technique between road cyclists of different competitive levels," *Journal of Sports Sciences*, vol. 34, no. 17, pp. 1619-1626, 2016/09/01 2016, doi: 10.1080/02640414.2015.1127987.
- [6] M. Rossato, R. Bini, F. Carpes, F. Diefenthäler, and A. Moro, "Cadence and workload effects on pedaling technique of well-trained cyclists," *International journal of sports medicine*, vol. 29, no. 09, pp. 746-752, 2008.
- [7] S. Leirdal and G. Ettema, "Pedaling technique and energy cost in cycling," *Medicine and science in sports and exercise*, vol. 43, no. 4, pp. 701-705, 2011.
- [8] B. Fonda and N. Sarabon, "Biomechanics of cycling," *Sport Science Review*, vol. 19, no. 1-2, p. 187, 2010.
- [9] J. Swart and W. Holliday, "Cycling Biomechanics Optimization—the (R) Evolution of Bicycle Fitting," *Current Sports Medicine Reports*, vol. 18, no. 12, pp. 490-496, 2019, doi: 10.1249/jsr.0000000000000665.
- [10] R. R. Bini, F. Diefenthäler, and F. P. Carpes, "Determining force and power in cycling: A review of methods and instruments for pedal force and crank torque measurements," *International SportMed Journal*, vol. 15, no. 1, pp. 96-112, 2014.
- [11] J. G. Webster, *Bioinstrumentation*. John Wiley & Sons, 2003.
- [12] S. Dorel, J.-M. Drouet, A. Couturier, Y. Champoux, and F. Hug, "Changes of pedaling technique and muscle coordination during an exhaustive exercise," *Medicine & Science in Sports & Exercise*, vol. 41, no. 6, pp. 1277-1286, 2009.
- [13] D. A. Winter, *Biomechanics and motor control of human movement*, 4th ed. Hoboken, N.J.: Wiley, 2009.
- [14] S. Park and G. E. Caldwell, "Muscle synergies are modified with improved task performance in skill learning," *Human Movement Science*, vol. 83, p. 102946, 2022.
- [15] R. R. Bini, "The need for a link between bike fitting and injury risk," *Journal of Science and Cycling*, vol. 5, no. 1, pp. 1-2, 2016.
- [16] J. M. Plaza-Bravo, M. Mateo-March, R. Sanchis-Sanchis, P. Pérez-Soriano, M. Zabala, and A. Encarnación-Martínez, "Validity and reliability of the leomo motion-tracking device based on inertial measurement unit with an optoelectronic camera system for cycling pedaling evaluation," *International Journal of Environmental Research and Public Health*, vol. 19, no. 14, p. 8375, 2022.

- [17] R. Thompson, R. Rico Bini, C. Paton, and K. Hébert-Losier, "Validation of LEOMO inertial measurement unit sensors with marker-based three-dimensional motion capture during maximum sprinting in track cyclists," *Journal of Sports Sciences*, vol. 42, no. 2, pp. 179-188, 2024.
- [18] A. Jung, *Machine learning: the basics*. Springer Nature, 2022.
- [19] O. Van Nierop, A. Blankendaal, and C. Overbeeke, "The evolution of the bicycle: a dynamic systems approach," *Journal of Design History*, vol. 10, no. 3, pp. 253-267, 1997.
- [20] Luc. "Lawson Bicycle." <https://openclipart.org/detail/174232/lawson-bicycle> (accessed).
- [21] J. R. L. "Through a Continent on Wheels ... Illustrated, etc." <https://www.flickr.com/photos/britishlibrary/11163099414/in/album-72157640537276514/> (accessed).
- [22] G. Millour, A. T. Velásquez, and F. Domingue, "A literature overview of modern biomechanical-based technologies for bike-fitting professionals and coaches," *International Journal of Sports Science & Coaching*, p. 17479541221123960, 2022.
- [23] G. Guerra-Filho, "Optical Motion Capture: Theory and Implementation," *RITA*, vol. 12, no. 2, pp. 61-90, 2005.
- [24] G. Wu *et al.*, "ISB recommendation on definitions of joint coordinate system of various joints for the reporting of human joint motion—part I: ankle, hip, and spine," *Journal of biomechanics*, vol. 35, no. 4, pp. 543-548, 2002.
- [25] J. I. Priego Quesada, Z. Y. Kerr, W. M. Bertucci, and F. P. Carpes, "The association of bike fitting with injury, comfort, and pain during cycling: An international retrospective survey," *European journal of sport science*, vol. 19, no. 6, pp. 842-849, 2019.
- [26] K. de Vey Mestdagh, "Personal perspective: in search of an optimum cycling posture," *Applied ergonomics*, vol. 29, no. 5, pp. 325-334, 1998.
- [27] S. Dorel, J. Drouet, F. Hug, P. Lepretre, and Y. Champoux, "New instrumented pedals to quantify 2D forces at the shoe-pedal interface in ecological conditions: Preliminary study in elite track cyclists," *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 11, no. S1, pp. 89-90, 2008.
- [28] F. Hug and S. Dorel, "Electromyographic analysis of pedaling: A review," *Journal of Electromyography and Kinesiology*, vol. 19, no. 2, pp. 182-198, 2009/04/01/2009, doi: <https://doi.org/10.1016/j.jelekin.2007.10.010>.
- [29] H. J. Hermens *et al.* "Seniam Project, Surface Electromyography for the Non-Invasive Assessment of Muscles." <http://www.seniam.org/> (accessed).
- [30] M. Besomi *et al.*, "Consensus for experimental design in electromyography (CEDE) project: Amplitude normalization matrix," *Journal of Electromyography and Kinesiology*, vol. 53, p. 102438, 2020.
- [31] G. R. Naik, "Computational intelligence in electromyography analysis: a perspective on current applications and future challenges," 2012.
- [32] A. M. Hunter, A. S. C. Gibson, M. Lambert, and T. D. Noakes, "Electromyographic (EMG) normalization method for cycle fatigue protocols," *Medicine & Science in Sports & Exercise*, vol. 34, no. 5, pp. 857-861, 2002.

- [33] Y. Albertus-Kajee, R. Tucker, W. Derman, and M. Lambert, "Alternative methods of normalising EMG during cycling," *Journal of Electromyography and Kinesiology*, vol. 20, no. 6, pp. 1036-1043, 2010.
- [34] M. Halaki and K. Ginn, "Normalization of EMG Signals: To Normalize or," *Computational intelligence in electromyography analysis: A perspective on current applications and future challenges*, p. 175, 2012.
- [35] P. Konrad, "The abc of emg," *A practical introduction to kinesiological electromyography*, vol. 1, no. 2005, pp. 30-5, 2005.
- [36] F. Hug, J. M. Drouet, Y. Champoux, A. Couturier, and S. Dorel, "Interindividual variability of electromyographic patterns and pedal force profiles in trained cyclists," *European journal of applied physiology*, vol. 104, no. 4, pp. 667-678, 2008.
- [37] F. Hug, N. A. Turpin, A. Guével, and S. Dorel, "Is interindividual variability of EMG patterns in trained cyclists related to different muscle synergies?," *Journal of applied physiology*, vol. 108, no. 6, pp. 1727-1736, 2010.
- [38] O. M. Blake, Y. Champoux, and J. M. Wakeling, "Muscle coordination patterns for efficient cycling," *Medicine and science in sports and exercise*, vol. 44, no. 5, pp. 926-938, 2012.
- [39] G. Mornieux, B. Stapelfeldt, A. Gollhofer, and A. Belli, "Effects of pedal type and pull-up action during cycling," *International Journal of Sports Medicine*, pp. 817-822, 2008.
- [40] M. O. Ericson, å. Bratt, R. Nisell, U. P. Arborelius, and J. Ekholm, "Power output and work in different muscle groups during ergometer cycling," *European journal of applied physiology and occupational physiology*, vol. 55, pp. 229-235, 1986.
- [41] R. R. Bini, P. A. Hume, and A. E. Kilding, "Saddle height effects on pedal forces, joint mechanical work and kinematics of cyclists and triathletes," *European journal of sport science*, vol. 14, no. 1, pp. 44-52, 2014.
- [42] F. Diefenthäeler *et al.*, "Assessment of the effects of saddle position on cyclists pedaling technique," *Med Sci Sports Exerc*, vol. 38, no. 5, p. S181, 2006.
- [43] M. Menard, M. Domalain, A. Decatoire, and P. Lacouture, "Influence of saddle setback on pedalling technique effectiveness in cycling," *Sports Biomechanics*, vol. 15, no. 4, pp. 462-472, 2016.
- [44] R. R. Bini, P. A. Hume, F. J. Lanferdini, and M. A. Vaz, "Effects of body positions on the saddle on pedalling technique for cyclists and triathletes," *European journal of sport science*, vol. 14, no. sup1, pp. S413-S420, 2014.
- [45] V. Ferrer-Roca, V. Rivero-Palomo, A. Ogueta-Alday, J. A. Rodríguez-Marroyo, and J. García-López, "Acute effects of small changes in crank length on gross efficiency and pedalling technique during submaximal cycling," *Journal of Sports Sciences*, vol. 35, no. 14, pp. 1328-1335, 2017/07/18 2017, doi: 10.1080/02640414.2016.1215490.
- [46] N. Jongerius, B. Wainwright, J. Walker, and A. Bissas, "The biomechanics of maintaining effective force application across cycling positions," *Journal of Biomechanics*, vol. 138, p. 111103, 2022.
- [47] T. Korff, L. M. Romer, I. Mayhew, and J. C. Martin, "Effect of pedaling technique on mechanical effectiveness and efficiency in cyclists," *Medicine & Science in Sports & Exercise*, vol. 39, no. 6, pp. 991-995, 2007.

- [48] G. Mornieux, B. Stapelfeldt, A. Gollhofer, and A. Belli, "Effects of pedal type and pull-up action during cycling," *International journal of sports medicine*, vol. 29, no. 10, pp. 817-822, 2008.
- [49] F. Hintzy and A. Belli, "INFLUENCE OF TOE CLIPS ON MECHANICAL CHARACTERISTICS OF SPRINT CYCLING," in *ISBS-Conference Proceedings Archive*, 1998.
- [50] P. Burt, *Bike Fit 2nd Edition: Optimise Your Bike Position for High Performance and Injury Avoidance*. Bloomsbury Publishing, 2022.
- [51] J. C. Holmes, A. L. Pruitt, and N. J. Whalen, "Lower extremity overuse in bicycling," *Clinics in sports medicine*, vol. 13, no. 1, pp. 187-205, 1994.
- [52] K. S. Nordean-Snyder, "The effect of bicycle seat height variation upon oxygen consumption and lower limb kinematics," *Medicine and science in sports*, vol. 9, no. 2, pp. 113-117, 1977.
- [53] D. J. Sanderson and A. T. Amoroso, "The influence of seat height on the mechanical function of the triceps surae muscles during steady-rate cycling," *Journal of Electromyography and Kinesiology*, vol. 19, no. 6, pp. e465-e471, 2009.
- [54] R. Bini, P. Hume, J. L. Croft, and A. Kilding, "Pedal force effectiveness in cycling: A review of constraints and training effects," 2013.
- [55] S. Dorel, A. Couturier, and F. Hug, "Influence of different racing positions on mechanical and electromyographic patterns during pedalling," *Scandinavian journal of medicine & science in sports*, vol. 19, no. 1, pp. 44-54, 2009.
- [56] U. Emanuele, T. Horn, and J. Denoth, "Influence of racing position on cycling patterns," in *ISBS-Conference Proceedings Archive*, 2011.
- [57] M. Kamba, H. Naito, H. Ozaki, S. Machida, and S. Katamoto, "Effect of Gear Ratio and Cadence on Gross Efficiency and Pedal Force Effectiveness during Multistage Graded Cycling Test Using a Road Racing Bicycle," *Sports*, vol. 11, no. 1, p. 5, 2022.
- [58] Ø. Sandbakk, H.-C. Holmberg, S. Leirdal, and G. Ettema, "Metabolic rate and gross efficiency at high work rates in world class and national level sprint skiers," *European journal of applied physiology*, vol. 109, pp. 473-481, 2010.
- [59] S. Leirdal and G. Ettema, "The relationship between cadence, pedalling technique and gross efficiency in cycling," *European journal of applied physiology*, vol. 111, pp. 2885-2893, 2011.
- [60] C. T. Candotti, J. Ribeiro, D. P. Soares, A. R. De Oliveira, J. F. Loss, and A. C. S. Guimarães, "Effective force and economy of triathletes and cyclists," *Sports Biomechanics*, vol. 6, no. 1, pp. 31-43, 2007.
- [61] J. Theurel, M. Crepin, M. Foissac, and J. Temprado, "Effects of different pedalling techniques on muscle fatigue and mechanical efficiency during prolonged cycling," *Scandinavian Journal of medicine & science in sports*, vol. 22, no. 6, pp. 714-721, 2012.
- [62] K. Zameziati, G. Mornieux, D. Rouffet, and A. Belli, "Relationship between the increase of effectiveness indexes and the increase of muscular efficiency with cycling power," *European journal of applied physiology*, vol. 96, pp. 274-281, 2006.

- [63] A. Alamdari and V. N. Krovi, "A review of computational musculoskeletal analysis of human lower extremities," *Human modelling for bio-inspired robotics*, pp. 37-73, 2017.
- [64] R. Bartlett, *Introduction to sports biomechanics: Analysing human movement patterns*. Routledge, 2014.
- [65] M. Hull and M. Jorge, "A method for biomechanical analysis of bicycle pedalling," *Journal of biomechanics*, vol. 18, no. 9, pp. 631-644, 1985.
- [66] P. R. Barratt, J. C. Martin, S. J. Elmer, and T. Korff, "Effects of pedal speed and crank length on pedaling mechanics during submaximal cycling," *Medicine and Science in Sports and Exercise*, vol. 48, no. 4, p. 705, 2016.
- [67] D. Hahn, W. Herzog, and A. Schwirtz, "Interdependence of torque, joint angle, angular velocity and muscle action during human multi-joint leg extension," *European journal of applied physiology*, vol. 114, pp. 1691-1702, 2014.
- [68] W. Herzog, "The biomechanics of muscle contraction: optimizing sport performance," *Sport-Orthopädie-Sport-Traumatologie-Sports Orthopaedics and Traumatology*, vol. 25, no. 4, pp. 286-293, 2009.
- [69] R. Molavian, A. Fatahi, H. Abbasi, and D. Khezri, "Artificial intelligence approach in biomechanics of gait and sport: a systematic literature review," *Journal of Biomedical Physics & Engineering*, vol. 13, no. 5, p. 383, 2023.
- [70] R. Molavian, A. Fatahi, H. Abbasi, and D. Khezri, "Artificial Intelligence Approach in Biomechanics of Gait and Sport: A Systematic Literature Review," (in eng), *J Biomed Phys Eng*, vol. 13, no. 5, pp. 383-402, Oct 2023, doi: 10.31661/jbpe.v0i0.2305-1621.
- [71] R. Dwivedi *et al.*, "Explainable AI (XAI): Core ideas, techniques, and solutions," *ACM Computing Surveys*, vol. 55, no. 9, pp. 1-33, 2023.
- [72] C. Richter, M. O'Reilly, and E. Delahunt, "Machine learning in sports science: challenges and opportunities," *Sports Biomechanics*, pp. 1-7, 2021.
- [73] A. Hilmkil, O. Ivarsson, M. Johansson, D. Kylenstierna, and T. van Erp, "Towards machine learning on data from professional cyclists," *arXiv preprint arXiv:1808.00198*, 2018.
- [74] B. Hollaus, J. C. Volmer, and T. Fleischmann, "Cadence detection in road cycling using saddle tube motion and machine learning," *Sensors*, vol. 22, no. 16, p. 6140, 2022.
- [75] A. A. Gatti, P. J. Keir, M. D. Noseworthy, M. K. Beauchamp, and M. R. Maly, "Equations to prescribe bicycle saddle height based on desired joint kinematics and bicycle geometry," *European journal of sport science*, vol. 22, no. 3, pp. 344-353, 2022.
- [76] A. A. Gatti, P. J. Keir, M. D. Noseworthy, M. K. Beauchamp, and M. R. Maly, "Hip and ankle kinematics are the most important predictors of knee joint loading during bicycling," *Journal of Science and Medicine in Sport*, vol. 24, no. 1, pp. 98-104, 2021.
- [77] J. Cordero-Sánchez, R. R. Bini, and G. Serrancoli, "Validity of Recurrent Neural Networks to Predict Pedal Forces and Lower Limb Kinetics in Cycling," Available at SSRN 4921186.
- [78] R. R. Bini, V. B. Nascimento, and A. Nibali, "Validity of neural networks in determining lower limb kinematics in stationary cycling," *Sport Sciences for Health*, vol. 20, no. 1, pp. 127-136, 2024.

- [79] R. R. Bini, G. Serrancoli, P. R. P. Santiago, A. Pinto, and F. Moura, "Validity of neural networks to determine body position on the bicycle," *Research quarterly for exercise and sport*, vol. 94, no. 4, pp. 905-912, 2023.
- [80] A. Viswakumar, V. Rajagopalan, T. Ray, and C. Parimi, "Human gait analysis using OpenPose," in *2019 fifth international conference on image information processing (ICIIP)*, 2019: IEEE, pp. 310-314.
- [81] S. Sudin, A. N. A. Aziz, F. S. A. Saad, N. S. Khalid, and I. I. Ibrahim, "Cycling performance prediction based on cadence analysis by using multiple regression," in *Journal of Physics: Conference Series*, 2021, vol. 2107, no. 1: IOP Publishing, p. 012058.
- [82] C. M. Eckhardt *et al.*, "Unsupervised machine learning methods and emerging applications in healthcare," *Knee Surgery, Sports Traumatology, Arthroscopy*, vol. 31, no. 2, pp. 376-381, 2023.
- [83] T. Horvat and J. Job, "The use of machine learning in sport outcome prediction: A review," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 10, no. 5, p. e1380, 2020.
- [84] C. Janiesch, P. Zschech, and K. Heinrich, "Machine learning and deep learning," *Electronic Markets*, vol. 31, no. 3, pp. 685-695, 2021.
- [85] A. Shrestha and A. Mahmood, "Review of deep learning algorithms and architectures," *IEEE access*, vol. 7, pp. 53040-53065, 2019.
- [86] B. Mahesh, "Machine learning algorithms-a review," *International Journal of Science and Research (IJSR). [Internet]*, vol. 9, no. 1, pp. 381-386, 2020.
- [87] T. Miller, "Explanation in artificial intelligence: Insights from the social sciences," *Artificial intelligence*, vol. 267, pp. 1-38, 2019.
- [88] M. M. Maad, "Artificial neural networks advantages and disadvantages," *Bagdad College of Economic Science University*, 2018.
- [89] S. Ray, "A quick review of machine learning algorithms," in *2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon)*, 2019: IEEE, pp. 35-39.
- [90] J.-X. Mi, A.-D. Li, and L.-F. Zhou, "Review study of interpretation methods for future interpretable machine learning," *IEEE Access*, vol. 8, pp. 191969-191985, 2020.
- [91] J. W. Osborne and E. Waters, "Four assumptions of multiple regression that researchers should always test," *Practical assessment, research, and evaluation*, vol. 8, no. 1, p. 2, 2019.
- [92] M. A. Schroeder, J. Lander, and S. Levine-Silverman, "Diagnosing and dealing with multicollinearity," *Western journal of nursing research*, vol. 12, no. 2, pp. 175-187, 1990.
- [93] M. Hindman, "Building better models: Prediction, replication, and machine learning in the social sciences," *The Annals of the American Academy of Political and Social Science*, vol. 659, no. 1, pp. 48-62, 2015.
- [94] W. Y. Loh, "Fifty years of classification and regression trees," *International Statistical Review*, vol. 82, no. 3, pp. 329-348, 2014.
- [95] M. Sheykhmousa, M. Mahdianpari, H. Ghanbari, F. Mohammadimanesh, P. Ghamisi, and S. Homayouni, "Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review,"

- IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 6308-6325, 2020.
- [96] P. Probst and A.-L. Boulesteix, "To tune or not to tune the number of trees in random forest," *Journal of Machine Learning Research*, vol. 18, no. 181, pp. 1-18, 2018.
 - [97] S. Altelbany, "Evaluation of ridge, elastic net and lasso regression methods in precedence of multicollinearity problem: A simulation study," *Journal of Applied Economics and Business Studies*, vol. 5, no. 1, pp. 131-142, 2021.
 - [98] J. Ranstam and J. A. Cook, "LASSO regression," *Journal of British Surgery*, vol. 105, no. 10, pp. 1348-1348, 2018.
 - [99] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 55-67, 1970.
 - [100] J. S. Kidwell and L. H. Brown, "Ridge regression as a technique for analyzing models with multicollinearity," *Journal of Marriage and the Family*, pp. 287-299, 1982.
 - [101] J. C. M. Pires, F. G. Martins, S. Sousa, M. C. Alvim-Ferraz, and M. Pereira, "Selection and validation of parameters in multiple linear and principal component regressions," *Environmental Modelling & Software*, vol. 23, no. 1, pp. 50-55, 2008.
 - [102] B. Li, J. Morris, and E. B. Martin, "Model selection for partial least squares regression," *Chemometrics and Intelligent Laboratory Systems*, vol. 64, no. 1, pp. 79-89, 2002.
 - [103] A. Vellido, J. D. Martín-Guerrero, and P. J. Lisboa, "Making machine learning models interpretable," in *ESANN*, 2012, vol. 12: Citeseer, pp. 163-172.
 - [104] J. M. Menezes Jr and G. A. Barreto, "A new look at nonlinear time series prediction with NARX recurrent neural network," in *2006 Ninth Brazilian Symposium on Neural Networks (SBRN'06)*, 2006: IEEE, pp. 160-165.
 - [105] G. Van Houdt, C. Mosquera, and G. Nápoles, "A review on the long short-term memory model," *Artificial Intelligence Review*, vol. 53, no. 8, pp. 5929-5955, 2020.
 - [106] R. Zaheer and H. Shaziya, "A study of the optimization algorithms in deep learning," in *2019 third international conference on inventive systems and control (ICISC)*, 2019: IEEE, pp. 536-539.
 - [107] M. I. Lourakis, "A brief description of the Levenberg-Marquardt algorithm implemented by levmar," *Foundation of Research and Technology*, vol. 4, no. 1, pp. 1-6, 2005.
 - [108] R. Tamura, K. Kobayashi, Y. Takano, R. Miyashiro, K. Nakata, and T. Matsui, "Best subset selection for eliminating multicollinearity," *Journal of the Operations Research Society of Japan*, vol. 60, no. 3, pp. 321-336, 2017.
 - [109] J. M. Cadenas, M. C. Garrido, and R. Martínez, "Feature subset selection filter-wrapper based on low quality data," *Expert systems with applications*, vol. 40, no. 16, pp. 6241-6252, 2013.
 - [110] R. May, G. Dandy, and H. Maier, "Review of input variable selection methods for artificial neural networks," *Artificial neural networks-methodological advances and biomedical applications*, vol. 10, no. 1, pp. 19-45, 2011.
 - [111] T. Hastie, R. Tibshirani, and R. Tibshirani, "Best subset, forward stepwise or lasso? Analysis and recommendations based on extensive comparisons," 2020.

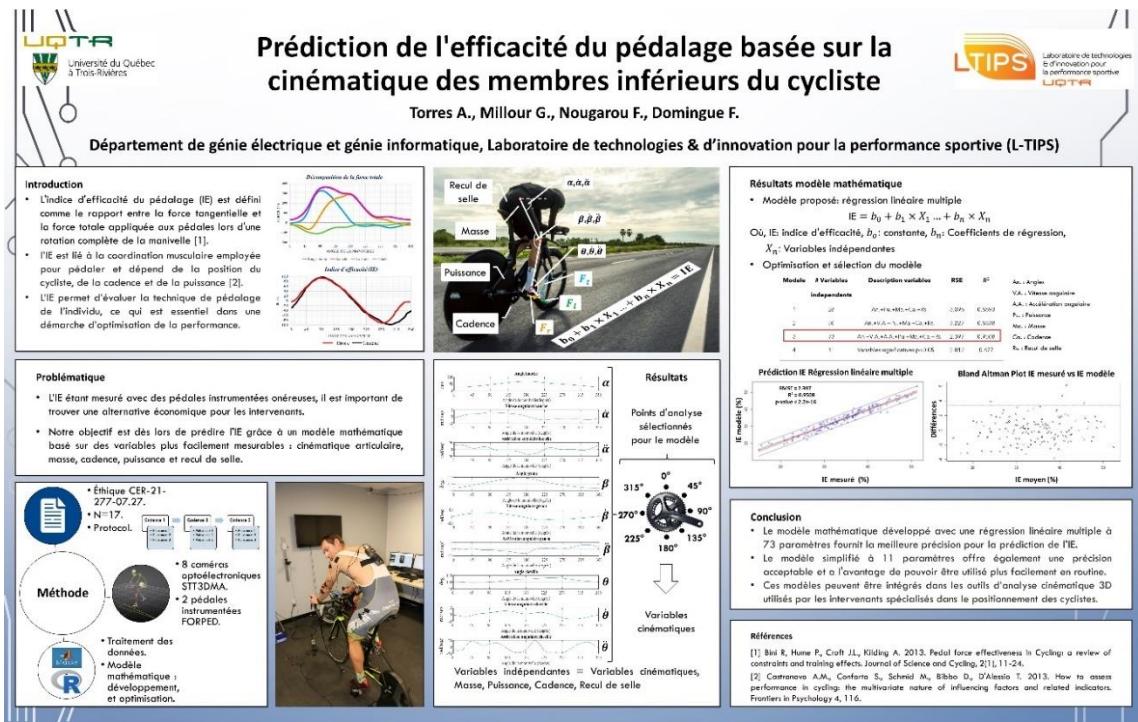
- [112] S. S. Gupta and H.-M. Yang, "Bayes-P* subset selection procedures for the best population," *Journal of statistical planning and inference*, vol. 12, pp. 213-233, 1985.
- [113] G. P. Brooks and P. Ruengvirayudh, "Best-subset selection criteria for multiple linear regression," *General Linear Model Journal*, 2016.
- [114] J. L. Castle, X. Qin, and W. R. Reed, "How to pick the best regression equation: A review and comparison of model selection algorithms," *University of Canterbury, Department of Economics and Finance Working Papers in Economics*, no. 09/13, 2009.
- [115] Y. Wang, L. Luangkesorn, and L. J. Shuman, "Best-subset selection procedure," in *Proceedings of the 2011 Winter Simulation Conference (WSC)*, 2011: IEEE, pp. 4310-4318.
- [116] L. Xu and W.-J. Zhang, "Comparison of different methods for variable selection," *Analytica Chimica Acta*, vol. 446, no. 1-2, pp. 475-481, 2001.
- [117] M. Karagiannopoulos, D. Anyfantis, S. Kotsiantis, and P. Pintelas, "Feature selection for regression problems," *Proceedings of the 8th Hellenic European Research on Computer Mathematics & its Applications, Athens, Greece*, vol. 2022, 2007.
- [118] J. Li *et al.*, "Feature selection: A data perspective," *ACM computing surveys (CSUR)*, vol. 50, no. 6, pp. 1-45, 2017.
- [119] D. Berrar, "Encyclopedia of Bioinformatics and Computational Biology. Cross-Validation," ed: Oxford: Academic Press, 2019.
- [120] H. Pham, "A new criterion for model selection," *Mathematics*, vol. 7, no. 12, p. 1215, 2019.
- [121] C. Lindsey and S. Sheather, "Variable selection in linear regression," *The Stata Journal*, vol. 10, no. 4, pp. 650-669, 2010.
- [122] É. Lebarbier and T. Mary-Huard, "Une introduction au critère BIC: fondements théoriques et interprétation," *Journal de la Société française de statistique*, vol. 147, no. 1, pp. 39-57, 2006.
- [123] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *Peerj computer science*, vol. 7, p. e623, 2021.
- [124] A. Gupta, T. S. Stead, and L. Ganti, "Determining a meaningful R-squared value in clinical medicine," *Academic Medicine & Surgery*, 2024.
- [125] B. Fonda, J. Babič, and N. Šarabon, "The Medial-Lateral Pedal Force Component Correlates with Q-Angle during Steady-State Cycling at Different Workloads and Cadences," *Applied Sciences*, vol. 11, no. 3, p. 1004, 2021.
- [126] Y. Fang, E. C. Fitzhugh, S. E. Crouter, J. K. Gardner, and S. Zhang, "Effects of workloads and cadences on frontal plane knee biomechanics in cycling," *Medicine & Science in Sports & Exercise*, vol. 48, no. 2, pp. 260-266, 2016.
- [127] G. Borg, "Psychological bases of physical exertion," *Med Sci Sports Exerc*, vol. 14, no. 5, pp. 377-81, 1982.
- [128] J. I. Priego Quesada, Z. Y. Kerr, W. M. Bertucci, and F. P. Carpes, "The categorization of amateur cyclists as research participants: findings from an observational study," *Journal of sports sciences*, vol. 36, no. 17, pp. 2018-2024, 2018.

- [129] L. Ansley and P. Cangley, "Determinants of "optimal" cadence during cycling," *European Journal of Sport Science*, vol. 9, no. 2, pp. 61-85, 2009.
- [130] P. J. M. Ali, R. H. Faraj, E. Koya, P. J. M. Ali, and R. H. Faraj, "Data normalization and standardization: a technical report," *Mach Learn Tech Rep*, vol. 1, no. 1, pp. 1-6, 2014.
- [131] T. L. Sczepanski, M. T. Gross, P. W. Duncan, and J. M. Chandler, "Effect of contraction type, angular velocity, and arc of motion on VMO: VL EMG ratio," *Journal of Orthopaedic & Sports Physical Therapy*, vol. 14, no. 6, pp. 256-262, 1991.
- [132] M. Hirashima, "Induced acceleration analysis of three-dimensional multi-joint movements and its application to sports movements," *Theoretical biomechanics*, pp. 303-318, 2011.
- [133] E. W. Faria, D. L. Parker, and I. E. Faria, "The science of cycling: factors affecting performance—part 2," *Sports medicine*, vol. 35, pp. 313-337, 2005.
- [134] R. D. De Veaux and L. H. Ungar, "Multicollinearity: A tale of two nonparametric regressions," in *Selecting models from data: artificial intelligence and statistics IV*: Springer, 1994, pp. 393-402.
- [135] E. Ostertagová, "Modelling using polynomial regression," *Procedia engineering*, vol. 48, pp. 500-506, 2012.
- [136] H. Han, X. Guo, and H. Yu, "Variable selection using mean decrease accuracy and mean decrease gini based on random forest," in *2016 7th ieee international conference on software engineering and service science (icsess)*, 2016: IEEE, pp. 219-224.
- [137] B. P. Lovatti, M. H. Nascimento, Á. C. Neto, E. V. Castro, and P. R. Filgueiras, "Use of Random forest in the identification of important variables," *Microchemical Journal*, vol. 145, pp. 1129-1134, 2019.
- [138] Z. Zhang, M. W. Beck, D. A. Winkler, B. Huang, W. Sibanda, and H. Goyal, "Opening the black box of neural networks: methods for interpreting neural network models in clinical applications," *Annals of translational medicine*, vol. 6, no. 11, 2018.
- [139] E. A. Hansen, B. R. Rønnestad, G. Vegge, and T. Raastad, "Cyclists' improvement of pedaling efficacy and performance after heavy strength training," *International Journal of Sports Physiology and Performance*, vol. 7, no. 4, pp. 313-321, 2012.
- [140] F. J. Lanferdini, T. C. Jacques, R. R. Bini, and M. A. Vaz, "Cyclists and triathletes do not differ in muscle volume, muscle recruitment or pedal force effectiveness," *Isokinetics and Exercise Science*, vol. 22, no. 4, pp. 279-286, 2014.
- [141] N. A. Turpin and B. Watier, "Cycling biomechanics and its relationship to performance," *Applied Sciences*, vol. 10, no. 12, p. 4112, 2020.
- [142] D. Achjari, "Partial least squares: another method of structural equation modeling analysis," *Jurnal Ekonomi dan Bisnis Indonesia*, vol. 19, no. 3, pp. 238-248, 2004.
- [143] M. L. Sawatsky, M. Clyde, and F. Meek, "Partial least squares regression in the social sciences," *The Quantitative Methods for Psychology*, vol. 11, no. 2, pp. 52-62, 2015.
- [144] H. B. Braiek and F. Khomh, "Machine Learning Robustness: A Primer," *arXiv preprint arXiv:2404.00897*, 2024.
- [145] R. Ahmadi *et al.*, "A Machine Learning Approach for Predicting Pedaling Force Profile in Cycling," *Sensors*, vol. 24, no. 19, p. 6440, 2024.

- [146] S. Afaq and S. Rao, "Significance of epochs on training a neural network," *Int. J. Sci. Technol. Res.*, vol. 9, no. 06, pp. 485-488, 2020.
- [147] B. Ding, H. Qian, and J. Zhou, "Activation functions and their characteristics in deep neural networks," in *2018 Chinese control and decision conference (CCDC)*, 2018: IEEE, pp. 1836-1841.
- [148] R. R. Bini and P. Hume, "A comparison of static and dynamic measures of lower limb joint angles in cycling: Application to bicycle fitting," *Human Movement*, vol. 17, no. 1, pp. 36-42, 2016.
- [149] A. Lucia *et al.*, "In professional road cyclists, low pedaling cadences are less efficient," *Medicine & Science in Sports & Exercise*, vol. 36, no. 6, pp. 1048-1054, 2004.
- [150] F. P. Carpes, C. B. Mota, and I. E. Faria, "On the bilateral asymmetry during running and cycling—A review considering leg preference," *Physical therapy in sport*, vol. 11, no. 4, pp. 136-142, 2010.
- [151] J. Cordero-Sánchez, R. Bini, and G. Serrancolí, "Validity of recurrent neural networks to predict pedal forces and lower limb kinetics in cycling," *Journal of biomechanics*, vol. 183, p. 112619, 2025.
- [152] P. R. Cavanagh and P. V. Komi, "Electromechanical delay in human skeletal muscle under concentric and eccentric contractions," *European journal of applied physiology and occupational physiology*, vol. 42, pp. 159-163, 1979.
- [153] R. Krätschmer, M. Stingl, D. Holzer, and F. K. Paternoster, "Neuromechanical Delay in the Quadriceps Muscle During Steady-Force Contractions at Different Intensities," Available at SSRN 4987657.

ANNEXE A - Contributions Affiche

Dans le cadre des activités académiques de l'UQTR, j'ai participé au concours de affiches 2022. Le nom de l'affiche était : « Prédiction de l'efficacité du pédalage à partir de la cinématique et de la cinétique du pédalage », cette affiche est liée à l'étude 1.



ANNEXE B - Autres Contributions

1. J'ai participé avec une présentation orale à l'événement « Expobiomedica 2022 », qui a eu lieu à Medellin, en Colombie, le 25 juillet 2022. Le titre de la présentation orale était: « Predicción de la eficiencia del pedaleo a partir de la cinemática de los miembros inferiores ». Cette présentation est liée à l'étude 1. Le nombre de participants était de 110 personnes.
2. J'ai participé avec une présentation orale à l'événement « Pedaleando ciencia » 2022, qui a eu lieu à Medellin, en Colombie, le 6 août 2022. Le titre de la présentation orale était: « Biomecánica, predicción de la eficiencia del pedaleo a partir de la cinemática de los miembros inferiores ». Cette présentation est liée à l'étude 1. Le nombre de participants était de 80 personnes.
3. J'ai participé en tant que conférencier au webinaire de la société Sensor Media sur la biomécanique appliquée au cyclisme. L'événement a eu lieu le 23 novembre 2022 et a été suivi par environ 30 personnes de différentes parties du monde. Le nom de la conférence était : « Analyse biomécanique appliquée au cyclisme, une perspective professionnelle et d'athlète ».
4. J'ai participé en tant que conférencier au webinaire de Suffolk University, Boston Massachusetts sur le thème : « Biomechanics Theory and Methodologies ». L'événement a eu lieu le 9 juin 2022 et a été suivi par 20 personnes.
5. J'ai participé à l'organisation du cours de formation de trois jours: Méthodes modernes pour le positionnement du cycliste, qui s'est déroulé au Québec et Colombie. Lien internet: www.uqtr.ca/fc.laboratoireltips.

6. J'ai organisé l'école d'été intitulée « La science des données au bénéfice du sport et de la santé » en Cali, Colombie, du 10 au 14 juillet 2023, avec une durée de 40 heures. Cet événement était organisé par les universités : Université du Québec à Trois-Rivières, Universidad Autonoma de Occidente et Escuela Nacional del Deporte. L'événement a été suivi par 25 personnes. Au cours de l'événement, un séminaire de recherche a été organisé sur le même sujet, au cours duquel une présentation orale du projet de recherche a été faite.
7. J'ai organisé une deuxième école d'été intitulée « La science des données au bénéfice du sport et de la santé » en Bogota, Colombie, du 29 juillet au 2 août 2024, avec une durée de 40 heures. Cet événement était organisé par les universités : Université du Québec à Trois-Rivières et Universidad Escuela Colombiana de Ingenieria Julio Garavito. L'événement a été suivi par 12 personnes. Au cours de l'événement, un séminaire de recherche a été organisé sur le même sujet, au cours duquel une présentation orale du projet de recherche a été faite.
8. Je participe en tant que conférencier au webinaire de Custom4.US : El arte de la fisiología y la Biomecanica, organisée par Jon Iberri, sur le thème : « Machine learning models to predict kinetic variables in cycling ». L'événement a eu lieu le 19 novembre 2024 et a été suivi par 15 personnes.
9. J'ai participé à l'événement organisé par l'UQTR en avril 2022, ma thèse en 180 secondes avec le travail intitulé : La science du pédalage.

ANNEXE C - Analyse de la dynamique inverse en pédalant

1) Concepts fondamentaux

L'analyse de la dynamique inverse en biomécanique est une méthodologie qui permet de déterminer les couples et les forces qui se produisent dans les articulations. Les données d'entrée dans le système sont les forces externes (ex. forces sur les pédales) qui se produisent dans un mouvement, c'est-à-dire la cinétique, ainsi que les positions, vitesses et accélérations de segments corporels, c'est-à-dire la cinématique. Des mesures anthropométriques sont également ajoutées au modèle (masse des segments, position du centre de gravité) en se basant sur des tables présentées dans la littérature.

La dynamique inverse est basé sur la dynamique classique qui utilise les principes et équations classiques de Newton-Euler. Ainsi, dans ce modèle, les segments corporels de la cuisse, de la jambe et du pied sont les objets mobiles et les articulations de la hanche, du genou et de la cheville sont les articulations autour desquelles les segments corporels tournent. Par conséquent, l'équation (1) est utilisée pour déterminer le moment de force dans une articulation, défini comme une mesure de la force qui peut faire tourner un objet autour d'un axe articulé.

$$M = F \times D \quad (1)$$

où, M est moment de force, F est la Force (N) et d est la distance perpendiculaire du centre de rotation à la force (m).

Aussi, en mouvement de rotation, un couple est nécessaire pour produire l'accélération angulaire d'un objet, alors l'équation générale est (2) :

$$\sum M_{articulation} = I_{segment} \times \alpha_{segment} \quad (2)$$

où, $\alpha_{segment}$ accélération angulaire du segment (rad/s^2), $I_{segment}$ moment d'inertie du segment ($\text{kg}\cdot\text{m}^2$).

De même, la deuxième loi de Newton doit être prise en compte, équation (3):

$$F = m \times a \quad (3)$$

où, F force résultante exercée sur un objet (N), m masse de l'objet (Kg), a accélération de l'objet (m/s^2).

En déterminant les forces de réaction sur le pied qui se produisent lors de l'application des forces sur les pédales, les données cinématiques du membre inférieur obtenues à partir du système de capture de mouvement et en utilisant les formules (1), (2) et (3), il est possible de trouver les couples à la cheville, au genou et à la hanche. Un script Matlab (MATLAB 2021a, MathWorks Inc, Natick, Massachusetts, États-Unis) a été développé pour calculer les couples et les puissances dans chaque articulation du membre inférieur, c'est-à-dire la hanche, le genou et la cheville.

De même, avec le même code Matlab, nous obtenons les variables cinématiques et cinétiques importantes dans l'analyse du mouvement utilisé dans les études proposées. Les variables cinématiques sont les angles, les vitesses angulaires et les accélérations de la hanche, du genou et de la cheville, en relation avec le plan sagittal. En matière de variables cinétiques sont la force efficace, la force radiale, la force medio latéral, la force total, l'indice d'efficacité (IE) et la proportion d'impulsions positives (PIP).

2) Définition des axes de coordonnées globales

L'orientation des axes de coordonnées globales est définie au moyen d'une procédure avec le logiciel STT3DMA dénommé « Set Ref. Frame », qui permet de définir le centre des

axes de coordonnées globales approximativement sous le cadre d'évaluation positionné à l'endroit où se trouve l'axe des manivelles. De cette façon, l'axe Y est défini comme vertical, l'axe X en direction horizontale aligné avec le cadre d'évaluation et l'axe Z transversal selon le trièdre de droite, figure 1.

Le signe des axes est défini comme positif selon la figure, c'est-à-dire en regardant le cadre d'évaluation, positif à droite sur l'axe X, positif vers le haut sur l'axe Y et positif en Z vers l'extérieur du cadre d'évaluation.



Figure 1 : Orientation des axes de coordonnées globales

3) Analyse de la dynamique inverse sur le pied

Pour réaliser la présente étude de la dynamique inverse le système STT3DMA sera utilisé pour collecter des données cinématiques. Ce système utilise un ensemble de 16 marqueurs placés sur différentes parties anatomiques du corps du cycliste à gauche et droite (8 de chaque côté) et trois sur le sol du laboratoire. Pour la cinétique de pédalage, le jeu de pédales instrumentées Forped sera utilisé.

- Définition des variables

Le tableau 1, ci-dessous définit les variables qui seront utilisés dans cette étude.

Tableau 1. Variables utilisées dans l'étude de dynamique inverse

Variable	Nom	Explication	Outils de mesure	Méthode requise	Unités
Couple	Couple	Une mesure de la force qui peut faire tourner un objet autour d'un axe	Système de capture de mouvement STT3DMA, Forped	Ecuation 1 générale du couple M $M = F \times d$ $F =$ Équation 2 générale du couple M : $\Sigma M = I \times \alpha$ I : est le moment d'inertie de l'objet en rotation	$N \times m$
I_{prox}	Moment d'inertie d'un segment du corps	Le moment d'inertie d'un objet est une valeur numérique qui peut être calculée pour tout corps rigide qui subit une rotation physique autour d'un axe fixe.	Données anthropométriques	$I_P = \sum_{i=1}^N m_i r_i^2$	$kg \times m^2$
α	Accélération angulaire d'un segment du corps	Le changement de vitesse angulaire par unité de temps	STT3DMA +Script Matlab	$\alpha = \frac{d\omega}{dt}$	$\frac{rad}{seg^2}$
m	Masse d'un segment du corps	Grandeur physique avec laquelle nous mesurons la quantité de matière qu'un corps contient	Données anthropométriques		kg
d_{cm}	Distance au centre de gravité	Distance du point de pivot au centre de gravité au pied	Données anthropométriques		mt
Θ, β, φ	Angle un segment d'un segment du corps	Angle du pied par rapport à l'horizontale	Données cinématiques STT3DMA		en degrés °
a	Accélération d'un segment du corps	Taux de changement de la vitesse d'un objet	STT3DMA +Script Matlab	$a = \frac{dv}{dt}$	$\frac{m}{seg^2}$

L	Longueur d'un segment du corps	Mesure anthropométrique du pied	Script Matlab	mt
F_{zp}	Forces d'entrée pédale	Force d'entrée dans la direction Z	Script Matlab +Forped	N
F_{xp}	Forces d'entrée pédale	Force d'entrée dans la direction Z	Script Matlab + Forped	N

- Analyse dans le plan sagittal du pied

Pour commencer l'analyse de la dynamique inverse dans le membre inférieur du cycliste est commencé avec le pied, la figure 2 montre le schéma d'un pied sur la pédale avec distances, forces et angles.

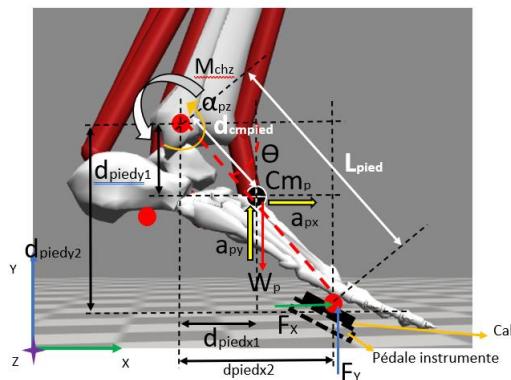


Figure 2 : Schéma de pied sur la pédale

Les points rouges représentent les trois marqueurs qui sont placés sur le pied, un sur la cheville ou la malléole, un autre sur la cinquième tête métatarsienne et un autre projetant une ligne verticale de la malléole au bord de la chaussure du cycliste, afin de former un triangle qui représente le pied dans le système STT3DMA.

Afin de simplifier les calculs qui seront présentés ci-dessous, le pied sera représenté par une ligne allant de la malléole de la cheville à la tête du cinquième métatarsien, sur cette même ligne sera déterminé le centre de masse du pied.

A partir des données obtenues de l'application Forped, un traitement des données est effectué avec un script Matlab obtenant la force dans la direction X Fx et la force dans la direction Y Fy, parallèles aux axes globaux de l'orientation X, Y.

- Accélération linéaire du pied

L'accélération linéaire dans la direction X et Y, au point où se trouve le centre de masse du pied. Cette accélération linéaire est obtenue à partir du logiciel STT3DMA et du script Matlab.

a_{px} : Accélération linéaire dans la direction X (m/seg²)

a_{py} : Accélération linéaire dans la direction Z (m/seg²)

- Accélération angulaire : accélération linéaire obtenue à partir du logiciel du script Matlab.

α_{pz} : Accélération angulaire du pied autour de l'axe z (rad/seg²)

- Distance au centre de masse

d_{x1} : distance dans la direction Z de la malléole au centre de gravité (m)

d_{y1} : distance dans la direction Y de la malléole au centre de gravité (m)

- Masse du pied : selon tables anthropométriques de Adolphe Quetelet.

m_p : masse du pied (kg) = masse du sujet \times 0,0145

- Moment d'inertie du pied : en général le moment d'inertie d'un objet est une valeur numérique qui peut être calculée pour tout corps rigide qui subit une rotation physique autour d'un axe fixe.

$$I = \sum_{i=1}^N m_i \times r_i^2 \quad (4)$$

Où m_i : est la masse de l'objet, r_i : distance de la particule de masse m_i à l'axe de rotation.

- Les moments d'inertie d'un corps sont différents dans les trois axes cartésiens, donc ce serait I_{xx} , I_{yy} et I_{zz} dans les axes X, Y et Z respectivement.
 - Distance au centre de gravité : distance du marqueur situé sur la cheville au centre de gravité du pied, selon les tables anthropométriques (à définir) :
- d_{cmp} : Distance de la cheville au centre de gravité du pied (m)
- Longueur du pied : Distance de la malléole au marqueur placé sur la cinquième tête métatarsienne (Figure 3).

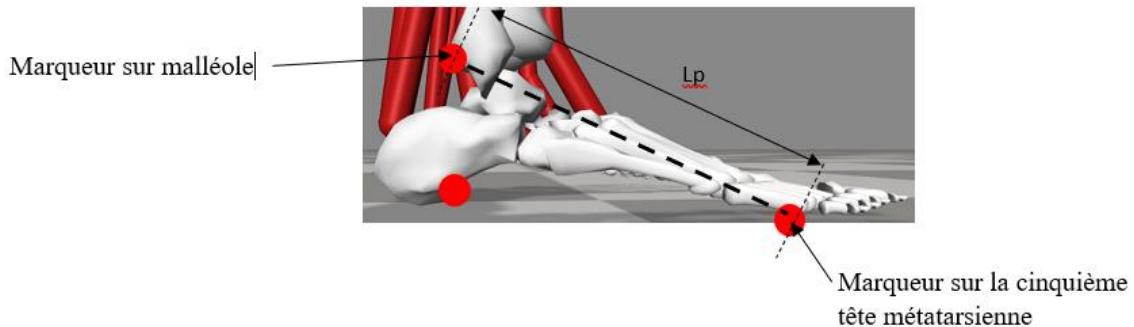


Figure 3 : Illustration longueur du pied

Connaissant la distance ab d_{ab} , la distance perpendiculaire d_y à Fx est :

$$d_{y2} = d_{ab} \times \cos \theta \quad (5)$$

Et la distance perpendiculaire d_z à Fy est :

$$d_z = d_{ab} \times \sin \theta \quad (6)$$

- Couple dans la cheville

En général, un couple est défini comme une mesure de la force qui peut faire tourner un objet autour d'un axe. Alors l'équation générale de couple M est :

$$M = F \times d \quad (7)$$

Où F : Force (N), d : distance perpendiculaire du centre de rotation à la force (m)

Aussi, dans le mouvement de rotation, un couple est nécessaire pour produire une accélération angulaire d'un objet, alors l'équation générale est:

$$\Sigma M = I \times \alpha \quad (8)$$

Où α : Accélération angulaire de l'objet (rad/s^2), I : moment d'inertie de l'objet ($\text{kg}\cdot\text{m}^2$)

Et selon la deuxième loi de Newton $F = m \times a$ (9)

Où m : masse d'objet (kg), a : accélération de l'objet (m/s^2)

Avec les équations, il est déterminé que:

$M_{Cheville}$: couple de cheville sur l'axe Z

$$\Sigma M_{Cheville} = I_{piedxx} \times \alpha_{pied} \quad (10)$$

$$M_{Cheville} + a_{piedx} \times m_{pied} \times d_{piedy1} + a_{piedy} \times m_{pied} \times d_{piedx1} - g \times m_{pied} \times d_{piedx1} + F_y \times d_{piedx2} + F_x \times d_{piedy2} = I_{piedxx} \times \alpha_{piedx}$$

$$M_{Cheville} = I_{piedxx} \times \alpha_{piedx} - a_{piedy} \times m_{pied} \times d_{cmpied} \times \cos \theta - a_{piedx} \times m_{pied} \times d_{cmpied} \times \sin \theta + g \times m_{pied} \times d_{cmpied} \times \cos \theta - F_y \times L_{pied} \times \cos \theta - F_x \times L_{pied} \times \sin \theta \quad (11)$$

- Détermination des forces de réaction dans la cheville

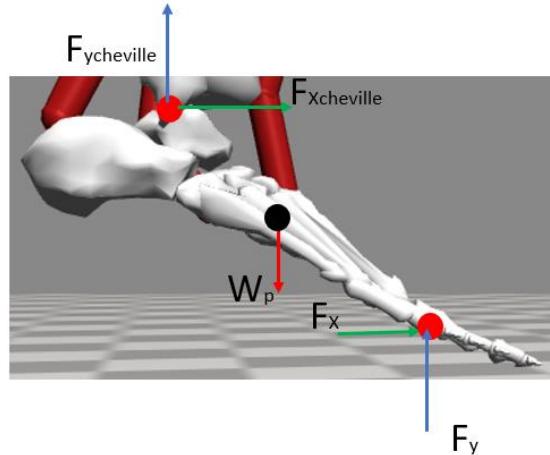


Figure 4 : Illustration de forces de réaction dans la cheville

$$\sum F_x = m_{pied} \times a_{piedx} \quad (12)$$

$$\sum F_y = m_{pied} \times a_{piedy} \quad (13)$$

Forces en axe X

$$F_x + F_{x\text{cheville}} = m_{pied} \times a_{piedx}$$

$$F_{x\text{cheville}} = m_{pied} \times a_{piedx} - F_x \quad (14)$$

Forces en axe Y

$$F_y + F_{y\text{cheville}} - m_{pied} \times g = m_{pied} \times a_{piedy}$$

$$F_{y\text{cheville}} = m_{pied} \times a_{piedy} + m_{pied} \times g - F_y \quad (15)$$

4) Analyse dynamique inverse dans le genou

En utilisant les mêmes équations 1,2,3 et les forces et les distances de la figure 5, le couple sur le genou est :

$$\sum M_{genou} = I_{tibiaxx} \times \alpha_{tibia} \quad (16)$$

$$\begin{aligned}
& M_{Genou} + M_{Cheville} + a_{tibiay} \times m_{tibia} \times d_{tibiay1} - a_{tibiay} \times m_{tibia} \times \\
& d_{tibiay1} + g \times m_{tibia} \times d_{tibiay1} - Fy_{cheville} \times d_{tibiay2} + Fx_{cheville} \times \\
& d_{tibiay2} = I_{tibiayx} \times \alpha_{Tibia} \\
M_{Genou} &= I_{Tibiayx} \times \alpha_{Tibia} - M_{Cheville} + Fy_{cheville} \times d_{x2} - Fx_{cheville} \times \\
d_{tibiay2} & \\
-g \times m_{tibia} \times d_{tibiay1} &+ a_{tibiay} \times m_{tibia} \times d_{tibiay1} - a_{tibiay} \times m_{tibia} \times \\
d_{tibiay1} & \tag{16}
\end{aligned}$$

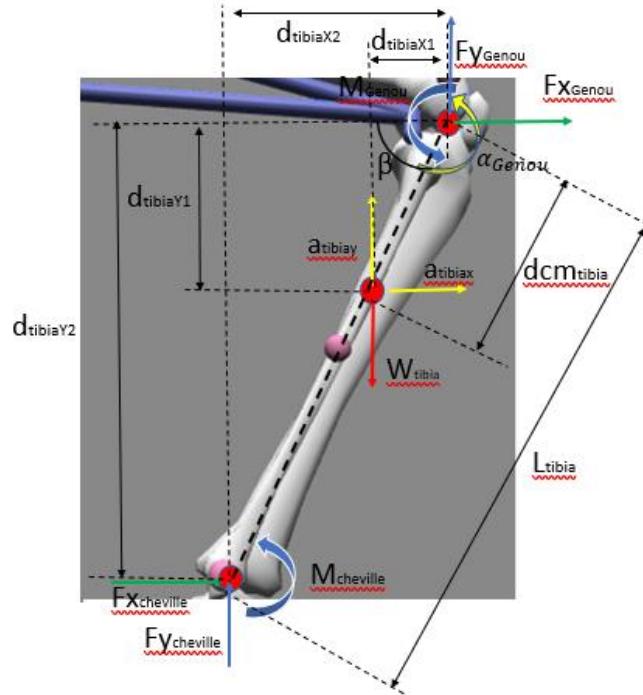


Figure 5 : Schéma du tibia

- Détermination des forces de réaction dans le genou

Équation générale des forces :

$$\sum F_x = m_{tibia} \times a_{tibiadx} \tag{17}$$

$$\sum F_y = m_{tibia} \times a_{tibiay} \tag{18}$$

Forces en axe X

$$F_{x_{cheville}} + F_{x_{genou}} = m_{tibia} \times a_{tibiay} \quad (19)$$

$$F_{x_{genou}} = m_{tibia} \times a_{tibiay} - F_{x_{cheville}} \quad (19)$$

Forces en axe Y

$$F_{y_{genou}} + F_{y_{cheville}} - m_{tibia} \times g = m_{tibia} \times a_{tibiay} \quad (20)$$

$$F_{y_{genou}} = m_{tibia} \times a_{tibiay} + m_{tibia} \times g - F_{y_{cheville}} \quad (20)$$

5) Analyse dynamique inverse dans la hanche

D'après la figure 6, les couples à la hanche sont obtenus à l'aide des mêmes formules

1,2,3 :

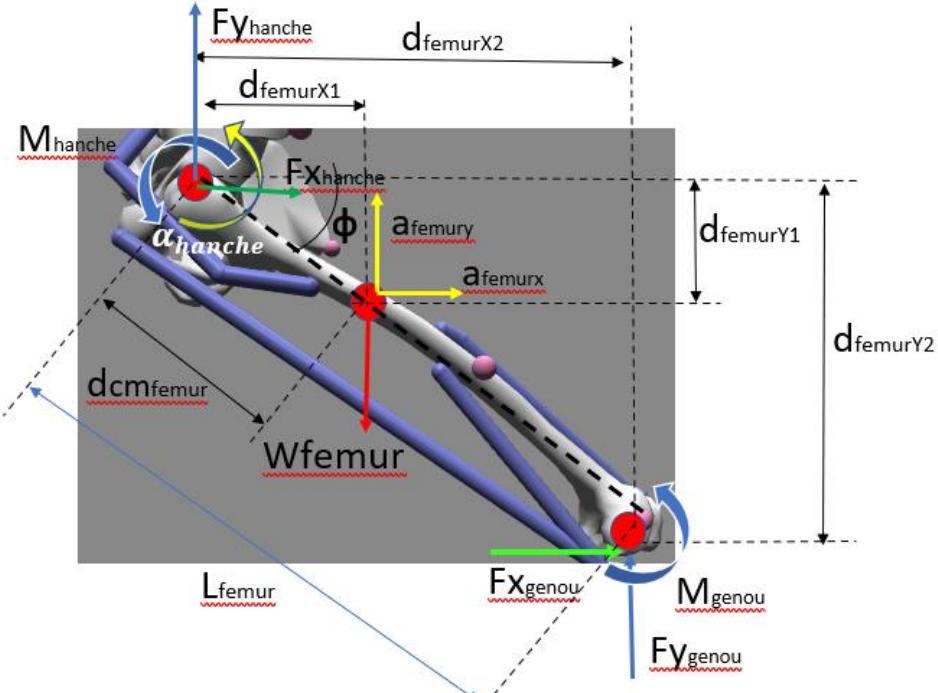


Figure 6 : Schéma du fémur

$$\sum M_{genou} = I_{femurxx} \times \alpha_{femur} \quad (21)$$

$$\begin{aligned}
& M_{Hanche} + M_{Genou} + a_{femurx} \times m_{femur} \times d_{femury1} + a_{femury} \times m_{femur} \times \\
& d_{femurx1} - g \times m_{femur} \times d_{femurx1} + Fy_{genou} \times d_{femurx2} + Fx_{genou} \times \\
& d_{femury2} = I_{femurxx} \times \alpha_{Femur} \\
& M_{Hanche} = I_{femurxx} \times \alpha_{Femur} - M_{Genou} - a_{femurx} \times m_{femur} \times d_{femury1} - \\
& a_{femury} \times m_{femur} \times d_{femurx1} + g \times m_{femur} \times d_{femurx1} - Fy_{genou} \times \\
& d_{femurx2} - Fx_{genou} \times d_{femury2} \tag{22}
\end{aligned}$$

6) Calcul de la puissance articulaire du membre inférieur

Pour calculer la puissance dans chaque joint, l'équation suivante est utilisée :

$$\text{Puissance} = M \times \omega \tag{23}$$

Où M : couple d'articulation, Nm, ω : Vitesse angulaire dans l'articulation, rad/seg, P : Nm/s , Watt

De cette manière, les équations suivantes sont définies :

$$\text{PuissanceCheville} = M_{Cheville} \times \text{Vitesse angulaire angle Tibia-Pied} \tag{24}$$

$$\text{PuissanceGenou} = M_{Genou} \times \text{Vitesse angulaire angle Fémur-Tibia} \tag{25}$$

$$\text{PuissanceHanche} = M_{Hanche} \times \text{Vitesse angulaire angle Fémur} \tag{26}$$

La puissance totale est définie comme :

$$\text{Puissance totale} = \text{PuissanceCheville} + \text{PuissanceGenou} + \text{PuissanceHanche} \tag{27}$$

Avec toutes les équations définies précédemment, les graphiques des couples de la hanche, du genou, de la cheville sont illustrés dans la Figure 7. Aussi, la puissance de la hanche, du genou et de la cheville sont illustrés dans la Figure 8. Ces graphiques représentent un cycliste qui produit une puissance de 250 Watts à une cadence de 90 RPM.

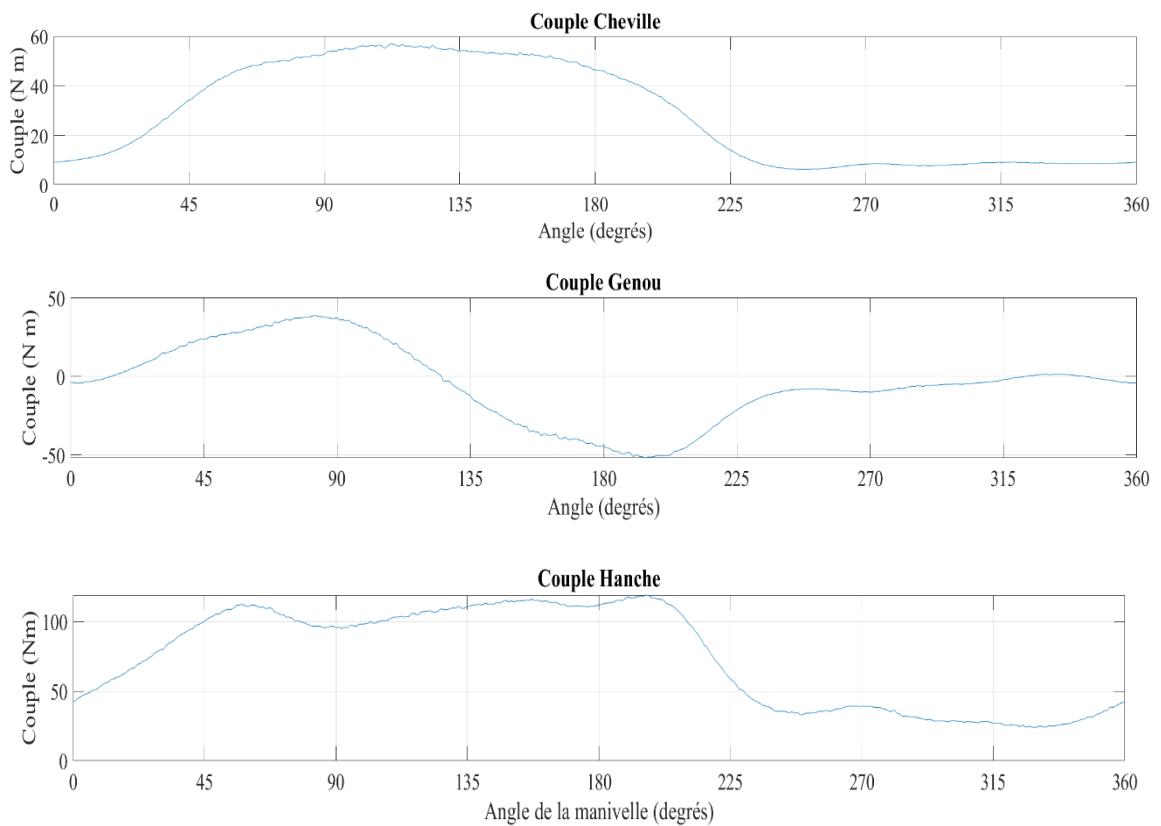


Figure 7 : Couples du membre inférieur

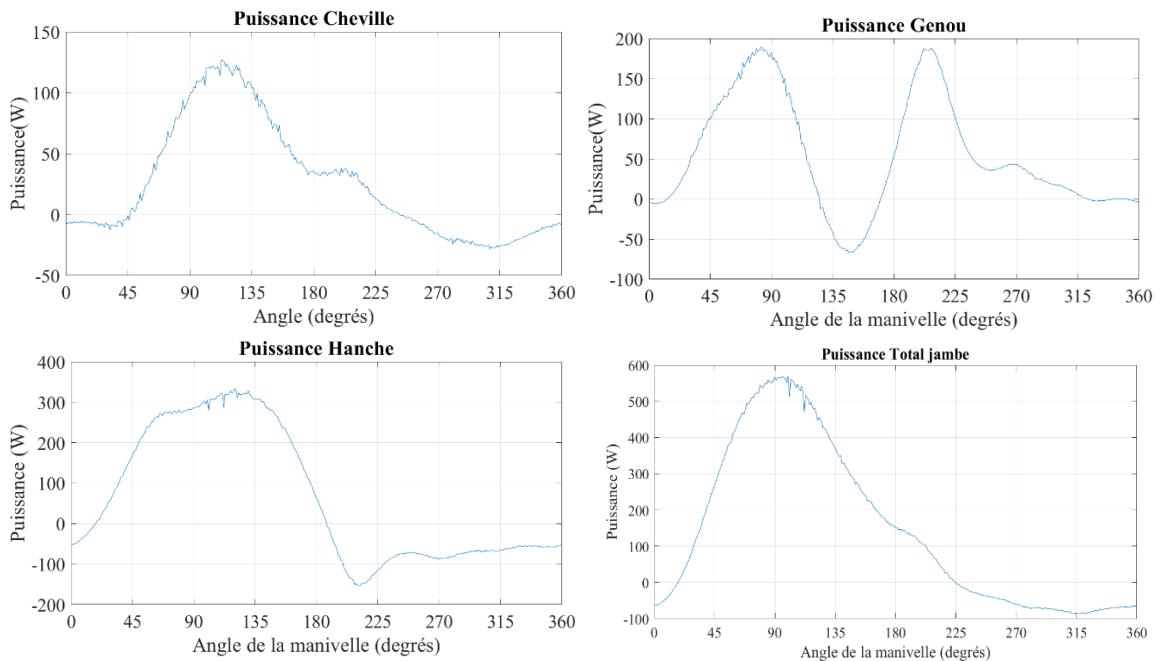


Figure 8 : Puissance des membres inférieurs

ANNEXE D - Équations cinématiques et cinétiques

Équations cinématiques		
<i>Variables</i>	<i>Équations</i>	<i>Description des termes de l'équation</i>
<i>Angle d'articulation</i>	$\alpha[n] = \tan^{-1} \frac{Y_a[n] - Y_b[n]}{X_b[n] - X_a[n]}$ (1)	n : Numéro de l'échantillon
	$\beta[n] = \alpha[n] + \tan^{-1} \frac{Y_b[n] - Y_c[n]}{X_b[n] - X_c[n]}$ (2)	Marqueur a : Hanche
	$\theta[n] = \tan^{-1} \frac{Y_b[n] - Y_c[n]}{X_b[n] - X_c[n]} + \tan^{-1} \frac{Y_c[n] - Y_d[n]}{X_d[n] - X_c[n]}$ (3)	Marqueur b : Genou Marqueur c : Cheville Marqueur d : Pied
<i>Vitesse angulaire de l'articulation</i>	$\dot{\alpha}[n] = \frac{\alpha[n] - \alpha[n-1]}{1/fs}$ (4)	$Y_a[n], Y_b[n], Y_c[n], Y_d[n]$: Coordonnées Y des marqueurs a, b, c, d
	$\dot{\beta}[n] = \frac{\beta[n] - \beta[n-1]}{1/fs}$ (5)	$X_a[n], X_b[n], X_c[n], Y_d[n]$: Coordonnées X des marqueurs a, b, c, d
	$\dot{\theta}[n] = \frac{\theta[n] - \theta[n-1]}{1/fs}$ (6)	$\alpha[n], \beta[n], \theta[n]$: Angle de la hanche, du genou et de la cheville
<i>Accélération angulaire de l'articulation</i>	$\ddot{\alpha}[n] = \frac{\dot{\alpha}[n] - \dot{\alpha}[n-1]}{1/fs}$ (7)	$\dot{\alpha}[n], \dot{\beta}[n], \dot{\theta}[n]$: Vitesse angulaire de la hanche, du genou et de la cheville
	$\ddot{\beta}[n] = \frac{\dot{\beta}[n] - \dot{\beta}[n-1]}{1/fs}$ (8)	$\ddot{\alpha}[n], \ddot{\beta}[n], \ddot{\theta}[n]$: Accélération angulaire de la hanche, du genou, et de la cheville
	$\ddot{\theta}[n] = \frac{\dot{\theta}[n] - \dot{\theta}[n-1]}{1/fs}$ (9)	fs : Fréquence d'échantillonnage

Kinetic equations		
<i>Total force</i>	$F_{tot} = \sqrt{(F_T^2 + F_R^2 + F_{ML}^2)}$ (10)	F_{tot} : total force (N) F_T : tangential force (N) F_R : radial force (N) F_{ML} : medio-lateral force (N)
<i>Index of effectiveness</i>	$IE = \frac{\int_0^{2\pi} F_T(\varphi) d\varphi}{\int_0^{2\pi} F_{tot}(\varphi) d\varphi} \cdot 100$ (11)	IE : index of effectiveness (%) φ : crank angle (degree)
<i>Positive impulse proportion</i>	$PIP = \frac{\int_0^{2\pi} P_+(\varphi) d(\varphi)}{\int_0^{2\pi} (P_+ + P_-)(\varphi) d\varphi} \cdot 100$ (12)	PIP : positive impulse proportion (%) φ : crank angle (degree) P_+ = Positive impulse (N) $ P_- $ = absolute value of the negative impulse (N)

ANNEXE E - Certificat d'éthique étude 1

	3698						
 <p>CERTIFICAT D'ÉTHIQUE DE LA RECHERCHE AVEC DES ÈTRES HUMAINS</p> <p>En vertu du mandat qui lui a été confié par l'Université, le Comité d'éthique de la recherche avec des êtres humains a analysé et approuvé pour certification éthique le protocole de recherche suivant :</p> <p>Titre : Détermination des facteurs qui influencent la production de force en cyclisme</p> <p>Chercheur(s) : Andres Torres Département de génie électrique et génie informatique</p> <p>Organisme(s) : Aucun financement</p> <p>N° DU CERTIFICAT : CER-21-277-07.27</p> <p>PÉRIODE DE VALIDITÉ : Du 07 juin 2021 au 07 juin 2022</p> <tr> <td colspan="2" style="padding: 10px;"> <p><u>En acceptant le certificat éthique, le chercheur s'engage à :</u></p> <ul style="list-style-type: none"> - Aviser le CER par écrit des changements apportés à son protocole de recherche avant leur entrée en vigueur; - Procéder au renouvellement annuel du certificat tant et aussi longtemps que la recherche ne sera pas terminée; - Aviser par écrit le CER de l'abandon ou de l'interruption prématuree de la recherche; - Faire parvenir par écrit au CER un rapport final dans le mois suivant la fin de la recherche. </td> </tr> <tr> <td style="text-align: center; vertical-align: middle;">  Me Richard LeBlanc Président du comité </td> <td style="text-align: center; vertical-align: middle;">  Fanny Longpré Secrétaire du comité </td> </tr> <tr> <td colspan="2" style="text-align: center; padding-top: 10px;"> <i>Décanat de la recherche et de la création</i> Date d'émission : 07 juin 2021 </td> </tr>		<p><u>En acceptant le certificat éthique, le chercheur s'engage à :</u></p> <ul style="list-style-type: none"> - Aviser le CER par écrit des changements apportés à son protocole de recherche avant leur entrée en vigueur; - Procéder au renouvellement annuel du certificat tant et aussi longtemps que la recherche ne sera pas terminée; - Aviser par écrit le CER de l'abandon ou de l'interruption prématuree de la recherche; - Faire parvenir par écrit au CER un rapport final dans le mois suivant la fin de la recherche. 		 Me Richard LeBlanc Président du comité	 Fanny Longpré Secrétaire du comité	<i>Décanat de la recherche et de la création</i> Date d'émission : 07 juin 2021	
<p><u>En acceptant le certificat éthique, le chercheur s'engage à :</u></p> <ul style="list-style-type: none"> - Aviser le CER par écrit des changements apportés à son protocole de recherche avant leur entrée en vigueur; - Procéder au renouvellement annuel du certificat tant et aussi longtemps que la recherche ne sera pas terminée; - Aviser par écrit le CER de l'abandon ou de l'interruption prématuree de la recherche; - Faire parvenir par écrit au CER un rapport final dans le mois suivant la fin de la recherche. 							
 Me Richard LeBlanc Président du comité	 Fanny Longpré Secrétaire du comité						
<i>Décanat de la recherche et de la création</i> Date d'émission : 07 juin 2021							

ANNEXE F - Certificat d'éthique étude 2

4131



CERTIFICAT D'ÉTHIQUE DE LA RECHERCHE AVEC DES ÉTRES HUMAINS

En vertu du mandat qui lui a été confié par l'Université, le Comité d'éthique de la recherche avec des êtres humains a analysé et approuvé pour certification éthique le protocole de recherche suivant :

Titre : Estimation du recrutement musculaire en cyclisme à partir d'une technique d'apprentissage automatique

Chercheur(s) : Andres Torres Velasquez
Département de génie électrique et génie informatique

Organisme(s) : Aucun financement

N° DU CERTIFICAT : CER-22-291-07.03

PÉRIODE DE VALIDITÉ : Du 12 octobre 2022 au 12 octobre 2023

En acceptant le certificat éthique, le chercheur s'engage à :

- Aviser le CER par écrit des changements apportés à son protocole de recherche avant leur entrée en vigueur;
- Procéder au renouvellement annuel du certificat tant et aussi longtemps que la recherche ne sera pas terminée;
- Aviser par écrit le CER de l'abandon ou de l'interruption prématurée de la recherche;
- Faire parvenir par écrit au CER un rapport final dans le mois suivant la fin de la recherche.

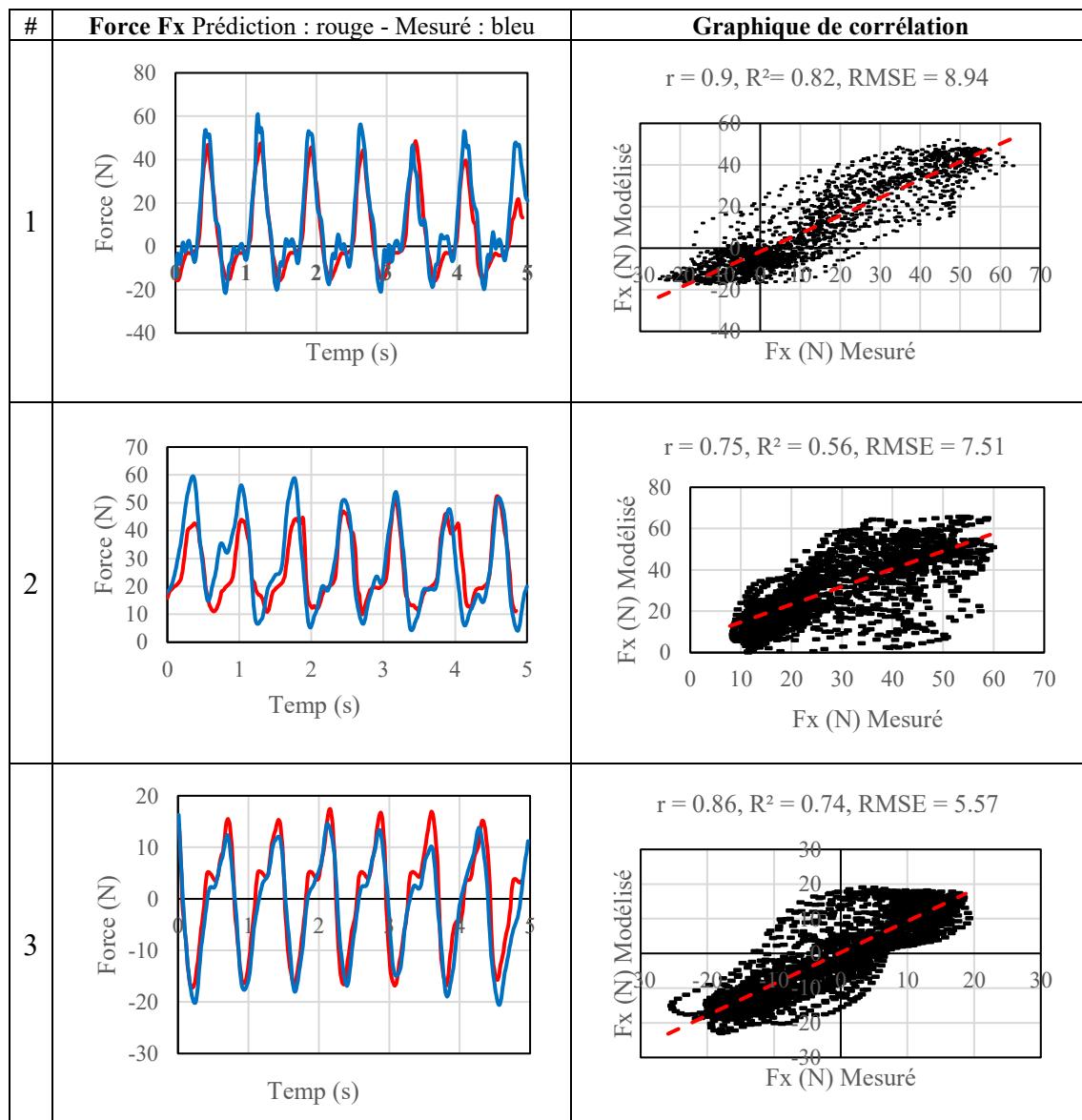
Me Richard LeBlanc
Président du comité

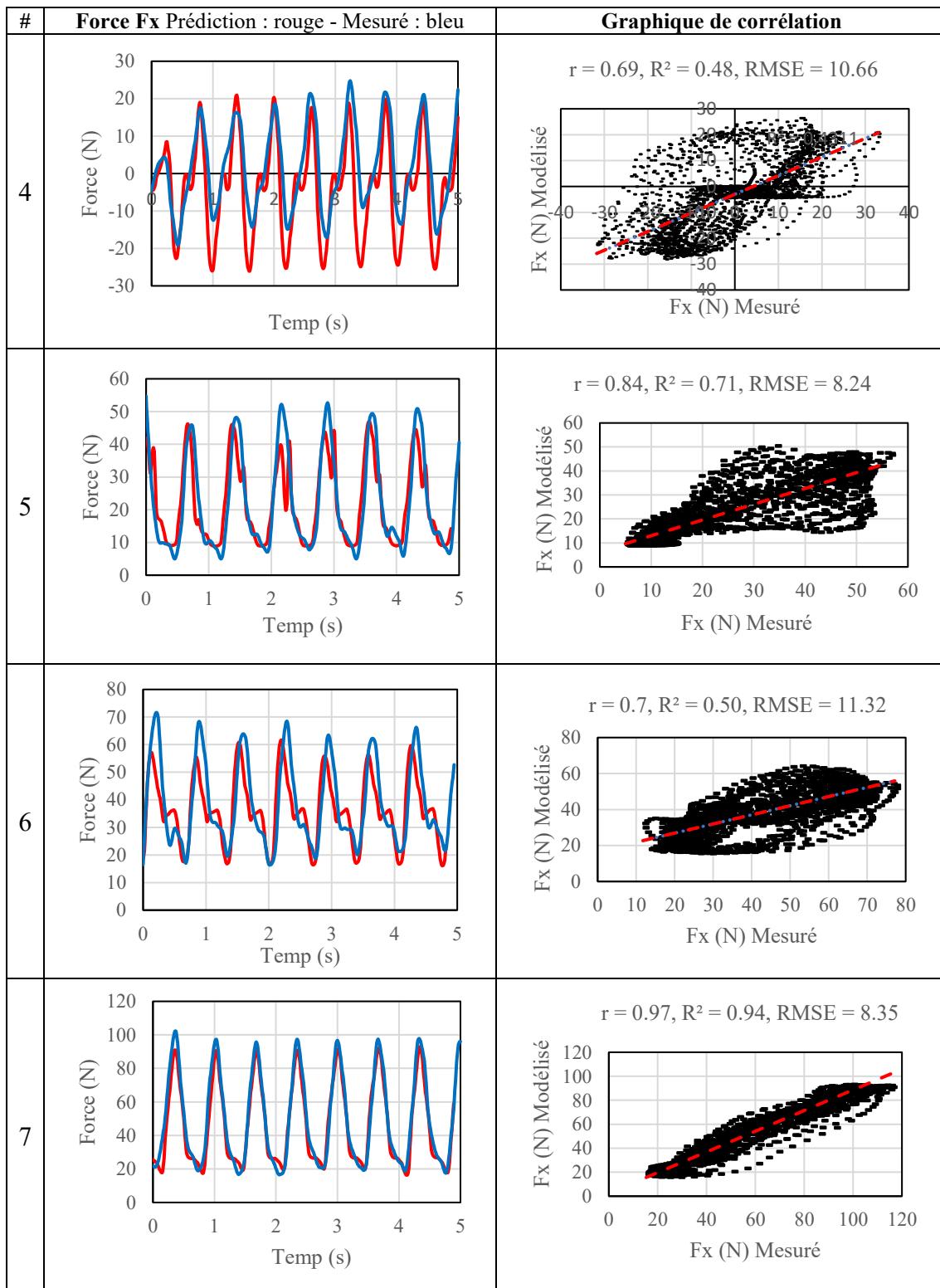
Fanny Longpré
Secrétaire du comité

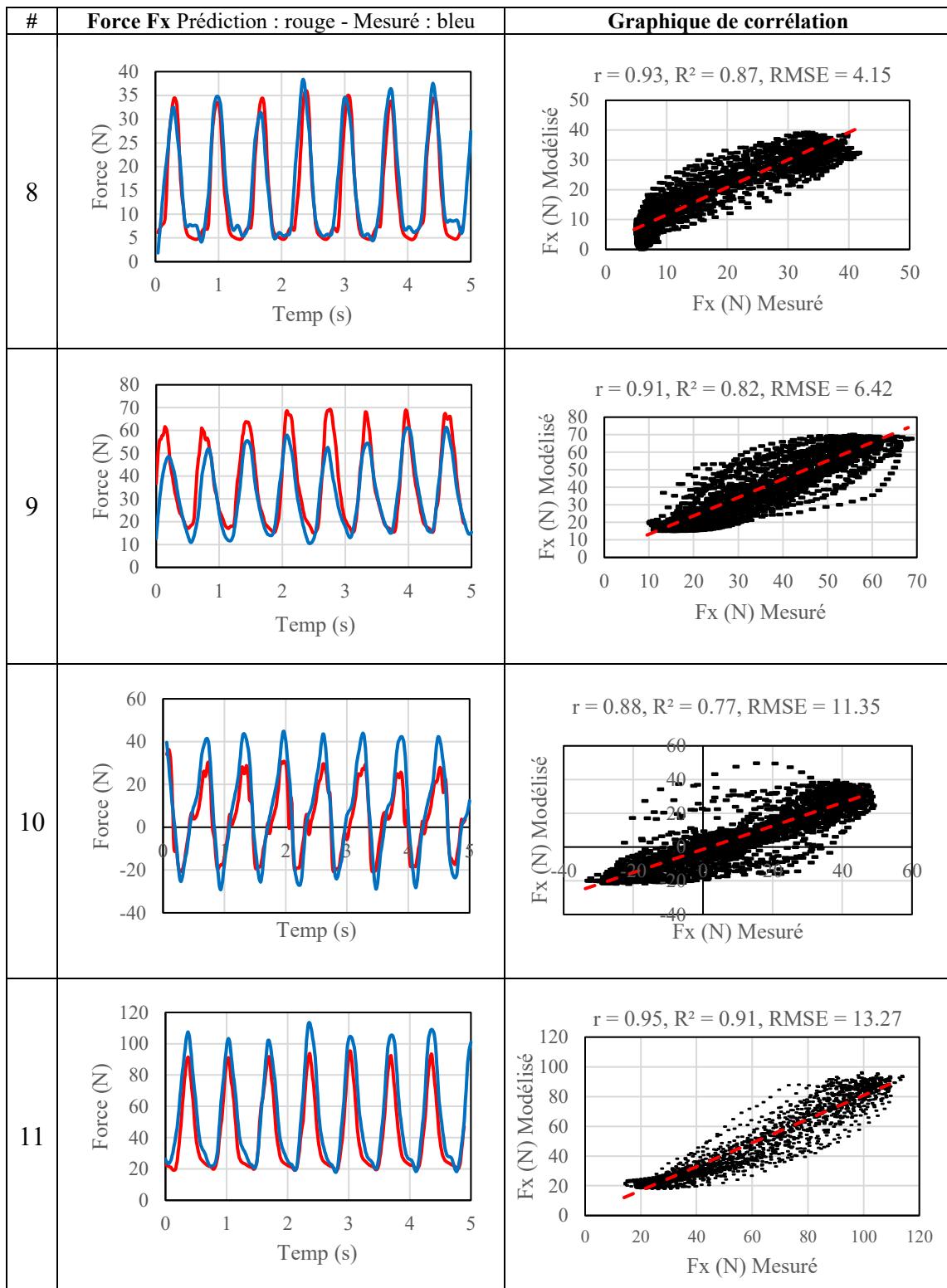
Décanat de la recherche et de la création

Date d'émission : 12 octobre 2022

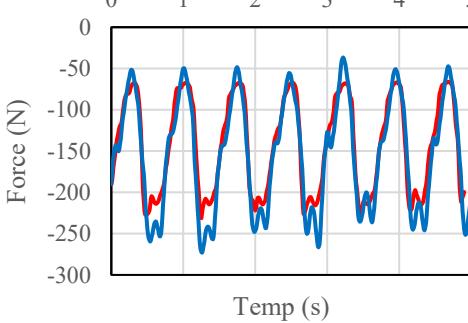
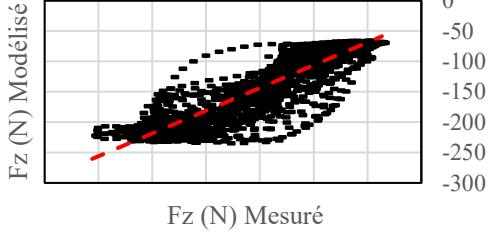
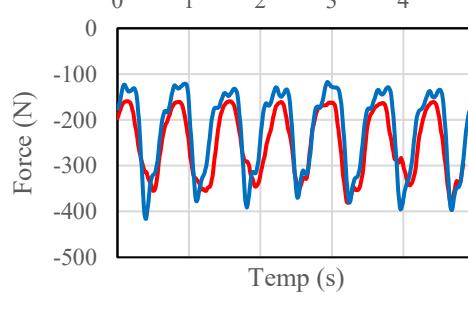
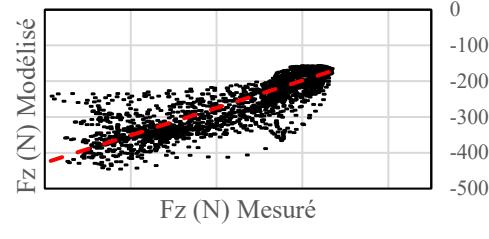
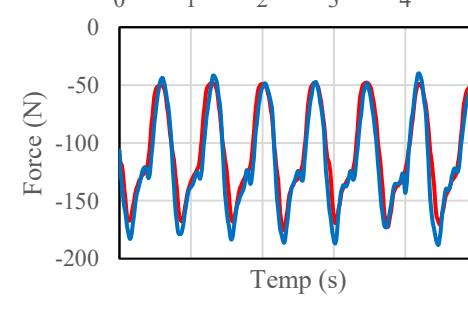
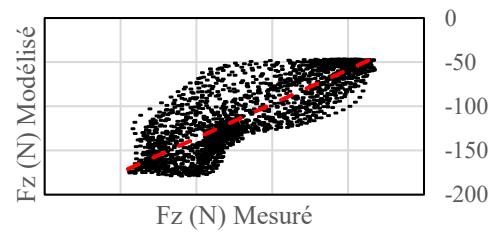
ANNEXE G - Prédiction de la force de pédalage Fx à 100 W

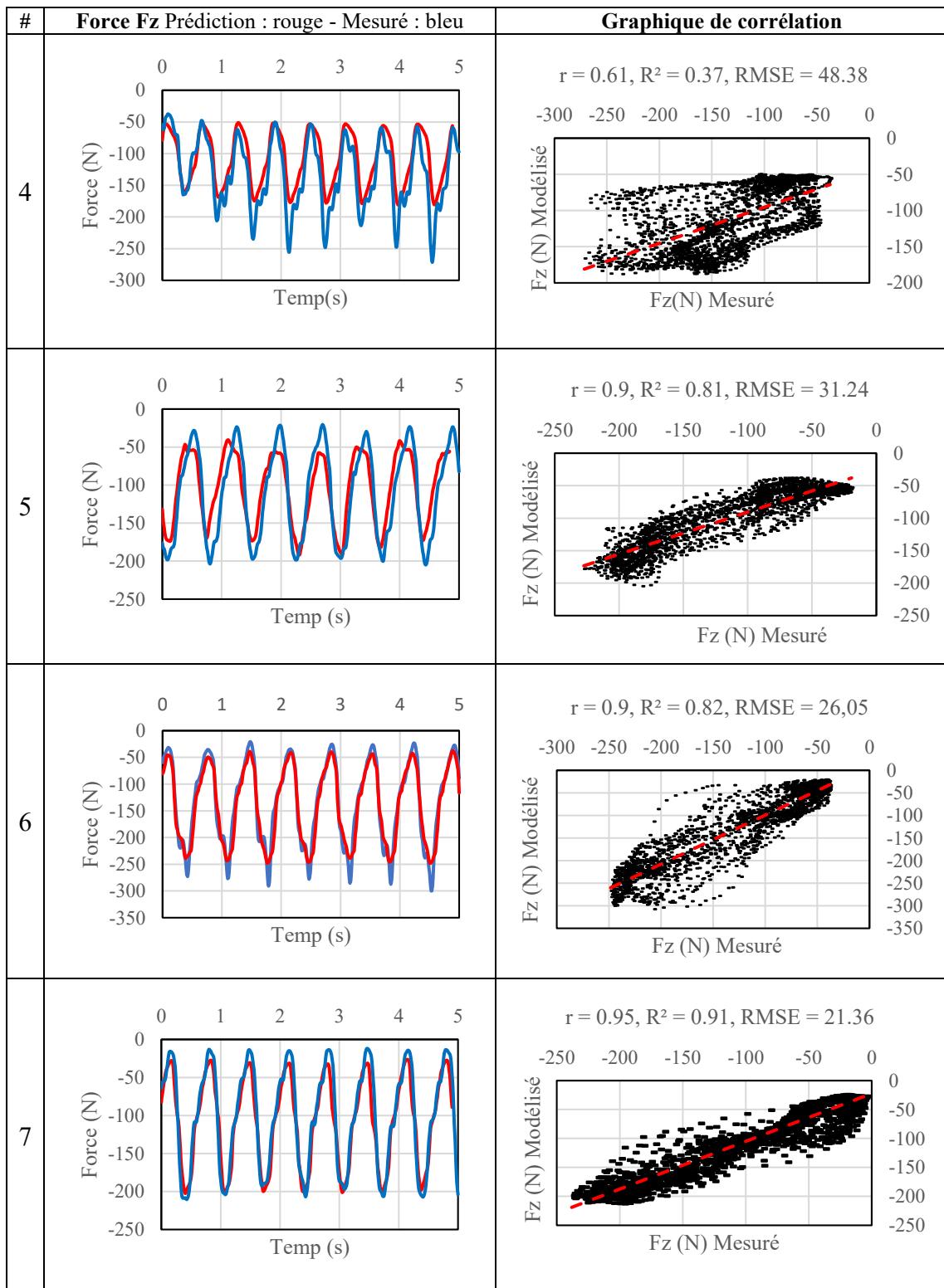


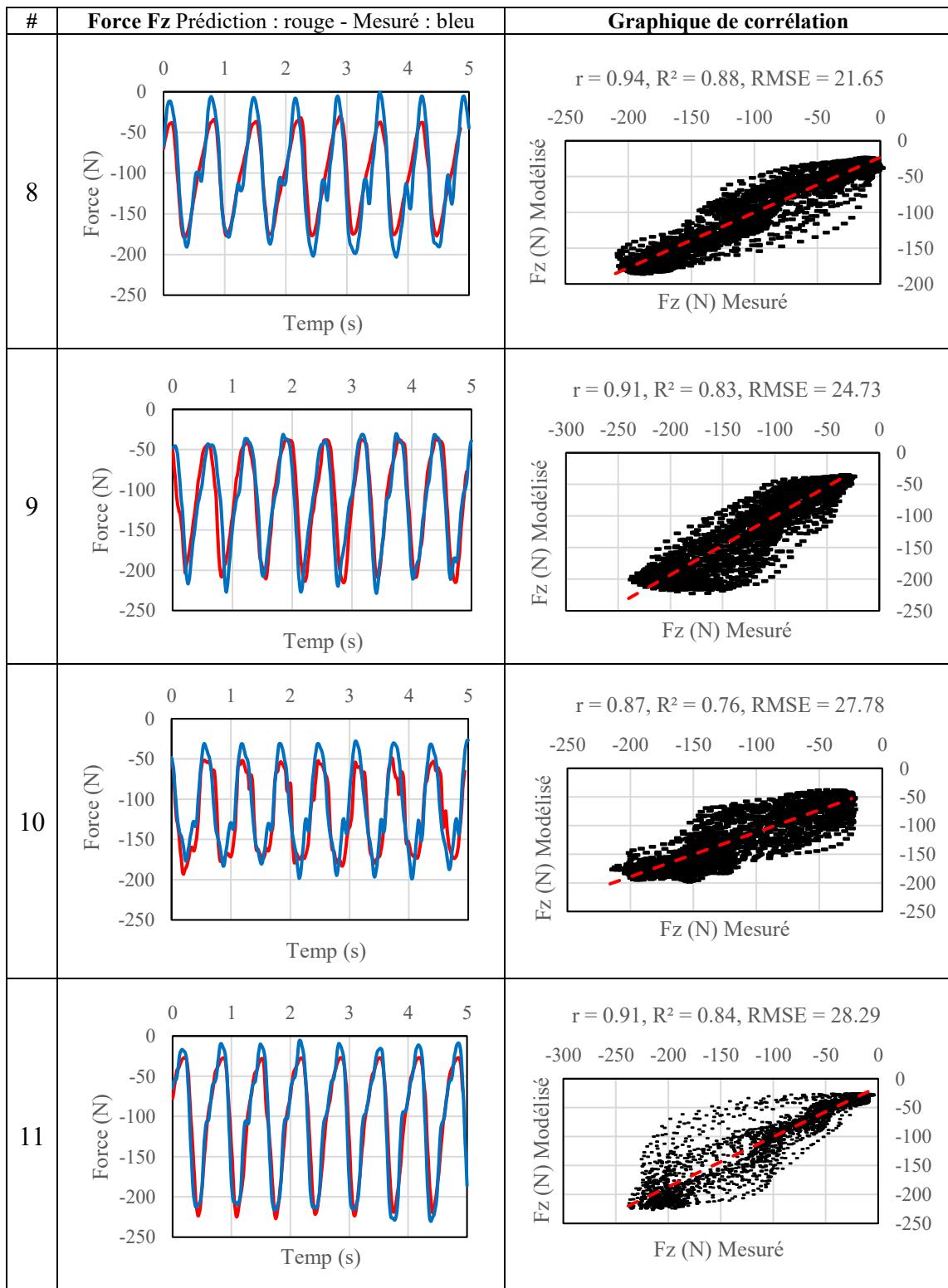




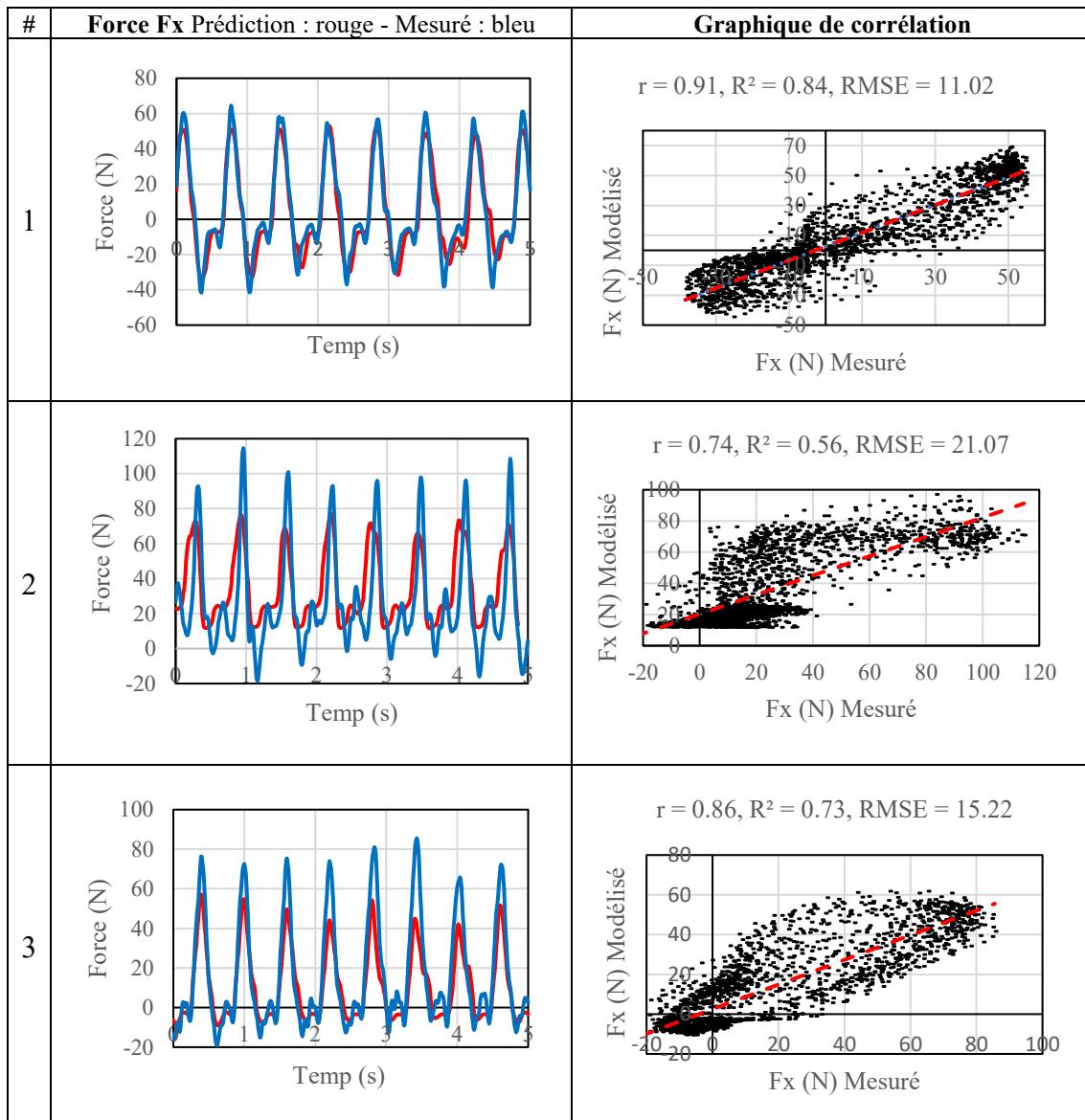
ANNEXE H - Prédiction de la force de pédalage Fz à 100 W

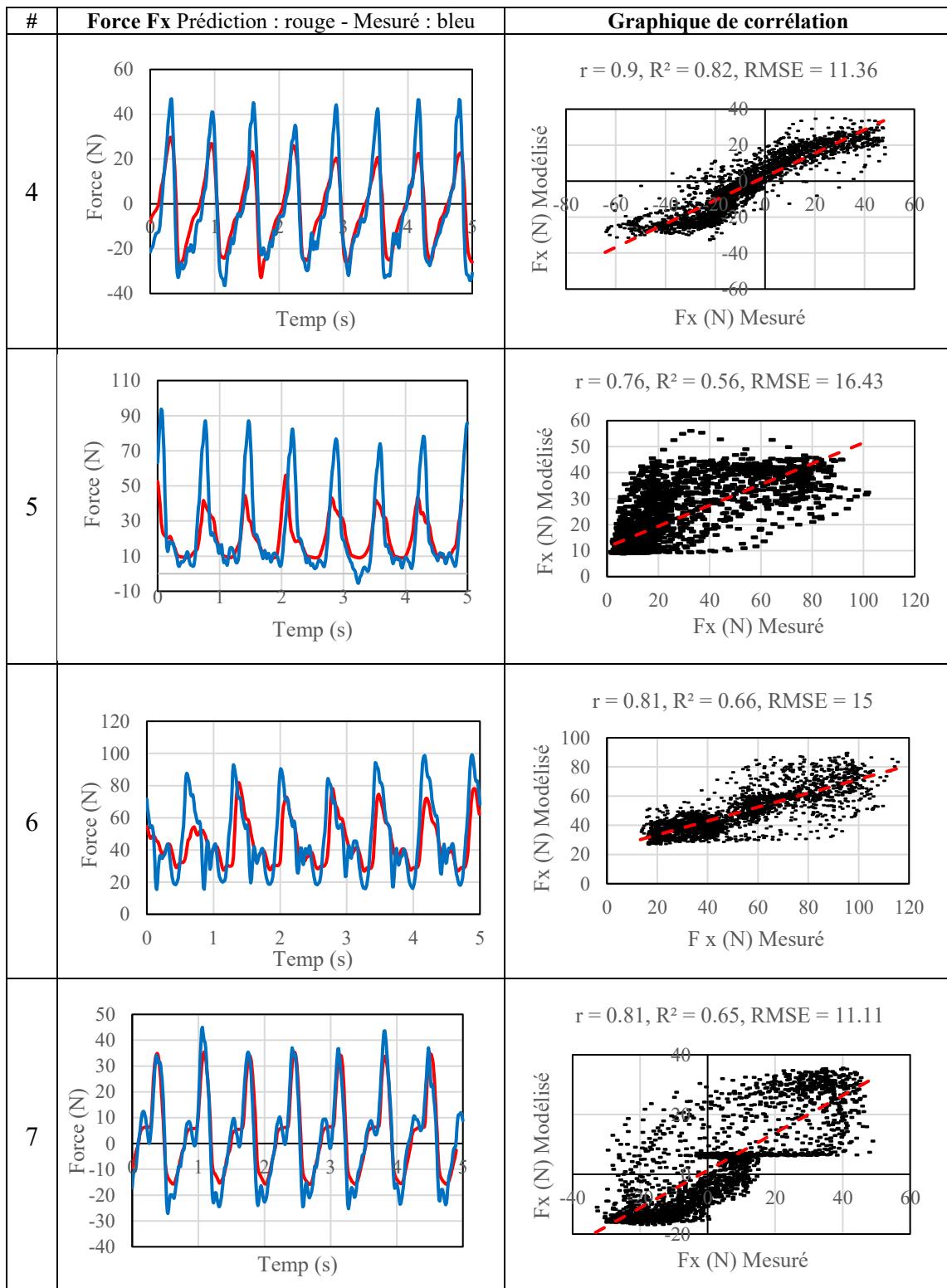
#	Force Fz Prédiction : rouge - Mesuré : bleu	Graphique de corrélation
1	 <p>Force (N)</p> <p>Temp (s)</p>	<p>$r = 0.91, R^2 = 0.83, \text{RMSE} = 30.29$</p>  <p>Fz (N) Modélisé</p> <p>Fz (N) Mesuré</p>
2	 <p>Force (N)</p> <p>Temp (s)</p>	<p>$r = 0.86, R^2 = 0.75, \text{RMSE} = 53.69$</p>  <p>Fz (N) Modélisé</p> <p>Fz (N) Mesuré</p>
3	 <p>Force (N)</p> <p>Temp (s)</p>	<p>$r = 0.85, R^2 = 0.72, \text{RMSE} = 24.28$</p>  <p>Fz (N) Modélisé</p> <p>Fz (N) Mesuré</p>

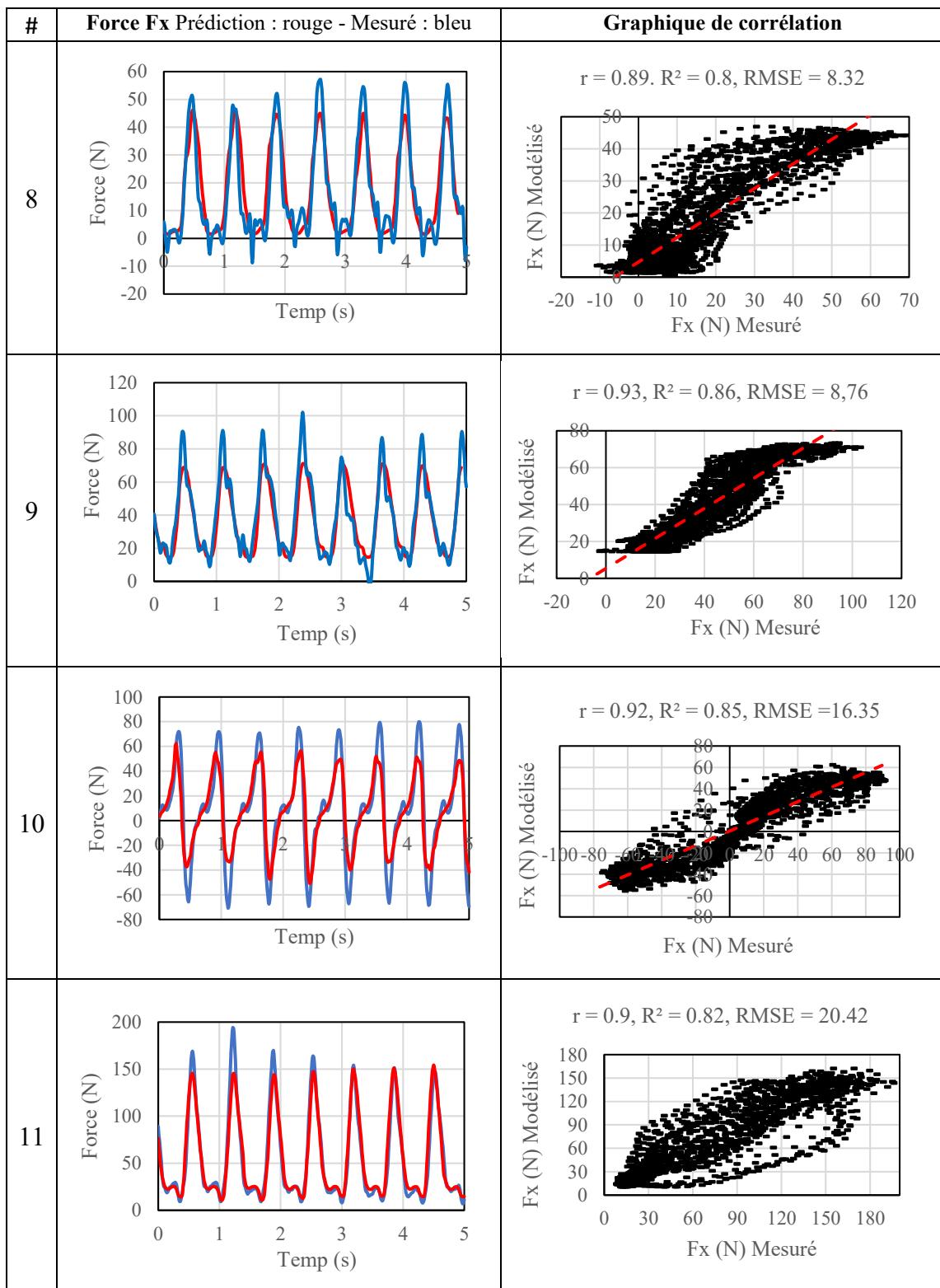




ANNEXE I - Prédiction de la force de pédalage Fx à 150 W

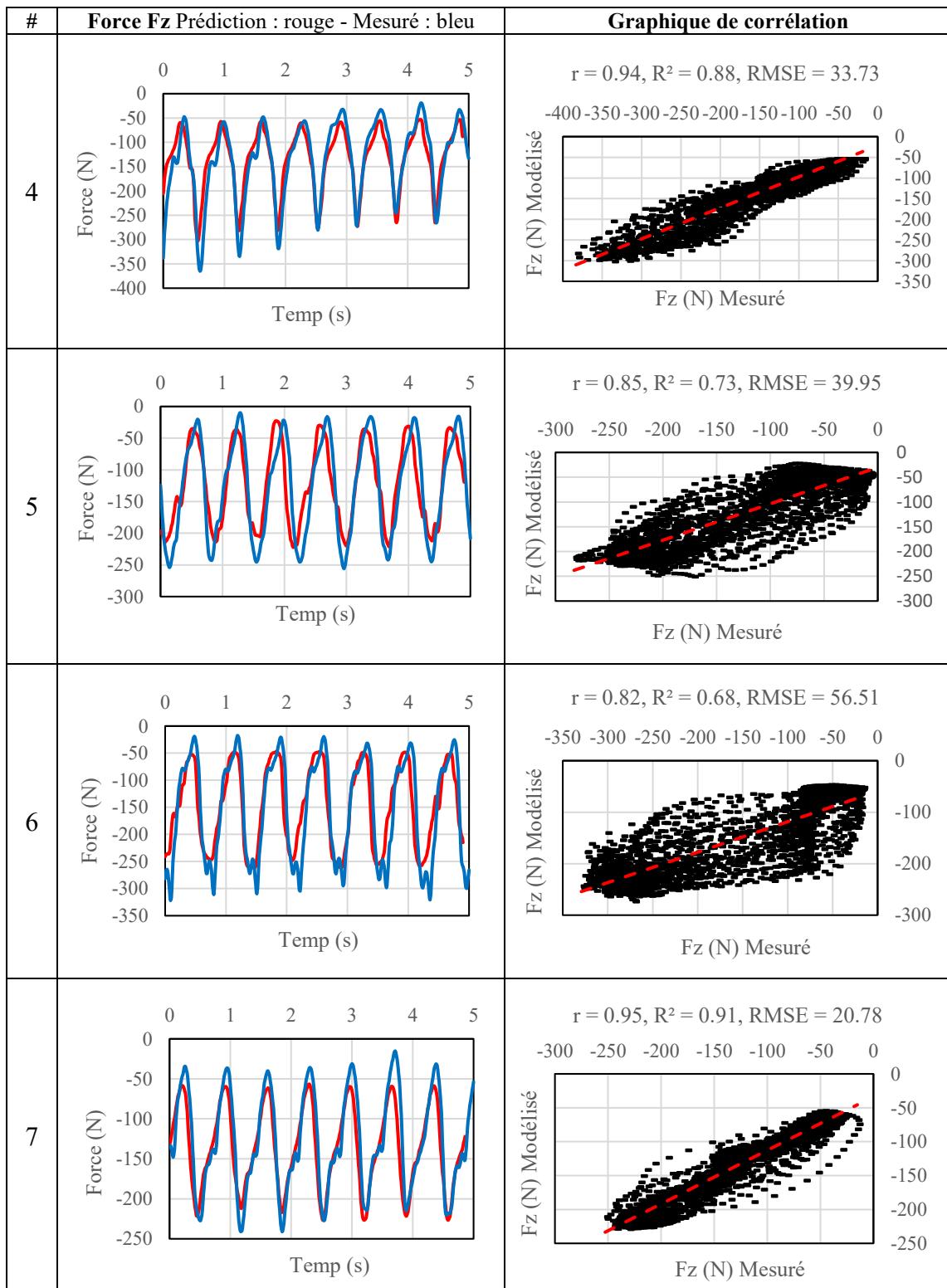


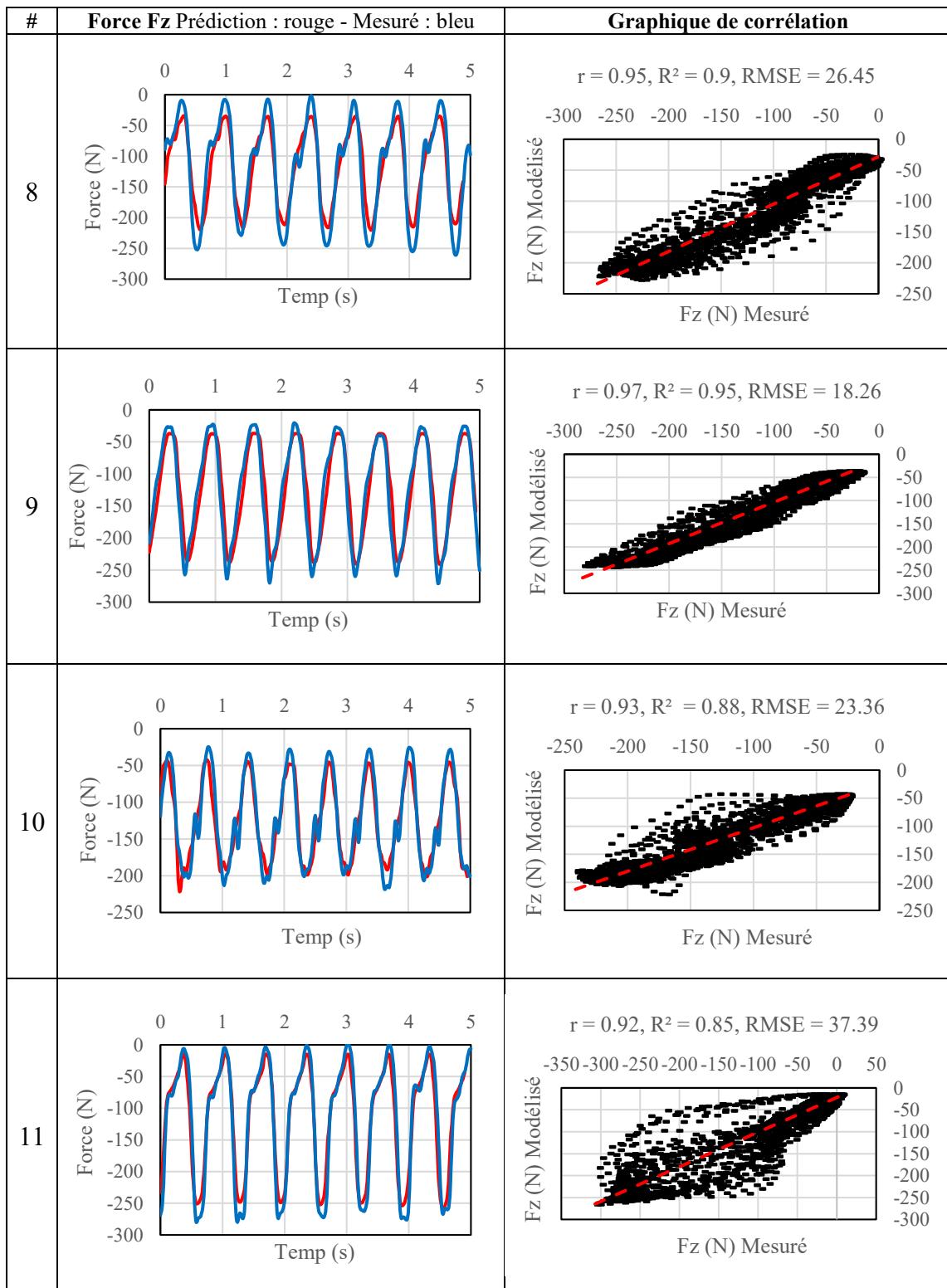




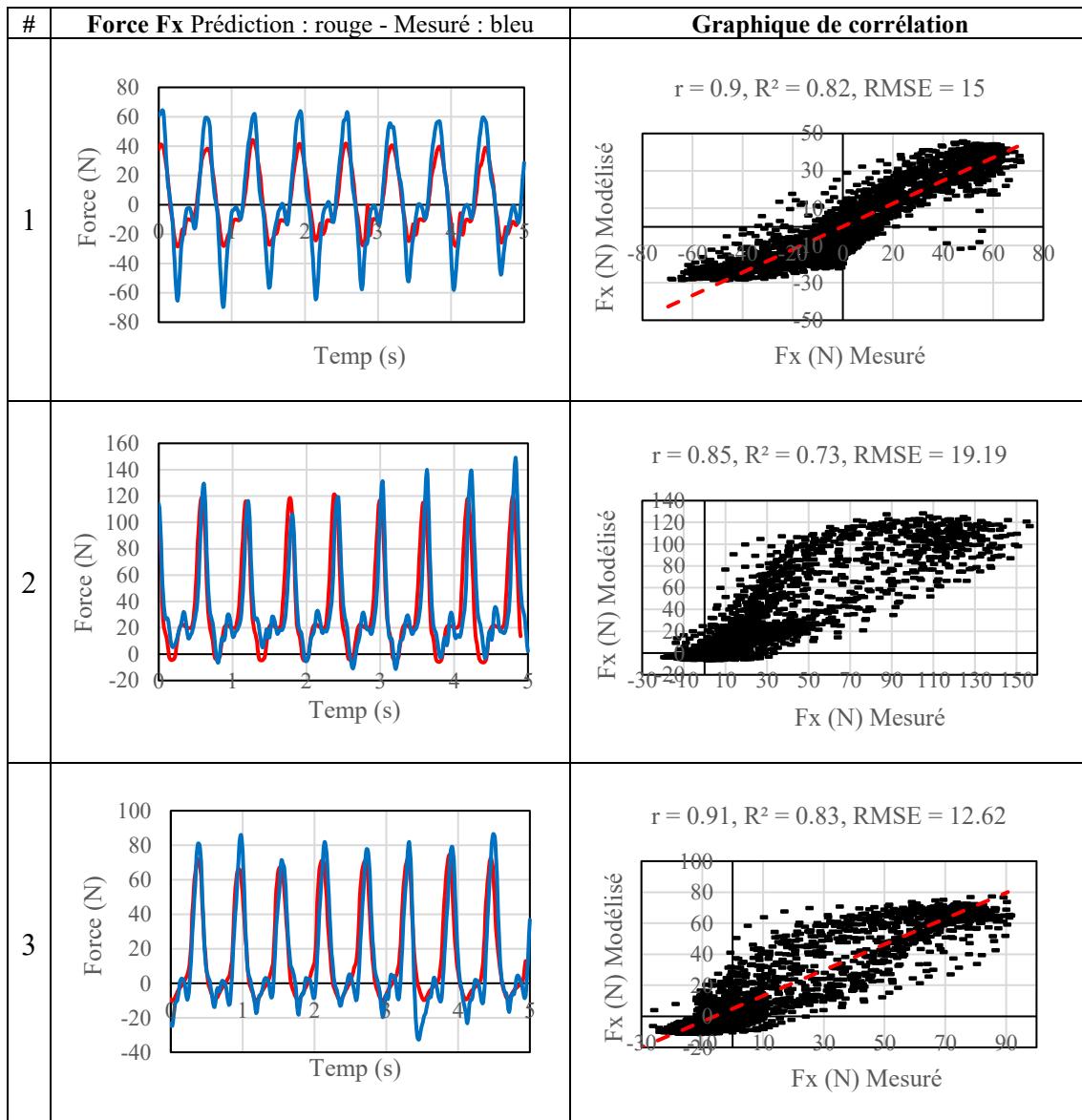
ANNEXE J - Prédiction de la force de pédalage Fz à 150 W

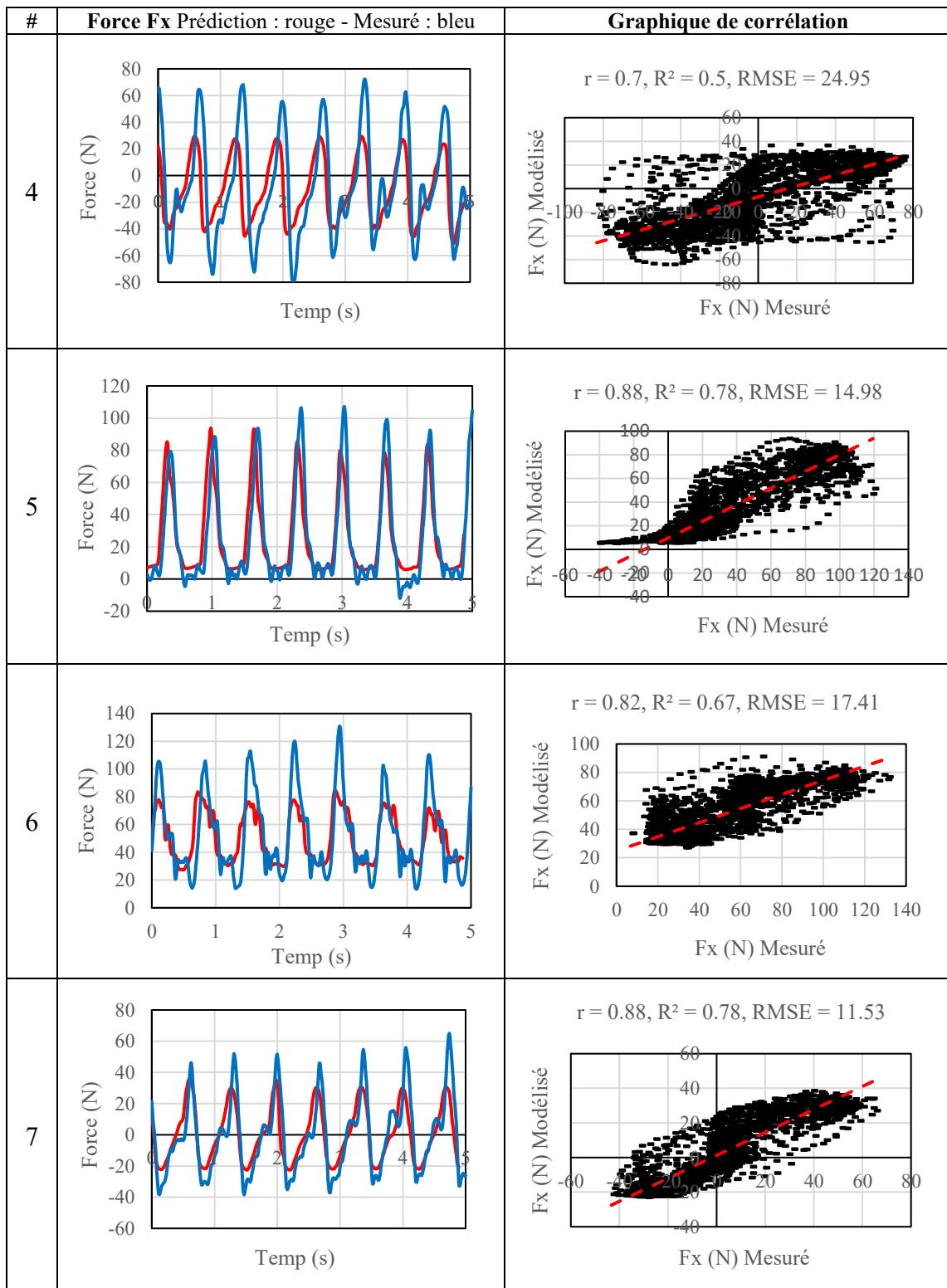
#	Force Fz Prédiction : rouge - Mesuré : bleu	Graphique de corrélation
1	<p>Force (N)</p> <p>Temp (s)</p>	$r = 0.97, R^2 = 0.95, \text{RMSE} = 31.29$ <p>Fz (N) Modélisé</p> <p>Fz (N) Mesuré</p>
2	<p>Force (N)</p> <p>Temp (s)</p>	$r = 0.89, R^2 = 0.8, \text{RMSE} = 47.18$ <p>Fz (N) Modélisé</p> <p>Fz (N) Mesuré</p>
3	<p>Force (N)</p> <p>Temp (s)</p>	$r = 0.8, R^2 = 0.65, \text{RMSE} = 54.2$ <p>Fz (N) Modélisé</p> <p>Fz (N) Mesuré</p>

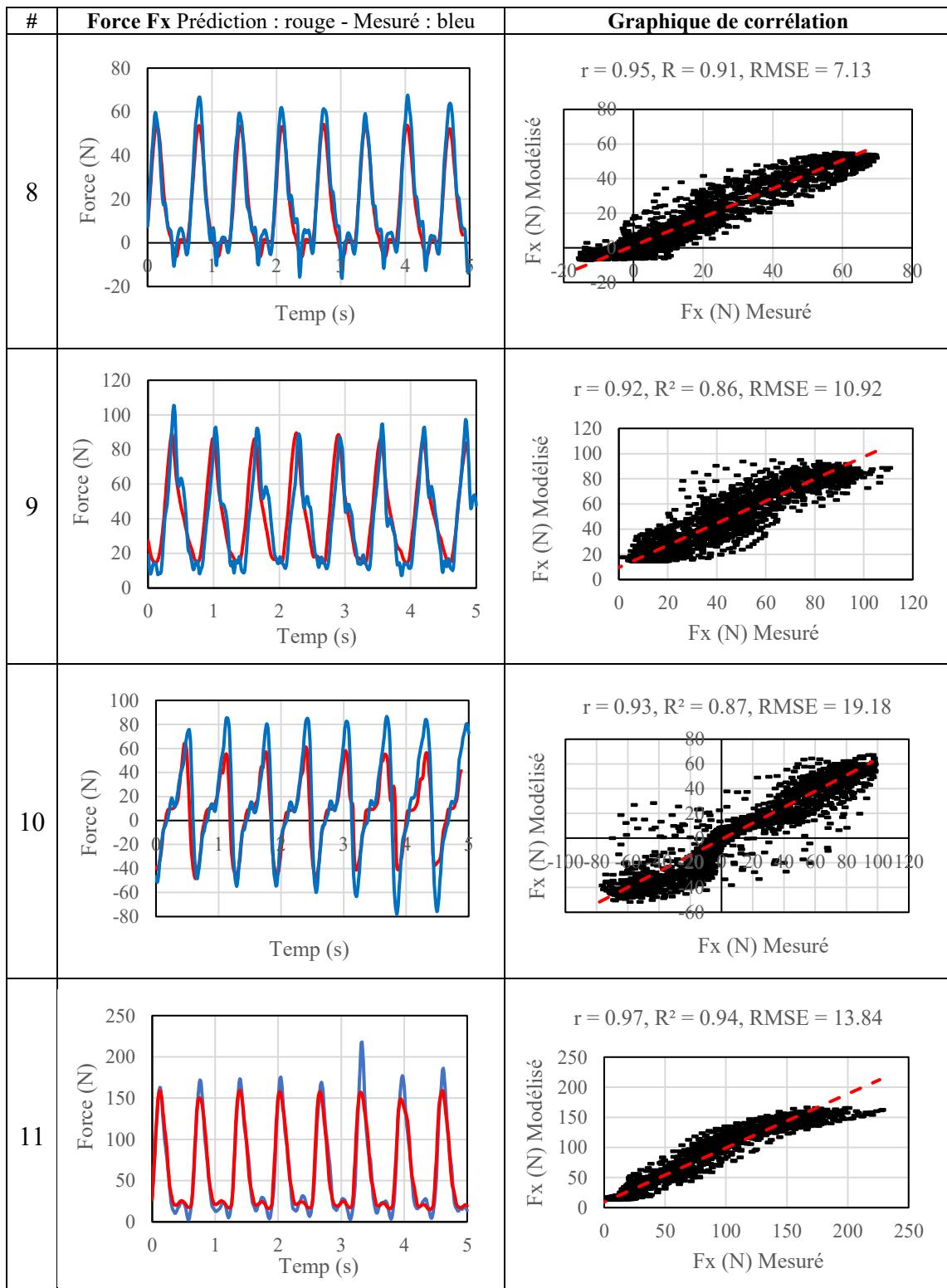




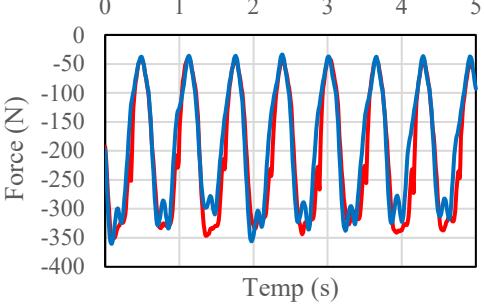
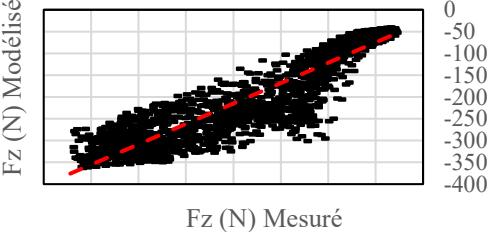
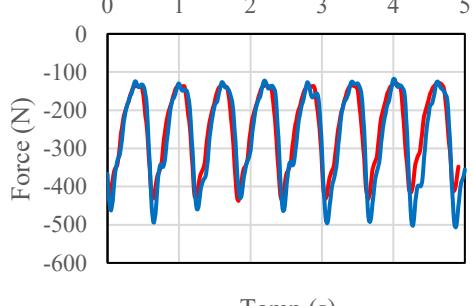
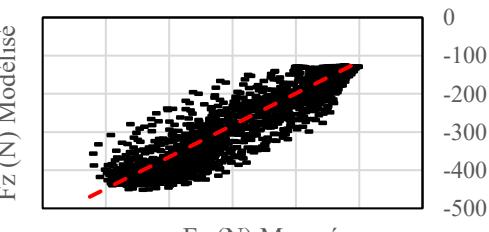
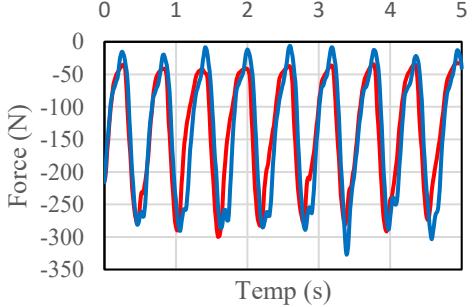
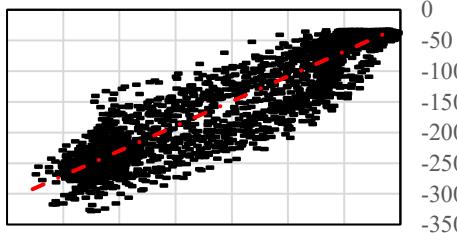
ANNEXE K - Prédiction de la force de pédalage Fx à 200 W

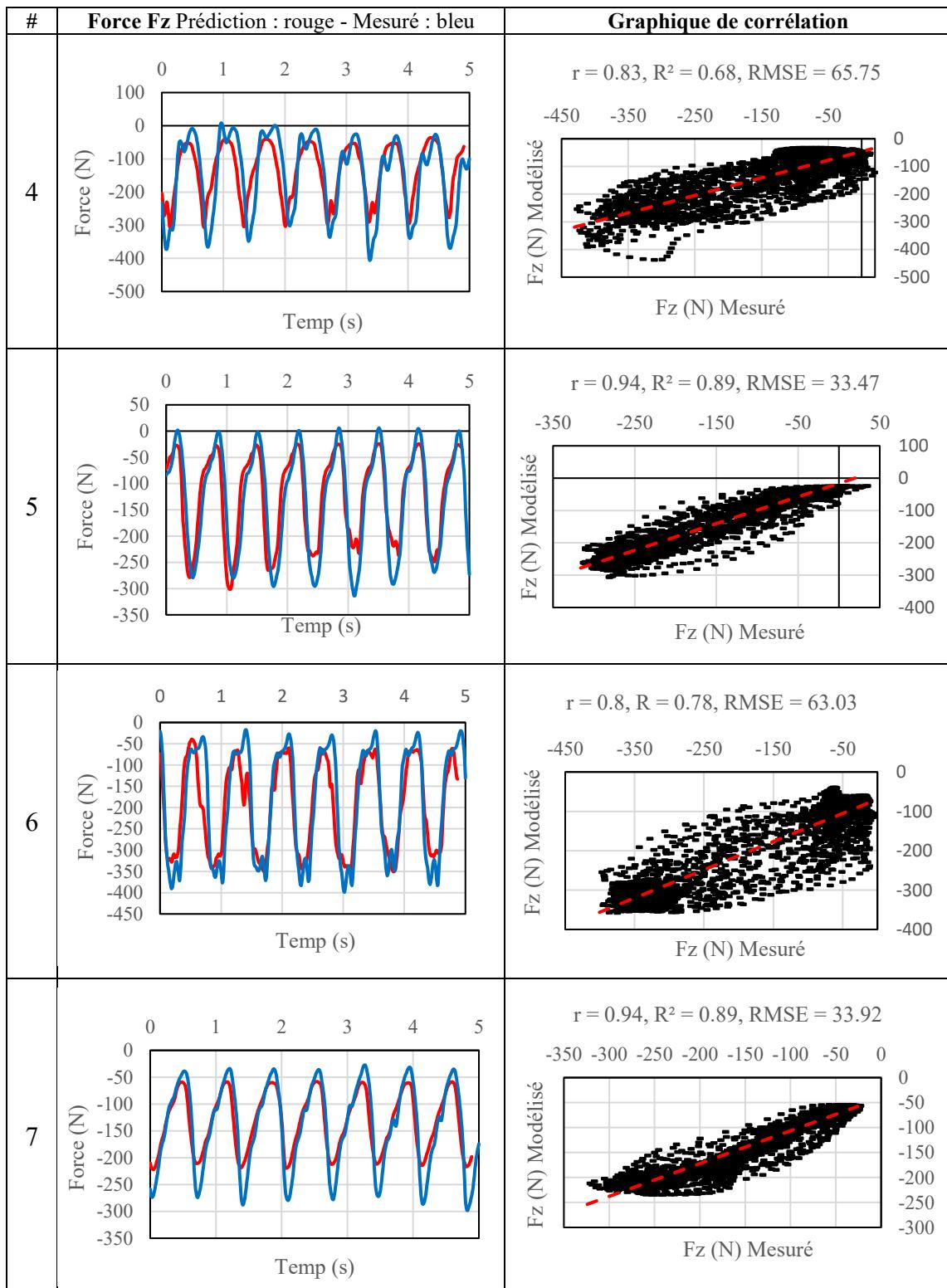


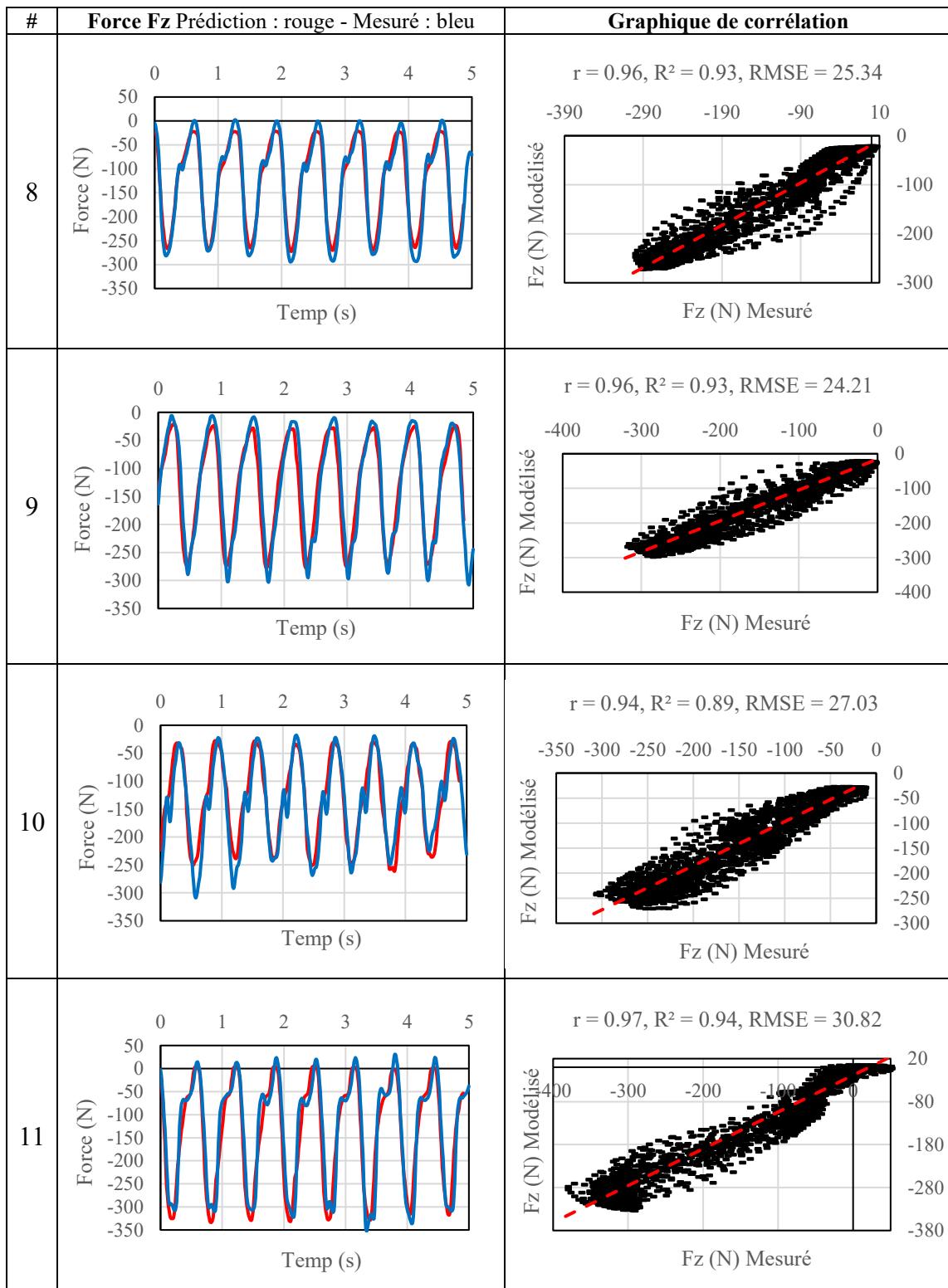




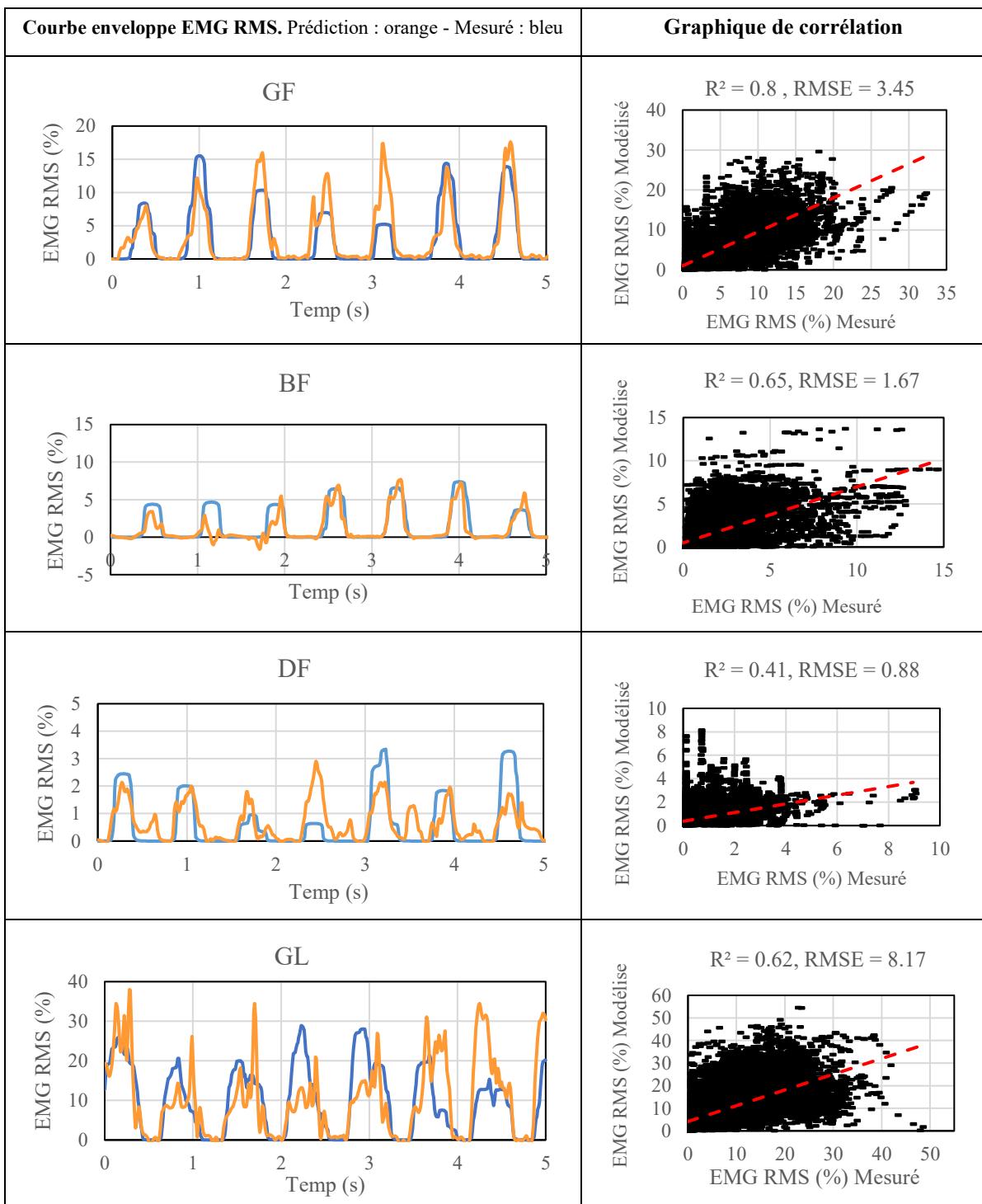
ANNEXE L - Prédiction de la force de pédalage Fz à 200 W

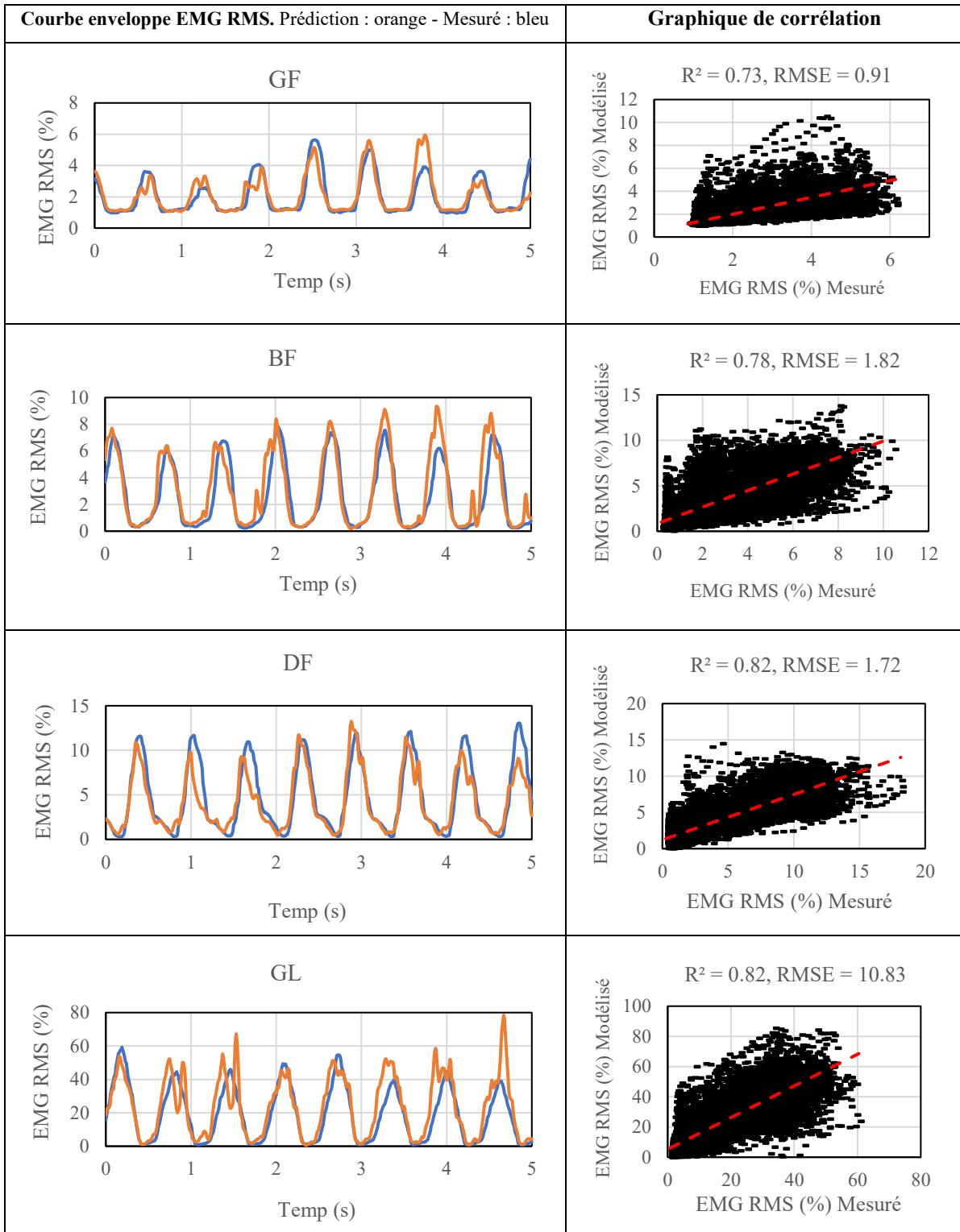
#	Force Fz Prédiction : rouge - Mesuré : bleu	Graphique de corrélation
1	 <p>Force (N)</p> <p>Temp (s)</p>	$r = 0.93, R^2 = 0.87, \text{RMSE} = 40.73$  <p>Fz (N) Modélisé</p> <p>Fz (N) Mesuré</p>
2	 <p>Force (N)</p> <p>Temp (s)</p>	$r = 0.93, R^2 = 0.88, \text{RMSE} = 41.65$  <p>Fz (N) Modélisé</p> <p>Fz (N) Mesuré</p>
3	 <p>Force (N)</p> <p>Temp (s)</p>	$r = 0.92, R^2 = 0.85, \text{RMSE} = 37.16$  <p>Fz (N) Modélisé</p> <p>Fz (N) Mesuré</p>

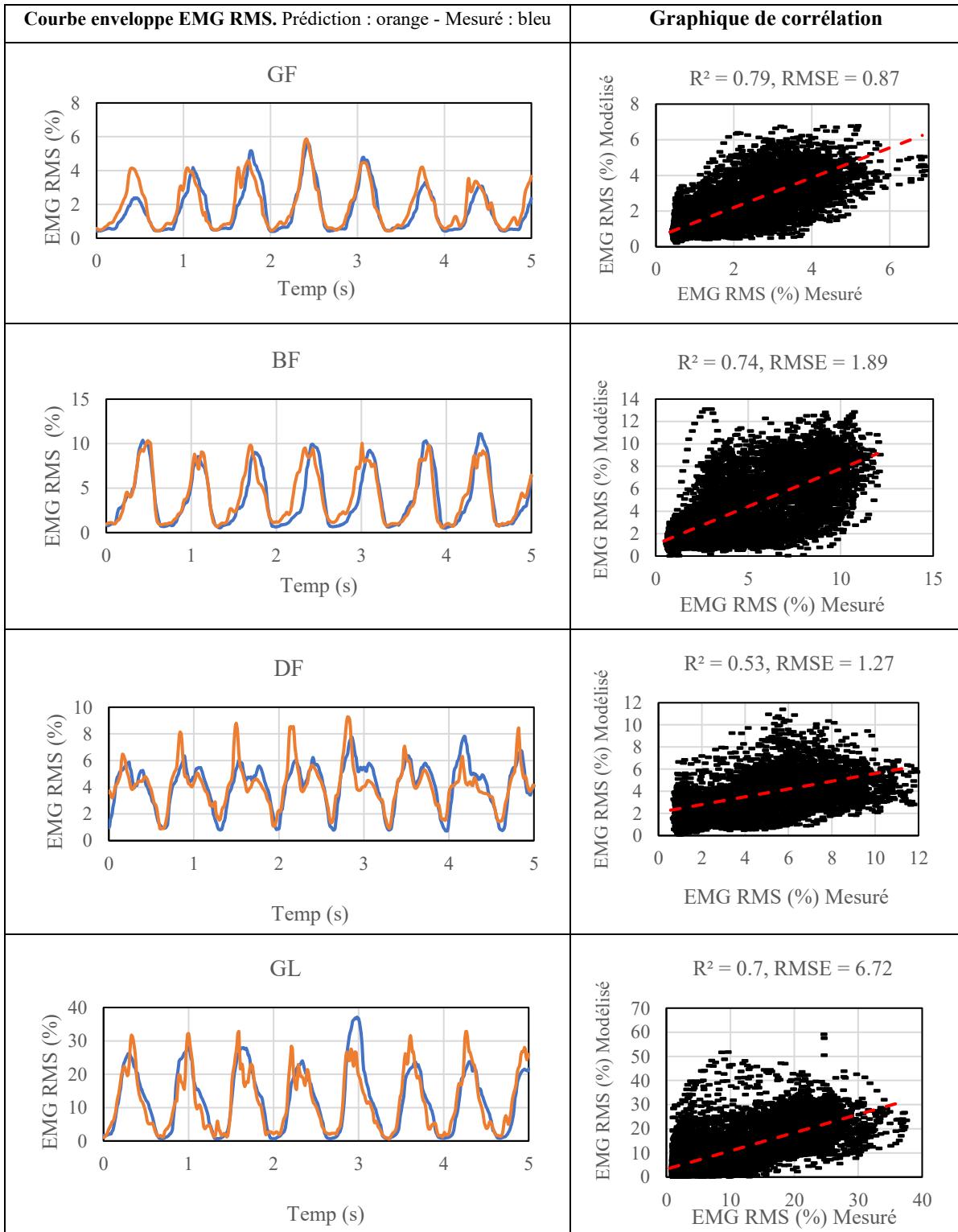


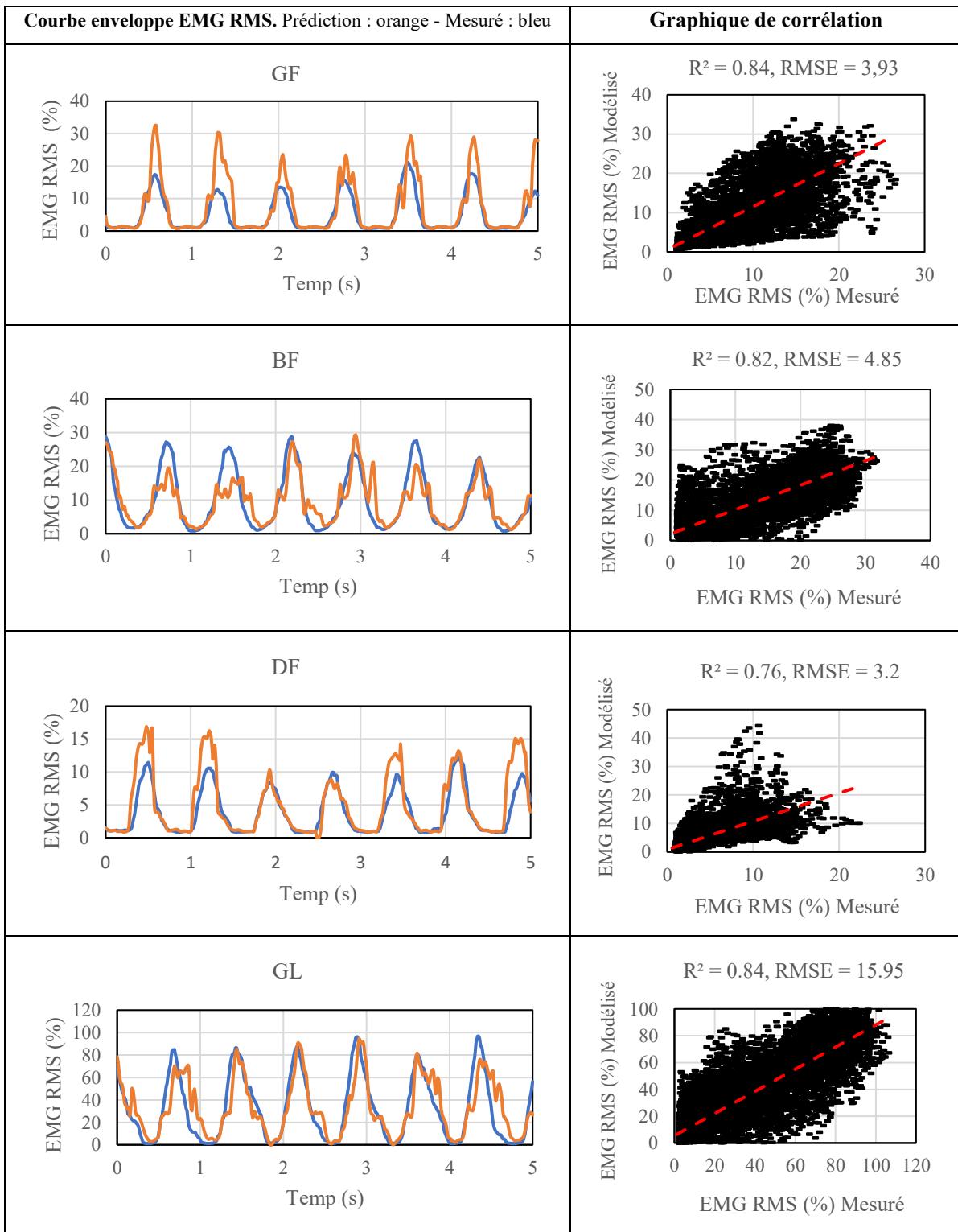


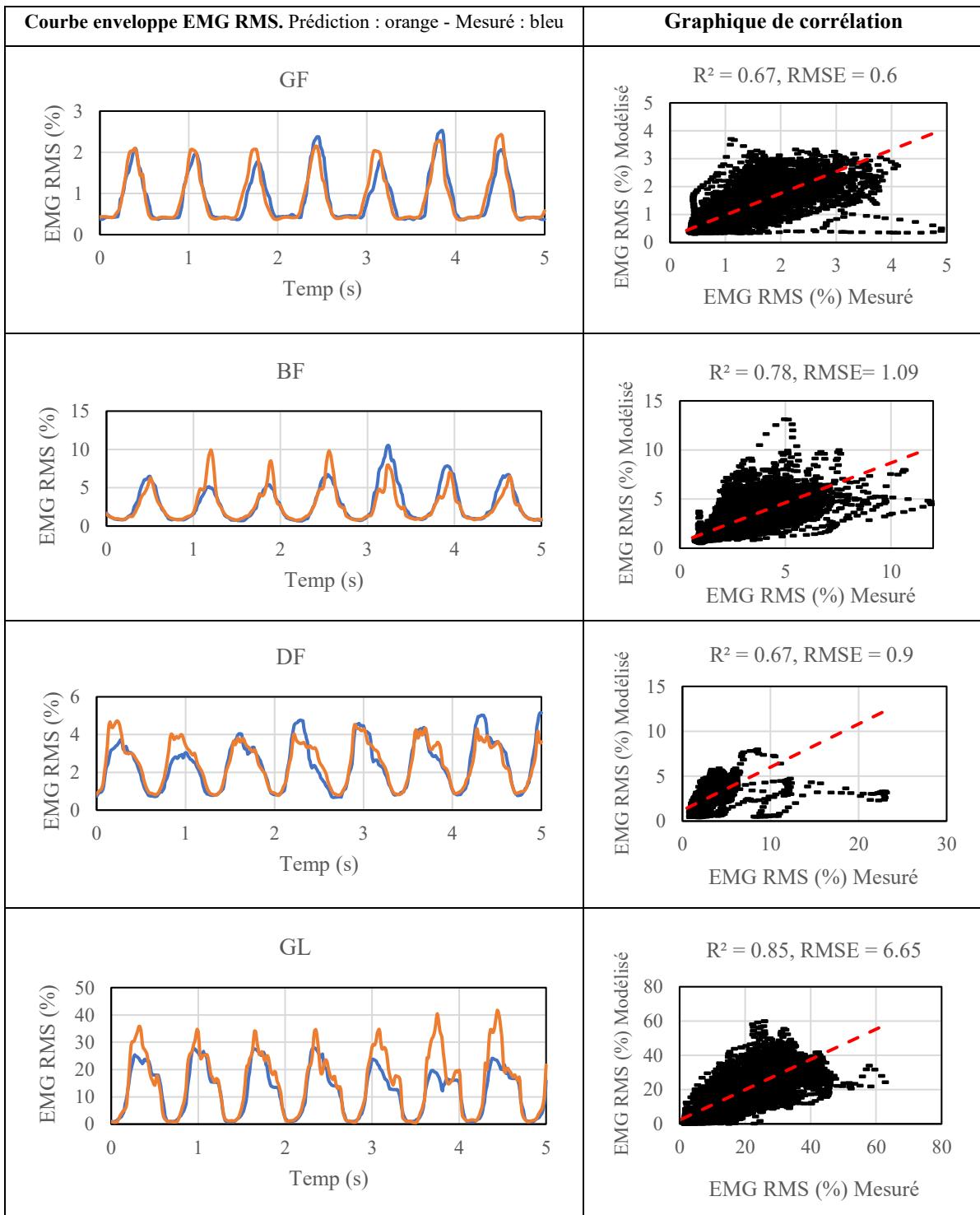
ANNEXE M - Prédiction courbe enveloppe EMG RMS 100 W
Sujet 1

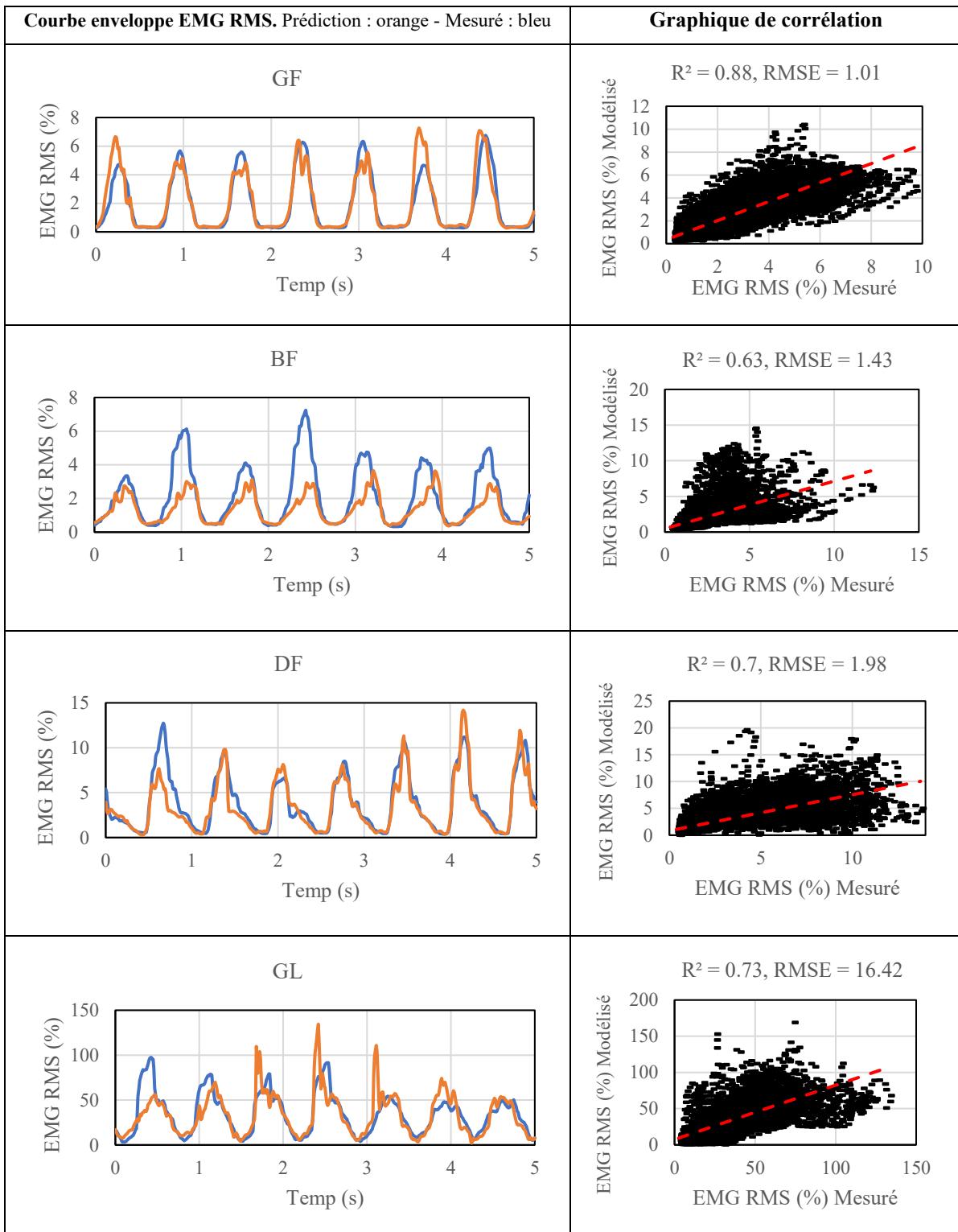


Sujet 2

Sujet 3

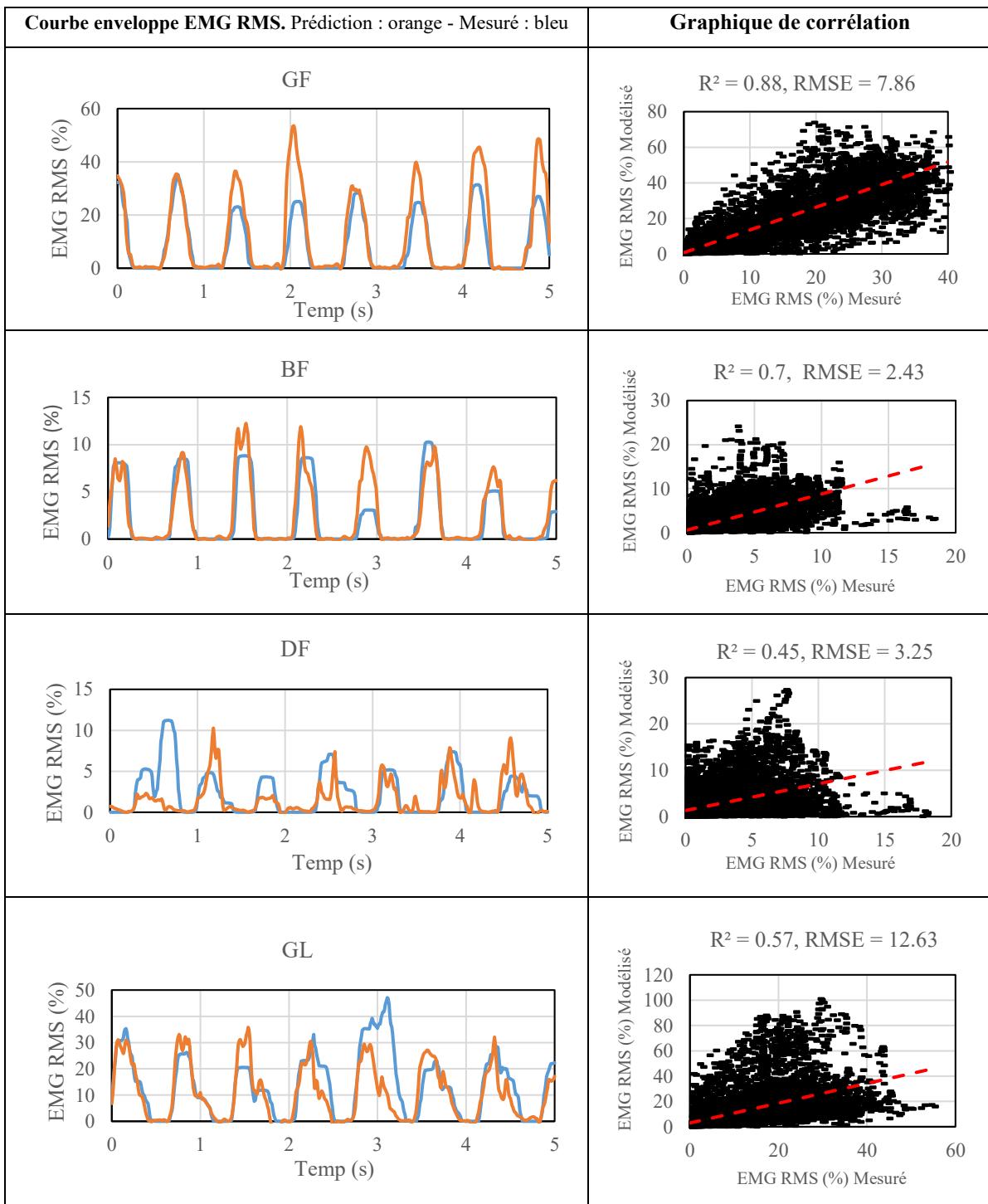
Sujet 4

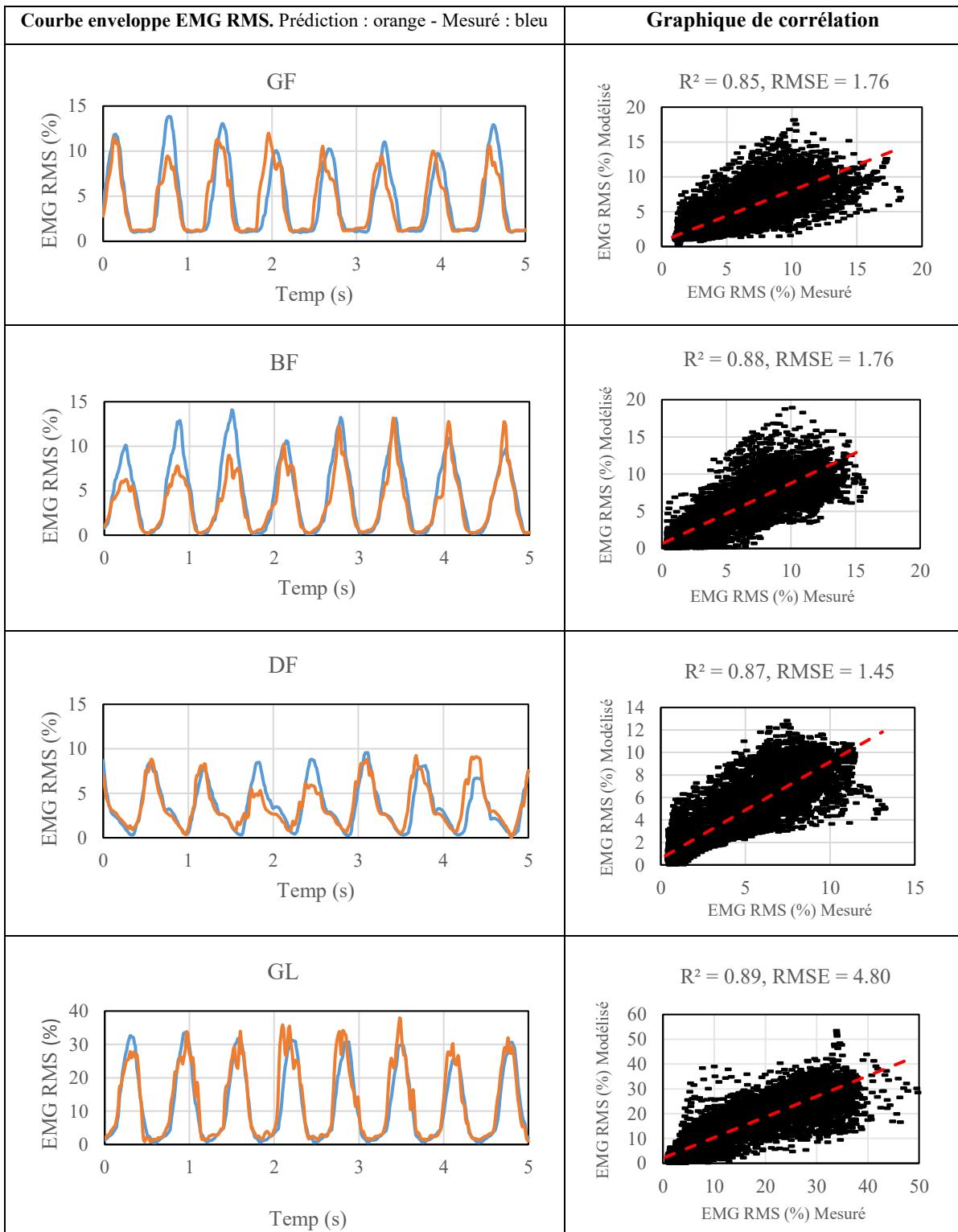
Sujet 5

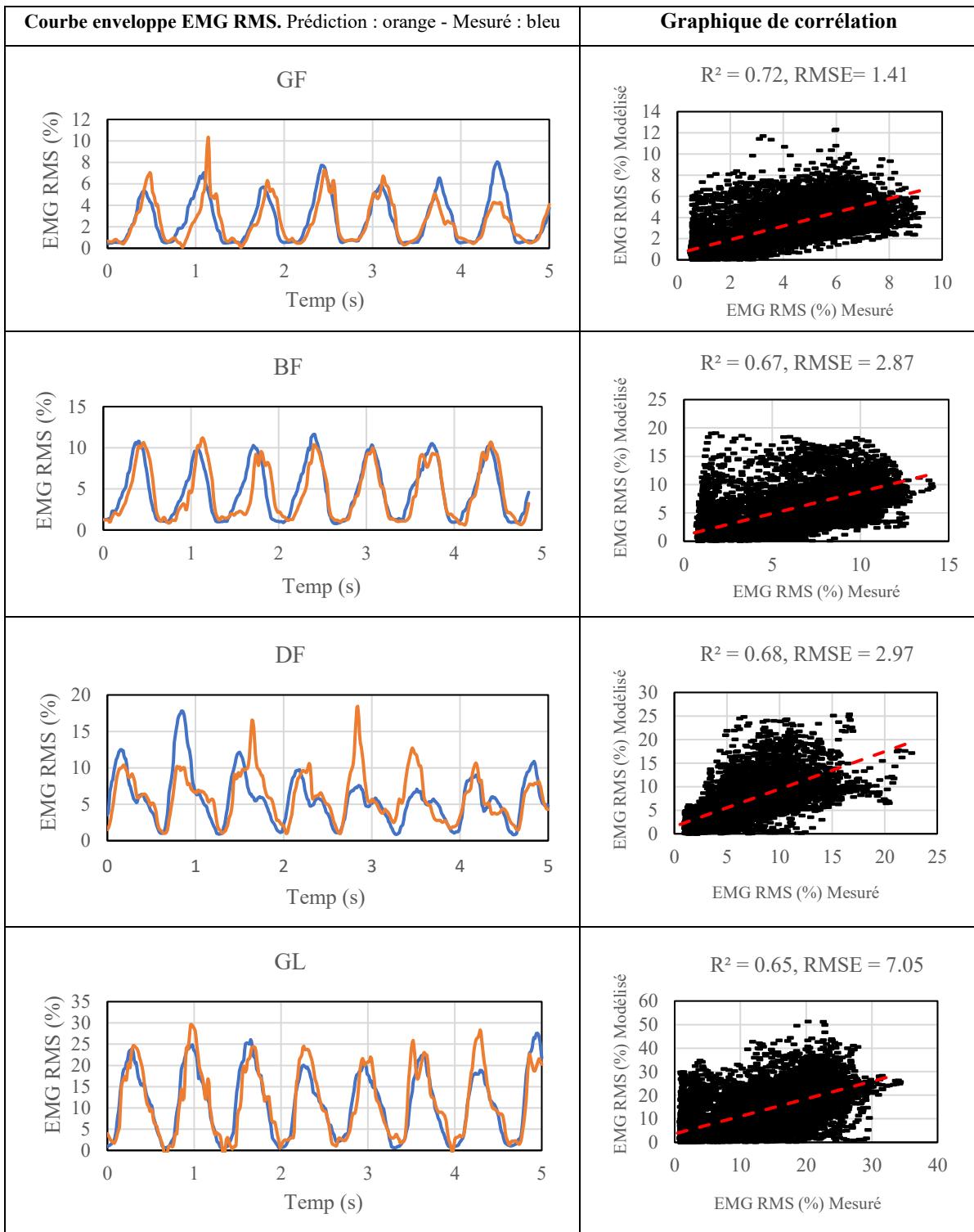
Sujet 6

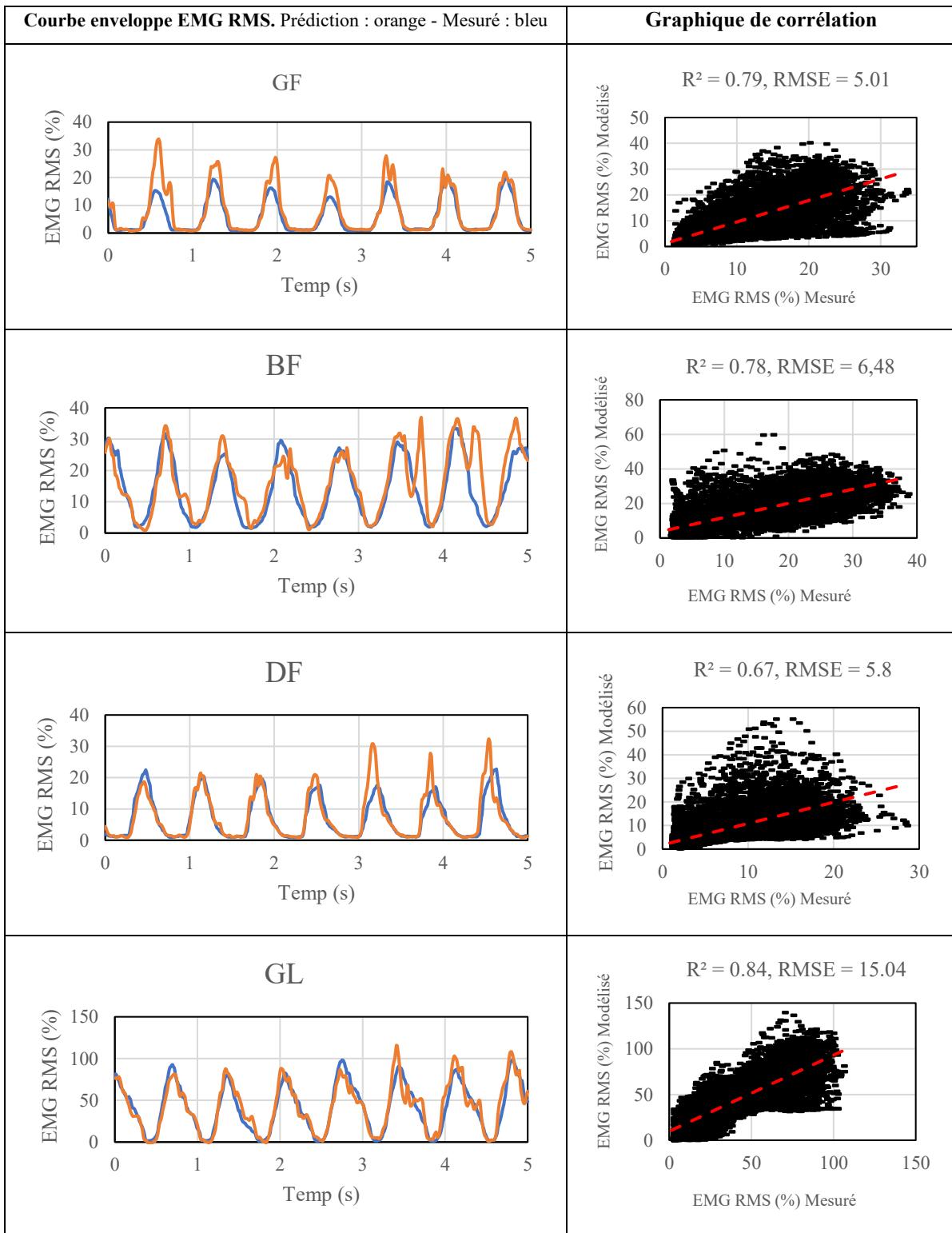
ANNEXE N - Prédiction courbe enveloppe EMG RMS 150 W

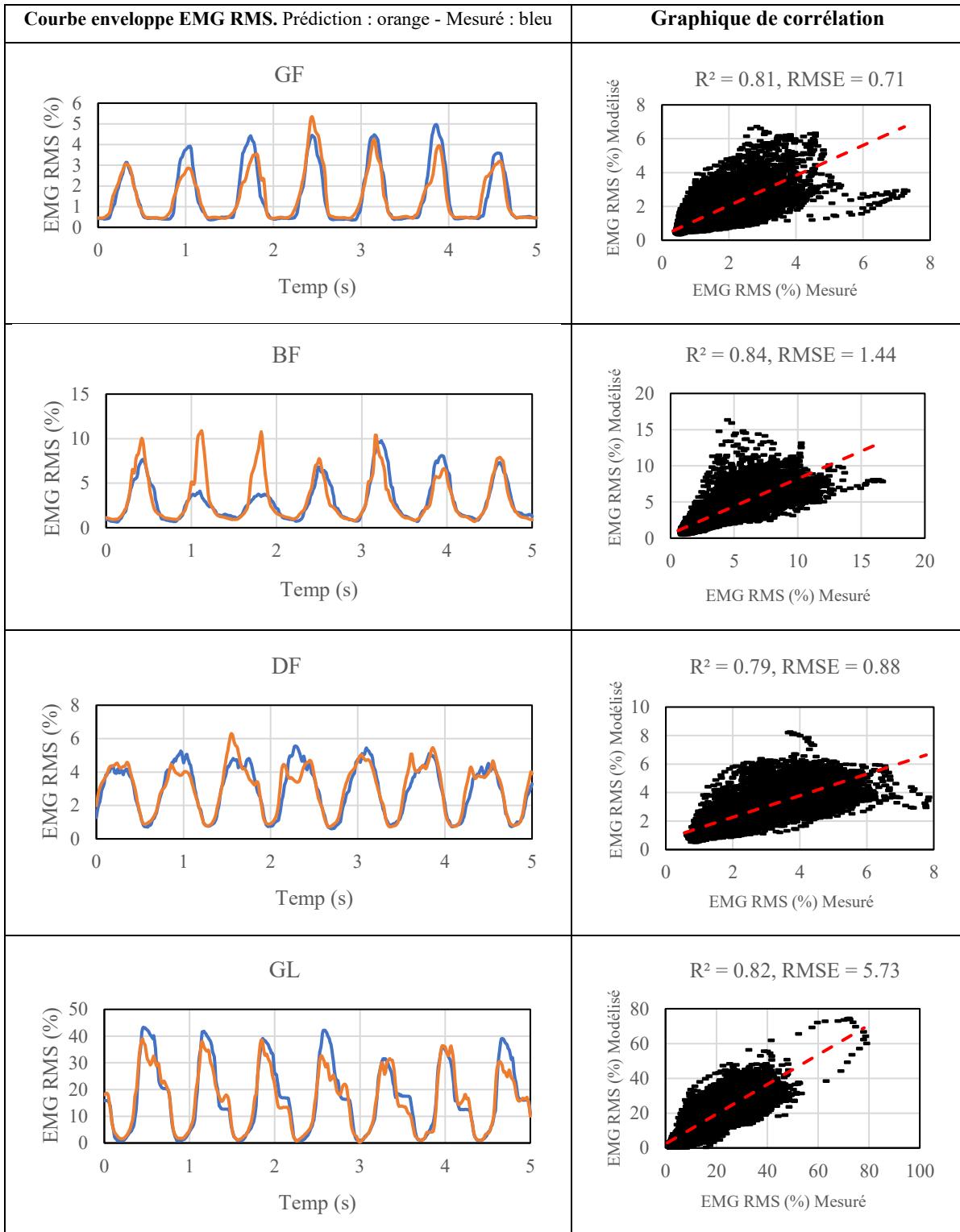
Sujet 1

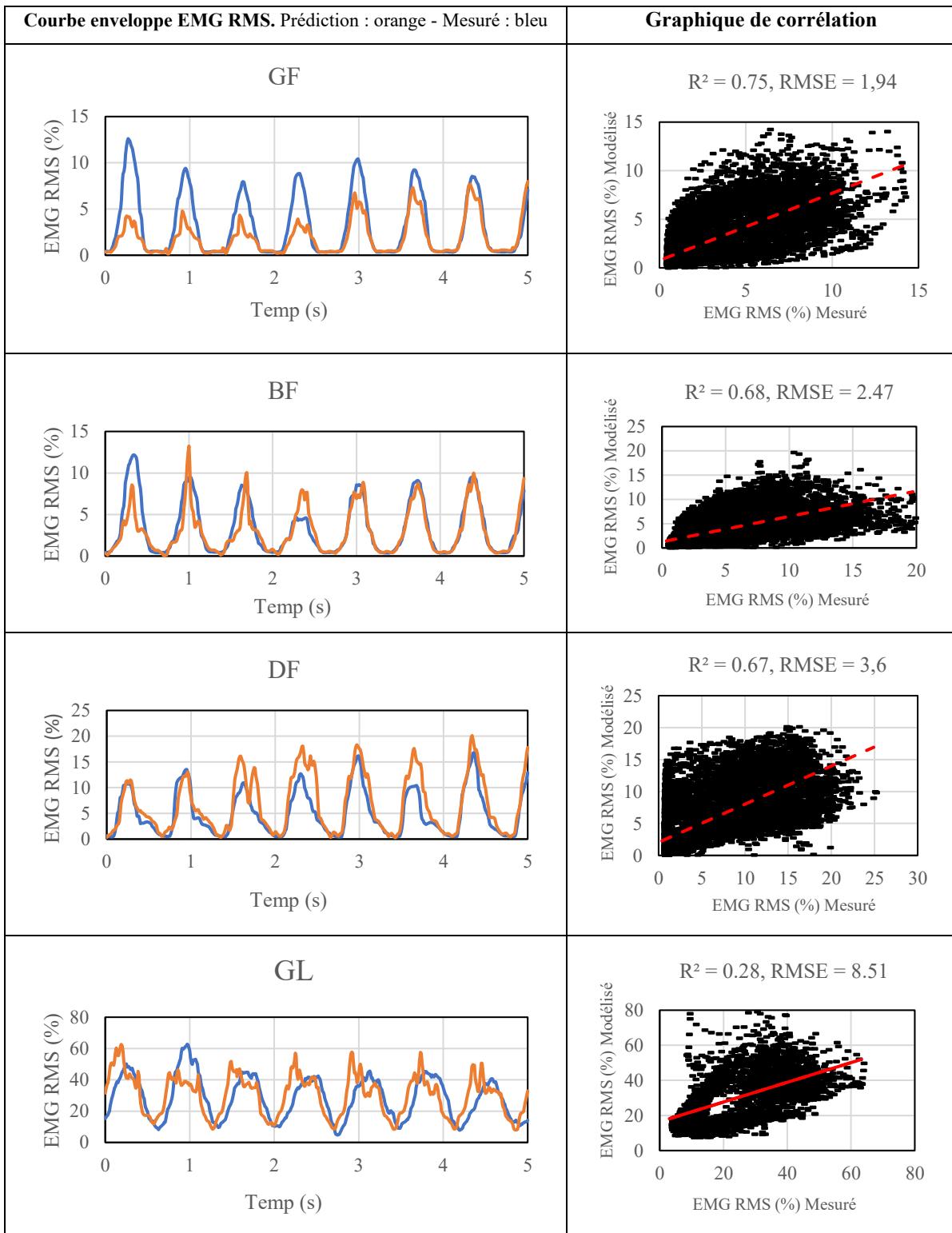


Sujet 2

Sujet 3

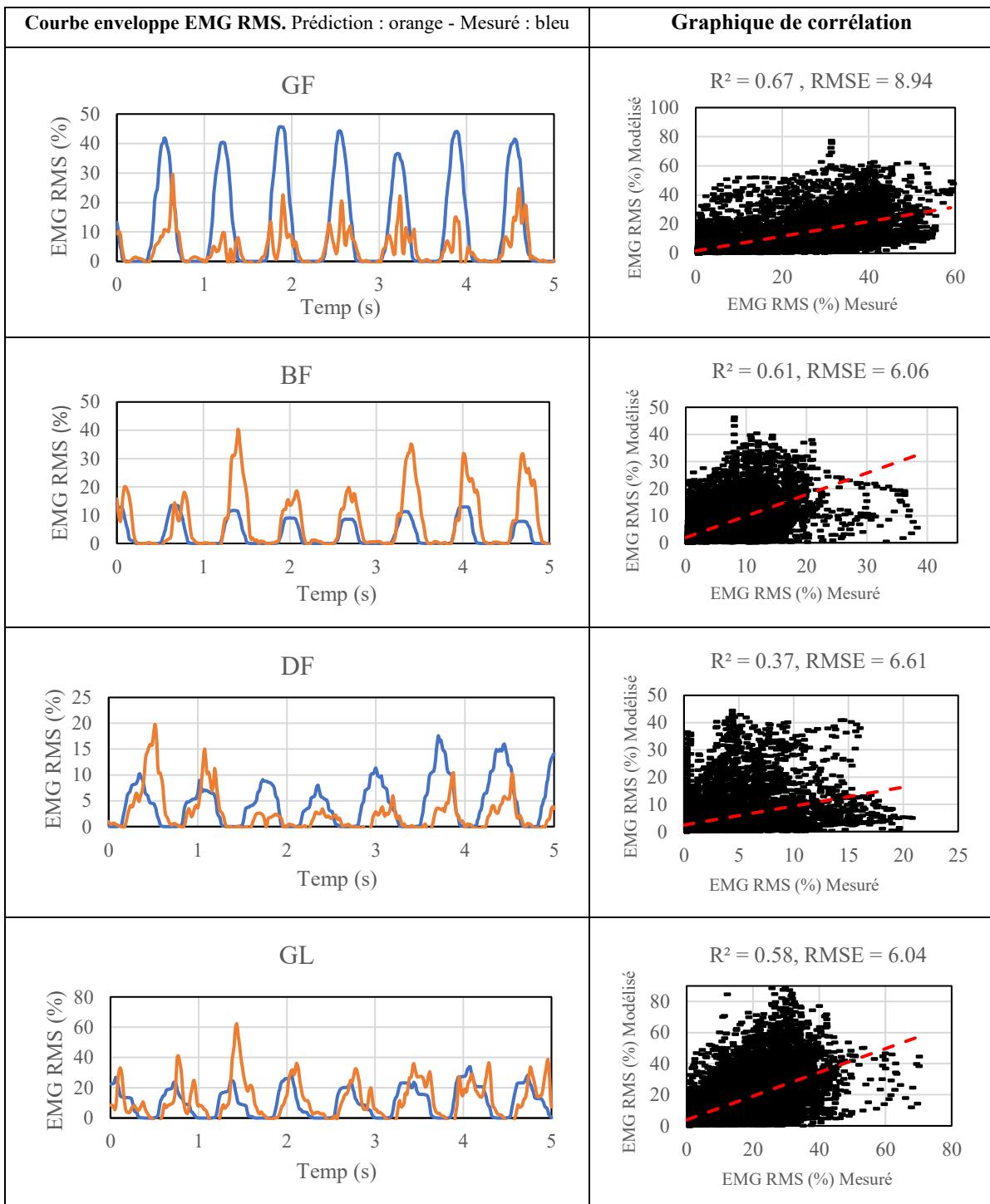
Sujet 4

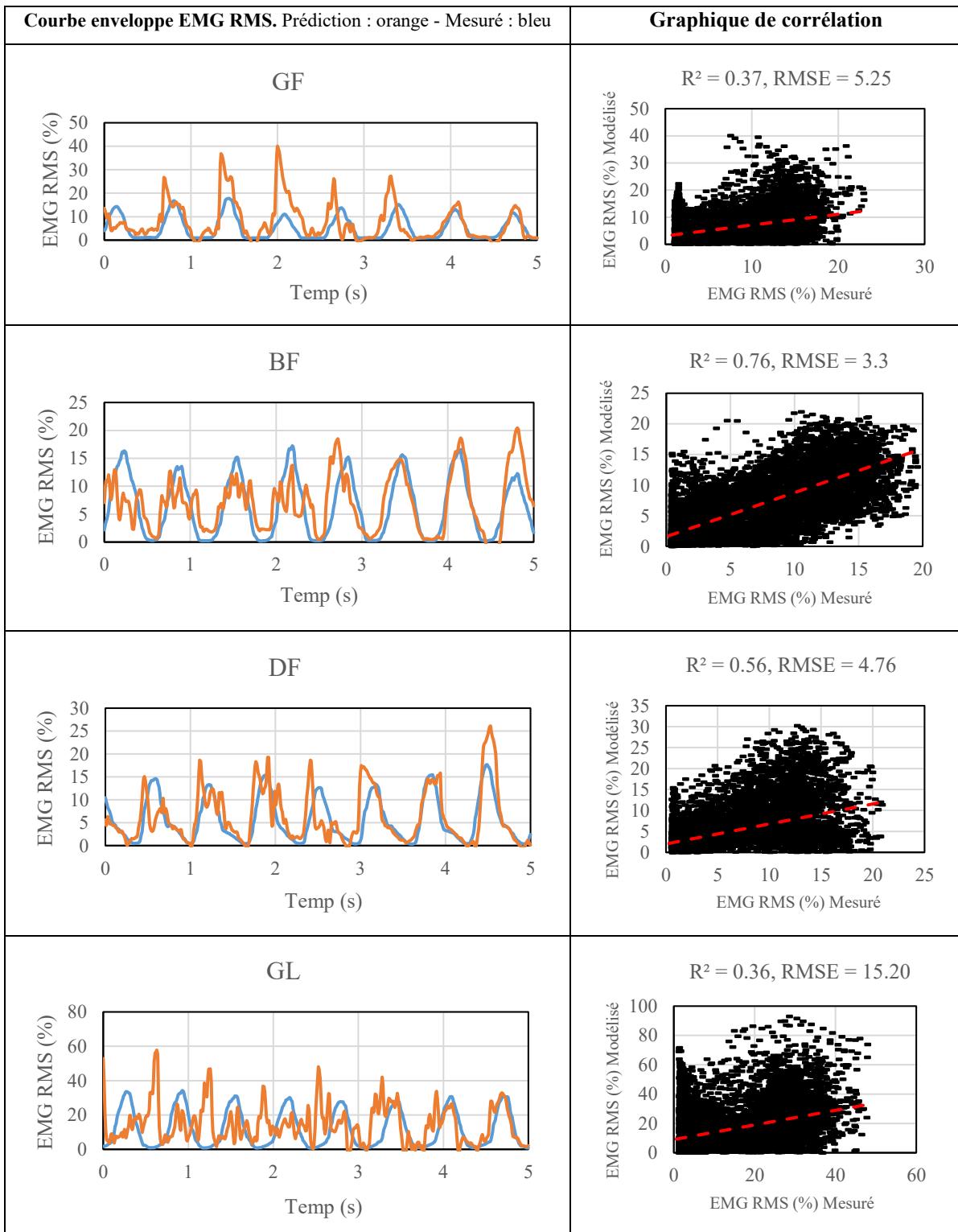
Sujet 5

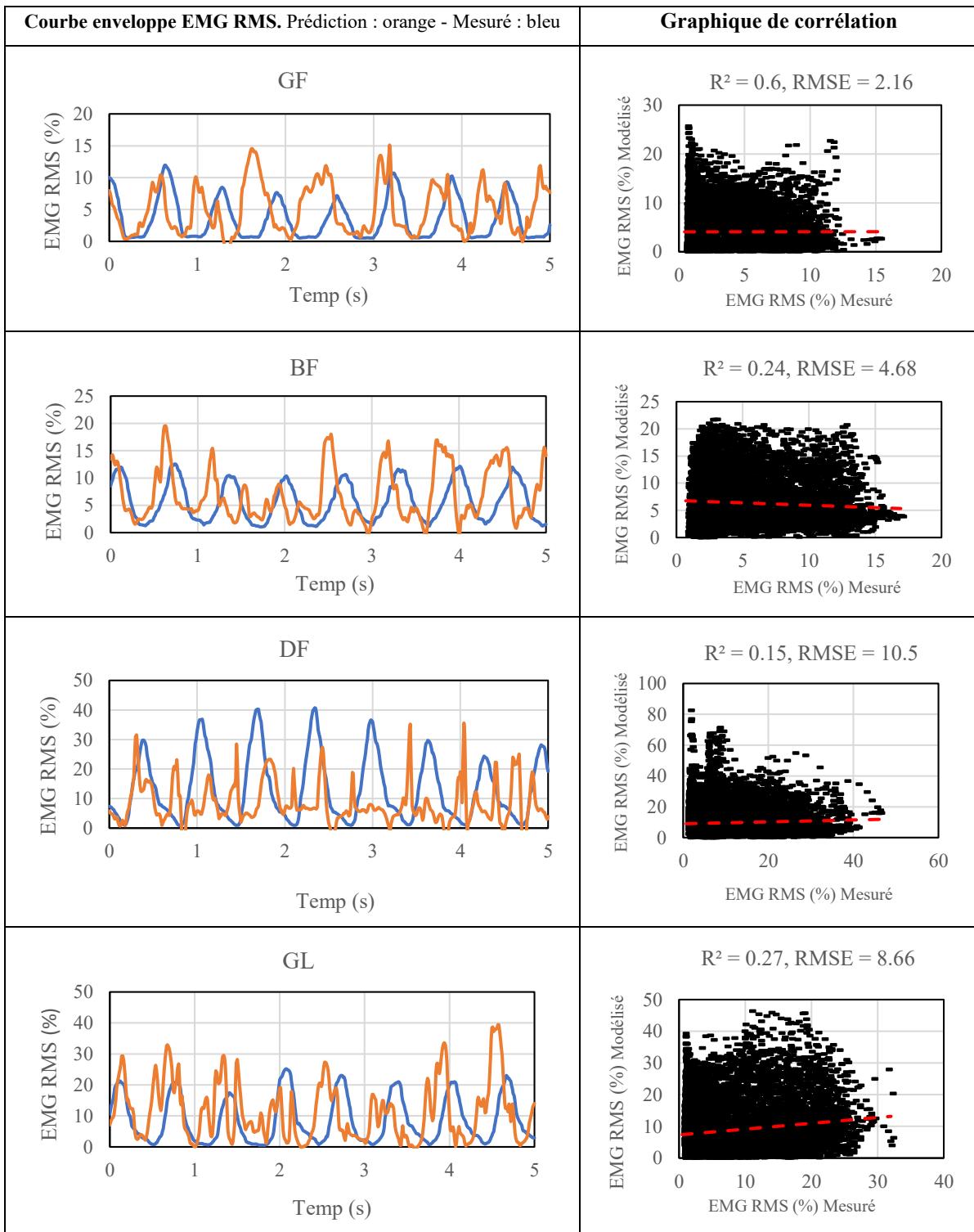
Sujet 6

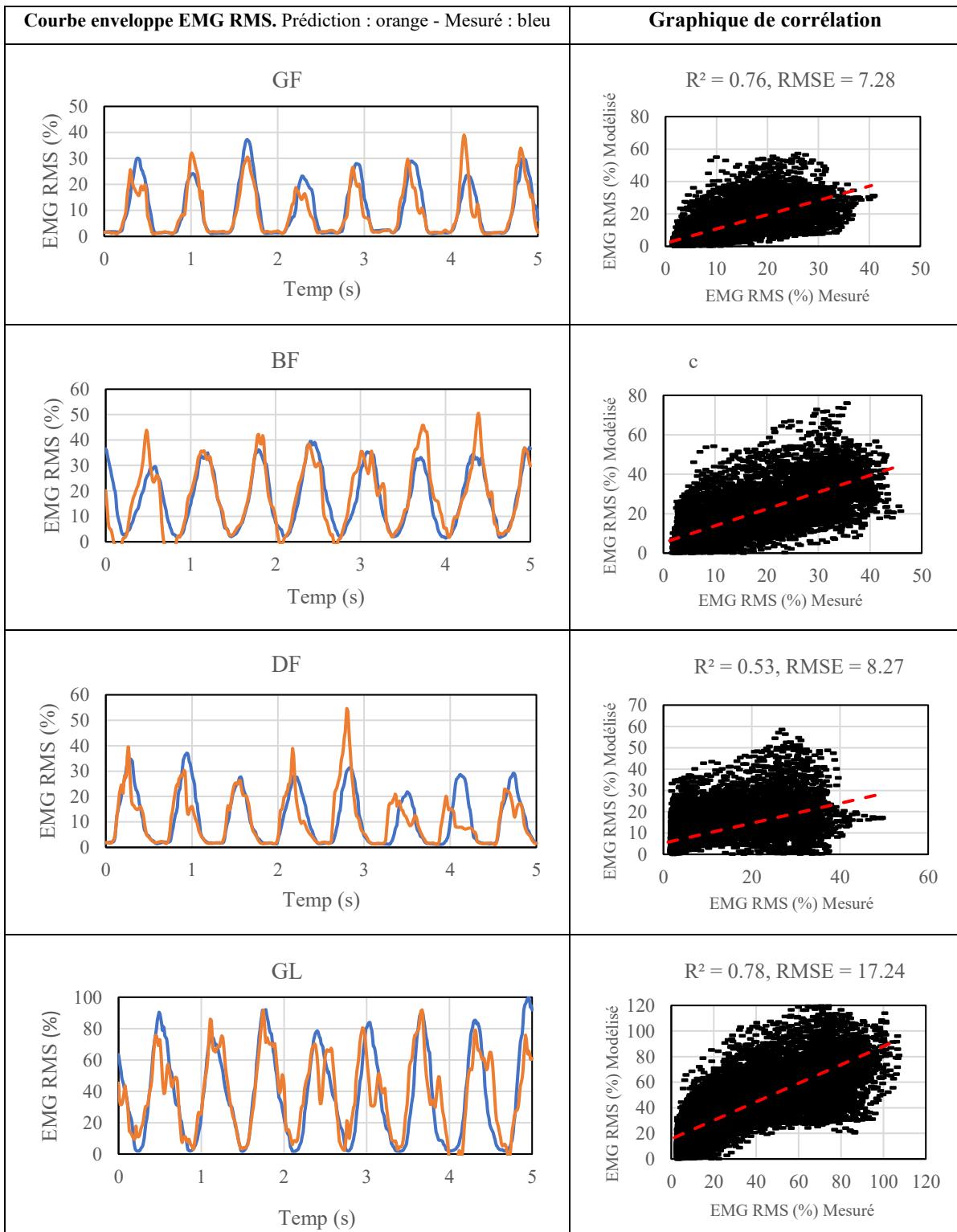
ANNEXE O - Prédition courbe enveloppe EMG RMS 200 W

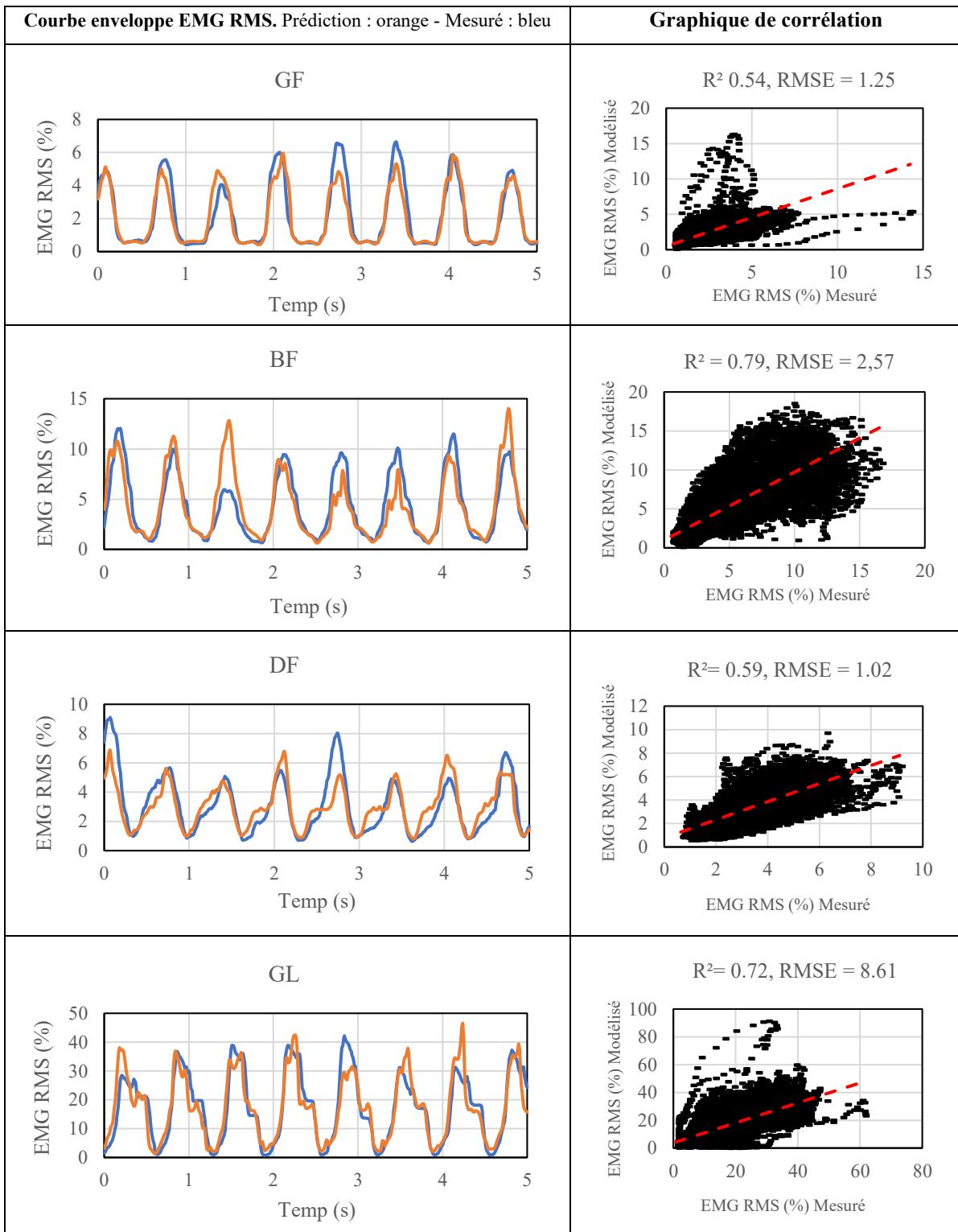
Sujet 1

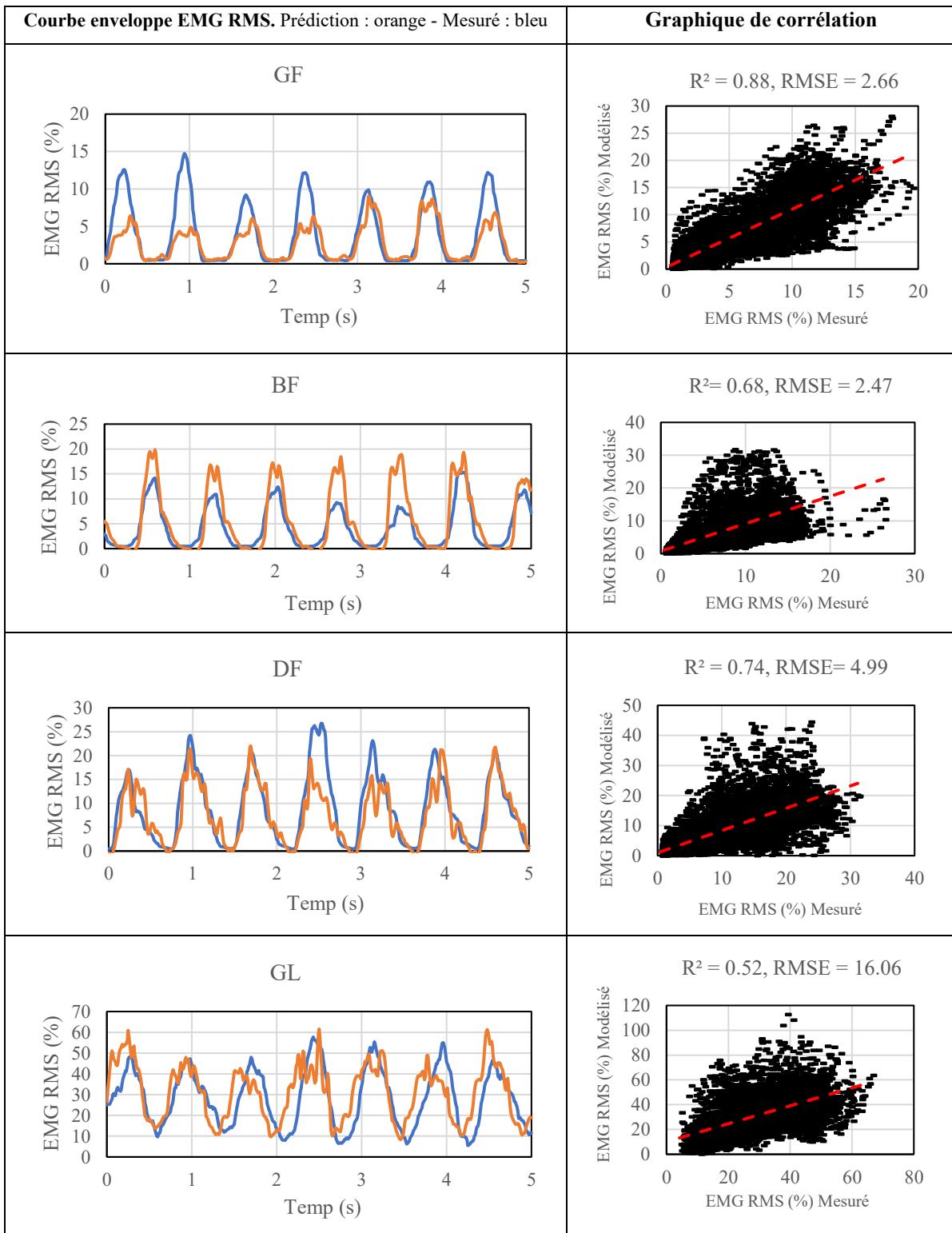


Sujet 2

Sujet 3

Sujet 4

Sujet 5

Sujet 6

ANNEXE P - Publication soumise**PREDICTING PEDALLING FORCES BASED ON LOWER LIMB
JOINT KINEMATICS WITH A FEED-FORWARD NEURAL
NETWORK**

Andrés Torres ^{1 3}, François Nougarou ^{2 3}, Frédéric Domingue ^{1 3}

¹ Laboratoire de technologies & d'innovation pour la performance sportive (L-Tips)

² Laboratoire de signaux et systèmes intégrés (LSSI)

³ Département de génie électrique et génie informatique, Université du Québec à Trois-Rivières, Trois-Rivières, Québec, Canada.

Abstract

This study aims to predict pedalling forces using a machine learning technique. Eleven cyclists pedalled for 5 minutes at power levels of 100 W, 150 W and 200 W. A feed-forward neural network per power value, with 5 hidden layers and 2000 neurons per layer, can predict the time series of pedalling forces. The lower limb joint kinematics in the sagittal plane, power, and pedalling cadence were used as input variables. The results demonstrate that using a machine learning technique, it is possible to predict with good accuracy and low error the time series pedalling forces.

Keywords: Cycling biomechanics; Modelling; Machine learning; Kinematics; pedalling forces

ANNEXE Q - Publication à soumettre**PREDICTING MUSCULAR RECRUITMENT DURING BICYCLE
PEDALLING WITH RECURRENT NEURAL NETWORK**

Andrés Torres ^{1 3}, Louis Coqueret ¹, François Nougarou ^{2 3}, Frédéric Domingue ^{1 3}

¹ Laboratoire de technologies & d'innovation pour la performance sportive (L-Tips)

² Laboratoire de signaux et systèmes intégrés (LSSI)

³ Département de génie électrique et génie informatique, Université du Québec à Trois-Rivières, Trois-Rivières, Québec, Canada.

Abstract

This study aimed to predict the muscular recruitment of four muscles of the lower limb during pedalling with kinematics data, forces on the pedal, power and cadence using a Long Short-Term Memory (LSTM) neural network. Six cyclists performed three submaximal tests of 5 minutes at three powers and free cadence. Joint kinematics, force on the pedals, power and cadence were used as the inputs for the neural network. The muscle activity envelope curve of gluteus major, biceps femoris, rectus femoris and gastrocnemius lateral were used as targets. The developed model can predict with good accuracy the muscle recruitment of the lower limb muscles during pedalling.

Keywords: Cycling biomechanics; Modelling; Machine learning; Motion capture; EMG; Pedalling forces; Neural Network; LSTM