

NNT : 2017SACLE006



THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PARIS-SACLAY
PRÉPARÉE À L'UNIVERSITÉ D'EVRY VAL D'ESSONNE
ET À LA FONDATION ELLEN POIDATZ

Ecole doctorale N°580
Sciences et Technologies de l'Information et de la Communication
Spécialité de doctorat: Traitement du signal et des images

par

M. OMAR ANTONIO GALARRAGA CASTILLO

Simulation of Surgery Effect on Cerebral Palsy Gait by Supervised
Machine Learning

Thèse présentée et soutenue à Evry, le 30 mars 2017.

Composition du Jury :

Mme ISABELLE BLOCH	Pr	Institut Télécom ParisTech	Présidente
M. PIERRE-YVES GUMÉRY	Pr	Polytech Grenoble	Rapporteur
M. ELMAR LANG	Pr	Universität Regensburg	Rapporteur
M. STÉPHANE ARMAND	CR HDR	Université de Genève	Examineur
M. VINCENT VIGNERON	MDC HDR	Université d'Evry Val d'Essonne	Directeur
Mme BERNADETTE DORIZZI	Pr	Institut Télécom SudParis	Co-directrice
M. ERIC DESAILLY	DR	Fondation Ellen Poidatz	Co-encadrant
M. NÉJIB KHOURI	PH	Hôpital univ. Necker-Enfants Malades	Invité

Acknowledgments

First of all I would like to thank my tutors whose advices from three totally different points of view, even though sometimes hard to consent and apprehend, crucially enriched this work, as well as my professional and personal education. I will be always grateful to them: Dr. Vincent Vigneron, for being always available for any kind guidance, and especially for sharing part of his vast knowledge in mathematics and statistical models. Prof. Bernadette Dorizzi, for having advised me since my master's studies and for her straight and incredibly accurate observations (sometimes tough) derived from many years of successful experience in machine learning research. Dr. Eric Desailly, director of the *Sim – PC²* project, for being always available and efficient, and for teaching me a small part of his great expertise in clinical gait and motion analyses.

A special thank to MD Néjib Khouri, who was also actively implicated in this work sharing his vast knowledge in surgery and gait analysis in cerebral palsy and his infinite energy.

I am grateful to Prof. Dr. Pierre-Yves Guméry, Prof. Dr. Elmar Lang, Prof. Dr. Isabelle Bloch and Dr. Stéphane Armand for having accepted to be members of the jury.

Thanks to the Ellen Poidatz Foundation for welcoming me as one of their own during the past three years (and more). I would like to thank particularly all the current and former medical and technical staff of the movement analysis laboratory: Daniel Yéprémian, Farid Hareb, Camille Thevenin-Lemoine, Lionel Lejeune, Djilali Bouchakour and Anne-Laure Guinet.

I am grateful to the Île de France region and Ellen Poidatz Foundation for funding this research, as well as to Bettencourt-Schueller Foundation, financial partner of the *Sim – PC²* project. I also thank the scientific partners of the project: University of Evry Val d'Essonne, Ellen Poidatz Foundation, TELECOM SudParis, Paris-Saclay University and Necker-Enfants Malades Hospital.

I would like to thank all the children and their families who accepted the usage of their data for this research: they are very courageous for fighting to improve their condition.

Thanks to my PhD colleagues Alexandra, Luc and Abdou, as well as to Latifa (RIP).

I am thankful to my friends in Paris and around, who are a big support for me being far from my family and friends outside France. I am sorry I was absent for a little while.

I am thankful to my family in Venezuela and around the world, that supported me from the distance: my aunts, my uncles, my cousins (old and young), my brother Tomás, my godsons Marcelo and Asdrúbal Aníbal, and my family in law. I am sorry I do not name all of them, but we are fortunately a really big family. And of course I'd like to thank my parents, Omar and Belkis, who even came to France to help me during the writing of this document. All I have achieved in life has been following their example and thanks to their support.

A big thank to my loving wife Valentina, who helped throughout my PhD studies. I know it was difficult to be at my side during this period, but she has remarkably managed to do it.

This PhD thesis represents much more than just the fruit of the past three years of work. In this regard, I would like to thank all of my former teachers, professors and colleagues, who have contributed in a way or another to my education.

Merci beaucoup à tous ! ¡Muchas gracias a todos!

Contents

1	Introduction	1
1.1	Objectives	3
1.2	Organization	3
I	Context and Data	5
2	State of the Art	7
2.1	Cerebral Palsy and Gait Troubles	7
2.2	Clinical Gait Analysis	9
2.3	Kinematics	12
2.4	Physical Examination	16
2.5	Single-Event Multilevel Surgery	18
2.6	Classification of Gait Patterns in Cerebral Palsy	20
2.7	Surgery outcome prediction	21
3	Data Description and Conditioning	23
3.1	Kinematic Data	24
3.2	Physical Examination Data	29
3.3	Surgical Data	32
3.4	Preoperative Kinematics and Physical Examination Data by Surgical Categories .	33
3.5	Surgery Outcome	35
3.6	General Discussion	37
4	Feature Extraction and Dimensionality Reduction	41
4.1	Curve Fitting for Kinematics	41
4.1.1	Harmonic approximation	43
4.1.2	Periodic splines approximation	45
4.1.3	Comparison between periodic splines and harmonic approximation	46
4.2	Variable Selection	48
4.3	Physical Examination Missing Data Imputation	49
4.4	Dimensionality Reduction by Principal Component Analysis	52
4.4.1	Method	55
4.4.2	Results	56
4.5	General Discussion	59

II Predicting Surgery Effect on Cerebral Palsy Gait using Supervised Machine Learning	65
5 Neural Networks for Initial Contact Prediction	69
5.1 Nonlinear regression with Neural Networks	69
5.2 Predicting postoperative knee flexion and pelvic tilt at initial contact of children with cerebral palsy	72
6 Predicting Postoperative Gait Using Multiple Linear Regression	79
6.1 Multiple linear regression	79
6.2 Predicting Postoperative Gait in Cerebral Palsy	81
7 Ensemble Learning With Linear and Nonlinear Regression	89
7.1 Arriving to Ensemble Learning Methods	89
7.2 Predicting Postoperative Gait in Cerebral Palsy with Ensemble Learning	91
8 Comparison of the Methods	109
8.1 Regression using Fourier gait parameters	109
8.2 Performance comparison	109
8.3 General Discussion	111
9 Conclusions and Perspectives	113
9.1 Limitations	116
9.2 Perspectives	117
References	119
List of Publications	127
Appendix A: Physical Examination Data Interface	130

List of Figures

1.1	Sketch of the simulation of the effect of the surgery on gait for children with cerebral palsy.	1
2.1	Topographical classification of CP	8
2.2	Treatments for gait troubles in CP.	9
2.3	Standard gait cycle	10
2.4	Clinical Gait Analysis laboratory	11
2.5	Rigid body gait model: segments and joints	13
2.6	Cardan angles rotation	14
2.7	Joint angles curves during walking.	15
2.8	Measurement of popliteal angle	16
2.9	Examples of outcome of orthopaedic surgery	19
3.1	Example of mean cycle computation	24
3.2	Preoperative Kinematic Data.	26
3.3	Postoperative Kinematic Data.	27
3.4	Considered preoperative kinematics	28
3.5	Preoperative height and weight distributions.	30
3.6	Preoperative hip physical examination distributions.	31
3.7	Preoperative knee physical examination distributions.	31
3.8	Preoperative ankle physical examination distributions	32
3.9	Preoperative kinematics per surgical procedure. 3.9a Average of limbs that had each surgical procedure (solid blue) and average of limbs that did not have the considered procedure (dashed blue). 3.9b Independence test between averages with and without each surgery. 0 (black) means no significance of the kinematic variable and 1 (white) is the maximal significance.	36
3.10	Statistical significance of physical examination variables on surgery selection. 0 (black) means no significance and 1 (white) is the maximal significance (labels in table 3.2).	37
3.11	Average variation and effect of surgical procedures. 3.11a Average kinematic variation (Postoperative - Preoperative) in a 2-standard deviation band. 3.11b Statistical effect of surgical procedure on kinematic variables. 1 (white) corresponds to a significant effect and 0 (black) corresponds to no statistical effect.	38
4.1	Example of trigonometric kinematic curves approximation for a random limb . .	44
4.2	Harmonic approximation MSE per subject and number of parameters	44
4.3	Optimal AIC_c models per subject for harmonic approximation	45
4.4	Example of periodic Spline fitting for a random limb	47
4.5	Gram-Schmidt orthogonalization.	49

4.6	Distribution of popliteal angle measure at V3 before and after imputation	52
4.7	Cumulative inertia of PCA projection	57
4.8	Projection of subjects (limbs) on the principal plane	57
4.9	Variable projection on principal planes	58
4.10	Average RMSE over all patients per PCA dimension.	59
4.11	Variance ratio of kinematics after PCA	60
4.12	Variance ratio of physical examination data after PCA	61
5.1	Multilayer Perceptron	70
A.1	Web-like interface for inputing physical examination data	130

List of Tables

2.1	Physiological interpretation of kinematic angles	14
2.2	A non-exhaustive list of physical examination measures	17
3.1	Age of patients at preoperative CGA, surgery and postoperative CGA.	23
3.2	List of the considered physical examination variables. For each limb, all the corresponding contralateral limbs' variables are also considered, except for height and weight.	29
3.3	Considered surgical procedures categories and their frequencies in the database. .	33
3.4	Examples of surgical procedures per surgical category.	34
4.1	Minimal number of approximation parameters per kinematic angle and average <i>MSE</i> thresholds for the harmonic approximation	45
4.2	Minimal number of approximation parameters per kinematic angle and average <i>MSE</i> thresholds for periodic Splines approximation	46
4.3	Comparison between periodic Splines and Fourier approximations.	48
4.4	Selected preoperative variables with probe technique for postoperative knee flex- ion at two consecutive points	50
8.1	Performance comparison of the prediction methods	110

Chapter 1

Introduction

This PhD thesis is part of the **SiM-PC²** project (*Simulation prédictive de la Marche et du bénéfice fonctionnel Post Chirurgical dans la Paralyse Cérébrale*, or predictive simulation of gait and functional benefit in cerebral palsy), which is financially supported by the Île de France region (PICRI2012), in association with Bettencourt-Schueller and Ellen Poidatz Foundations. The SiM-PC² project, directed by Dr. Eric Desailly, is a collaborative work associating academic partners (Paris-Saclay University by University of Evry Val d'Essonne and Télécom SudParis Institute, as well as Paris VI Pierre-and-Marie-Curie University), clinical partners (Necker-Enfants malades Hospital, Ellen Poidatz Foundation) and representatives of the civil society (Fondation Motrice, SESEP, Institut de Motricité Cérébrale).

This multidisciplinary project, which involves skills in statistical signal processing, classification, feature selection, robotics, biomechanics, clinical gait analysis and orthopaedics, aims to develop gait simulations and regressions under constraints (see figure 1.1). These tools will operate a multimodal database (kinematics, kinetics, EMG, imaging, clinical data) to allow the simulation of the effect of surgery on gait and function of children and adolescents with cerebral palsy. It represents a continuation of several previous works of CGA-based therapeutic decision support methods that have been implemented in the Ellen Poidatz Foundation (Desailly, Yepremian, Khouri, Hareb, Lejeune, Bouchakour, Sardain & Lacouture, 2009; Desailly, Sebsadji, Yepremian, Djemal, Hoppenot & Khouri, 2013).

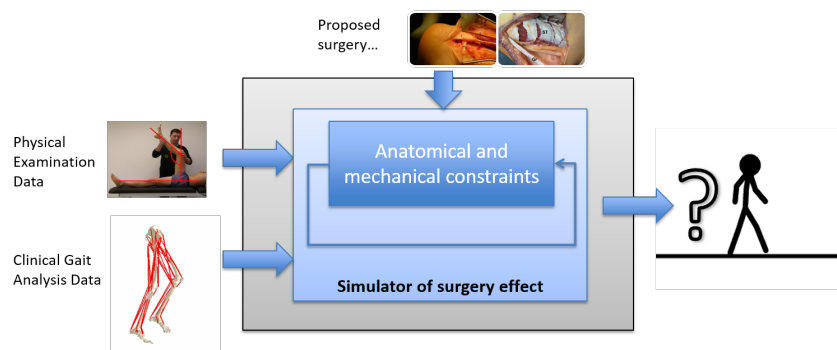


Figure 1.1 – Sketch of the simulation of the effect of the surgery on gait for children with cerebral palsy. [Courtesy of E. Desailly. Modified]

This thesis corresponds to the statistical simulation of surgery on gait of children with cerebral palsy (CP), using machine learning techniques, starting from the database organization to the prediction reporting and performance assessment, passing by the data conditioning, modeling, feature extraction and regression analysis. Machine learning refers to the utilization of an algorithm to modify the parameters of a mathematical model based on some training data

(Duda, Hart & Stork, 2000, p.16). There are several types of machine learning, the principal ones being supervised learning and unsupervised learning or clustering. In this study we will use supervised machine learning, which refers to the classification or regression when classes or labels are known for each training sample, such that the learning algorithm seeks to minimize the error between target labels and model outputs. In this case, target outputs correspond to postoperative gait descriptors. Stages of supervised machine learning are preprocessing, feature extraction and classification or regression. All of these stages will be developed during this document. Trained models can then be used for automatic classification, estimation of a variable value with respect to other variables, or prediction of the future value of a variable.

This work follows previous collaborations between the Ellen Poidatz Foundation UNAM laboratory (*Unité d'Analyse du Mouvement*, or Motion Analysis Unit) and IBISC laboratory (*Informatique, Biologie Intégrative et Systèmes Complexes*, or Information, Integrative Biology and Complex Systems) of the University of Evry Val d'Essonne (Sebsadji, Khouri, Djemal, Yepremian, Hareb, Hoppenot & Desailly, 2012; Sebsadji, 2011; Desailly et al., 2013; Desailly, Sebsadji, Yepremian, Hareb & Khouri, 2012) on the evaluation of surgery effect on cerebral palsy gait. IBISC is a laboratory with large expertise in medical applications of signal and image processing (Vigneron, Kodewitz, Tome, Lelandais & Lang, 2016; Sacco, Suda, Vigneron & Sartor, 2015; Kodewitz, Lelandais, Montagne & Vigneron, 2013). The Ellen Poidatz Foundation manages medical and medico-social institutions and promotes research in pediatric care for different types of handicaps. UNAM is a clinical and research gait laboratory specialized in decision making and assessment of surgery effect on gait of children and adolescents with cerebral palsy. The SAMOVAR laboratory (*Services répartis, Architectures, MOdélisation, Validation, Administration des Réseaux*, or Distributed Services, Architectures, Modeling, Validation and Network Administration) completes the collaborative team of this specific thesis to addition its expertise on statistical pattern recognition, especially on biometrics systems (Allano, Dorizzi & Garcia-Salicetti, 2012; Houmani, Garcia-Salicetti & Dorizzi, 2012; El-Yacoubi, Shaiek & Dorizzi, 2011). In this work, mechanical constraints are not considered, except for the intrinsic physical constraints that are present in the analyzed data. Throughout this work, we confront scientific problems such as pathological gait modeling, surgery modeling, statistical small data, dimensionality reduction, missing data, feature extraction, regression, three dimensional computer graphics representation and performance assessment.

The choice of a surgery in cerebral palsy is a complex task that is usually based on clinical gait analysis (CGA) and physical examination. Three dimensional gait analysis is a technique that allows to quantitatively characterize gait. It has been proven that CGA is of great help for clinicians in treatment assessment (Wren, Gorton III, Öunpuu & Tucker, 2011). Even though the outcome of the orthopaedic surgery in CP is generally good and sometimes outstanding, it is difficult to have an a priori estimation of the most likely functional outcome. Given the high costs and risks of surgery, characterization of treatment outcome based on CGA data has been of great interest in the scientific field, especially in medical applications. The advances in movement analysis and the development of CGA databases, plus the advances in computational power, machine learning methods and data mining, make possible to apply advanced statistical techniques for decision-making. In this regard, recently some decision-making tools have been developed to predict the outcome (Desailly et al., 2013; Reinbolt, Fox, Schwartz & Delp, 2009; Schwartz, Rozumalski, Truong & Novacheck, 2013; Arnold, Liu, Schwartz, Öunpuu & Delp, 2006; Hicks, Delp & Schwartz, 2011; Niler, Richards & Miller, 2007), however the estimations are whether qualitative or the number of predicted parameters or considered surgical procedures is insufficient to preview the most likely resulting gait.

Simulations of the surgery effect on CP gait would be of great help for both the clinician and the patient. For the clinician, it would help to adjust the surgery in order to find an optimal

treatment, avoid unnecessary surgeries and better communicate with the patients with respect to the treatment itself and its likely outcome. For the patient, it would help to better understand the treatment and its likely outcome, avoiding false expectations. All of this would be of great societal impact given that CP is one of the most common motor disability.

1.1 Objectives

In this above described context, the general objective of this work is to use statistical machine learning techniques to conceive a system able to predict postoperative kinematic curves of children with cerebral palsy based on preoperative physical examination and 3-D gait analysis, and a proposed surgery plan. The purpose of such system is to have an a priori estimation of the most likely outcome of the treatment in order to improve outcome comprehension and to be used as decision-making tool for orthopaedic surgery in cerebral palsy.

The specific objectives include:

- Organizing a database of clinical gait data, physical examination and surgical data of children with cerebral palsy that have been operated.
- Selecting and conditioning the different types of data issued from CGA, physical examination and surgery.
- Conceiving, developing and comparing statistical models to simulate the effect of surgery on cerebral palsy gait.
- Determining the most important variables among all the available data utilizing different feature extraction methods.
- Developing of different regression techniques in order to find a mathematical relation between postoperative and preoperative data, given the surgical data.
- Assessing the performances of all the tested prediction methods in order to compare them.
- The system output (prediction) should represent the most likely postoperative gait with quantitative variables, and be suitable for clinical interpretation and for 3-D animation.

1.2 Organization

This document is divided in 2 principal parts plus the conclusion and the present introduction. Part I is dedicated to the context, background and data description and conditioning, and is composed of three chapters. Chapter 2 is about the state of the art in cerebral palsy gait analysis: gait troubles, treatment assessment and orthopaedic surgery, as well as related work for predicting surgery outcome in cerebral palsy. In chapter 3, the different data that were considered for the experiments are presented: kinematics, physical examination and surgery model. The preprocessing stage for kinematic data is also presented. Chapter 4 refers to the different feature extraction methods that were considered for the experiments: curve fitting, variable selection and principal component analysis. In addition, the imputation of physical examination missing data is presented.

Part II is dedicated to the different experiments for predicting postoperative gait, as well as the different supervised machine learning methods that were considered for regression. This second part is composed of 4 chapters. Chapter 5 is about the prediction of postoperative gait at initial

contact using neural networks with variable selection. Chapter 6 presents the prediction of postoperative kinematic curves using multiple linear regression and principal component analysis. The prediction of kinematic curves using ensemble learning with principal component analysis and linear and nonlinear regression is presented in chapter 7. The above methods, as well as other methods, are compared in chapter 8.

Finally, conclusions and perspectives are presented in chapter 9.

At the end of this document, references and related publications are listed.

Part I

Context and Data

Chapter 2

State of the Art

Contents

2.1	Cerebral Palsy and Gait Troubles	7
2.2	Clinical Gait Analysis	9
2.3	Kinematics	12
2.4	Physical Examination	16
2.5	Single-Event Multilevel Surgery	18
2.6	Classification of Gait Patterns in Cerebral Palsy	20
2.7	Surgery outcome prediction	21

2.1 Cerebral Palsy and Gait Troubles

According to the international definition in (Rosenbaum, Paneth, Leviton, Goldstein, Bax, Damiano, Dan & Jacobsson, 2007):

“Cerebral palsy describes a group of permanent disorders of the development of movement and posture, causing activity limitation, that are attributed to nonprogressive disturbances that occurred in the developing fetal or infant brain. The motor disorders of cerebral palsy are often accompanied by disturbances of sensation, perception, cognition, communication, and behavior; by epilepsy, and by secondary musculoskeletal problems.”

Cerebral Palsy (CP) refers to a group of neurological disorders from brain damage during development (pregnancy, birth or first years of life), that affect human movement, balance and posture. CP generally entails muscle and bone deformities and typically manifests by gait troubles, where some patients are even unable to walk. The prevalence of CP has remained constant in recent years at 2.11 per 1000 live births (Oskoui, Coutinho, Dykeman, Jetté & Pringsheim, 2013), making it the most common motor disability in childhood (Accardo, 2007, p. 17).

CP can be categorized according to several criteria, such as physiology, etiology, topography, severity, level of treatment, among others. From a topographical point of view, CP is classified according to the limbs that are affected by the disease and the categories are: monoplegia, diplegia, hemiplegia, triplegia, quadriplegia and double hemiplegia (Öunpuu, Thomason, Harvey & Graham, 2009) [see Figure 2.1]. In this work we will only analyze lower limbs, thus there will be two general topographical categories for patients:

1. Unilateral: which includes both hemiplegic and monoplegic patients.

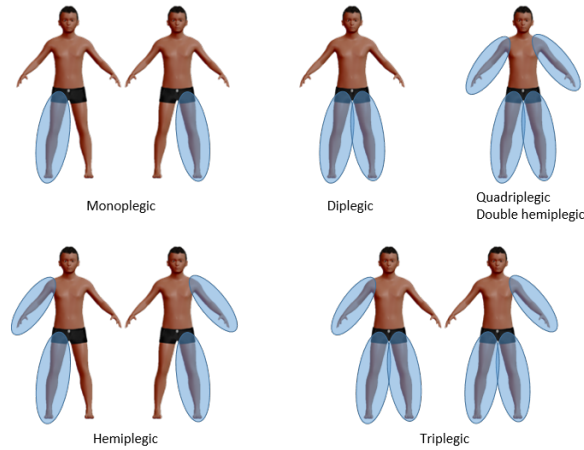


Figure 2.1 – Topographical classification of CP (depending on the affected limbs). The affected limbs are covered by the blue oval. The difference between quadriplegic and double hemiplegic is that in the first case lower limbs are more affected than upper limbs and the opposite on the second case.

2. Bilateral: which refers to diplegic, triplegic and quadriplegic or double hemiplegic patients.

On the other hand, CP can also be classified physiologically in: spastic, athetotic, rigid, ataxic, tremor, mixed and unclassified (Öunpuu et al., 2009). Since spastic CP is the most common physiological type with around 77% of CP cases (Yeargin-Allsopp, Van Naarden Braun, Doernberg, Benedict, Kirby & Durkin, 2008), in this work we will be especially interested in spasticity, which refers to excessive muscle tone (Peacock, 2009). Muscle tone or *tonus* is the constant tension of muscles that increases when passively stretched (Jennett, 2008). Spastic muscles stay contracted even at rest and have exaggerated reflexes when stretching. Spasticity produces muscle stiffness (Olsson, Krüger, Meyer, Ahnlund, Gransberg, Linke & Larsson, 2006) and can prevent joints to attain their maximum range during movement.

Another widely used classification of CP is by its functional severity and is called the Gross Motor Functional Classification System (GMFCS) (Palisano, Rosenbaum, Bartlett & Livingston, 2008). This system has five categories from I to V, where I represents the less severe condition (close to non-pathological movement) and V the most severe situation (extremely pathological). Since patients with GMFCS levels of IV and V are unable to walk on their own, in this work we will be only interested in patients with GMFCS levels from I to III.

Regarding walking, there are numerous different patterns that can be observed in CP. There are many gait pattern classifications that have been proposed according to clinical data or cluster analysis of quantified gait data (Öunpuu et al., 2009; Sutherland & Davids, 1993; Rodda, Graham, Carson, Galea & Wolfe, 2004; Armand, Watelain, Mercier, Linsel & Lepoutre, 2006; Carriero, Zavatsky, Stebbins, Theologis & Shefelbine, 2009; Winters, Gage & Hicks, 1987). Some of these classification systems will be discussed in section 2.6.

At the present time, there is no cure for this neurological condition. However, there are several possible treatments to improve movement, balance and posture on patients with CP. The orthopaedic treatment for CP is generally rehabilitation, orthoses, plaster casts, walking aids (simple canes, tripod canes, quad canes, crutches, K-walker, rollator), botulinum toxin and especially surgery [Figure 2.2]. There are some neurological treatments too, but they will not be discussed in this report. Conversely, section [section 2.5] is dedicated to orthopaedic surgery in cerebral palsy.

Treatment selection depends on many factors such as patient's medical history, observation of gait, physical examination, among others. In the physical examination, maximum passive

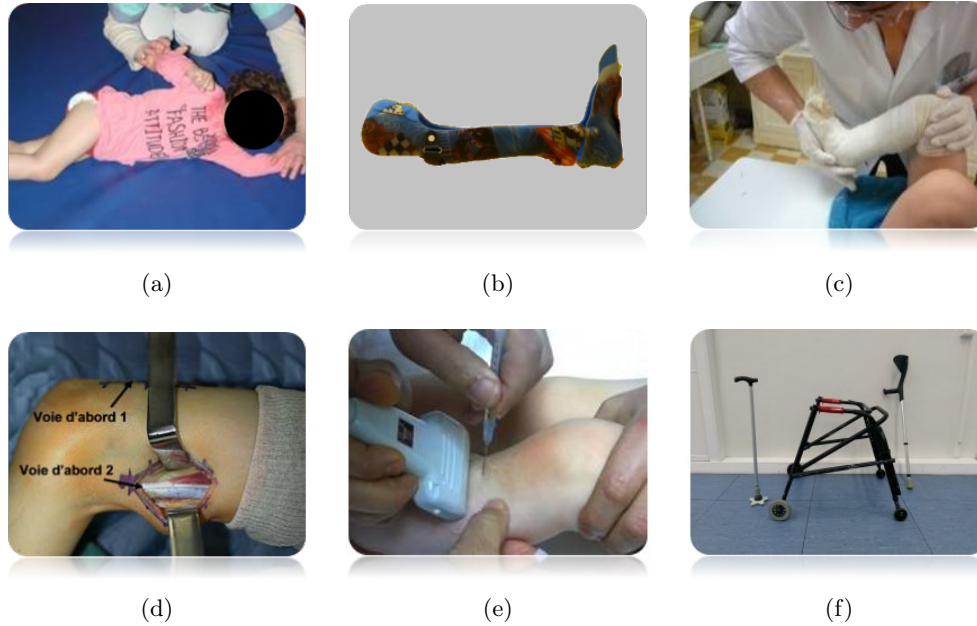


Figure 2.2 – Treatments for gait troubles in CP. 2.2a) Rehabilitation. 2.2b) Orthoses. 2.2c) Plaster cast. 2.2d) Orthopaedic surgery. 2.2e) Toxin injection. 2.2f) Walking aids. [Ellen Poidatz Foundation]

joint angles are measured, as well the muscle strength, muscle selectivity and bone deformities (Trost, 2009). Physical examination will be discussed in section 2.4. Additionally, a clinical gait analysis (CGA) is performed in order to achieve a better diagnosis and a more adapted treatment. This exam allows to identify and quantify causes of gait troubles and thereby choose a suitable treatment. CGA will be detailed in the next section [section 2.2].

2.2 Clinical Gait Analysis

The study of human movement and especially gait (or walking) has interested scientists at least from Aristotle (Baker, 2013, p. 1). Thanks to the advances in biomechanics, computer science and image processing, nowadays human movement analysis has numerous applications, in areas such as sports (Chow & Knudson, 2011; Elliott, 2006), animation and video games (Menache, 2000), person recognition (security) (Boulgouris, Hatzinakos & Plataniotis, 2005; Rani & Arumugam, 2010) and medicine (Gage, Schwartz, Koop & Novacheck, 2009; Perry & Burnfield, 2010; Baker, 2013). The purpose of clinical gait analysis is to observe and quantify walking for medical applications. Walking is a cyclic process that alternates both lower limbs to produce movement [see Figure 2.3]. If only one lower extremity is considered, we identify a stance phase (from initial contact to toe off) and a swing phase (from toe off to initial contact). On the other hand, if both sides are considered, we perceive a double support phase (both feet touch the ground) and a single support stage (only one foot touches the floor). CGA is based on the above description of the gait cycle (Baker, 2013, p. 8).

A typical CGA laboratory is equipped with video-cameras, a motion capture system, force platforms and surface electromyography electrodes [Figure 2.4]. Recorded data in CGA includes videography, kinematics, kinetics and surface electromyography. Occasionally it also includes implanted electromyography, pedobarography and oxygen consumption (Gage & Stout, 2009). The motion capture system may be optical (active markers, passive markers or markerless), inertial, mechanic or magnetic. For the optical systems with markers, the patient wears markers

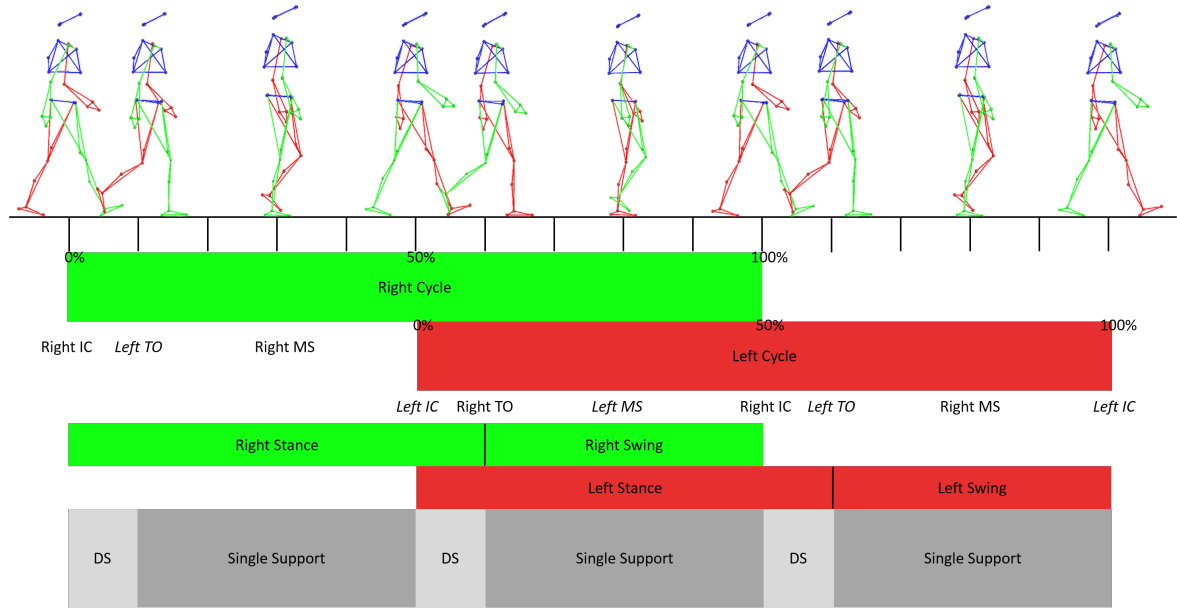


Figure 2.3 – Standard gait cycle. IC, TO and MS stand for gait events respectively for Initial Contact, Toe-Off and Mid-Stance. Blue color corresponds to the right lower limb and its gait phases, while red color corresponds to left lower limb and its gait phases. Similarly, gait events in normal style correspond to the right lower limb and events written with italic style correspond to the left lower limb. Gray blocks represent phases considering both lower limbs, with DS standing for double support.

that are placed in specific points of the body according to a biomechanical model (Robertson, Caldwell, Hamill, Kamen & Whittlesey, 2013, pp. 39-41) [Figure 2.4]. The subject also wears electrodes in the proximity of certain muscles. The patient then walks within the motion capture range and passes over the force platforms until enough gait cycles are recorded. Thanks to the calibrated motion capture system, we are able to know the position of all the markers in the laboratory 3-D coordinate system at every time step. The platforms record 3-D forces and torques when the person passes over them and the electrodes allow to measure muscle activity during gait.

Markers 3-D coordinates are used to compute spatial-temporal parameters and kinematics. Force platform's data is combined with markers' positions to compute kinetics and the electrodes are used for electromyography. Summarizing, data principally derived from CGA are:

- Videography: usually synchronized frontal and profile video of the person walking taken by the video-cameras.
- Spatial-temporal parameters: Among the spatial-temporal information derived from CGA we find: step length, step width, cadence (number of steps per minute), stride and step time and walking speed (Baker, 2013, pp. 8-11).
- Kinematics: In physics, the study of movement without considering its causes is called kinematics. In CGA, kinematics typically refer to lower limb joint angles during walking [see next section for more details]. This is the principal CGA information we will consider throughout this manuscript. Kinematics can also include joint angular velocities, which are the derivatives of the joint angles.
- Kinetics: It refers to the study of the causes of movement. Kinetics describe forces and torques that generate kinematics. Kinetic signals are the result of inverse dynamics computations. This process consists in indirectly determining joint forces and torques from

kinematics, inertial properties and the external forces and torques. Among this external forces and torques are those resulting from the foot-floor contacts that are recorded by the force platforms. In order to compute kinetics, the person needs to pass over the force platforms without any perturbation. Just one foot can touch the platform at a time and the whole contact zone must be within the platform. For example, if the person uses a walking aid (*i.e.* a cane), the aid might perturb the measures and kinetics cannot be calculated. Since this information is not always available, kinetics are not considered in the experiments of this report.

- **Electromyography (EMG):** EMG is the measure of muscle behavior through their electrical activities. Muscle contractions are triggered by electrical signals coming from the brain. EMG enables to detect abnormal activity of certain muscles that helps to characterize the pathology. Muscle activity is usually recorded through probes attached to the skin, thus it is not invasive. However, to record deeper muscles, it is necessary to introduce a fine wire probe to the muscle itself. It is extremely difficult to interpret the amplitude of the EMG signals, thus the interpretation is usually limited to muscle activation and deactivation of muscles (Bouisset, 1999, p. 291). EMG is rarely done postoperatively and the muscles that are measured at CGA may vary according to the patient and the analyst. For this reason, this information is not considered for the experiments in this work.

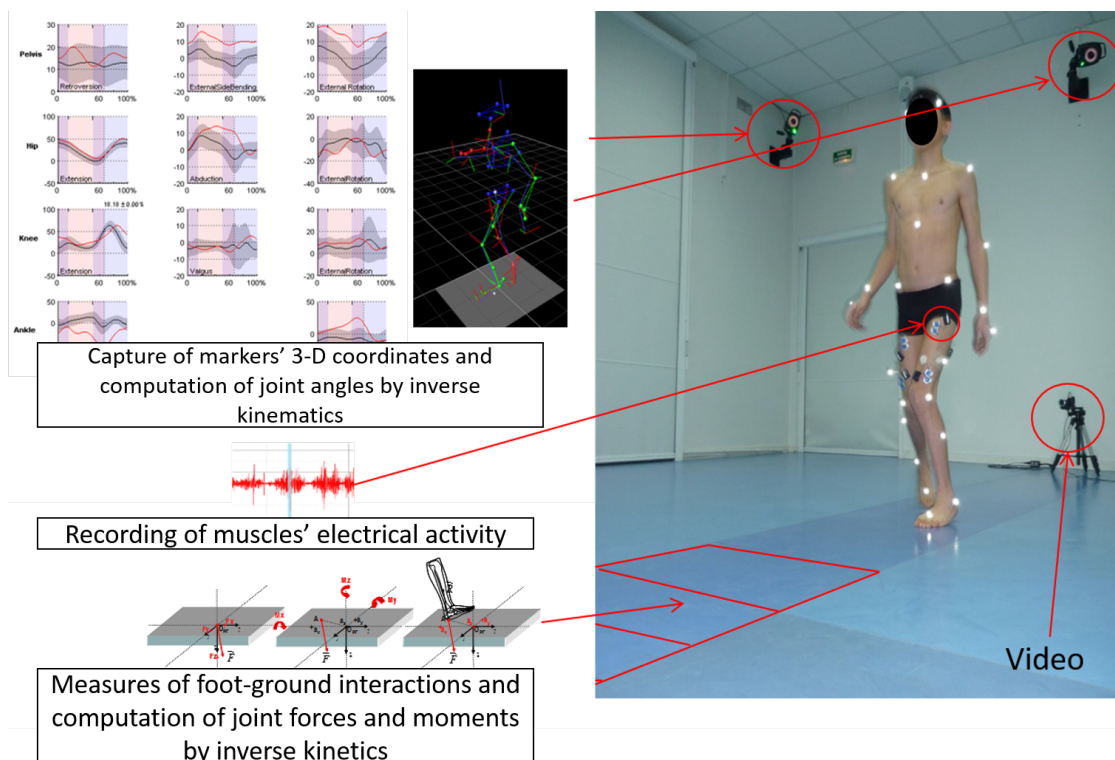


Figure 2.4 – Clinical Gait Analysis laboratory at Ellen Poidatz Foundation. CGA includes video, 3-D kinematics, kinetics and electromyography. In this picture, the motion capture system consists of infrared cameras and reflective passive marker. [Courtesy of E. Desailly (modified)]

2.3 Kinematics

Kinematics refer to the description of the movement itself without considering the forces that caused it. Thus, kinematics basically consider positions, velocities and accelerations. In CGA, kinematics principally consist of 3-D joint or segment angular positions during walking, especially lower limb joints or segments from the pelvis to the feet. In order to compute kinematics of a body, one needs a model of the body and tracking of some landmark points of the body. The tracked points' positions are then matched to the model to obtain the segments' positions. The landmark points are usually markers that are tracked by an optical system (*i.e.* infrared cameras or video cameras).

The multi-sensor system is calibrated using Direct Linear Transform (DLT) (Abdel-Aziz & Karara, 1971; Shapiro, 1978) by using control points. For each camera i , the DLT parameters $L^i = (L_1^i, \dots, L_{11}^i)^T$ are found to relate 3-D coordinates (X, Y, Z) to the 2-D camera image coordinates (x^i, y^i) by the following equations (Robertson et al., 2013, p. 38):

$$x^i + L_1^i X + L_2^i Y + L_3^i Z + L_4^i + L_9^i x_i X + L_{10}^i x_i Y + L_{11}^i x_i Z = 0 \quad (2.1)$$

$$y^i + L_5^i X + L_6^i Y + L_7^i Z + L_8^i + L_9^i y_i X + L_{10}^i y_i Y + L_{11}^i y_i Z = 0 \quad (2.2)$$

Once the DLT parameters are obtained from the calibration and the laboratory coordinate system is defined (origin and axes), the 3-D coordinates of a marker in Equation 2.2 can be computed if that marker is in the field of view of at least two cameras at the same instant.

Based on a kinematic model, the markers' positions are processed to construct the local coordinate system of each body segment at each time step. In the conventional gait model, the lower body is divided in seven segments: pelvis, two thighs/femurs, two shanks/tibias and two feet [see Figure 2.5]. The links between contiguous segments are the joints: *hips* between pelvis and thighs, *knees* between thighs and shanks, and *ankles* between shanks and feet [see Figure 2.5]. Body joints are classically modeled by ball and socket joints of three degrees of freedom. The origin of the local coordinate system of each segment is located at the center of the joint to the parent segment (Baker, 2013, ch.3). For example, the origin of the local coordinate system of the shank is located at the center of the knee. The kinematic data of this work are derived from a modified version of the Helen Hayes kinematic model (Davis III, Öunpuu, Tyburski & Gage, 1991; Kadaba, Ramakrishnan, Wootten & others, 1990). Details on the computation of each local coordinate system and markers' placement for this kinematic model can be found in (Desailly, 2008).

Once the local coordinate systems are defined, it is possible to compute the rotation matrix of each segment j of local coordinate system $(\vec{x}, \vec{y}, \vec{z})$ with respect to the laboratory or to its parent segment k of coordinate system $(\vec{X}, \vec{Y}, \vec{Z})$ as:

$$R_{k,j} = \begin{pmatrix} \vec{x}\vec{X} & \vec{y}\vec{X} & \vec{z}\vec{X} \\ \vec{x}\vec{Y} & \vec{y}\vec{Y} & \vec{z}\vec{Y} \\ \vec{x}\vec{Z} & \vec{y}\vec{Z} & \vec{z}\vec{Z} \end{pmatrix} \quad (2.3)$$

Hence we are able to obtain the rotation matrix of each joint. For instance, the rotation matrix of the shank with respect to the thigh $R_{right_thigh, right_shank}$ or $R_{left_thigh, left_shank}$ corresponds to the rotation matrix of the knee joint. For the pelvis, rotations are considered with respect to the laboratory (global coordinate system).

Since the rotation matrices cannot be interpreted by clinicians, the Cardan angles or Euler angles (Robertson et al., 2013, pp. 45-48) are computed from these matrices. The principle is that the rotation of any segment with respect to another can be expressed by the three successive

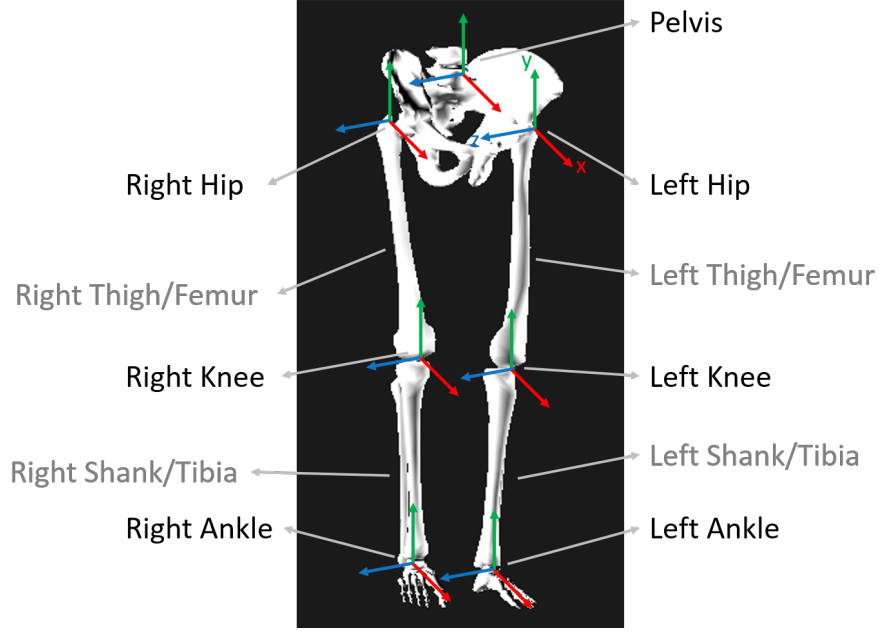


Figure 2.5 – Rigid body gait model: segments and joints that describe the lower body from the pelvis to the feet. Each segment has its local coordinate system's origin at the joint with its parent segment. In this work, x-axis (red) go forwards, y-axes (green) go upwards and z-axes (blue) go sideways.

rotations (Cardan sequence) of the known axes. A rotation of β about the X-axis [see 2.6a] can be expressed as:

$$R_x = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\beta) & -\sin(\beta) \\ 0 & \sin(\beta) & \cos(\beta) \end{pmatrix} \quad (2.4)$$

Similarly, rotations of γ about the Y-axis [figure 2.6b] and of α about the Z-axis [figure 2.6c] are respectively:

$$R_y = \begin{pmatrix} \cos(\gamma) & 0 & \sin(\gamma) \\ 0 & 1 & 0 \\ -\sin(\gamma) & 0 & \cos(\gamma) \end{pmatrix} \quad (2.5)$$

$$R_z = \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) & 0 \\ \sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (2.6)$$

Rotation about the Z-axis is a rotation of the X-Y plane, which corresponds to the profile of the segments or sagittal plane [see Figure 2.5]. Similarly, rotations about the X-axis and Y-axis correspond to rotations of Y-Z and X-Y planes respectively, which are called the frontal (segments seen from the front) and the transverse (segments seen from the top) planes. There are 12 possibilities of Cardan sequences, but the convention is to rotate in the order of first sagittal plane, second frontal plane and third transverse plane; thus Z-axis, X-axis and Y-axis. From equations 2.4, 2.6 and 2.6 the conventional Cardan sequence for equation 2.3 is:

$$R_{k,j} = R_z R_x R_y$$

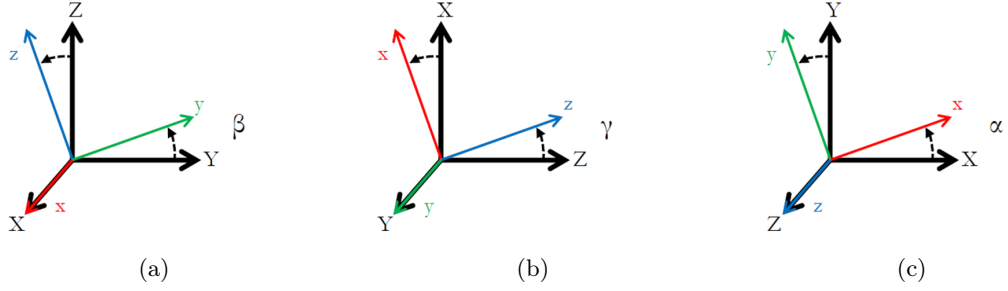


Figure 2.6 – Cardan angles rotation. 2.6a) Rotation about X-axis. 2.6b) Rotation about Y-axis. 2.6c) Rotation about Z-axis.

Table 2.1 – Physiological interpretation of kinematic (Cardan) angles according to joint and plane. The cardan angle corresponding to the ankle in the transverse plane is usually replaced by the foot progression angle, which is the rotation of the foot with respect to the walking axis.

	Sagittal Plane	Frontal Plane	Transverse Plane
Pelvis	Anteversion/Retroversion	Internal/External Bending	Rotation (Internal/External)
Hip	Flexion/Extension	Adduction/Abduction	Rotation (Int./Ext.)
Knee	Flexion/Extension	Varus/Valgus	Rotation (Int./Ext.)
Ankle	DorsiFlexion/PlantarFlexion	Varus/Valgus	Adduction/Abduction (Sohrweide, 2009)

Hence

$$R_{k,j} = \begin{pmatrix} \cos(\alpha) \cos(\gamma) - \sin(\alpha) \sin(\beta) \sin(\gamma) & -\sin(\alpha) \cos(\beta) & \cos(\alpha) \cos(\gamma) + \sin(\alpha) \sin(\beta) \cos(\gamma) \\ \cos(\alpha) \sin(\beta) \sin(\gamma) + \sin(\alpha) \cos(\gamma) & \cos(\alpha) \cos(\beta) & \sin(\alpha) \sin(\gamma) - \cos(\alpha) \sin(\beta) \cos(\gamma) \\ -\cos(\beta) \sin(\gamma) & \sin(\beta) & \cos(\beta) \cos(\gamma) \end{pmatrix}$$

Thus $\beta = \sin^{-1}(\vec{y}\vec{Z})$ and then it is possible to also find the values of α and γ . The advantage of the cardan angles is that they can be interpreted as physiological movements according to the joint and the plane, as shown in Table 2.1. These are the signals that are presented and analyzed to detect the gait deviations.

Kinematic curves are in general presented in cycles within two consecutive initial contacts and are normalized with respect to the cycle duration as shown in Figure 2.7. Also, the different gait phases are expressed in percentage of the gait cycle (Sebsadji et al., 2012). In addition, standard non-pathological gait patterns are presented in the graphics in order to facilitate comparison and detection of gait deviations. Usually, the ankle varus-valgus curve is not presented and the transverse plane ankle curve is replaced by the *foot progression angle*, which is the angle between the foot and the walking direction axis. In this work, we will not consider either the knee varus/valgus, the knee rotation and the ankle varus/valgus. On the other hand, the foot progression angle and sometimes the foot-ground angle (chapter 5) will be considered. The foot-ground angle is not usually represented, but corresponds to the angle between the foot and the laboratory floor.

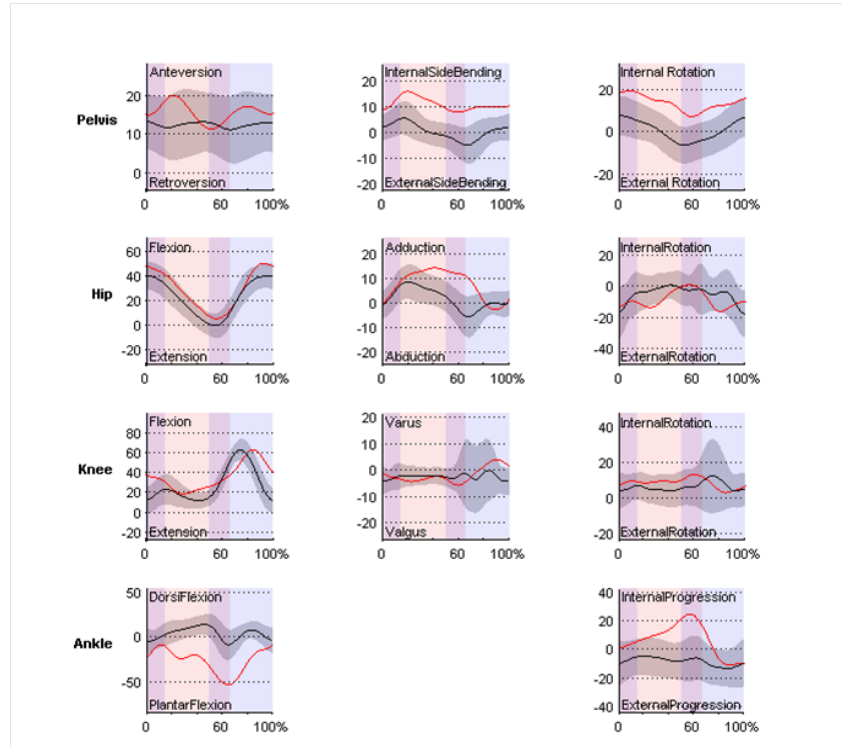


Figure 2.7 – Joint angles curves during walking for a lower limb from two consecutive initial contacts. The graphs of the left column represents the sagittal plane angles, the middle and right columns correspond to the frontal and transverse plane respectively. Each row corresponds to a joint or segment. The red lines correspond to the patient's signals. The black lines and gray zones correspond respectively to the average standard gait signals (non-pathological) and their 2-standard-deviation band. Red backgrounds represent the single support phase, purple backgrounds correspond to double support phases and blue backgrounds represent the swing phase. [courtesy of E. Desailly (original in (Desailly, 2008))]

2.4 Physical Examination

CGA is associated to a physical examination (PhEx), where several measurements are made on table (not during walking) in order to elaborate a neuro-orthopaedic profile of the patient. Physical examination principally searches information about joint ranges of motion, muscle contracture, muscle strength, spasticity and bone deformity (Trost, 2009) [see Table 2.2]. It may also include information about balance, posture, sensitivity, muscle selectivity, general body characteristics, among others.

PhEx data are not necessarily correlated to the CGA data (Desloovere, Molenaers, Feys, Huenaerts, Callewaert & Walle, 2006) and thus represent an important complementary information for treatment decision. PhEx data might confirm some gait deviations that are detected from CGA data or might give hints of possible causes of gait abnormalities. For example, the popliteal angle [Figure 2.8] gives information about the hamstring length or the maximal extension capacity of the knee when the hip is flexed. This capacity might be, but not necessarily, related to the maximum extension of the knee during walking at initial contact. If there is an abnormal kinematic range of motion that does not correspond to the passive range of motion, the cause of this gait abnormality may be abnormal muscle tone (spasticity), deficient muscle strength or a compensation of any other phenomenon occurring in another joint of the same limb or its contralateral limb. Following the previous example, if it exists an excessive knee flexion at initial contact and the popliteal angle is close to the standard, hamstring length might not be the cause of this deviation but hamstring spasticity or knee flexum (negative knee extension) might be other plausible causes. In the first case, there would probably be an important difference between popliteal angle at V1 and V3. In the second case, the passive knee extension value would be negative.

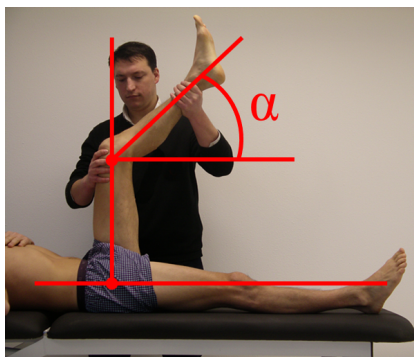


Figure 2.8 – Measurement of popliteal angle. Popliteal angle is measured as the angle α between the shank axis and the orthogonal to the thigh axis when the hip is flexed at 90° and the knee goes to its maximal passive extension (the foot goes upwards). The contralateral hip and knee are extended. There are other ways of measuring popliteal angle [see for example (Baker, 2013, p. 110)].

There are numerous neuro-orthopaedic variables that can be made in PhEx and several ways of performing the different tests. For details on the different tests and the way of performing them, see (Cleland et al., 2011) or (Baker, 2013, ch. 8). Physical examination is performed by a trained clinician, who adapts the tests to each patient, hence the variables that are measured change depending on the clinician and the patient. However, there are some tests that are almost systematically performed. To statistically exploit the PhEx information, in this work we have selected a restricted number of physical examination variables [see Appendix 9.2] that are most frequently measured [see section 3.2].

Table 2.2 – A non-exhaustive list of physical examination measures. For a more detailed list, see Appendix 9.2 (in French). Joint ranges of motion, deformity and spasticity features are measured in degrees. Muscle strength testings are measured in a scale of 0 to 5. Unipedal stance is measured in seconds, weight in kilograms and height in centimeters. For details on the techniques for taking the measures, see (Cleland et al., 2011) or (Baker, 2013, ch. 8).

Segment/joint	Measure	Type
General	Height	-
	Weight	-
Spine	Scoliosis	Deformity
	Hyperlordosis	Deformity
	Hyperkyphosis	Deformity
	Abdominals testing	Muscle strength
Pelvis/Hip	Femoral anteversion	Deformity
	Hip extension	Passive range of motion
	Hip abduction, flexed hip, V1*	Passive range of motion
	Hip abduction, flexed hip, V3**	Spasticity
	Hip abduction, extended hip, V1	Passive range of motion
	Hip abduction, extended hip, V3	Spasticity
	Hip abduction, flexed knee, V1	Passive range of motion
	Hip abduction, flexed knee, V3	Spasticity
	Hip external rotation	Passive range of motion
	Hip internal rotation	Passive range of motion
	Hip flexion	Passive range of motion
Thigh/Knee	Knee extension	Passive range of motion
	Knee flexion	Passive range of motion
	Quadriceps/Knee extensor lag	Muscle strength
	Popliteal angle V1	Passive range of motion
	Popliteal angle V3	Spasticity
	Bilateral popliteal angle	Passive range of motion
	Hamstrings test	Muscle strength
	Duncan-Ely test	Spasticity
Shank/Ankle	Ankle plantar-flexion	Passive range of motion
	Ankle dorsiflexion (extended knee) at V1	Passive range of motion
	Ankle dorsiflexion (extended knee) at V3	Spasticity
	Ankle dorsiflexion (flexed knee) at V1	Passive range of motion
	Ankle dorsiflexion (flexed knee) at V3	Spasticity
	Triceps testing	Muscle strength
	Tibial torsion	Deformity
	Relievers testing	Muscle strength
	Inversors testing	Muscle strength
	Eversors testing	Muscle strength
Foot	Forefoot pronation	Passive range of motion
	Forefoot supination	Passive range of motion
	Hindfoot varus	Deformity
	Hindfoot valgus	Deformity
	Unipedal stance test (time)	Balance

*At the slowest stretching velocity that is possible (Scholtes, Becher, Beelen & Lankhorst, 2006).

**The angle of catch when stretching quickly (Scholtes et al., 2006).

2.5 Single-Event Multilevel Surgery

In order to improve gait patterns, orthopaedic surgery may be performed on patients with cerebral palsy. The purpose of this treatment is to ameliorate abnormal gait patterns by surgically modifying the patient's anatomy (bones, muscles and tendons). These modifications might either try to match the patient's deformed anatomy to a typical developing child's anatomy, whether to give a non-typical anatomy that could help improving the patient's gait. In any case, the cerebral lesion is not modified. However, it is possible that new motion schemes appear after the surgery.

Multiple bone and soft tissue deformities can be surgically modified at different levels of the lower limbs during the same operation. For this reason, it is called single-event multilevel surgery (SEMLS). The goal and the actual outcome of SEMLS depends on the condition of the patient and the severity of gait troubles or anatomical deformities [see Figure 2.9]. Sometimes SEMLS gives outstanding outcomes where the postoperative gait resembles standard walking [see for example figures 2.9a and 2.9b]. Some other times, gait deviations are remarkably improved even if gait abnormalities remain [see for example figures 2.9c and 2.9d]. In some cases, the objective of SEMLS is to avoid the loss of walking ability in order to preserve or achieve patient's autonomy in interiors (*i.e.* walk only at home) [see figures 2.9e and 2.9f]. In figure 2.9f, it can also be seen that the surgical treatment was reinforced by orthoses and walking aids.

SEMLS is followed by a recovering stage and a rehabilitation stages that are crucial for good outcomes. In this work, the effect of SEMLS on gait is studied by analyzing the effect of different surgical procedures and their combinations. In all cases, we assumed that the recovery and the rehabilitation stages were successfully completed.

There is a large number of surgical procedures proposed for CP according to the functional objective, the technique applied, the parts of the body that are modified or secondary affected, etc. Different surgical procedures might have the same impact on gait, but they differ on the causes of abnormality. For instance, the functional goal of hamstring lengthening surgery (HL) is to decrease knee flexion at initial contact. HL represents one possible treatment for crouch gait (Stout, Novacheck, Gage & Schwartz, 2009), which is associated to excessive knee flexion at stance. Ma *et al.* (2006) reported an average improvement of 17° of knee flexion at initial contact on 19 patients (38 limbs) with increased knee flexion. If the cause of crouch gait is hamstring shortness, a hamstring lengthening surgery seems to be the most adequate operative treatment. However, if crouch gait is not caused by hamstring shortness, then hamstring lengthening would have null effect on gait (Sebsadji *et al.*, 2012). In such case, other surgical procedures should be considered after clear identification of the cause of the gait trouble. Another important consideration is the possible side effects that the surgical procedure may entail. Continuing with the previous example, it has been reported that HL may increase pelvic tilt in some cases (DeLuca, Ounpuu, Davis & Walsh, 1998), hence this surgical procedure should also be avoided when there is excessive pelvic tilt.

Another typical surgical procedure is rectus femoris transfer, whose functional goal is to improve knee flexion at swing phase (Novacheck, 2009). The different functional goals of most of the surgical treatments of gait troubles in CP, as well as their recommended indications, can be found in (Gage *et al.*, 2009, sec.5).

The combination of different surgical procedures is believed to act more or less independently in order to improve all the detected gait troubles. For example, hamstring lengthening and rectus femoris transfer might be performed at the same operative time hoping to improve knee flexion at both initial contact and swing phase. However, given the variety of surgical procedures and gait or physical examination patterns, the indication of SEMLS is not straightforward. For a same patient, different medical teams usually suggest different SEMLS composition. To overcome



(a)



(b)



(c)



(d)



(e)



(f)

Figure 2.9 – Examples of outcome of orthopaedic surgery. 2.9a) Preoperative gait of a patient with equinus gait. The hind-feet never touch the ground. 2.9b) Postoperative gait of patient in 2.9a. The hindfoot touches the floor at stance. The surgery consisted of bilateral gastrocnemius fasciotomy, bilateral hamstring lengthening, left tibialis posterior lengthening and left rectus femoris transfer. 2.9c) Preoperative gait of a patient with crouch gait. The knee flexion is excessive at stance. 2.9d) Postoperative gait of the patient in 2.9c. The knee flexion has considerably decreased. 2.9e) Preoperative gait of a patient with severe gait troubles. Knee flexion and hip internal rotation are excessive. 2.9f) Postoperative gait of the patient in 2.9e. Both knee flexion and hip internal rotation have decreased. The patient still needs technical aids to walk.

this problem, classification of patients and gait patterns have been developed, as well as other decision-making tools. These points will be discussed in the next two sections.

2.6 Classification of Gait Patterns in Cerebral Palsy

Multiple gait patterns can be observed in children with CP. Therefore, a variety of classification systems and quality scores have been proposed in order to better assess the medical care of the patients. In section 2.1, we have seen that CP can be classified according to topography, physiology and functional severity. This section is dedicated to some gait indices or scores and classifications of CP that are based on CGA data, especially on kinematic data.

Among the widely used gait indices, we find Gillete Gait Index (GGI) (Schutte, Narayanan, Stout, Selber, Gage & Schwartz, 2000), Gait Deviation Index (GDI) (Schwartz & Rozumalski, 2008) and Gait Profile Score (GPS) (Baker, McGinley, Schwartz, Beynon, Rozumalski, Graham & Tirosh, 2009). All three indices measure distance between gait variables of a limb and the average of a non-pathological sample, but they differ on the gait variables they consider or the distance computation. GGI considers 16 discrete gait variables including spatio-temporal parameters and kinematic variables and the distance is computed by the mean-squared error (MSE). The higher the GGI value, the more important are the gait deviations. GDI considers 9 time-normalized kinematic curves (the same postoperative signals considered in this work, see section 3.1) and computes root mean-squared error (RMSE) in a 15-feature singular value decomposition (SVD) space. The RMSE value is then scaled such that a $GDI \leq 100$ corresponds to non-pathological gait and every 10 points distance below 100 corresponds to 1-standard-deviation distance from the non-pathological gait average (Schwartz & Rozumalski, 2008). GDI has moderated correlation to GGI (IDEM). On the other hand, GPS considers the same 9 kinematic curves and directly computes the RMSE to the average non-pathological gait. GPS is intimately related to GDI, since they both consider the same kinematic variables and because of the SVD decomposition property of preserving distances in a lower-dimensional space [see section 4.4 for more details on SVD]. The difference is that GPS allows to separately compute a gait index for every kinematic signal, which is called the Gait Variable Index (GVS) (Baker et al., 2009). GVS, GPS and GDI will be used as descriptors in some of the experiments in Part II.

Sutherland and Davids (Sutherland & Davids, 1993) identified four major patterns in knee flexion during gait:

- Crouch knee: increased knee flexion during walking, especially during stance.
- Stiff knee: limited range of motion of the knee during gait, especially during swing.
- Jump knee: excessive knee flexion at early stance and then normal knee flexion during late stance.
- Recurvatum knee: hyperextension of the knee at mid-stance.

The above classification is based on observation of the knee curve and no quantitative information was given to systematically utilize it. Some efforts have followed to quantify these patterns. For example, Goldberg *et al.* (2006) proposed a score for identifying stiff knee, as well as not stiff knee and borderline (between stiff and not stiff) based on four kinematic variables of the knee flexion/extension curve: maximal knee flexion in swing phase, knee range of motion in early swing phase (from toe-off to maximal flexion), total knee range of motion during gait and percentage of cycle of maximal knee flexion during swing phase. This time the classification is quantitative, but the classes are arbitrary. Other classification methods utilize unsupervised

learning techniques in order to automatically determine the classes. For instance, Armand *et al.* (2006) used Fuzzy K-means clustering (Duda et al., 2000, pp. 528-530) on the ankle dorsal/plantar flexion curve for patients with toe-walking and identified three main patterns. The previous classifications consider only one kinematic signal and not an ensemble of gait curves. Rodda *et al.* (2004) identified five groups of sagittal plane kinematics (pelvis, hip, knee and ankle) for bilateral involvement of CP:

- I True equinus: foot contact is first done by toes, with normal or excessive knee extension, thus with ankle plantar flexion [figure 2.9a].
- II Jump gait: equinus (foot contact by toes) with excessive knee flexion and ankle plantar flexion.
- III Apparent equinus: excessive knee flexion and hip flexion with normal ankle dorsal/plantar flexion, but foot contact done by toes.
- IV Crouch gait: crouch knee (excessive flexion), with excessive hip flexion and ankle dorsiflexion [figure 2.9c].
- V Asymmetrical gait: combination of any two of the above.

Classification and scores of gait patterns help to adapt the medical care of the patients but are not enough for making decisions about the surgical treatment. For this reason, some outcome-predictive decision-making tools have been recently developed.

2.7 Surgery outcome prediction

It exists a large number of methods for assessing treatment in CP, but only a few of them predict treatment outcome. These methods are based on mechanical models or statistical models. The mechanical models are based on musculoskeletal simulations to estimate muscle length. For example, hamstrings length during gait can be estimated from kinematic data in order to determine if a hamstring lengthening surgery would be efficient or not (Arnold et al., 2006; Desailly et al., 2009).

Other predictive tools are based on purely statistical models. Reinbolt *et al.* (2009) predicted “good” and “poor” outcomes of rectus femoris transfer on 62 patients with stiff knee (originally 81 patients, but 19 were excluded) using Linear Discriminant Analysis (LDA). Best cross-validation results were 88% of correct predictions using five input variables: hip flexion and hip power after initial contact, knee power at peak knee extension in stance, knee flexion velocity at toe-off, and hip internal rotation in early swing. Desailly *et al.* (2012) obtained a 80% accuracy rate using Support Vector Machines (SVM) to predict “positive” and “not-positive” outcomes of hamstring lengthening over 60 lower limbs with crouch gait. Schwartz *et al.* (2013) used Random Forests for predicting “good” and “poor” outcomes of psoas lengthening and obtained 78% of prediction accuracy over 210 limbs.

The decision supports above mentioned are able to indicate if a surgical procedure is suitable or unsuitable for a patient, but they do not give any quantified information of the probable outcome. The binary qualitative classification is too subjective, and prediction rate varies with different criteria. In fact, these three works utilize different criteria for labeling “good” and “not-good” outcomes. For example, a predicted “good” outcome, might be interpreted by the patient as a total eradication of gait abnormalities. Since the neurological damage is not cured, such outcome cannot be reasonable achieved in reality.

Other methods predict some gait parameters. Kay *et al.* (2001) predicted walking speed (and

evaluated other spatio-temporal parameters) using univariate and multiple linear regression for 47 operated patients. Hicks *et al.* (2011) used Multiple Linear Regression to predict average knee flexion during stance for patients with crouch gait and with or without different treatments. 10-fold cross-validation gave $R^2 = 0.44$ over 353 limbs. The predicted value was then used for predicting “improved” and “unimproved” limbs, with a classification accuracy of 71%. Unfortunately, results in terms of knee flexion degrees are not presented and the tool basically works as a binary predictor, such as those presented before. In this work, multiple linear regression is considered in chapter 6.

Sullivan *et al.* (1995) used regression analysis to predict postoperative knee flexion curve after rectus femoris transfer on a 25-limb database of patients with stiff knee. Unfortunately, there is no detail about the regression method or the performance of the model, which is described as “imperfect but encouraging”. Hersh *et al.* (1997) utilized feedforward neural networks for predicting hip flexion and knee flexion curves during gait. However, model performance was not reported. Niiler *et al.* continued Hersh’s work (rectus femoris transfer and stiff knee) on 24-patient database and obtained average root-mean squared errors (RMSE) of respectively 8.1° , 9.7° and 6.7° for hip flexion, knee flexion and ankle dorsiflexion during gait. Niiler (2001) then reported 6.4° and 9.2° of average RMSE for respectively hip flexion and knee flexion during gait among 50 patients with stiff knee. In this work, regression with feedforward neural networks is considered for several experiments [see sections 5.2, 7.2 and section 8.1].

Niiler *et al.* (2007) considered some concurrent surgeries (Achilles lengthening, gastrocnemius lengthening and hamstring lengthening) in combination with rectus femoris transfer for predicting knee range of motion during gait using multiple linear regression. The database was composed of 94 limbs with stiff knee, but there were four groups and a model was associated to each group, thereby giving a maximum of 43 limbs for training a model.

Despite those previous works surgery planning remains difficult and outcome prediction still lack of accuracy. The previous works have several drawbacks:

- Most of the decision-making methods are qualitative and binary (“good” or “not-good” outcomes), which may be too subjective to interpret.
- Most of them are based on a single surgical procedure, whilst SEMLS is often composed of combinations of surgical procedures [see section 3.3].
- Most of them are based on a single gait pattern (*i.e.* crouch gait, stiff knee), whilst there are numerous gait pattern in cerebral palsy [see, for example, section 2.6].
- Quantitative predictions only output few gait parameters.
- Some of the tools were evaluated in a small number of patients and lack of further validation.
- The prediction accuracy, especially for quantitative variables, is far from being ideal.

With qualitative or few quantitative parameters, it is impossible to have a clear preview of the most likely postoperative gait. Moreover it is difficult to discuss about the expected outcome with the patient and the patient’s family. If an improvement is sincerely expected by the medical team, an accurate prediction of this improvement remains uncertain. In this work, quantitative gait parameters will be predicted considering several surgical procedures and multiple combinations of these procedures, and for a large number of gait patterns.

Chapter 3

Data Description and Conditioning

Contents

3.1	Kinematic Data	24
3.2	Physical Examination Data	29
3.3	Surgical Data	32
3.4	Preoperative Kinematics and Physical Examination Data by Surgical Categories	33
3.5	Surgery Outcome	35
3.6	General Discussion	37

The data used for all the experiments of this report were acquired in the Laboratory of Movement Analysis of the Ellen Poidatz Foundation. All the patients in the Ellen Poidatz Foundation are children with different kinds of handicaps. A database was created with walking patients with CP (GMFCS level from I, II and III (Palisano, Rosenbaum, Walter, Russell, Wood & Galuppi, 1997)) that have undergone single-event multilevel surgery and that have had at least a CGA and physical examination before surgery and a CGA after surgery. Since after surgery there are recovery and rehabilitation stages, postoperative exams within one-year period after the surgery were excluded. Patients with problems during recovery or rehabilitation were also excluded, *i.e.* infections or other complications during recovery, or incomplete rehabilitation program. For patients that utilize different walking aids, trials with the same walking aid were chosen in both preoperatively and postoperatively (*i.e.* For a patient that used a K-Walker for the preoperative CGA, only postoperative trials with K-walker were considered). 10.5% of the patients were unilaterally affected and the rest were bilaterally affected. The average ages of patients at preoperative CGA, surgery and postoperative CGA are respectively 12, 13 and 15 years old [Table 3.1].

Table 3.1 – Age of patients at preoperative CGA, surgery and postoperative CGA.

	Ages (years)			
	Average	Standard Deviation	Minimum	Maximum
Preoperative CGA	12.38	3.26	5.84	22.91
Surgery	13.20	3.22	6.71	24.19
Postoperative CGA	15.25	3.21	8.70	26.77

With the continuing arrival of new patients in the Ellen Poidatz Foundation, this database is in constant grow. For this reason, some experiments during the three-year thesis project have different number of patients. In section 5.2, the experiment was conducted with 99 patients. In

section 6.2, the database comprised 115 patients, whilst in section 7.2 the study was conducted with 134 patients, which corresponds to the last version (retrieval date: September 25th 2015) of this database in this report. A comparison of all these methods and some others with the 134-patient database is presented in chapter 8. All the retrieved data were anonymized.

3.1 Kinematic Data

In the Ellen Poidatz Foundation motion analysis laboratory, there are records of all CGA exams that have been performed in the Ellen Poidatz Foundation since 2004. From 2008, kinematic data have been stored in C3D format and have been recorded with a Vicon motion system consisting of eight infrared cameras at 100 frames per second (fps). The sample frequency of force platforms and electromyography is 2000 Hz. Before 2008, the camera system was a SAGA 3RT Biogesta recording at 50 fps and raw data were stored in text format. In both systems lower limb marker placements were identical and kinematic data were computed with the same software based on a modified Helen Hayes model (Davis III et al., 1991; Kadaba et al., 1990) with anatomical markers on the femoral condyles and the medial-malleolus (Desailly, 2008). Even if walking is a cyclic process, kinematic signals are not exactly periodical due to the variability of biomedical data (Boudaoud, 2006). In general two different gait cycles slightly differ in time and amplitude, even if it is the same person who walks. In order to give a unique representation, an average gait cycle is computed for every lower limb. For this purpose, kinematic signals are segmented into gait cycles, where a lower limb's cycle is defined from two consecutive initial contacts [Figure 2.3]. In this work, foot contacts were automatically detected utilizing the high pass algorithm (HPA) (Desailly, Daniel, Sardain & Lacouture, 2009). HPA localizes minima of the vertical component of the heel marker, and then uses a high pass filter on the horizontal displacement of the forefoot marker and the heel marker to determine both initial contacts and toe-offs (the beginning of the swing phase). Every gait cycle was then resampled to incrementations of 2% of the cycle (51 points per cycle) as in (Schwartz & Rozumalski, 2008). Resampling was done by B-Splines interpolation (De Boor, 2001). Next, the average of the resampled cycles was computed [Figure 3.1].

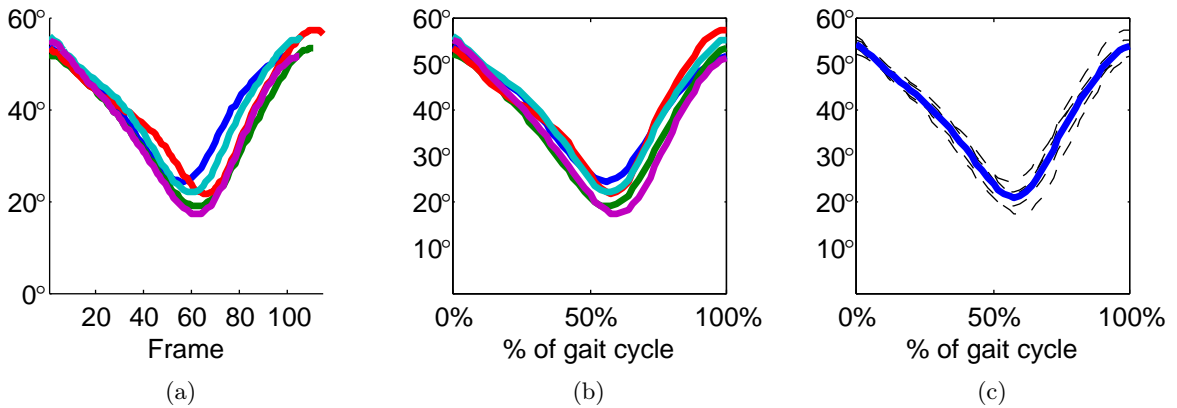


Figure 3.1 – Example of mean cycle computation. 3.1a Original segmented cycles. 3.1b Resampled cycles. 3.1c Mean cycle (blue) and resampled cycles (dashed black).

To characterize the considered kinematic population, we computed the average kinematic signals over all patient-average kinematics and their standard deviations [Figure 3.2]. These average preoperative kinematics are in general in the standard gait (non-pathological) two-standard-deviation (2-SD) band, except for the knee flexion and the hip rotation. The average

preoperative knee flexion is first excessive in the stance phase (up to 60% of the cycle [Figure 2.3]) and then insufficient in early swing phase until becoming again excessive at late swing phase. The average preoperative hip rotation corresponds to an excessive internal rotation during the whole gait cycle. On the other hand, the average preoperative hip flexion is somewhat excessive at mid-stance phase. For all the considered kinematic angles, the preoperative signal variability is much higher than the standard gait (preoperative 2-SD band is much wider than the 2-SD standard gait band)[Figure 3.2]].

In postoperative stage, the average kinematics are similar to the preoperative signals [Figures 3.2 and 3.3]. However, the postoperative knee flexion is within the standard gait band, but still somewhat excessive at initial contact and late swing, and somewhat insufficient at early swing phase [Figure 3.3]. On the other hand, the postoperative hip rotation decreases on stance phase and enters in the standard gait band, but is still excessive on swing phase. Thus, the postoperative averages tend somewhat closer towards the standard gait averages than preoperative averages. This can be interpreted as a general reduction of gait deviation in operated patients. In addition, the postoperative signal variability is smaller than preoperative variability (narrower 2-standard deviation bands and minimum-maximum difference). This postoperative changeability is still higher than the standard gait variability, although close for frontal plane angles (pelvic obliquity and hip adduction), hip rotation and ankle dorsiflexion.

Except when stated, the preoperative vectors in this work comprised ipsilateral and contralateral information. For each limb, the preoperative input vector was composed of the nine ipsilateral gait angles and hip, knee and ankle/foot gait angles of the contralateral limb taken from ipsilateral initial contacts [see figure 3.4 for an example]. This implies a vector of $N_{pre}^{ang} = 15$ time series of $N_{points} = 51$ points each, which gives a total of $N_{pre}^{kine} = N_{pre}^{ang} \times N_{points} = 15 \times 51 = 765$ kinematic variables.

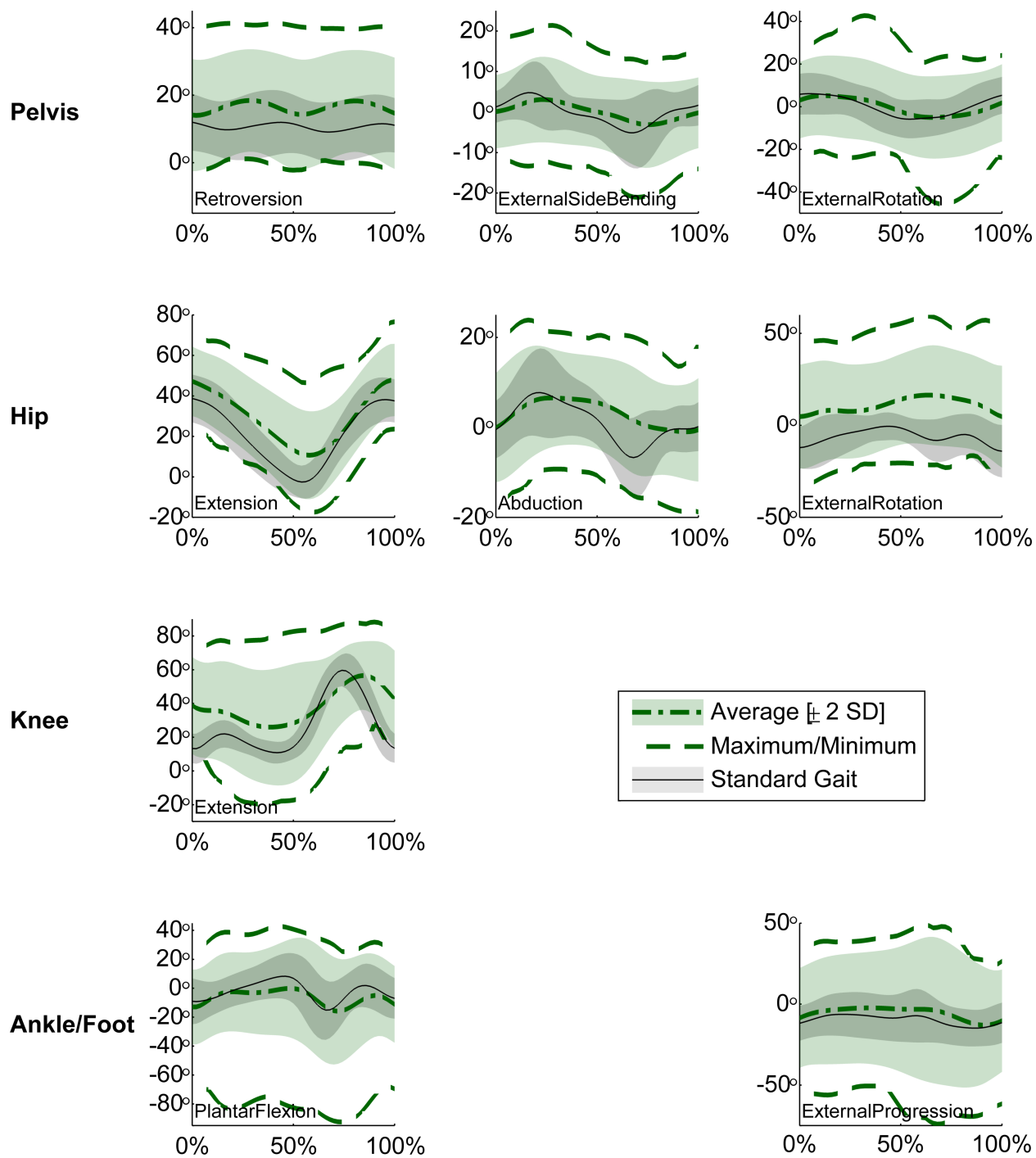


Figure 3.2 – Preoperative Kinematic Data.

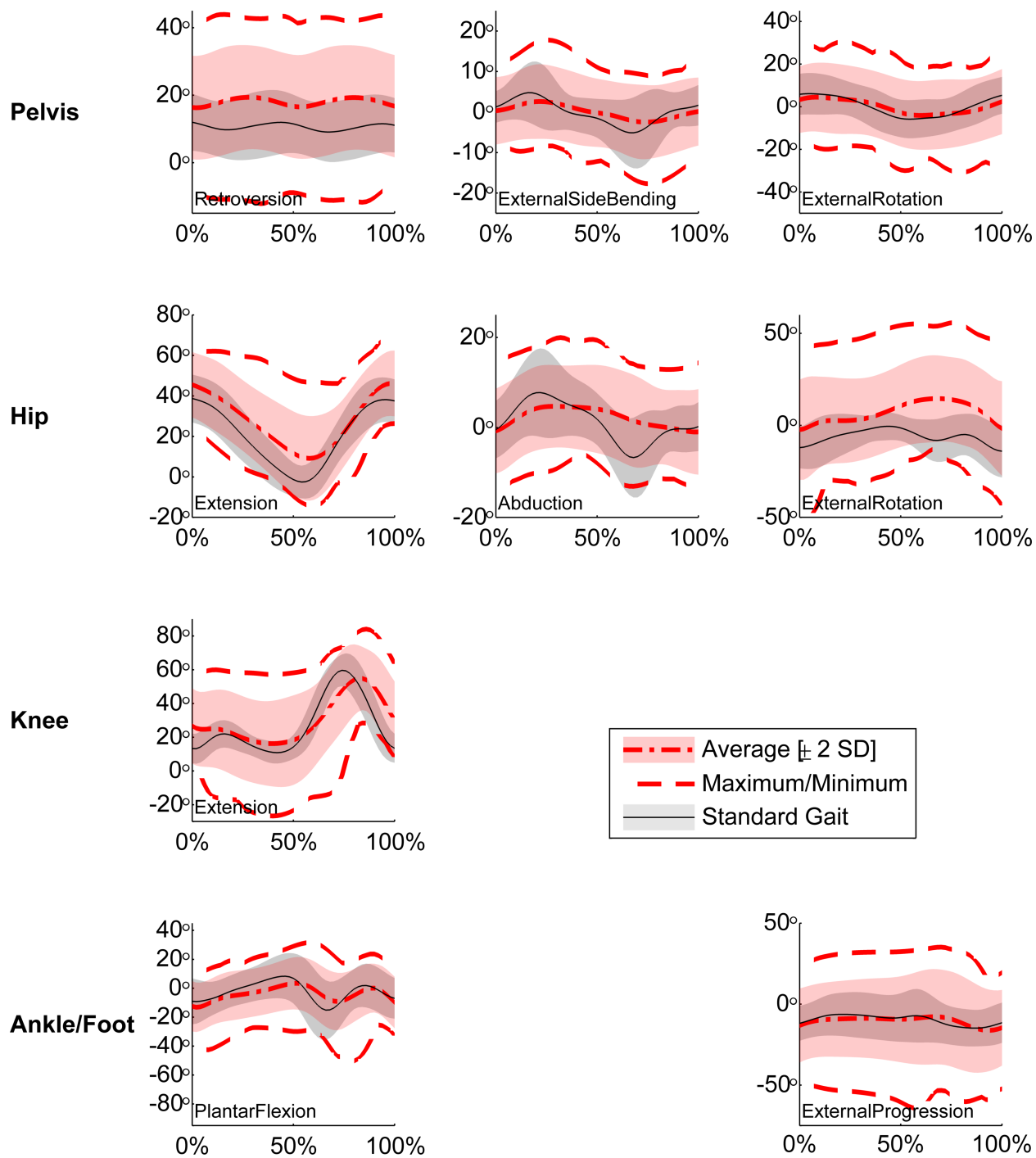


Figure 3.3 – Postoperative Kinematic Data.

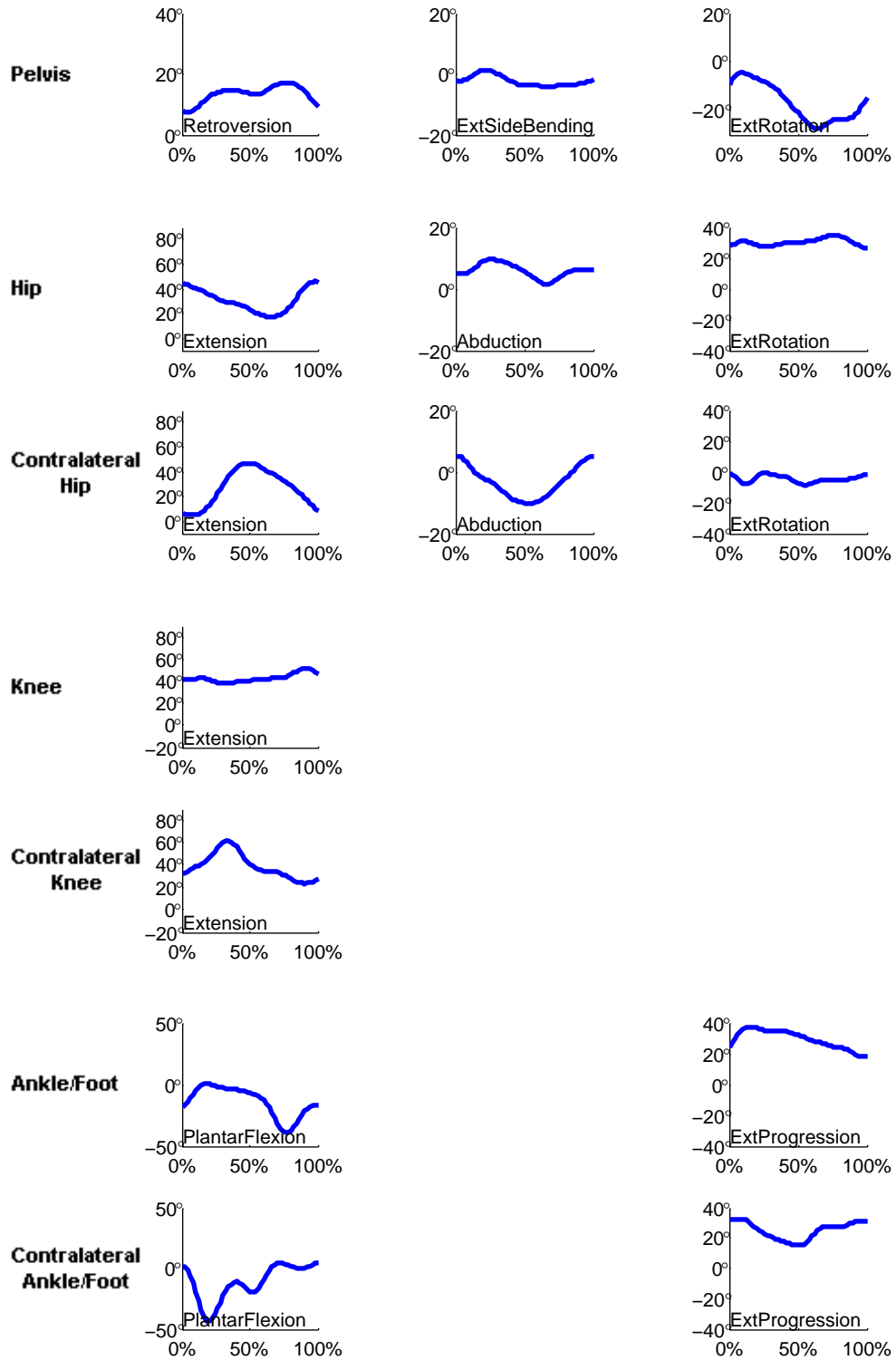


Figure 3.4 – Considered preoperative kinematics for a random patient. The preoperative kinematic vector is composed of the concatenation of the mean per patient of all these signals taken from left to right and from up to down: it starts with pelvic tilt, pelvic obliquity and pelvic rotation. Then continues with hip angles, contralateral hip angles, knee angle, contralateral knee angle and finally ipsilateral and contralateral ankle/foot angles.

3.2 Physical Examination Data

In Table 2.2 [section 2.4], examples of measures in physical examination forms were presented [see also Appendix 9.2]. For some clinicians, certain measures are unnecessary in some cases. For instance, if a patient is unilaterally affected (*i.e.* hemiplegic or monoplegic), many measures of the non-affected limb will be missing. Thus, the completion of the form depends on the clinician judgment and the patient that is being examined.

Since data from several clinicians were considered, measurements of physical examination may vary depending on the patient and the clinician that performed the exam. For this reason, only 19 variables measured at a minimum rate of 80% were considered (Tufféry, 2012, p. 48) [see Table 3.2]. These variables include information about size and weight; hip, knee and ankle ranges of motion; muscle force; and spasticity. Within the selected parameters, measures about spasticity and muscle force tend to have lower completion rate, *e.g.* measures at V3 and knee extensor lag [Table 3.2].

Likewise kinematic data, except when stated differently, preoperative physical examination was considered ipsilaterally and contralaterally (except for height and weight). Hence, the total of preoperative physical examination variables considered for each lower limb is $N_{pre}^{phex} = 36$. These physical examination data was appended to the 765-element preoperative kinematic vector [section 3.1], making a preoperative limb vector of $N_{pre}^{tot} = N_{pre}^{kine} + N_{pre}^{phex} = 801$ elements.

Table 3.2 – List of the considered physical examination variables. For each limb, all the corresponding contralateral limbs' variables are also considered, except for height and weight.

Label	Physical Examination Variable	Completion Rate
Height	Height	100%
Weight	Weight	100%
HipAbd_HF_V1	Hip abduction (hip flexed) at V1*	89.93%
HipAbd_HF_V3	Hip abduction (hip flexed) at V3**	80.28%
HipAbd_HE_V1	Hip abduction (hip extended) at V1	89.18%
HipAbd_HE_V3	Hip abduction (hip extended) at V3	80.49%
KneeExt	Passive knee extension	80.25%
PoplA_V1	Popliteal angle at V1	97.01%
PoplA_V3	Popliteal angle at V3	83.21%
AnkDorF_KF_V1	Ankle dorsiflexion (knee flexed) at V1	89.55%
AnkDorF_KF_V3	Ankle dorsiflexion (knee flexed) at V3	80.03%
AnkDorF_KE_V1	Ankle dorsiflexion (knee extended) at V1	89.18%
AnkDorF_KE_V3	Ankle dorsiflexion (knee extended) at V3	80.01%
TibTors	Tibial torsion	87.69%
QuadLag	Quadriceps/knee extensor lag	80.05%
HipIntRot	Hip internal rotation (passive)	97.76%
HipExtRot	Hip external rotation (passive)	95.15%
FemAnt	Femoral anteversion	95.15%
HipExt_KE	Passive hip extension (knee extended)	88.43%

*V1 Passive range of motion, at the slowest stretching velocity possible

**V3 The angle of catch when stretching quickly

All the considered physical examination variables are visually measured by the clinician (except for height and weight). Most of the physical examination measures in degrees are multiples of 5° [figures 3.6, 3.7 and 3.8].

The average preoperative height and weight are respectively 145 cm (values between 105 and 185 cm) and 39.4 kg (between 15 kg and 80 kg) [figure 3.5]. The average preoperative hip abduction measures vary from 15° and 37° , whereas the average hip extension is -8° and the average femoral anteversion is 36° [figure 3.6]. The average preoperative hip internal and external rotations are respectively 59° and 14° , with some outliers very far from the boxes [figure 3.6].

The average preoperative knee extension, popliteal angles (V1 and V3) and knee extensor lag are respectively -5° , 20° , 8° and 7° [figure 3.7]. Outliers in popliteal angles correspond to typical developing children values and are most likely from non-affected limbs of the hemiplegic children in the database. The average tibial torsion is 22° . On the other hand, preoperative ankle dorsiflexion measures averages vary from 7° to -10° [figure 3.8].

Missing data in physical examination data is a crucial challenge for predicting postoperative gait. Surgeons base surgical plans partially on this data and they know which variables need to be considered depending on the patient. They may consider missing values as normal values of typical developing children, or may simply ignored some values for decision whether or not these values are available. This missing data problem will be treated in section 4.3.

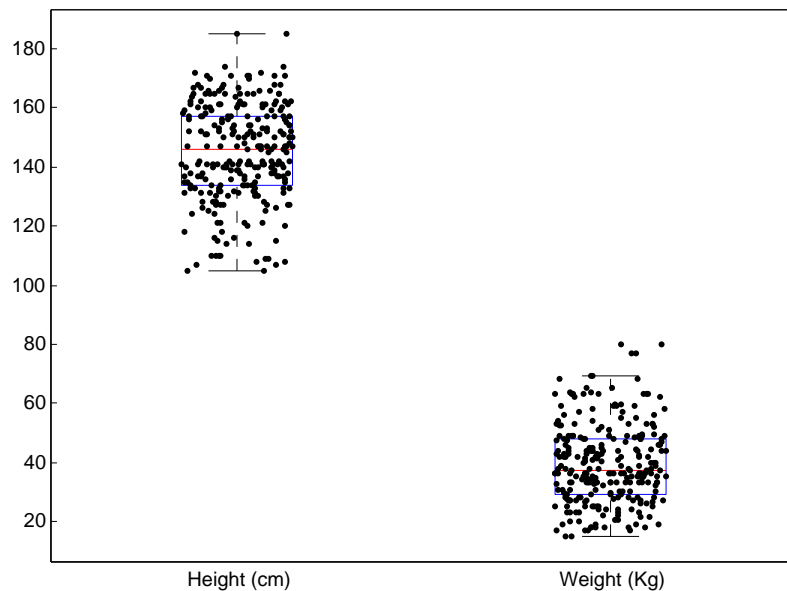


Figure 3.5 – Preoperative height and weight distributions.

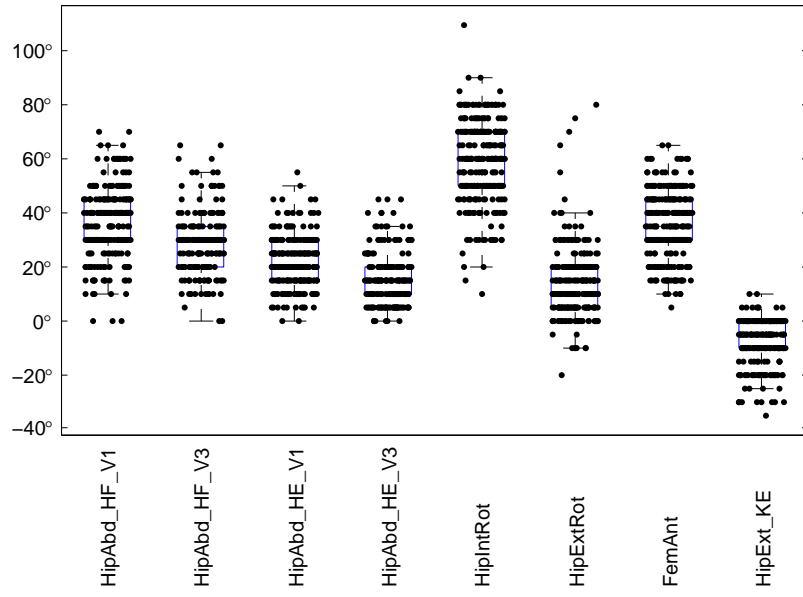


Figure 3.6 – Preoperative hip physical examination distributions (labels in Table 3.2).

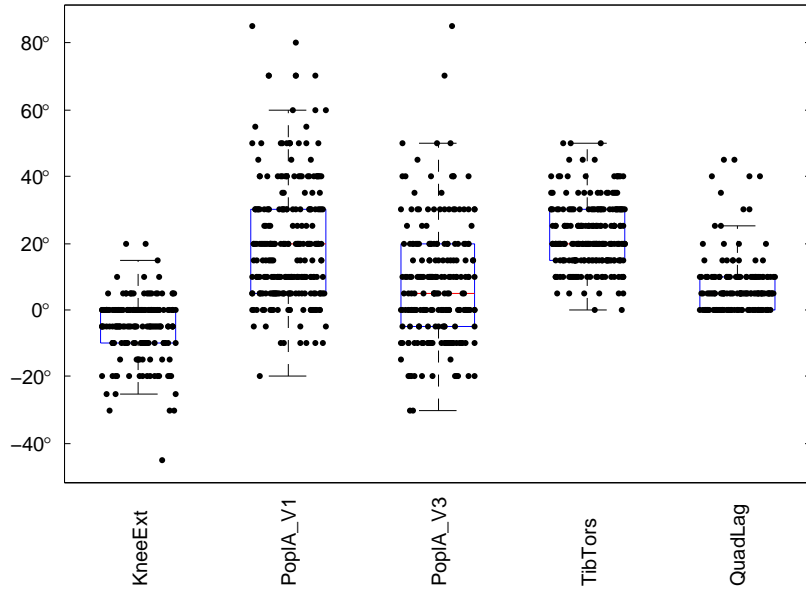


Figure 3.7 – Preoperative knee physical examination distributions (labels in Table 3.2).

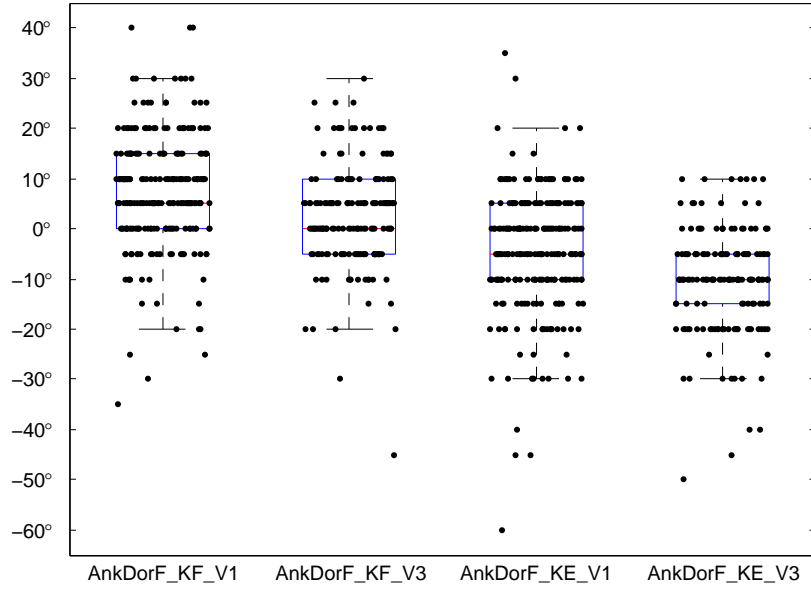


Figure 3.8 – Preoperative ankle physical examination distributions (labels in Table 3.2).

3.3 Surgical Data

In section 2.5 we have seen that the orthopaedic surgery in cerebral palsy includes a high number of different surgical procedures that vary in technique, functional objective, muscles or bones modified, etc. This introduces an important variability associated to the surgery. In addition, criteria for selecting specific surgical procedures may vary depending on the medical team, which increases this variability even more.

We have also seen in section 2.5 that the surgery was, in general, a combination of several surgical procedures performed during the same operation. $N_s = 9$ categories of surgical procedures have been established depending on their functional objective and joint or lower limb segment that is operated [table 3.3]. In these categories, some different surgical procedures are grouped in the same class if their functional objective and the affected joints or segment are the same or alike. For instance, rectus femoris transfer (Khouri & Desailly, 2013) and rectus femoris transposition (Hemo, Aiona, Pierce, Dorociak & Sussman, 2007) have different surgical techniques, but they both modify the same muscle (rectus femoris) and have the same objective: to increase knee flexion during swing for stiff knee patients. Hence they are both grouped in the rectus femoris surgery category [see Table 3.4 for other examples]. Once the surgical categories were established, surgeries of every limb in the database were expressed in terms of combinations of these classes. From now on, in this report, the term *surgical procedure* will refer to a considered surgical procedure category, even if it comprises distinct surgical procedure techniques. For example, a patient that had a rectus femoris transfer associated to a patella lowering procedure and a triceps lengthening, would be considered as having rectus femoris surgery, patella lowering and muscle ankle/foot surgery. For each lower limb j , this is expressed by a nine-element binary vector:

$$S^j = (s_1^j, \dots, s_{N_s}^j)^T \quad (3.1)$$

where

$$s_i^j = \begin{cases} 1 & \text{if procedure } i \text{ was conducted on limb } j \\ 0 & \text{if procedure } i \text{ was not conducted on limb } j \end{cases} \quad (3.2)$$

with $i = 1, \dots, N_s$ and T is the transpose operator.

In the database there are $C = 80$ distinct surgical combinations over $SC = 2^9 - 1 = 511$ (without considering a zero vector, which corresponds to no surgery) mathematical possibilities, even if in practice one probably never finds all the possible combinations of these surgical procedures. The most frequent combination has been performed on 13 lower limbs and are two different surgeries:

- Bony hip (isolated)
- Bony hip surgery with rectus femoris surgery, hamstring lengthening and muscle ankle/foot surgery.

On average, every distinct surgical combination has been performed on 3 lower limbs. On the other hand, three surgical procedures are always associated to other procedures, thus they are never alone: rectus femoris surgery, patella lowering and shank bony surgery [see table 3.3].

Table 3.3 – Considered surgical procedures categories and their frequencies in the database.

Joint/Segment	Category	Number of limbs		Number of isolated
Hip/Pelvis	Bony Hip Surgery	89	33%	13
	Muscle Hip Surgery	62	23%	2
Knee/Thigh	Rectus Femoris Surgery	129	48%	0
	Hamstring Lengthening	134	50%	5
	Patella Lowering	50	19%	0
	Distal Femoral Osteotomy	27	10%	4
Leg	Bony Shank Surgery	12	4%	0
Ankle/Foot	Muscle Ankle/Foot Surgery	131	49%	12
	Bony Foot Surgery	63	24%	5

3.4 Preoperative Kinematics and Physical Examination Data by Surgical Categories

Intuitively, one would think that different surgeries imply different kinematic patterns or physical measures. Indeed, surgery plans are based on CGA and PE data, but the correspondence between these patterns and the surgery is far from being trivial. First, criteria for surgery selection are constantly evolving thanks to the numerous scientific studies that evaluate them and the new decision-making tools. Second, surgical techniques are sometimes evolving too.

Looking at preoperative kinematics and physical examination of all the limbs that had a certain surgical procedure and those limbs that did not have the same procedure, one can see that they are similar in average [figure 3.9a], but there are some specific variables that can differentiate them [figures 3.9 and 3.10]. Applying a t-test (Stuart, 1994) to compare kinematics of limbs with and without a certain surgical procedure, the variables that can differentiate those groups can be identified [figure 3.9b]. The following interpretations can be extracted from these results:

Table 3.4 – Examples of surgical procedures per surgical category.

Category	Surgical Procedure
Bony Hip	Pelvic osteotomy
	Hip open reduction
	Proximal femoral derotational osteotomy
Muscle Hip	Hip flexor release
	Psoas lengthening
	Adductor release
Rectus Femoris	Rectus femoris transfer
	Rectus femoris transposition
Hamstring Lengthening	Hamstring lengthening
Patella Lowering	Patella lowering
Distal Femoral Osteotomy	Distal femoral derotational osteotomy
	Distal femoral extension osteotomy
Bony Shank	Tibial derotational osteotomy
	Other shank osteotomies
Muscle Ankle/Foot	Tendo-Achilles lengthening
	Soleus lengthening
	Gastrocnemius fasciotomy
Bony Foot	Calcaneum lengthening osteotomy
	Arthrodesis (double, triple)
	Other foot osteotomies

- Pelvic tilt kinematics were significantly different on limbs that had rectus femoris transfer or hamstring lengthening.
- Also limbs that had rectus femoris transfer differentiated from the others on knee flexion at early swing phase.
- Limbs that had hamstring lengthening differentiated on knee flexion at stance and late swing phases.
- Limbs that had muscle ankle/foot surgery had different pelvic rotation and ankle dorsi-flexion from limbs that had other types of surgery.
- Limbs that had bony hip surgery were distinct principally in hip rotation and foot progression.
- Limbs that underwent muscle hip surgery were different mainly on hip flexion.
- The significant preoperative kinematic characteristics of limbs that had patella lowering were excessive knee and hip flexion, as well as an excessive ankle dorsiflexion at initial contact and external foot progression.
- Distal femoral osteotomy was performed on limbs that had different knee flexion at initial contact and hip rotation and adduction.
- Limbs that underwent bony shank surgery, although very few, were mainly different on hip rotation. 42% of these limbs also had bony hip surgery.

If we look at the differences of preoperative physical examination variables per surgical procedures [figure 3.10], it can be noticed that:

- Limbs that had bony shank procedure differentiated on tibial torsion.
- Limbs that underwent muscle ankle/foot had different ankle dorsiflexion measures.
- For bony hip surgery, differences were found on hip internal and external rotations and femoral anteversion.
- For muscle hip surgery, there were differences on hip abduction and extension.
- Limbs that underwent hamstring lengthening were different on knee extension and popliteal angles.
- For distal femoral osteotomy, there were significant differences on hip abduction at extended hip, knee extension, popliteal angles and femoral anteversion.
- For patella lowering, the differences were found on hip and knee extensions, tibial torsion, quadriceps lag, ankle dorsiflexion with extended knee at V3 and hip abduction with extended hip at V1.
- Conversely, there is no significance difference on any physical examination variable for limbs that had rectus femoris transfer or bony foot surgery. For rectus femoris surgery, this may be explained by not considering the Duncan-Ely test, which is related to rectus femoris dysfunction at swing phase (Marks & Chambers, 2003). In the case of foot surgery, the explanation may be on the omission of some foot deformity variables, such as hindfoot varus-valgus.

All these kinematic and physical examination relations are globally coherent with the purposes of each surgical procedure.

3.5 Surgery Outcome

The purpose of the surgery is to improve abnormalities on kinematic patterns. Each surgical procedure is supposed to modify certain patterns and the surgical combination should globally improve kinematics by approaching them to the standard gait patterns.

In section 2.5, we have seen some theoretical kinematic improvements of some surgical procedures. In order to statistically identify the effect of each considered surgical procedure, kinematic variation (postoperative - preoperative kinematics) of limbs that had a certain surgical procedure were compared to the variations of limbs that did not have the surgical procedure in question [figure 3.11]. From this analysis, the following interpretations are worth to be highlighted:

- Bony hip surgery has significant effect principally on hip rotation and foot progression at stance phase, as well as pelvic obliquity at swing phase, pelvic rotation at mid-stance, hip flexion at swing and hip adduction at mid-stance.
- Muscle hip surgery affects pelvic rotation and hip adduction, as well tenuously hip flexion, especially at late swing.
- Rectus femoris transfer increases knee flexion at swing, and has also an effect on pelvic tilt at mid-stance and late swing.
- Hamstring lengthening decreases considerably knee flexion at initial contact, stance and late swing and also has an increasing effect on pelvic tilt.

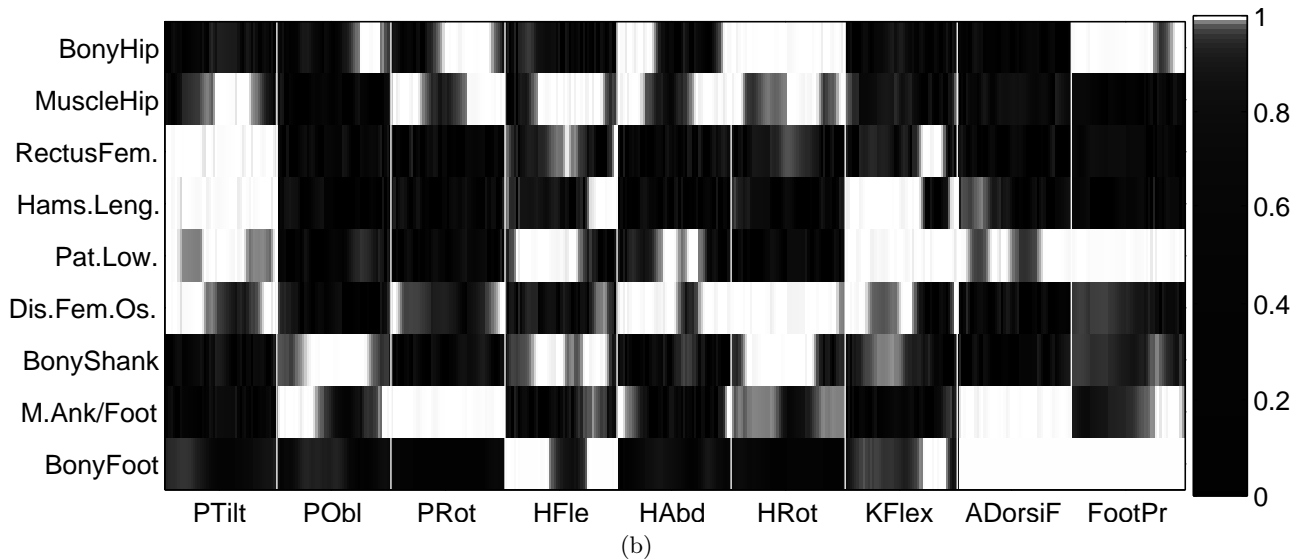
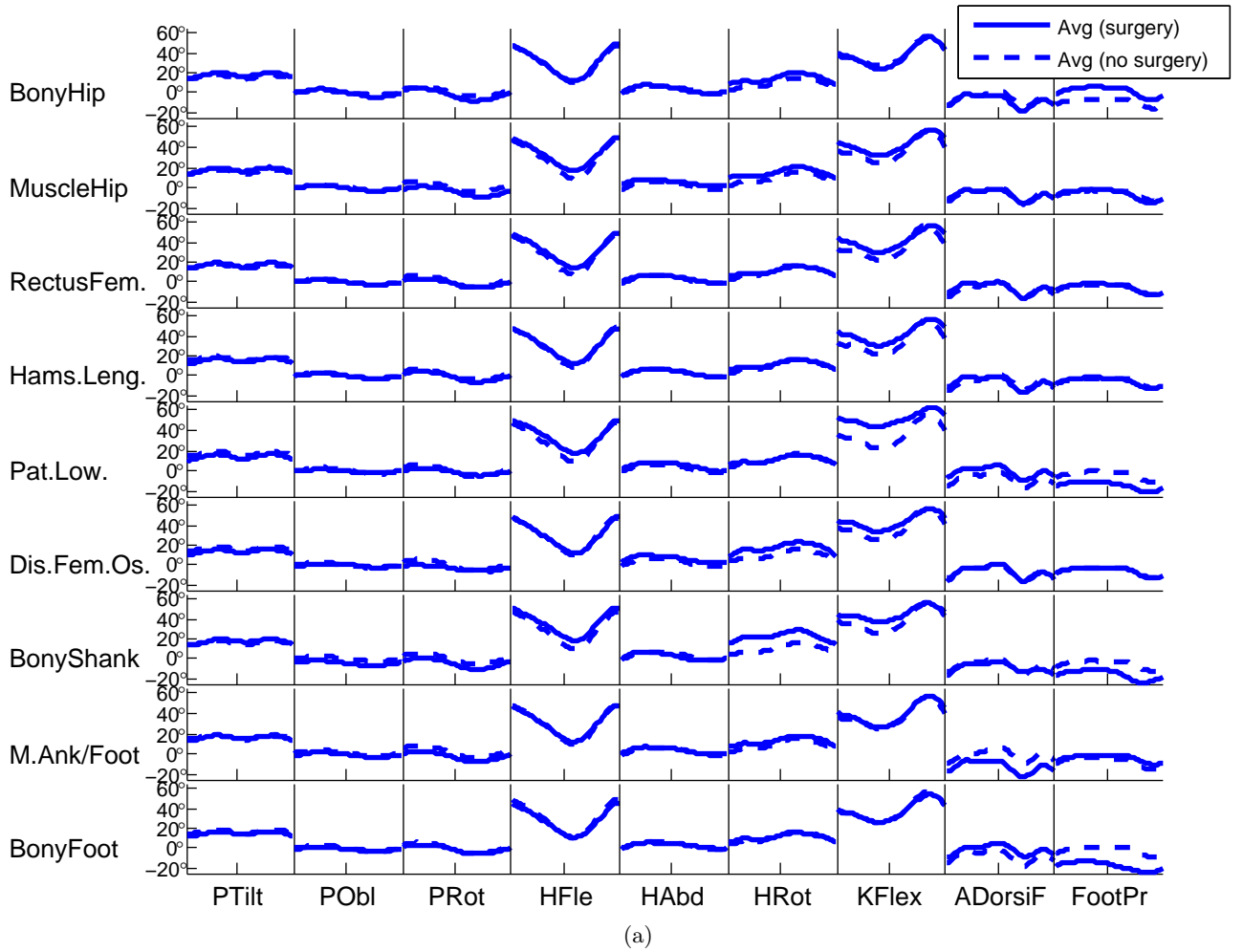


Figure 3.9 – Preoperative kinematics per surgical procedure. 3.9a Average of limbs that had each surgical procedure (solid blue) and average of limbs that did not have the considered procedure (dashed blue). 3.9b Independence test between averages with and without each surgery. 0 (black) means no significance of the kinematic variable and 1 (white) is the maximal significance.

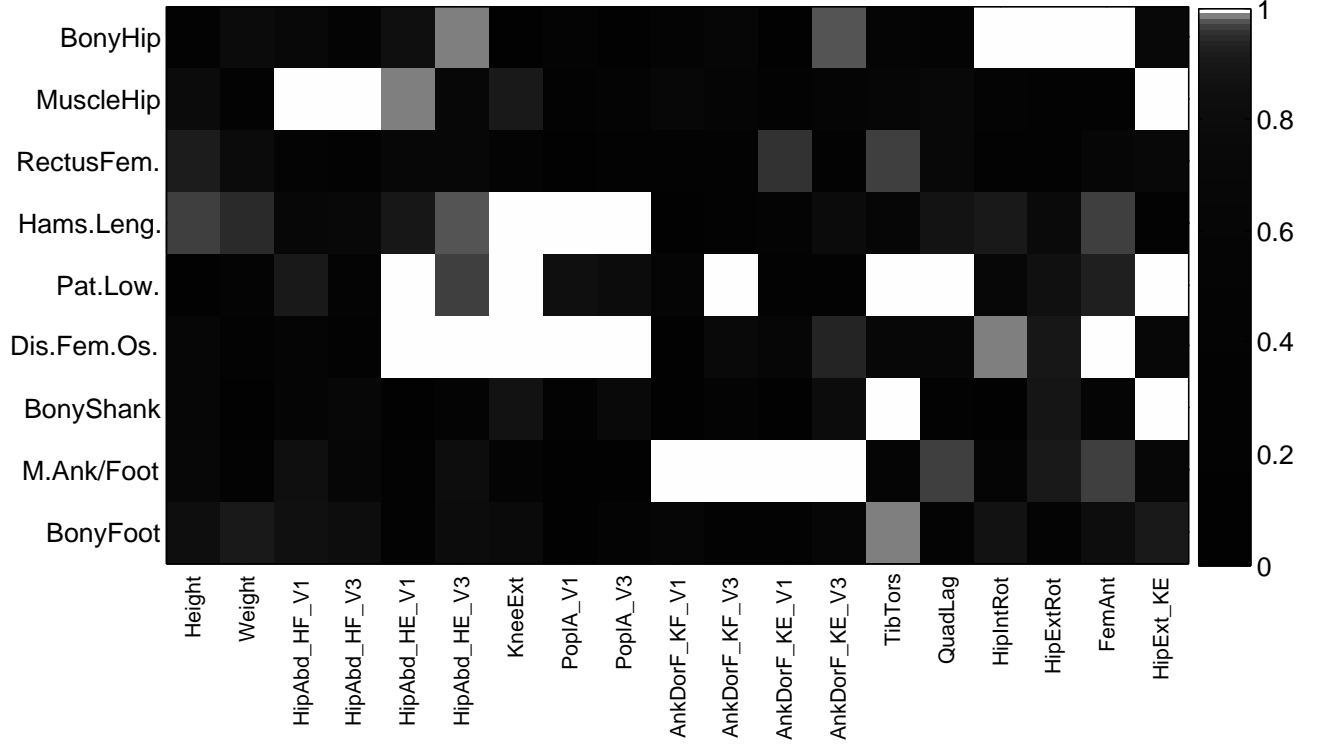


Figure 3.10 – Statistical significance of physical examination variables on surgery selection. 0 (black) means no significance and 1 (white) is the maximal significance (labels in table 3.2).

- Patella lowering also has significant effects on pelvic tilt and on knee flexion (during the whole gait cycle).
- Distal femoral osteotomy also has an effect on pelvic tilt and on knee flexion at stance phase.
- Muscle ankle/foot surgery modifies pelvic rotation and ankle dorsiflexion, as well as foot progression at swing phase.
- Bony foot surgery has its main effect on foot progression, but also affects ankle dorsiflexion at pre-swing and knee flexion at mid-stance.
- Conversely, there is almost no effect found for bony shank surgery, except at initial contact for pelvic rotation and foot progression, as well as hip rotation at early swing phase.

These found effects are closely related to the differences or abnormalities found in the pre-operative data [section 3.4], which is totally logical, as well as most of the scientific literature [see section 2.5]. Nevertheless, there are some unexpected results. For example, muscle hip surgery did not have a considerable effect on hip flexion and bony shank surgery has an effect on very few kinematic variables. The second example might be explained by the lack of limbs having undergone bony shank surgical procedure. Another unexpected result is the insignificant statistical effect of distal femoral osteotomy on hip rotation.

3.6 General Discussion

Given the complexity of the surgery with all the possible combinations and the number of examples per each surgery, the statistical learning task seems really challenging. Since there

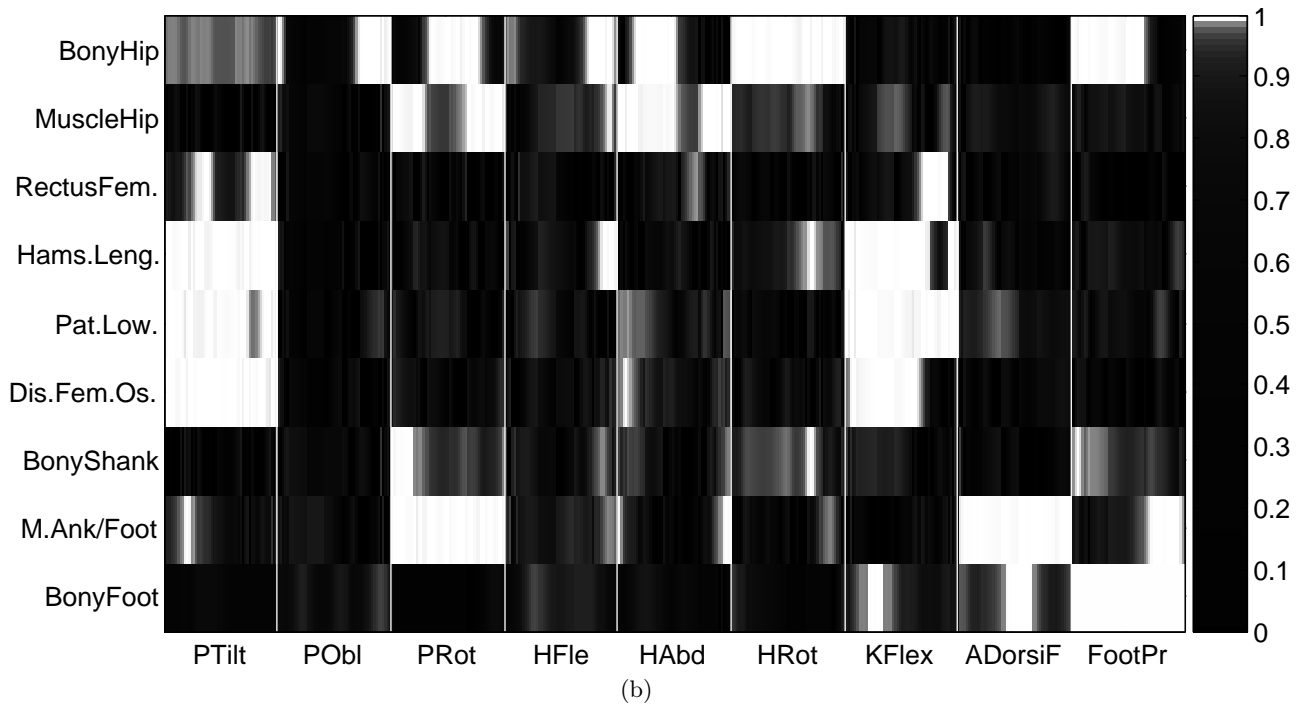
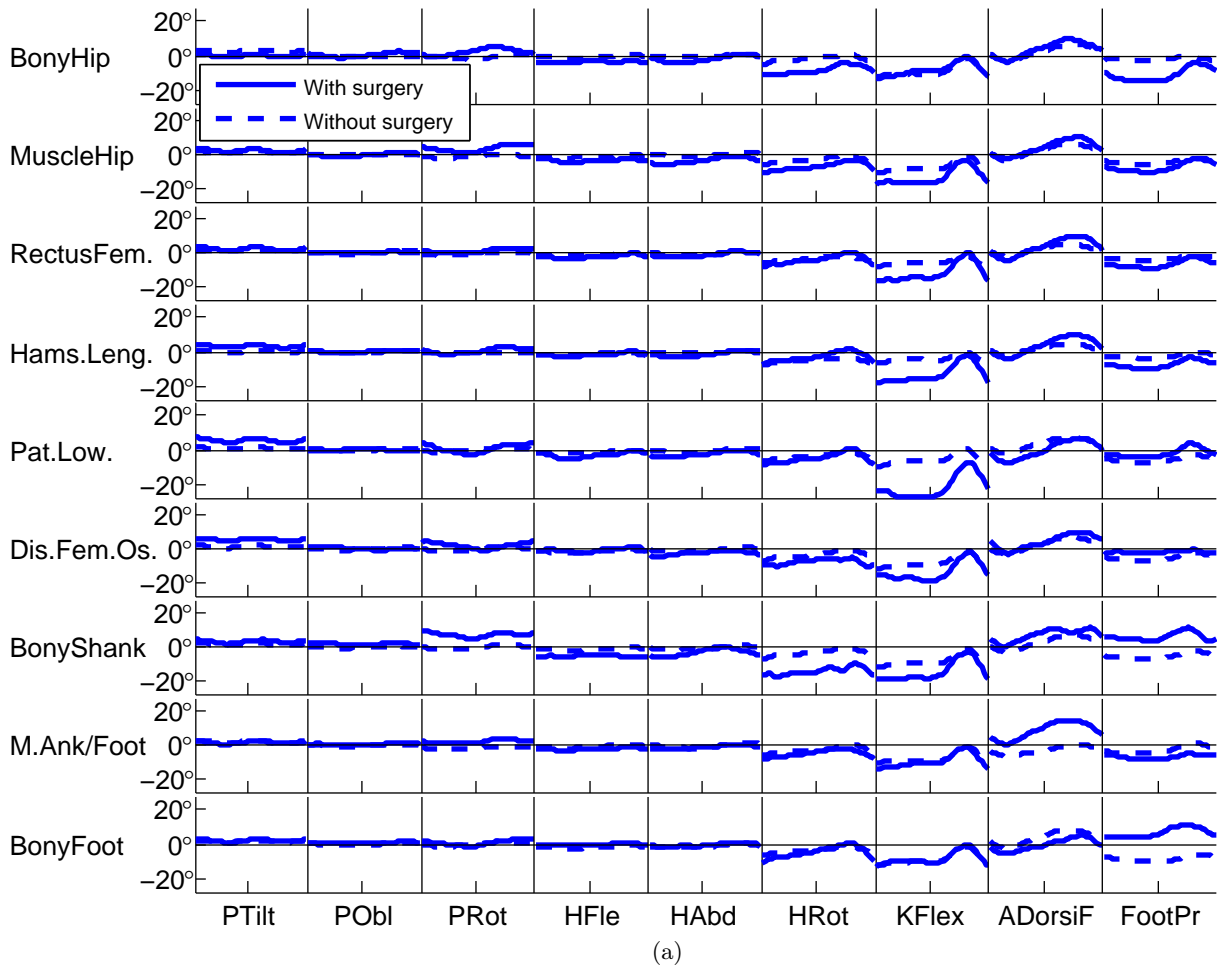


Figure 3.11 – Average variation and effect of surgical procedures. 3.11a Average kinematic variation (Postoperative - Preoperative) in a 2-standard deviation band. 3.11b Statistical effect of surgical procedure on kinematic variables. 1 (white) corresponds to a significant effect and 0 (black) corresponds to no statistical effect.

are numerous combinations (80 over 511 possibilities) and too few examples per combination (3 on average per surgical combination and 13 maximum), it is impossible to construct surgery-specific statistical models as, for example, Niiler et al. (Niiler et al., 2007) did for rectus femoris transfer and concurrent surgeries. It would also be too complicated to introduce other probable factors such as contralateral surgery, which would increase the number of possible treatments and would decrease the number of examples per treatments (ipsilateral and contralateral surgery). In addition, the number of preoperative variables (801 in total when adding kinematics and physical examination) is too large compared to the total number of examples (268), which complicates even more the learning task. This introduces the necessity of a dimensionality reduction stage, that will be discussed in the next chapter [chapter 4]. Moreover, the missing data problem of physical examination data introduces another complication into the learning task. This is also a problem we try to circumvent in the next chapter.

On the other hand, we have found some important statistical relations between preoperative variables and surgical procedures, which might simplify the problem. Moreover, we have seen that surgical procedures only significantly affect some kinematic variables. This may also simplify the predictive model design. We have also seen that, after surgery, kinematics tend towards standard gait and are less variable, although are still more variable than standard gait. This gives a hint on how postoperative kinematics should be in general.

Notwithstanding, there are still several questions to answer:

1. What can be done with missing data of physical examination?
2. Is it possible to find a good lower-dimensional representation of the preoperative vector?
3. Which is the optimal regression method for predicting postoperative kinematics?

There are at least three options to answer the first question. One option is to discard the physical examination information, although clinicians give importance to it. The second option is to utilize learning methods that can deal with missing data, such as Hopfield networks (Wang, 2005). The third option is to replace the missing data with some values in order to have a complete imputed data and then be able to use any learning method. This third option, as well as the second question, are discussed in chapter 4. The answer to the third question will be progressively discussed in the subsequent chapters of this report.

Chapter 4

Feature Extraction and Dimensionality Reduction

Contents

4.1	Curve Fitting for Kinematics	41
4.1.1	Harmonic approximation	43
4.1.2	Periodic splines approximation	45
4.1.3	Comparison between periodic splines and harmonic approximation	46
4.2	Variable Selection	48
4.3	Physical Examination Missing Data Imputation	49
4.4	Dimensionality Reduction by Principal Component Analysis	52
4.4.1	Method	55
4.4.2	Results	56
4.5	General Discussion	59

After preprocessing, preoperative kinematic data for any given subject consist of fifteen time-normalized concatenated time series, thus $N_{pre}^{kine} = 765$ elements. Given the quantity of available samples $N_{limb} = 268$, it is a hard task to effectively learn too many parameters.

For this reason, a dimensionality reduction stage is needed for feature extraction, so that we have a reduced number of parameters to describe every sample. In addition, dimensionality reduction helps to improve learning performance by reducing noise and data redundancy. Dimensionality reduction methods include linear subspaces (Jiang, 2011), dictionary learning (Tosic & Frossard, 2011), pruning on neural networks (Dorizzi, Pellioux, Jacquet, Czernichow & Munoz, 1996; Leray & Gallinari, 2001), among others. We have chosen to characterize gait cycles (and sometimes physical examination data) by three different approaches: data fitting (Motulsky & Christopoulos, 2004; Jamshidi, Kirby & Broomhead, 2011), variable selection (Guyon & Elisseeff, 2003) and principal component analysis (Jolliffe, 2002).

4.1 Curve Fitting for Kinematics

The principle of curve fitting is to adjust parametric functions by least mean squared error optimization (Bishop, 2006, pp.4-5), where the mean squared error is

$$MSE = \frac{\sum_{i=1}^N (y_i - f(x_i))^2}{N} \quad (4.1)$$

with $\{(x_i, y_i)\}_{i=1}^N$ the set of input-output vectors, $f(x_i)$ the estimated value, N the number of points and f a mathematical function to be designated. Then the original data points will be replaced by the optimized parameters.

Given any data set, it is always possible to quickly and easily fit thousands of functions, but it is difficult to know which one is the best for a given problem. With so many candidate models, overfitting is a real danger. It is important to recall that we want to preserve the information of the kinematic curves and, at the same time, reduce their dimensionality and complexity. Moreover, as the final application will include an animation of the motion, it would be interesting to keep continuity and to force periodicity of the gait cycle. Hence, only continuous and periodic functions were tested. Specifically, both periodic Splines (De Boor, 2001) and trigonometric approximation were utilized as fitting functions.

For this experiment, 236 CGA recordings of 94 children with CP were considered independently on their preoperative or postoperative nature. 65% (153) of the trials correspond to male patients and 35% (83) correspond to female patients. The average age of the patients was 13.81 years old ($MIN = 5$, $MAX = 26$, $\sigma = 3.76$). The purpose here was to find generic optimal models per kinematic angle and not a mathematical relation between preoperative and postoperative walking signals.

In order to avoid overfitting and find a compromise between approximation error and number of parameters, Akaike's information criterion (Akaike, 1998) was evaluated. This criterion is widely used in order to determine complexity of optimal models in numerous different applications (Motulsky & Christopoulos, 2004, p.143). We have actually considered the corrected version of Akaike's information criterion (Hurvich & Tsai, 1989), which considers the size of the sample and is given by the following equation:

$$AIC_c = N \ln(MSE) + 2K + \frac{2K(K+1)}{N-K-1} \quad \forall N > K+1 \quad (4.2)$$

N is the number of data points, MSE is given in Equation 4.1 and K is the number of parameters of the model.

Alternatively to Akaike's information criterion, thresholding on MSE and correlation (r) was also examined to determine optimal models. In this regard, we have computed MSE [Equation 4.1], root-mean-squared error ($RMSE$) and r with different numbers of parameters by:

$$r = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}} \quad (4.3)$$

and from Equation 4.1

$$RMSE = \sqrt{MSE} = \sqrt{\frac{\sum_{i=1}^N (y_i - f(x_i))^2}{N}} \quad (4.4)$$

Then optimal models were those with the minimal number of parameters that satisfy the th conditions:

$$\begin{aligned} & \arg \min (M(K)) \\ & \text{with} \\ & \overline{RMSE}_M \leq th_{rmse} \\ & \overline{r}_M \geq th_r \end{aligned} \quad (4.5)$$

with $M(K)$ is the model with number of parameters K , th_{rmse} and th_r respectively error and correlation thresholds and $\bar{(\cdot)}$ is the mean operator over all subjects, where

4.1.1 Harmonic approximation

The harmonic fitting function tested is a truncated Fourier series decomposition (Kawata, 1972, pp.49-51):

$$f(x) = c_0 + \sum_{k=1}^{m_f} a_k \sin(k\omega x) + b_k \cos(k\omega x) \quad (4.6)$$

where ω is the fundamental angular frequency and m_f is the order of the model.

Since we consider a complete gait cycle (from 0 to 2π), it can be assumed that the sinusoidal model has the same fundamental frequency as the mean cycle frequency, so $\omega = 1$ for the time-normalized cycle. In this case, the number of needed parameters to approximate a curve becomes:

$$K = 2m_f + 1 \quad (4.7)$$

Then, Equation 4.6 becomes:

$$f(x) = c_0 + \sum_{k=1}^{m_f} a_k \sin(k x) + b_k \cos(k x) \quad (4.8)$$

The parameters c_0 , a_k and b_k are directly estimated as:

$$c_0 = \frac{1}{N} \sum_{i=1}^N y_i \quad (4.9)$$

$$a_k = \frac{2}{N} \sum_{i=1}^N y_i \sin(k x_i) \forall k \in \mathbb{N}^* \quad (4.10)$$

$$b_k = \frac{2}{N} \sum_{i=1}^N y_i \cos(k x_i) \forall k \in \mathbb{N}^* \quad (4.11)$$

where $\{y_i\}_{i=1}^N$ are the measured data points and x_i the normalized cycle time step (from 0 to 2π).

The main advantage of this model is that sinusoids are naturally continuous and periodic, which is perfectly adapted to our problem.

The harmonic approximations closely follow the kinematic curves for all the considered angles [Figure 4.1], but one wants to find the best (possible) representation with the minimal number of parameters. In Figure 4.2, the *MSE* of the trigonometric approximation can be observed for all the ankle dorsiflexion signals with respect to the number of parameters of the model. The higher we go through the figure, the larger the number of parameters is. Then, a good compromise between number of parameters and error can be found in the lowest horizontal line for which all the subjects have a somewhat small error (blue). For this dorsiflexion angle, the nine-parameter model ($K = 9$) can be considered as optimal.

Following this error-complexity compromise approach for the same angle and approximation function (sinusoidal), the corrected Akaike's information criterion (AIC_c , see Equation 4.2) optimal models for each subject are shown in Figure 4.3. There is no uniformity in the selected models (it varies from 17 to 41 parameters). Since we previously observed that nine parameters suffice for a good approximation, and the minimal number of parameters by AIC_c is 17, AIC_c criterion seems to be unsuitable for this approximation problem. For this reason, the thresholding on *MSE* and r [Equation 4.5] will be considered for determining optimal models.

If we consider the minimal number of parameters of sinusoidal approximation that satisfy Equation 4.5, between 3 (hip flexion) and 11 (hip rotation) parameters are needed to approximate a kinematic curve with $th_{rmse} = 4^\circ$ and $th_r = 0.95$, depending on the considered angle

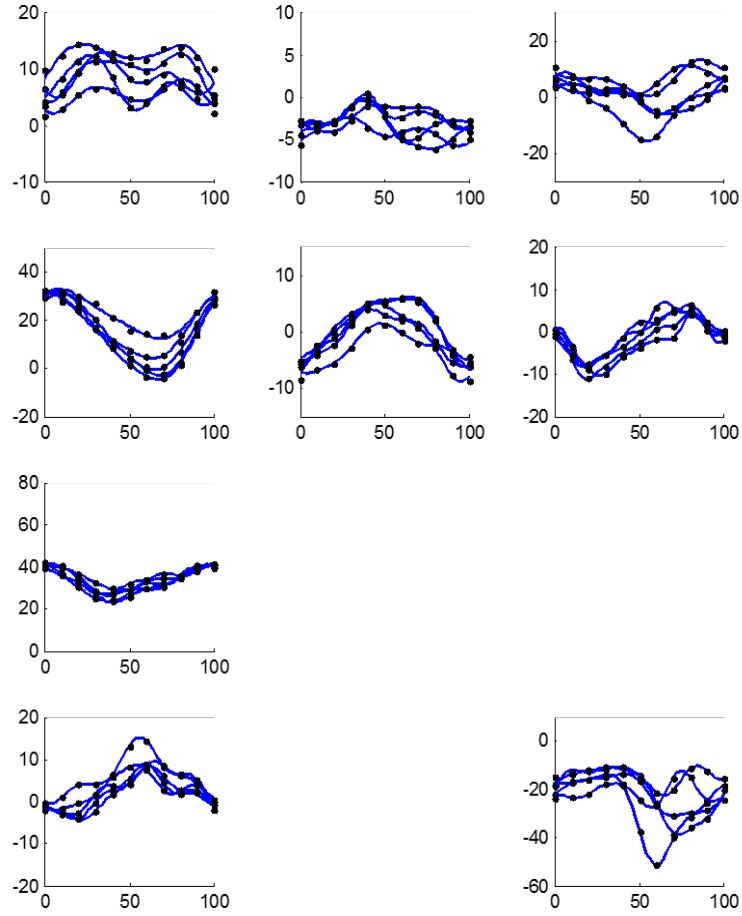


Figure 4.1 – Example of trigonometric kinematic curves approximation for different gait cycles of a random limb. Blue lines are the trigonometric approximations and black crosses correspond to the measured points.

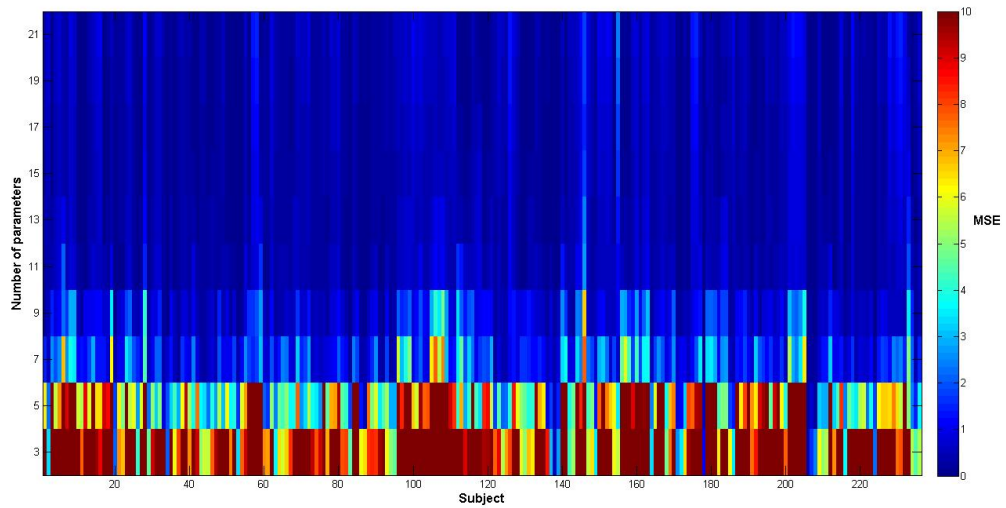


Figure 4.2 – Harmonic approximation MSE per subject and number of parameters. Example for ankle dorsiflexion. Blue color corresponds to a small MSE.

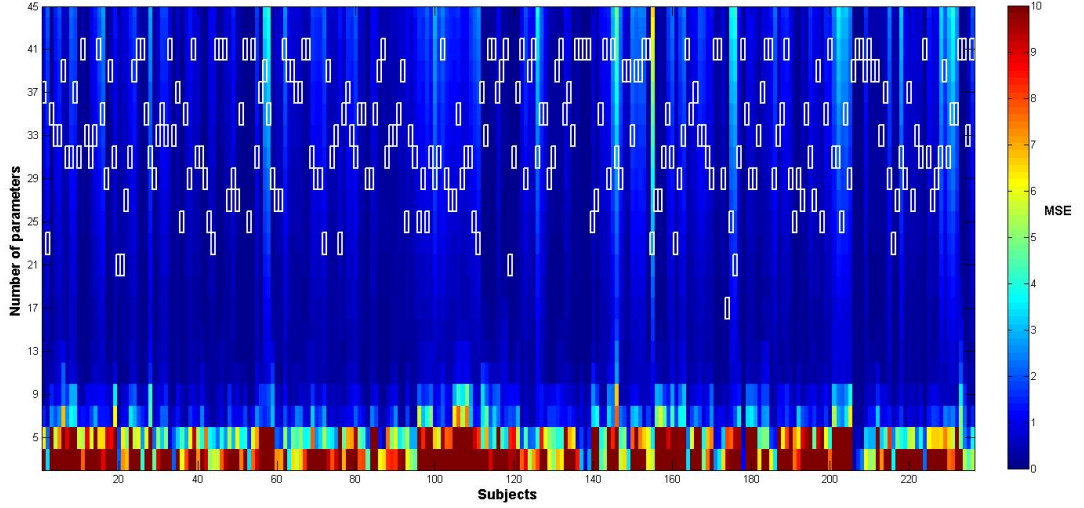


Figure 4.3 – Optimal AIC_c models per subject for harmonic approximation. Example for ankle dorsi-flexion angle. White boxes correspond to optimal models selected with AIC_c . Blue color corresponds to a small MSE.

[Table 4.1]. In total, 101 parameters are necessary to approximate the fifteen kinematic curves with the same thresholds.

Table 4.1 – Minimal number of approximation parameters per kinematic angle and average MSE thresholds for harmonic approximation [see equation 4.1]. Correlation threshold is fixed to $th_r = 0.95$. The number of parameters was computed with Equation 4.7. The total number of parameters considers both lower limbs (parameters per angles count twice, except for pelvis angles).

$th_{rmse}^2 (deg^2)$	1	2	3	4	16
$th_{rmse} (^\circ)$	1	1.41	1.73	2	4
Pelvic tilt	5	5	5	5	5
Pelvic obliquity	7	7	7	7	7
Pelvic rotation	7	5	5	5	5
Hip flexion	7	7	5	5	3
Hip adduction	7	7	7	7	7
Hip rotation	19	13	11	11	11
Knee flexion	13	9	9	7	5
Ankle DorsiFlexion	11	9	7	7	7
Foot progression	19	9	9	9	9
Total	171	125	113	109	101

4.1.2 Periodic splines approximation

A spline is piecewise polynomial function, *e.g.*

$$f(x) = \begin{cases} x^2 + 2, & \text{if } 0 \leq x \leq 1 \\ -x^2 + 4x, & \text{if } 1 < x \leq 2 \end{cases}$$

is a spline of order $m = 2$ and $s = 2$ segments.

Spline fitting consists in dividing the data series in segments by selecting knots and then adjusting

a polynomial to each part of the curve. For continuity, the derivatives of adjacent segments polynomials must be the same at the knots. For one curve, the number of parameters

$$K_{spline} = s \times (m + 1) \quad (4.12)$$

where s is the number of segments and m the order of the polynomials (order of the Spline). Different combinations of number of segments and polynomial order were studied, varying between $s = 2, \dots, 10$ and $m = 2, \dots, 9$.

A spline is periodic if it satisfies the following boundary condition:

$$f^{(j)}(\underline{x}) = f^{(j)}(\bar{x}), j = 0, \dots, m - 1 \quad (4.13)$$

where \underline{x} and \bar{x} are respectively the lower and the upper bounds of the interval $[x^{(j)}]$, thus respectively the beginning and the end of the gait cycle. This boundary condition imposes that function values must be the equal at the ending points as well as the values of the their derivatives [Equation 4.13] (De Boor, 2001, p.282).

The implemented periodical splines are basis splines, also named B-Splines, which allow to have a periodic basis directly [IDEM, p.87].

To illustrate the approximation capability of periodic splines, Figure 4.4 shows the results of curve fitting for a random limb. Likewise the trigonometric approximation [Figure 4.1], the splines follow closely the kinematic curves, making it a good approximation, except for certain points such as the ankle dorsiflexion minimum.

When considering the minimal number of parameters with thresholds $th_{rmse} = 4$ and $th_r = 0.95$ [see equation 4.5], between 7 (hip flexion) and 50 (hip rotation) are needed to model a kinematic curve depending on the considered angle [Table 4.2]. In total, for these same thresholds, 286 parameters are needed to approximate the fifteen kinematic angles.

Table 4.2 – Minimal number of approximation parameters per kinematic angle and average MSE thresholds for periodic Splines approximation [see equation 4.1], with $th_r = 0.95$. The number of parameters was computed with Equation 4.12. The total number of parameters considers both lower limbs (parameters per angles count twice, except for pelvis angles).

$th_{mse} (deg^2)$	1	2	3	4	16
$th_{rmse} (^\circ)$	1	1.41	1.73	2	4
Pelvic tilt	14	14	14	14	14
Pelvic obliquity	10	10	10	10	10
Pelvic rotation	14	10	10	10	10
Hip flexion	18	14	14	14	7
Hip adduction	14	14	14	14	14
Hip rotation	100	100	50	50	50
Knee flexion	50	27	27	21	14
Ankle dorsiflexion	50	27	21	21	14
Foot progression	27	27	27	27	27
Total	556	452	340	328	286

4.1.3 Comparison between periodic splines and harmonic approximation

Since we know that both splines and harmonic series are able to approximate joint angles (figures 4.4 and 4.1), a good compromise between error and number of parameters must be searched. The most suitable fitting function will be the one that gives the best compromise between number of parameters and approximation error.

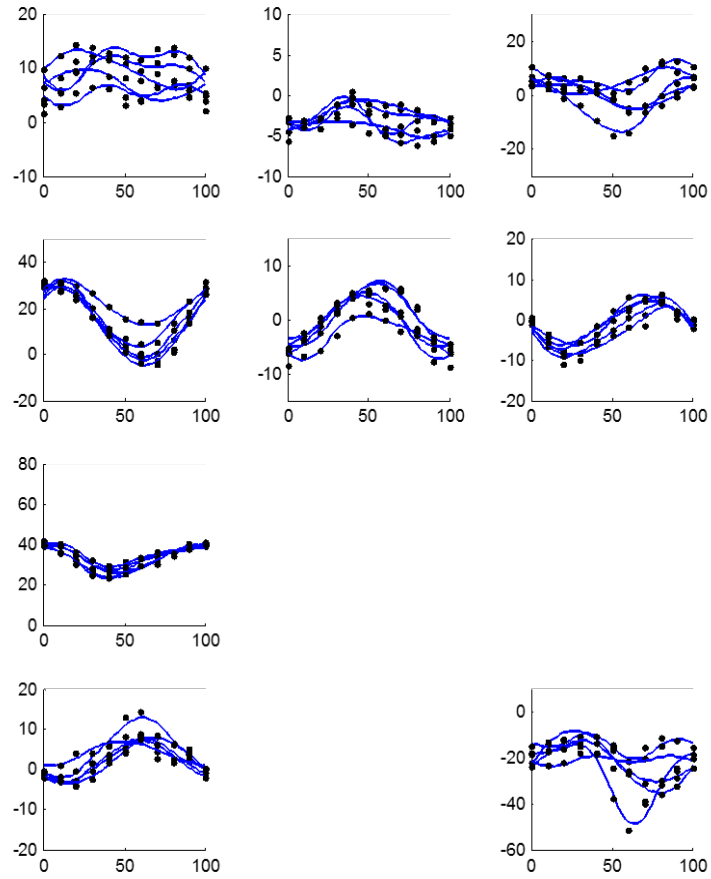


Figure 4.4 – Example of periodic Spline fitting for a random limb. Exclusively the left limb angles are shown (9 angles instead of 15 in total). Splines of 4 segments and of order 3.

For an average approximation error smaller or equal to $th_{mse} = 5$ and a average correlation greater or equal to $th_r = 0.95$ [see Equation 4.5], Fourier approximation needs 101 parameters in total, whilst periodic Splines approximation needs 286 parameters [see Table 4.3]. For a lower average correlation of $th_r = 0.85$, Fourier needs 67 parameters and periodic Splines needs 87. With these second conditions, the trigonometric approximation attains a dimensionality reduction rate of 91.24%, and the Splines approximation obtains a reduction rate of 88.63%. Even if the number of parameters decreases more for periodic Splines, in both cases the trigonometric approximation needs less parameters to achieve the same performance. For this reason, the only curve fitting function that will be further considered is the trigonometric approximation.

Furthermore, the angle that needs the largest number of parameters is hip rotation, making it the most difficult signal to approximate. On the other hand, the easiest angle to approximate is hip flexion, followed by hip adduction, pelvic rotation and pelvic obliquity.

Table 4.3 – Comparison between periodic Splines and Fourier approximations. Minimal number of parameters for an average error of $th_{mse} = 5$ and different values of th_r .

th_r	Fourier		Splines	
	0.95	0.85	0.95	0.85
Pelvic tilt	5	5	14	7
Pelvic obliquity	7	3	10	5
Pelvic rotation	5	3	10	5
Hip flexion	3	3	7	4
Hip adduction	7	3	14	5
Hip rotation	11	7	50	7
Knee flexion	5	5	14	7
Ankle dorsiflexion	7	5	14	5
Foot progression	9	5	27	7
Total	101	67	286	87

4.2 Variable Selection

Another possible approach is to select some of the available variables without doing any kind of global approximation or projection. This would allow us to select at the same time local information of the kinematic curves and some physical examination variables.

It exists several techniques for variable selection [see (Guyon & Elisseeff, 2003)]. In this work we have considered a variable selection technique based on the “probe” technique described in (Stoppiglia, Dreyfus, Dubois & Oussar, 2003). This algorithm is able to select discriminant variables when there are more candidate variables than examples in the training database (Dreyfus, Martinez, Samuelides & Collectif, 2008).

The variable selection is done by a ranking technique adding a probe candidate variable. The variables ranked higher than the probe are selected and the variables that are ranked lower than the probe are rejected. The probe is a completely random variable unrelated to the target output.

The ranking technique consists in a Gram-Schmidt orthogonalization (Chen, Billings & Luo, 1989). If $\vec{p}_1, \dots, \vec{p}_k$ are the k column vectors corresponding to the candidate variables, \vec{r} the probe and \vec{y} the target output vector. This iterative procedure is described by algorithm 1, where $\langle \vec{a}, \vec{b} \rangle$ is the inner product of vectors \vec{a} and \vec{b} .

Algorithm 1 Gram-Schmidt variable ranking with probe.

```

1:  $\vec{p}_{k+1} \leftarrow \vec{r}$ 
2: for  $i = 1$  to  $k + 1$  do
3:    $\vec{z}_i \leftarrow \arg \max_{\vec{p}_j, j=1, \dots, k+1} \frac{\text{Cov}(\vec{p}_j, \vec{y})}{\sigma_{p_j} \sigma_y}$ 
4:   for  $j = 1$  to  $k + 1$  do
5:      $\vec{p}_j \leftarrow \vec{p}_j - \frac{\langle \vec{p}_j, \vec{z}_i \rangle}{\langle \vec{z}_i, \vec{z}_i \rangle} \vec{z}_i$ 
6:   end for
7:    $\vec{y} \leftarrow \vec{y} - \frac{\langle \vec{y}, \vec{z}_i \rangle}{\langle \vec{z}_i, \vec{z}_i \rangle} \vec{z}_i$ 
8: end for

```

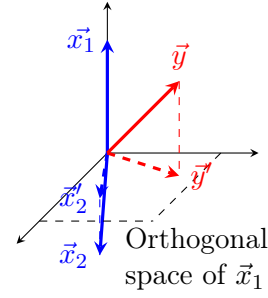


Figure 4.5 – Gram-Schmidt orthogonalization.

The inconvenient of such approach is that the number of input variables may differ when several output variables are considered. For instance, for a 51-point kinematic curve, two consecutive points may be associated to different selected inputs (and of different dimensions). This is, for example, the case for postoperative pelvic obliquity at initial contact and at 2% of the gait cycle, where 35 and 16 preoperative variables are respectively selected [see table 4.4]. In general, for both postoperative variables, some points from the beginning to the end of the preoperative pelvic obliquity are selected. The reason of this is that it is the same kinematic angle before and after the surgery. In addition, some points from the other preoperative kinematic variables are selected, with less importance for hip flexion, foot progression, and contralateral knee flexion and ankle dorsiflexion. The selected preoperative instants are close but in general different, *i.e.* preoperative pelvic obliquity respectively at 8% and 10% of the cycle for postoperative pelvic obliquity at initial contact and at 2% of the gait cycle. On the other hand, it is interesting that in both cases the contralateral ankle dorsiflexion (knee extended) is selected from physical examination data. This relation between passive ankle dorsiflexion and kinematic pelvic obliquity at the beginning of the cycle may be explained as follows: at initial contact, the contralateral foot is in stance phase and the contralateral knee tends to the extension. In such case, the pelvic obliquity may adapt depending on the ankle dorsiflexion capacity.

Despite the limitations mentioned above, this variable selection technique was utilized in the experiment reported in section 5.2, where two single outputs were considered for prediction: knee flexion and pelvic tilt at initial contact.

Since curve fitting is insufficient in terms of dimensionality reduction and since variable selection is not well adapted to multiple outputs, another feature extraction method should be considered. Moreover, if we want to include the available (or a good part of) physical examination data, we need to circumvent the missing data problem. In the next two sections, the physical examination missing data imputation will be discussed and then we will discuss a considered dimensionality reduction technique using principal component analysis.

4.3 Physical Examination Missing Data Imputation

Most of the regression and classification techniques work with complete data. If we want to exploit the available physical examination information, it is important to circumvent the missing data problem described in section 3.2. Physical examination is widely used for treatment decision-making in cerebral palsy (Trost, 2009) and complements 3-D gait quantified analysis (Desloovere et al., 2006).

Missing data might have different natures such as missing completely at random (*MCAR*), missing at random (not completely, *MAR*) and not missing at random (*NMAR*) (Allison, 2009).

Table 4.4 – Selected preoperative variables with probe technique for postoperative pelvic obliquity at two consecutive points: initial contact (IC) and at 2% of the gait cycle. Percentages correspond to cycle instant (phase) of the gait signals.

Preoperative variable	Postoperative Pelvic Obliquity	
	at initial contact	at 2% of gait cycle
Pelvic tilt	38%	36%
	62%	
	100%	100%
Pelvic obl.	8%	10%
	46%	44%
	62%	78%
	100%	100%
Pelvic rot.	2%	2%
	28%	
	42%	
Hip flexion	80%	
Hip add.	58%	58%
	80%	
	92%	
Hip flex. cont.	62%	64%
Hip add. cont.	IC-2%	
	10%	8%
	20%	42%
	100%	
Knee Flexion	100%	96%
Knee flex. cont.	52%	
	70%	
Ankle DorFl.	16%	
	58%	100%
Ankle DorFl. cont.	10%	
	58%	
Foot Prog.	-	-
Foot Prog. cont.	6%	
	56%	56%
	100%	100%
Physical exam.	Ankle DorFl. (ext. knee) cont.	Ankle DorFl. (ext. knee) cont.
	Tibial torsion	
	Hip external rotation	
	Hip ext. rotation cont.	
Total number of param.	35	16

MCAR assumption means that the probability of having a missing value for a variable Z is independent of the value that is missing and the other available variables X . If R_z is a Bernoulli variable such that:

$$R_z = \begin{cases} 1 & \text{if } Z \text{ is missing,} \\ 0 & \text{if } Z \text{ is observed.} \end{cases}$$

Thus, the probability of missing for *MCAR* is:

$$P(R_z = 1/X, Z) = P(R_z = 1)$$

For instance, *MCAR* can be assumed if it exists a complete data table and some values are randomly removed or lost.

MAR is when the assumption is that the missing value is independent of the missing value itself but dependent of the other observed variables, hence the probability of missing is:

$$P(R_z = 1/X, Z) = P(R_z = 1/X)$$

For example, a specific test or measure might be performed on subjects depending on their ages. In such case, measures of subjects outside the age range of interest are most likely to be missing. On the other hand, *NMAR* when the probability of missing of variable Z depends of the value that is missing and the observed variables X . In such case, the probability of missing can only be estimated as:

$$P(R_z = 1/X, Z) = f(X, Z)$$

Where f is a bounded function between 0 and 1 of variables X and Z .

For example, suppose you want to conduct a survey of people's income in a certain place: people with high incomes may be less likely to report their incomes (Allison, 2009), thus income data would likely be missing if the value is high. In addition, people within a certain age range are most likely to have higher incomes, thus the missing probability depends on the value itself and on the observed variable (age).

As mentioned in section 3.2, in physical examination data, the probability of having missing values generally depends on the clinician and the patient. For example, for a certain patient, if a measure is assumed to be close or equal to a typical developing child, it may be not be reported. It has been shown that children with GMFCS of level I (Palisano et al., 1997), which are the most similar to typical developing children, are most likely to have missing values in their neuro-orthopedic assessment (Hedström & Carlberg, 2015). Hedström and Carlberg [IDEM] also showed that adolescents with GMFCS of level III tend to have more missing data in their physical examination. This can be interpreted as if the patient has a general severe state, missing data are more likely.

In addition, spasticity measures are more likely to have missing values than passive ranges of motion (Hedström & Carlberg, 2015). This is confirmed by our database, where considered spasticity measures have an average completion rate of 80.8% (only 5 variables over 80% completion rate) while the average completion rate for the considered passive ranges of motion is 90.72% (9 variables over 80% of completion) [see Table 3.2].

In order to complete the data without heavily truncating the original data table, an imputation approach has been chosen. Imputation consists in replacing the missing values with a reasonable guess of those unavailable values (Donders, van der Heijden, Stijnen & Moons, 2006). Imputation techniques have been used for medical machine learning applications of discrete incomplete data, for instance for breast cancer data (García-Laencina, Abreu, Abreu & Afonso, 2015; Jerez, Molina, García-Laencina, Alba, Ribelles, Martín & Franco, 2010). Moreover, it

has been proven that imputed data give better classification performance than incomplete data (Farhangfar, Kurgan & Dy, 2008).

There are numerous algorithms for imputation of discrete data, passing by mean methods, hot deck methods, multiple imputation, among others [see (Donders et al., 2006)]. One of the easiest approach consists in replacing missing values by the mean value of the variable. However, this deforms the distribution of the variable (Tufféry, 2012) [see figures 4.6a and 4.6b]. Moreover, we have seen that in our problem, extreme values (close to typical development or severe in general) are most likely to be missing. For these reasons, in this work we have chosen a linear regression algorithm, which is called the iterative robust model-based imputation algorithm (IRMI) (Templ, Kowarik & Filzmoser, 2011).

The IRMI algorithm is based on the IVEWARE algorithm (Raghunathan, Lepkowski, Van Hoewyk & Solenberger, 2001), but is more robust and can be used when no variable has 100% completion rate. The technique consists in initializing missing values and then iteratively perform linear regression of one variable with respect to the others [see algorithm 2]. There is no formal proof that the algorithm converges, but several experiments with real and artificial data show that the algorithm always converges after few iterations (Templ et al., 2011). In our experiment, the initialization stage was done by k -Nearest-Neighbors algorithm (Duda et al., 2000, pp. 174-192) with $k = 5$ followed by computing the median of the nearest neighbors. For determining the nearest neighbors, only physical examination data were considered. Variable distributions were considered as continuous and the linear regression was performed with the robust MM-estimation (Yohai, 1987). It can be noticed that the IRMI algorithm allows a much better distribution preservation than the mean imputation [see Figure 4.6].

After imputation, we are able to utilize most of the statistical techniques, which are commonly used with complete data, hence to integrate kinematic and physical examination data. The problem now is that the dimension of the limb vector, already too large, which increases with the inclusion of the physical examination variables. In the next section, we will treat this problem by applying a dimensionality reduction technique.

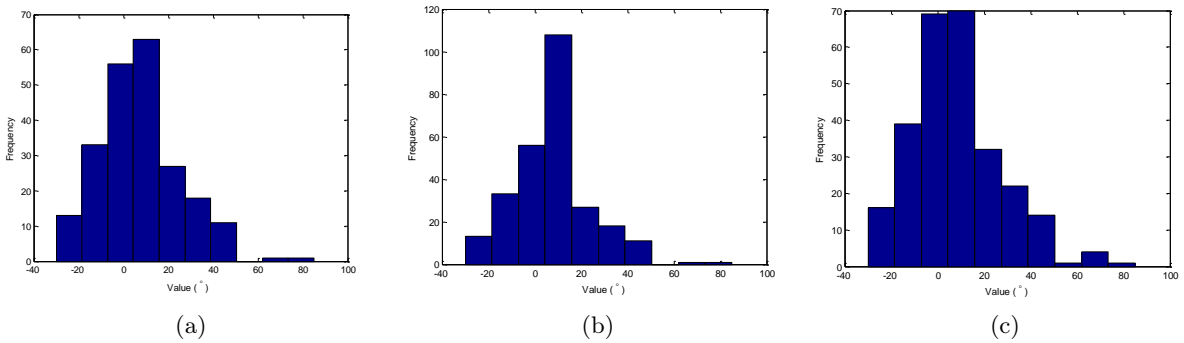


Figure 4.6 – Distribution of popliteal angle measure at V3 before and after imputation. 4.6a Original distribution before imputation. 4.6b Distribution after mean imputation. The shape of the distribution changes with respect to the original distribution. 4.6c Distribution after IRMI imputation. The shape is similar to the original distribution.

4.4 Dimensionality Reduction by Principal Component Analysis

When it comes to dimensionality reduction, one of the most widely used methods is Principal Component Analysis (PCA), also called Karhunen-Loève transform (Bishop, 2006, p. 561).

Algorithm 2 IRMI imputation (Templ et al., 2011).

- 1: Sort variables according to the amount of missing values, so that

$$\mathcal{M}(x_1) \geq \mathcal{M}(x_2) \geq \dots \geq \mathcal{M}(x_p)$$

where $\mathcal{M}(x_j)$ denotes the number of missing values of variable x_j , $\forall j \in I = \{1, \dots, p\}$.

- 2: Initialize missing values using a simple imputation technique (*i.e.* mean, median, k-nearest neighbors, etc.).

3: **repeat**

4: **for** $l = 1$ **to** p **do**

- 5: Denote $m_l \subset \{1, \dots, n\}$ and $o_l = \{1, \dots, n\} \setminus m_l$ respectively the indices of the observations that were originally missing and available in variable x_l .

- 6: Denote $X_{I \setminus \{l\}}^{o_l}$ and $X_{I \setminus \{l\}}^{m_l}$ the matrices with the all the other variables.

- 7: Consider the linear regression problem $x_l^{o_l} = X_{I \setminus \{l\}}^{o_l} \beta + \epsilon$ with unknown regression coefficients β and error ϵ .

- 8: Let d be the distribution of $x_l^{o_l}$

- 9: **if** d is continuous **then**

- 10: Apply robust regression method to estimate β , such as least trimmed squares (LTS) (Rousseeuw & Van Driessen, 2006) or MM-estimation (Yohai, 1987).

- 11: **else if** d is categorical **then**

- 12: Apply generalized linear regression (Maronna, Martin & Yohai, 2006, pp. 239-244) or robust generalized linear regression (Cantoni & Ronchetti, 2001) to estimate β .

- 13: **else if** d is binary **then**

- 14: Apply logistic linear regression (Maronna et al., 2006, pp. 229-233) to estimate β .

- 15: **else if** d is semi-continuous **then**

- 16: Apply logistic regression.

- 17: **if** No constant is imputed **then**

- 18: Apply robust regression on the continuous (non-constant) part of the response to estimate β .

- 19: **end if**

- 20: **else if** d is count **then**

- 21: Apply robust generalized linear regression of family Poisson (Cantoni & Ronchetti, 2001) to estimate β .

- 22: **end if**

- 23: Impute missing values of x_l as

$$\hat{x}_l^{m_l} = X_{I \setminus \{l\}}^{m_l} \hat{\beta}$$

where $\hat{\beta}$ and $\hat{x}_l^{m_l}$ are the estimations of β and $x_l^{m_l}$ respectively. Notice that it is possible to use a subset of variables $L \subset I \setminus \{l\}$. In such case $X_{I \setminus \{l\}}^{m_l}$ should be replaced by $X_L^{m_l}$.

- 24: **end for**

- 25: **until** imputed values stabilize, so that:

$$\sum_i \hat{x}_{l,i}^{m_l} - \tilde{x}_{l,i}^{m_l} < \delta \forall i \in m_l \text{ and } l \in I$$

where δ is small constant, $\hat{x}_{l,i}^{m_l}$ and $\tilde{x}_{l,i}^{m_l}$ are respectively the i -th imputed value of the current and previous iteration.

The objective of PCA is to find a linear transformation of the data into a lower-dimensional space that best preserves the distance between the original points. Such space is the one that maximizes the variance. The first principal direction gives the maximal variance or covariance of the data, the second principal direction is orthogonal to the first one and corresponds to the second biggest variance, and so on for the rest of the principal components. Thereby, the last principal components are susceptible of being dismissed and the discriminant information can be preserved in a lower-dimensional space that eliminates data redundancy. Let x be a vector of p elements that correspond to different variables, first we want to find a linear transformation

$$u_1^T x = \sum_{i=1}^p u_{1,i}^T x_i$$

that maximizes the variance, where T is the transpose operator and $u_1^T u_1 = 1$. The variance is given by:

$$Var[u_1^T x] = u_1^T \Sigma_x u_1 \quad (4.14)$$

where Σ_x is the covariance matrix of the variables in x . Since $\|u_1\| = 1$, to maximize equation (4.14), we need to introduce the Lagrange multiplier λ_1 , so:

$$u_1^T \Sigma_x u_1 + \lambda_1 (u_1^T u_1 - 1) \quad (4.15)$$

By setting to zero the derivative of equation (4.15) with respect to u_1 , we get:

$$\Sigma_x u_1 - \lambda_1 u_1 = 0 \Rightarrow \Sigma_x u_1 = \lambda_1 u_1 \quad (4.16)$$

Thus λ_1 is the first eigenvalue of Σ_x and u_1 is the associated eigenvector. If we left multiply equation 4.16 by u_1^T , we obtain

$$u_1^T \Sigma_x u_1 = Var[u_1^T x] = \lambda_1 \quad (4.17)$$

So, to maximize the variance we need to get λ_1 the eigenvalue of Σ_x with the highest value and choose u_1 its corresponding eigenvector.

Then, we want to find $u_2^T x, \dots, u_d^T x$, where $d < p$ and $u_k^T x$ is orthogonal with respect to $u_1^T x, \dots, u_{k-1}^T x$. By induction, u_2, \dots, u_d correspond to eigenvectors associated to the eigenvalues of Σ_x ordered from second highest to lowest values (Bishop, 2006, p. 563). PCA has been applied in face recognition (Turk & Pentland, 1991) through eigenfaces representation described by Sirovich and Kirby (1987). This work from Sirovich and Kirby (IDEM) has also inspired the Gait Deviation Index (GDI) (Schwartz & Rozumalski, 2008), which is a popular measure in clinical gait analysis [2.6]. GDI is based on Singular Value Decomposition (SVD), which is intimately related to PCA. In fact, SVD provides an computational efficient method for finding principal components (Jolliffe, 2002, pp. 44-45). The relation between SVD and PCA is at follows: Let X be an $(n \times p)$ matrix of n observations of p variables. The SVD of X gives:

$$X = USV^T \quad (4.18)$$

where S is a diagonal matrix of the singular values of X , $U^T U = I_n$ and $V^T V = I_p$, with I_n and I_p are the identity matrix of size $n \times n$ and $p \times p$ respectively. If X is centered (row vector of means $\bar{X} = 0$), the sample covariance matrix is given by:

$$C = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^T (X_i - \bar{X}) = \frac{1}{n-1} X^T X \quad (4.19)$$

Replacing X of equation (4.18) in equation (4.19):

$$C = \frac{1}{n-1} V S U^T U S V^T = V \frac{S^2}{n-1} V^T \quad (4.20)$$

In equation 4.20, the columns of V are the right singular vectors of X and at the same time the eigenvectors of C . The corresponding eigenvalues of C are related to the singular values of X by $\lambda_i = \frac{s_i^2}{n-1}$. This relation between SVD and PCA is only true when X is centered.

One of the advantages of PCA is that it allows to reconstruct estimations of the original variables from the lower-dimensional representation

$$\tilde{x} = \sum_{k=1}^d u_k^T x u_k + \sum_{k=d+1}^p b_k u_k$$

where b_k are constants.

In medical applications PCA has been applied, for example, for dimensionality reduction, to assess the amount of variability and to discriminate types of dementia (Stühler & Merhof, 2012). PCA has been also used for CGA data. For example, Federolf *et al.* (2013) used PCA to differentiate healthy and medial knee-osteoarthritic gait (kinematics and kinetics), and Schweizer *et al.* (2012) proposed a method for selecting the most representative trial (kinematics) based on PCA. Carriero *et al.* (2009) did cluster analysis after PCA to identify gait patterns in cerebral palsy, whereas Schwartz *et al.* (2000) determined a hip flexor function index based on PCA to evaluate hip surgery in cerebral palsy. Specifically for surgery effect prediction, Sebsadji *et al.* (2012) classified good or bad results of hamstring lengthening surgery in cerebral palsy using Support Vector Machines (Burgess, 1998) after PCA.

PCA can be applied on the covariance matrix or the correlation matrix. If the variance has no clear meaning or the considered variables have different scales or too different variances, correlation matrix is preferable (Vigneron, 1997). On the converse case, covariance matrix might be more suitable for PCA and the results are more easily interpretable.

There are several methods for selecting an optimal value of d (number of principal components). The simplest is to choose the smallest d that guarantees a minimal cumulative inertia. For instance, in most cases we can choose the minimum d that contains between 70% and 90% of the total variance, although depending on the data set we might need more or less contained variation (Jolliffe, 2002, p.113). Other techniques consist in looking at each eigenvalue, inertia I_i , $\log(CI_d)$, or are based on hypothesis tests or cross-validation [see (Jolliffe, 2002, ch.6) for more information on these methods].

4.4.1 Method

PCA was applied on the variance-covariance matrix C of the $(n \times p)$ matrix of the preoperative data of all the available limbs:

$$X = \begin{pmatrix} x_{kine,1}^1 & \cdots & x_{kine,1}^{N_{pre}^{kine}} & \tilde{x}_{phex,1}^1 & \cdots & \tilde{x}_{phex,1}^{N_{pre}^{phex}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ x_{kine,n}^1 & \cdots & x_{kine,n}^{N_{pre}^{kine}} & \tilde{x}_{phex,n}^1 & \cdots & \tilde{x}_{phex,n}^{N_{pre}^{phex}} \end{pmatrix}$$

where $x_{kine,j}^k$ is the k -th kinematic point of limb j , $\tilde{x}_{phex,j}^i$ is the imputed i -th physical exam value of limb j , $n = N_{limb} = 268$ and $p = N_{pre}^{tot} = N_{pre}^{kine} + N_{pre}^{phex} = 801$ [see chapter 3 for details]. If $n < p$, which is the case in this work, the covariance matrix maximum rank is n , hence there

are $n < p$ non-zero eigenvalues. It is then possible to compute only $n - 1 = 267$ principal components (Legendre & Legendre, 1998, p. 450).

In order to efficiently reduce dimension, it is crucial to concentrate the most of the original variation in a much smaller subset of variables or principal components. To evaluate the variation contained in each principal component, the percentage of inertia is computed as:

$$I_i = 100 \frac{\lambda_i}{\sum_{j=1}^p \lambda_j} \quad (4.21)$$

Since $n < p$ and there are at least $p - n$ null eigenvalues, we can sum until n instead of p . For a certain PCA dimension d , its cumulative percentage of inertia is:

$$CI_d = \sum_{i=1}^d I_i = 100 \frac{\sum_{i=1}^d \lambda_i}{\sum_{j=1}^p \lambda_j} \quad (4.22)$$

After reconstruction, the average root-mean-square error ($RMSE$) is:

$$\overline{RMSE}_{rec} = \frac{1}{n} \sum_{j=1}^n \sqrt{\frac{1}{p} (\tilde{x}_i - x_i)^T (\tilde{x}_i - x_i)} \quad (4.23)$$

In order to know which variables are mostly penalized by the dimensionality reduction, the variance loss after reconstruction of each preoperative variable has been evaluated for different PCA dimension projections [see Figure 4.11 and page 61]. To measure the variance loss, we have computed the reconstruction-original variance ratio as

$$VR = \frac{\text{Variance after reconstruction}}{\text{Original variance}} \quad (4.24)$$

The PCA dimension was empirically optimized by testing different values of d , that have cumulative variation between 25% to 99%, and then selecting the value of d that gives the best results for regression [see section 6.2 and section 7.2].

4.4.2 Results

The first eight principal components contain almost 80% of the total variance, whereas 90% of the inertia is obtained with the first sixteen PCA dimensions, 95% with the first twenty-seven principal components and 98% with the forty-seven principal components [Figure 4.7]. Most of the variance is contained in a lower-dimensional space than the original variable space and far from null eigenvalues: $d < n - 1 < p$.

Most of the pairs of limbs of a same patient are closely localized in the principal plane (pairs of blue and red cercles in figure 4.8), with a right-left correlation of $r_{r-l} = 0.99$ for both the first and the second principal dimension. This suggests that the first principal components correspond to a global characterization of the patients, such as the size effect. In addition, the projection of ipsilateral and contralateral variables are also close in the first principal plane, which reinforces that intuition [see figure 4.9a]. On the other hand, the third principal dimension separates symmetrically ipsilateral and contralateral variables [see figure 4.9b]. The first principal component is closely related to kinematics of ankle dorsiflexion (both ipsilateral and contralateral) and foot progression (also bilaterally) [figure 4.9]. The second PCA dimension is closely related to kinematic knee flexion (bilaterally) and moderately related to some bilateral physical examination data. The third principal dimension, besides symmetrically dividing ipsilateral and contralateral variables, is related to kinematic hip rotation (bilaterally) and pelvic rotation.

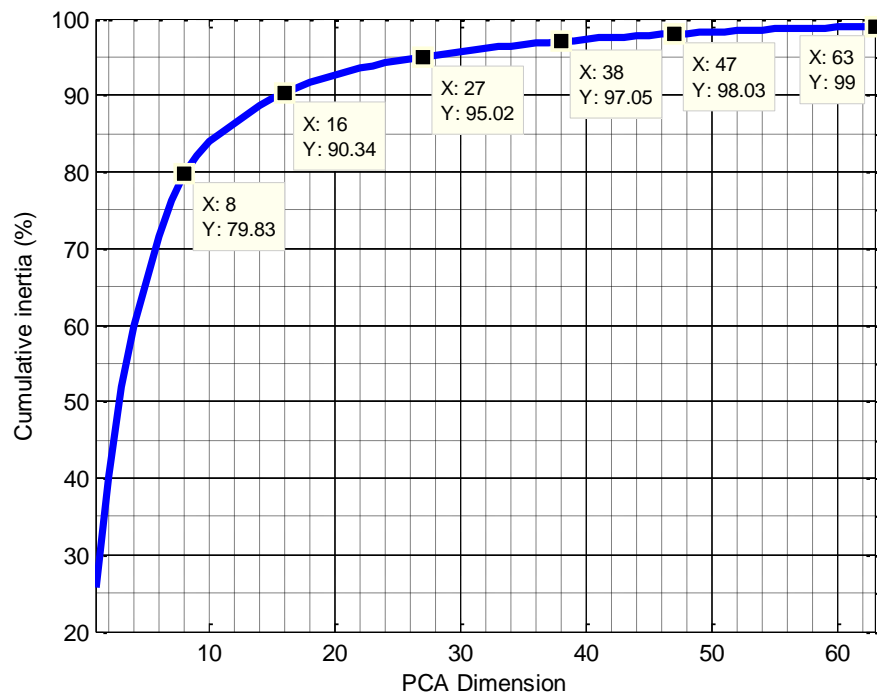


Figure 4.7 – Cumulative inertia of PCA projection

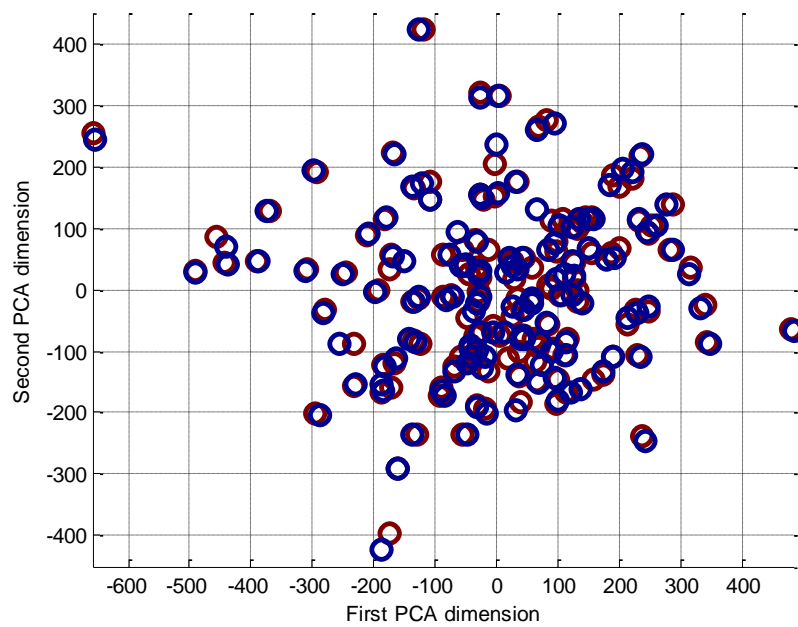
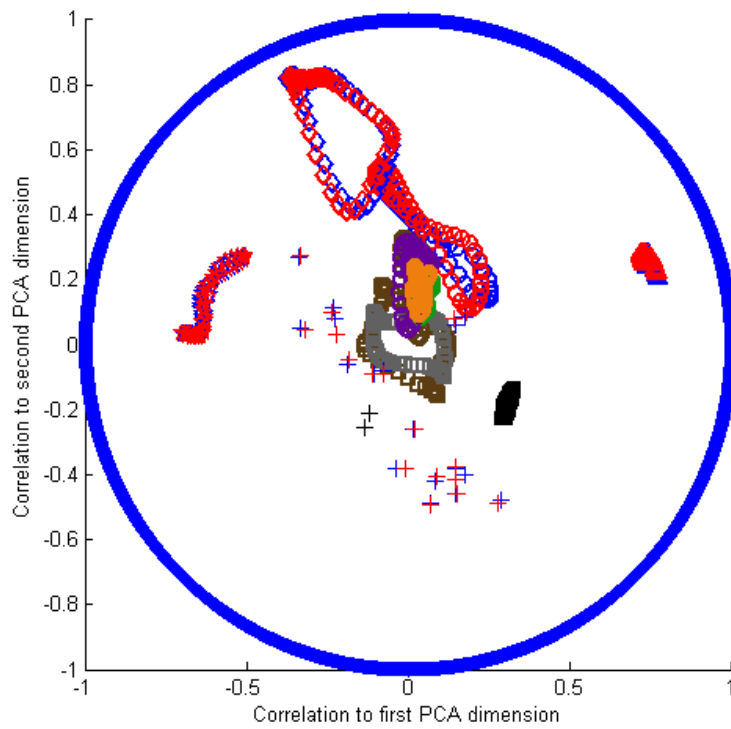
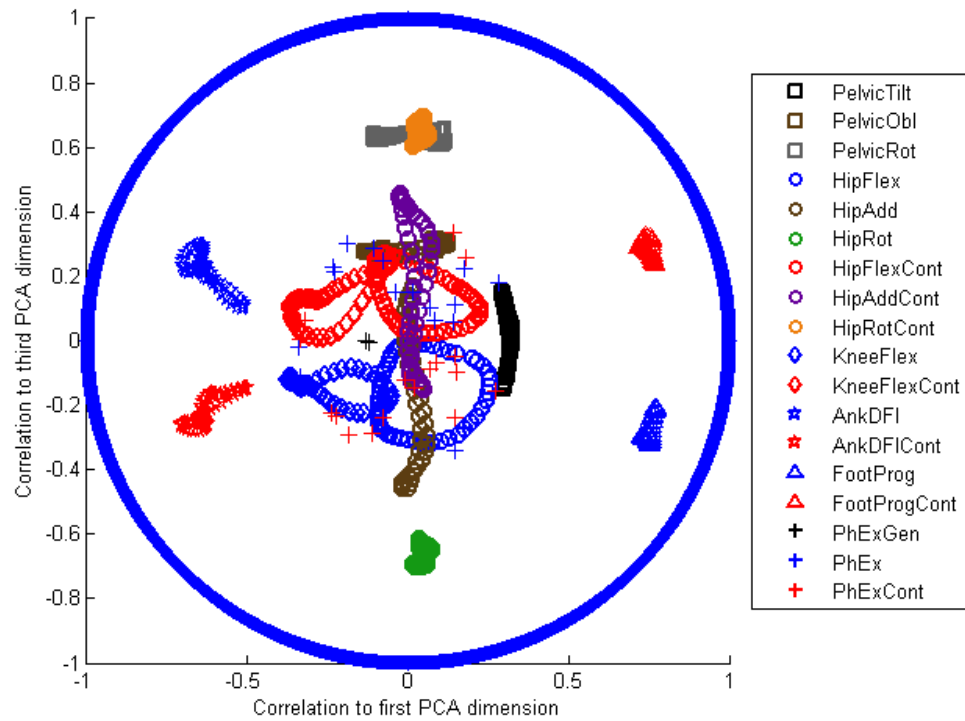


Figure 4.8 – Projection of subjects (limbs) on the principal plane. Blue and red circles correspond respectively to right and left limbs.



(a)



(b)

Figure 4.9 – Variable projection on principal planes. 4.9a) Correlation of original variables and principal plane. 4.9b) Correlation of original variables and first and third PCA dimensions.

From $d = 10$ and higher, the average reconstruction $RMSE$ given by equation (4.23) $\overline{RMSE}_{rec} < 5^\circ$ [see Figure 4.10]. $d = 9$ gives an average $RMSE$ slightly higher than 5° , hence at around 80% of the total variation, the reconstruction error is about 5° .

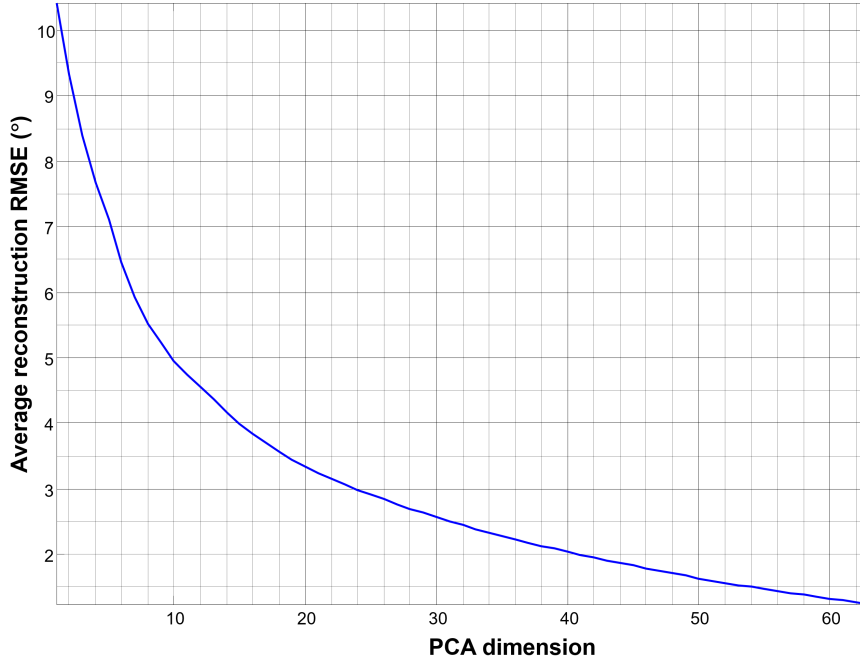


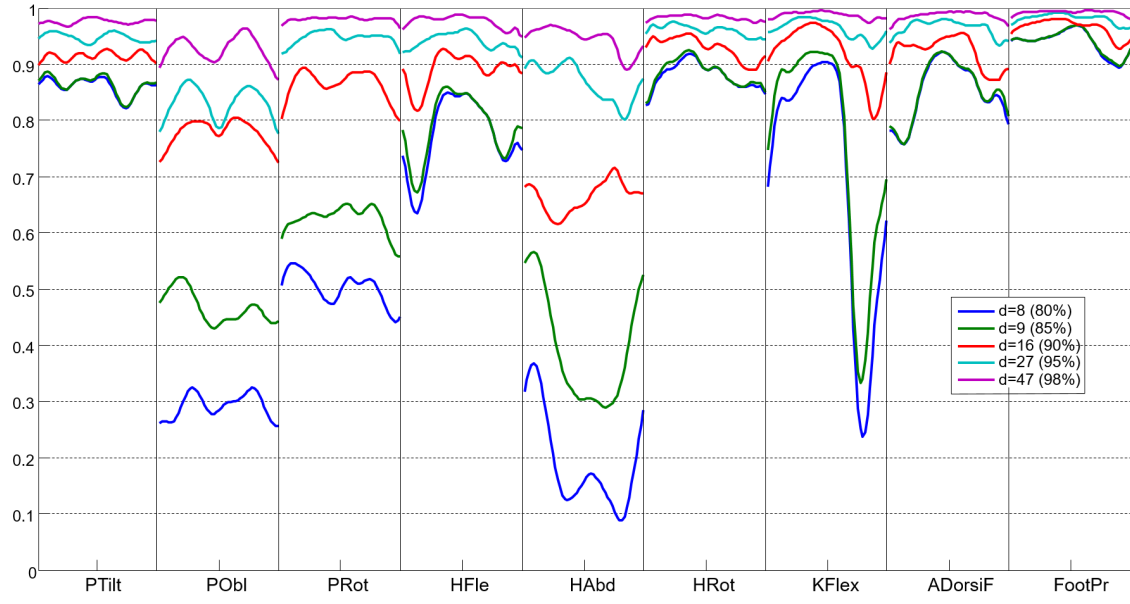
Figure 4.10 – Average RMSE over all patients per PCA dimension.

All the variance losses of ipsilateral kinematic variables are under 40% except for pelvic obliquity, pelvic rotation, hip adduction and knee flexion at swing phase, which have lost half of their variance at $d = 9$ or $CI_d = 85\%$ [see figure 4.11a]. For contralateral kinematics, variance ratio is always over 0.68 except for contralateral hip adduction and contralateral hip flexion at ipsilateral mid-stance (contralateral swing) [see figure 4.11b]. Regarding physical examination data, the variance ratios are more variable depending on the value of the PCA dimension d . At $d = 16$ or less, the variables that at least lost half of their variance are ankle ranges of motion (ankle dorsiflexion with extended or flexed knee at V1 or V3), tibial torsion, quadriceps lag, hip internal and external rotation, femoral anteversion and hip extension [see figure 4.12a]. Similarly, this phenomenon occurred identically for contralateral physical examination data [see figure 4.11b].

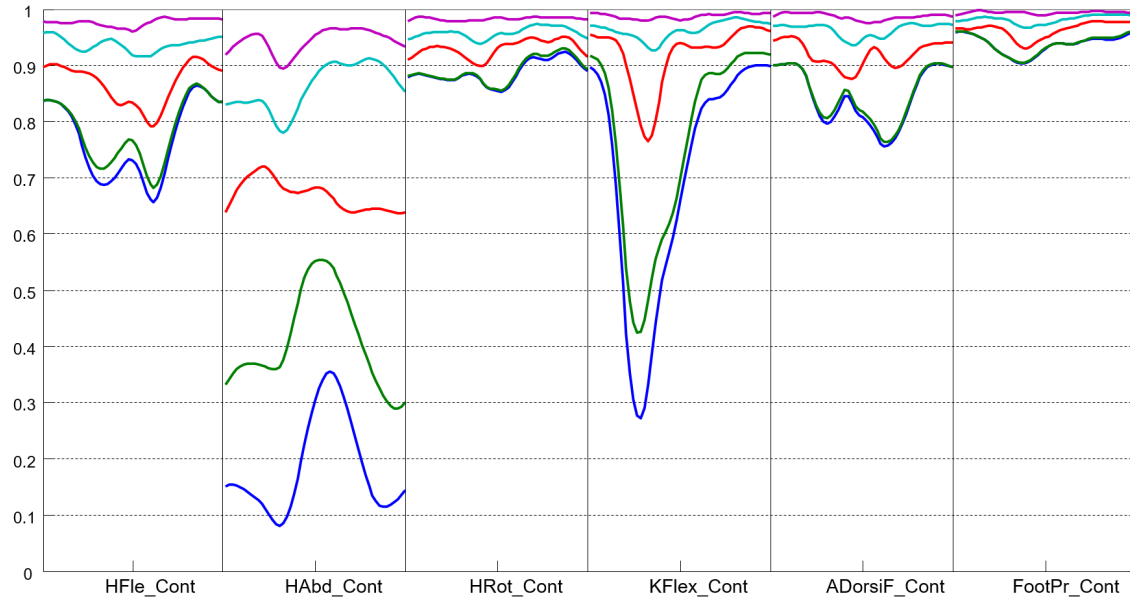
4.5 General Discussion

Feature extraction has been applied by three different methods: curve fitting for kinematic signals, preoperative variable selection (kinematics and physical examination) and principal component analysis (kinematics and imputed physical examination). In order to integrate physical examination data and utilize its maximum potential, an imputation of missing values has been performed.

Kinematic curves have been successfully approximated by B-Splines and harmonic functions. The principal advantage of the tested functions is that they preserve the continuity and the periodicity of the kinematic signals, making them, for instance, suitable for 3D gait animation. Harmonic approximations achieved more compact representations of these curves, needing

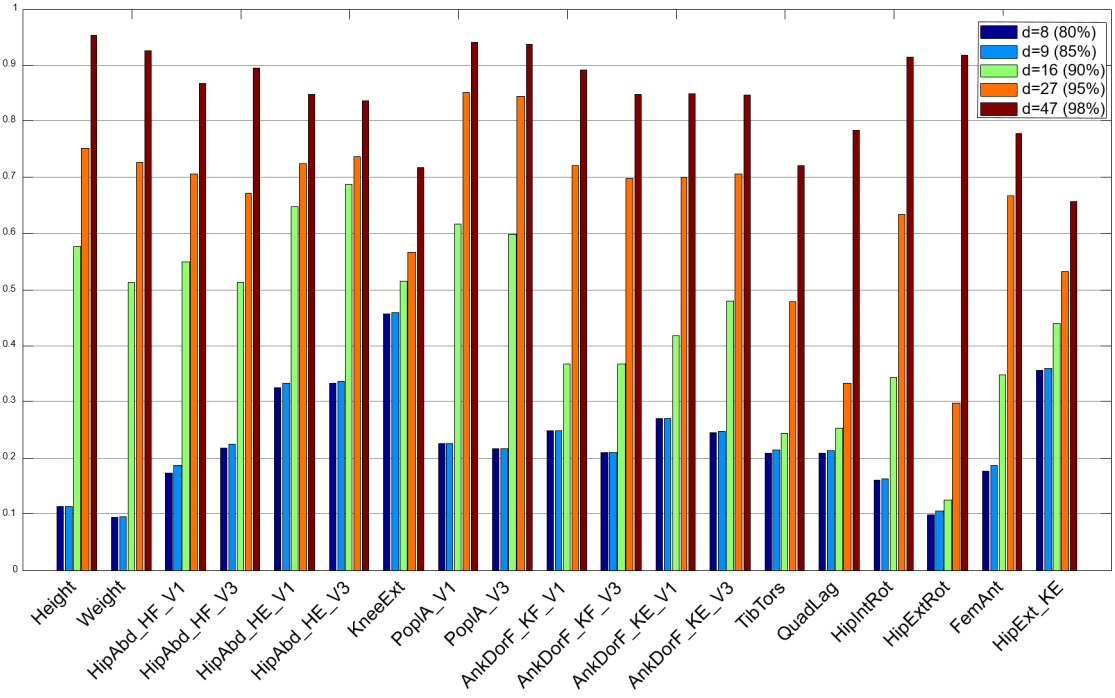


(a)

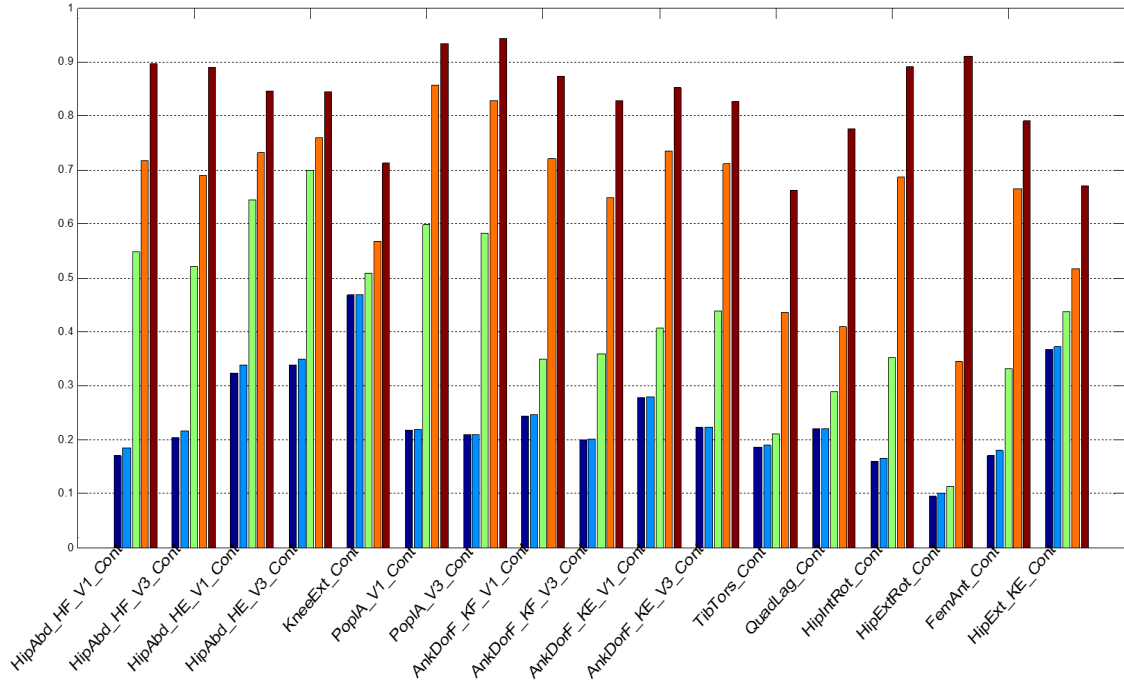


(b)

Figure 4.11 – Variance ratio of preoperative kinematics for different PCA dimensions d (associated cumulative inertia between parenthesis. Variance ratio as in Equation 4.24). 4.11a) Variance ratio for ipsilateral preoperative kinematics. 4.11b) Variance ratio for contralateral preoperative kinematics.



(a)



(b)

Figure 4.12 – Variance ratio of preoperative physical examination data for different PCA dimensions d (associated cumulative inertia between parenthesis. Variance ratio as in Equation 4.24). 4.12a) Variance ratio for ipsilateral preoperative physical examination data. 4.11b) Variance ratio for contralateral preoperative physical examination data (labels in Table 3.2).

always less parameters than the B-Splines approximations. Best curve fitting results approximates the fifteen time-normalized kinematic signals (765 variables) with only 101 parameters using trigonometric approximation and thresholding on average mean-squared error and correlation. This thresholding obliges the approximation to maintain both a similar form and a small distance from the original curve while taking the minimal number of parameters, and it has achieved good results for less parameters than the corrected Akaike's information criterion.

Variable selection gives coherent and correlated local preoperative variables (both kinematics and physical examination data) given a local postoperative kinematic variable, for example the same kinematic variable in the preoperative vector. However, the size of the selected variables' vector may drastically change from one postoperative kinematic variable to another, which complicates the use of regression techniques with fixed input vector (i.e. neural networks). In addition, the selected preoperative variables are in general different for different postoperative kinematic variables, which complicates the interpretation of the predictions that would be estimated from different input variables every time.

Missing data imputation allows to exploit the available physical examination data. Even if the imputation introduces an error, it is better to be able to include this information than not consider it at all. On the other hand, even if the original distributions are more or less preserved after imputation, it is unsure that the imputed values are good guesses of the missing values. In the first place, by nature, we cannot have access to those missing values in order to compare them to the imputed values. On the second place, it is difficult to establish a test procedure of the imputation, given that in this case the missing values are not random. For example, a typical test procedure might be to randomly remove known values, then impute the removed values and finally comparing to the real values. In such case, we would have been testing the performance of the imputation algorithm on values missing at random, while we actually want to measure its performance on not randomly missing values.

Curve fitting obtained a best compression percentage rate of 86.80% for $\overline{RMSE} \leq 5^\circ$, while PCA obtained a best compression rate of 98.75% for the same average error and also considering physical examination data. Schwartz and Rozumalski (2008) had a compression rate of 96.73% for reconstruction errors "on the order of 1° ". Schwartz and Rozumalski (IDEM) claimed a compression rate 98% of the information with 15 features, thus 32 features less than us for the same cumulative variance, but they did not consider contralateral kinematics and physical examination data ($p = 459$ for them).

As Federolf *et al.* (2013), we encountered the small data size problem (Moran, Richter & O'Connor, 2014) with an example-variable ratio $\frac{n}{p} = \frac{268}{801} \approx \frac{1}{3}$. However, PCA could be efficiently applied when $n < p$ depending on the correlations of the variables and if maximum $n - 1$ principal components are considered. For instance, PCA has been successfully applied in genetics for much smaller ratios ($\frac{n}{p} = \frac{3}{500} \approx \frac{1}{167}$). Since an important part of the considered data corresponds to kinematic time series, where consecutive points are heavily correlated, PCA can be effectively applied in our problem. This can be observed after PCA in the projection of the original variables on the three principal components shown in Figure 4.9, where points of a same kinematic angle are very close and create closed segments due to their periodicity. In addition, 99% of the total variance is contained in the first 63 dimensions, which is more than four times lower from $n - 1 = 267$ principal components. Most of the above cited works that apply PCA on gait data deal with small example-variable ratios, with some having $n < p$ and lower ratios than in this work: $\frac{n}{p} = \frac{1}{414}$ (Federolf, Boyer & Andriacchi, 2013; Moran et al., 2014) and $\frac{n}{p} = \frac{1}{51}$ (PCA applied on each limb and angle separately) (Schweizer, Cattin, Brunner, Müller, Huber & Romkes, 2012). For the other cited works, ratios are: $\frac{n}{p} = \frac{31}{25} \approx \frac{1}{1}$ (31 patients and maximum

25 points per vector) (Sebsadji et al., 2012; Sebsadji, 2011), $\frac{n}{p} = \frac{40}{27} \approx \frac{4}{3}$ (Carriero et al., 2009), $\frac{n}{p} = \frac{29}{5} \approx \frac{6}{1}$ (Schwartz, Novacheck & Trost, 2000) and $\frac{n}{p} = \frac{6702}{459} \approx \frac{15}{1}$ for the SVD in (Schwartz & Rozumalski, 2008).

The most penalized kinematic variables after PCA dimensionality reduction are frontal plane kinematics (pelvis obliquity and bilateral hip adduction), which rarely drastically vary in CP gait after surgery; as well as bilateral knee flexion at swing. In addition, physical examination data loses more variance than kinematic data, especially ankle ranges of motion, hip extension, internal and external rotations, tibial torsion and femoral anteversion.

The advantage of PCA is that it reduces the preoperative data dimension of both kinematics and physical examination at the same time, with a better compression rate than the other tested methods and resulting of less input parameters for regression. PCA gives an average 5-degree error order for 10 parameters while curve fitting needs 101 parameters. However the physical examination data, that was treated by imputation in order to optimally integrate it, is more penalized than the preoperative kinematic variables. A better preservation of physical examination data might be achieved by performing dimensionality reduction separately for kinematics and physical examination data. In such case, techniques such as probabilistic PCA (PPCA) (Tipping & Bishop, 1999), which is too complex for taking all the variables (kinematics and physical examination), can be used. Moreover, PPCA allows to compute PCA with missing data, so previous imputation would be unnecessary. However, separating kinematics and physical examination would introduce another optimization problem for sizes of dimensionality-reduced kinematic vector with respect to dimensionality-reduced physical examination vector. Now that we have reduced the dimension of the preprocessed preoperative data, we are able to proceed to the regression analysis for predicting most likely postoperative kinematics given the preoperative. As mentioned before, given the available number of examples, this regression stage would be impossible without the dimensionality reduction stage. Variable selection is considered in section 5.2, while PCA dimensionality reduction is considered in experiments in section 6.2 and section 7.2.

Part II

Predicting Surgery Effect on Cerebral Palsy Gait using Supervised Machine Learning

Supervised learning can be used for classification or, especially for regression. Regression is the process of learning mathematical relation between inputs X and continuous outputs Y from example data (Stulp & Sigaud, 2015):

$$Y = g(X) + \epsilon \quad (4.25)$$

Where g is the parametric regression function and ϵ is the vector of approximation errors. In classification, the output is discrete instead of continuous. The model parameters of g are adjusted such that the probability of Y given X ($P(Y/X)$) is maximized, which is to minimize the error ϵ . The optimization process consists in minimizing a cost function J , which is usually, but not necessarily, the mean squared error (MSE) over the training set:

$$J = \frac{1}{2n} \sum_{i=1}^n (y_i - g(x_i))^2 \quad (4.26)$$

Where n is number of training examples and (x_i, y_i) is the input-target output pair. The $\frac{1}{2}$ is a constant that is only there for convenience. It will disappear when deriving J [see section 5.1]. The learning process can be *sequential* or by *batch*. Sequential learning, also called incremental or on-line learning, consists in presenting each example at a time and updating model parameters after each such presentation (Bishop, 2006, ch.3). In batch learning, all examples are presented at the same time in order to adapt the parameters. Some models can be trained using both sequential and batch learning methods.

Once the model parameters are trained, it is possible to estimate values of the output \hat{Y}_t for novel input values X_t , and to predict future values of the output:

$$\hat{Y}_t = g(X_t) \quad (4.27)$$

To be able to test for new inputs, it is only necessary to know the model g and its optimal parameters, but the crucial task is to find these optimal parameters. There are some models where some parameters need to be defined by the user and cannot be optimized by the learning algorithm. These user-defined parameters are called *hyperparameters*. For example, in feedforward neural network, the number of hidden layers and hidden units must be defined by the user in order to use a learning algorithm to optimize the weights. In this case, number of hidden layers and units are hyperparameters and weights are parameters (optimized by the learning algorithm). Feedforward neural networks will be discussed in section 5.1.

Families of regression or classification methods are derived from different fields of mathematics and computer science, such as statistics, data mining or artificial intelligence (Fernández-Delgado, Cernadas, Barro & Amorim, 2014). Most widely used regression methods can be based on connectionist approaches, *i.e* artificial neural networks, or ensembles, *i.e* regression trees or mixture of experts.

It exists a large number of regression algorithms, each of them assumes a certain type of model (Stulp & Sigaud, 2015). For example, linear regression corresponds to linear combinations of linear functions and parameters are usually optimized by linear least mean squares (LMS). However, linear LMS can be also used in nonlinear regression for linear combinations of a fixed set of nonlinear functions (Bishop, 2006, ch.3). Linear regression will be discussed in section 6.1. Other regression algorithms are regularized least squares (Bishop, 2006, ch.3), receptive field weighted regression (Atkeson & Schaal, 1995), radial basis function networks (Park & Sandberg, 1993), regression trees (Breiman, Friedman, Stone & Olshen, 1984, ch.8), extreme learning machine (Huang, Zhu & Siew, 2006), among others.

Stulp and Sigaud (2015) identified two principal categories of regression models used in numerous regression algorithms: mixture of linear models and weighted sum of basis functions. They even

proposed that the latter can be seen as a specific case of the former (IDEM), making regression algorithms specific cases of a unique regression model.

In this part we present several regression experiments for predicting postoperative gait in cerebral palsy. The utilized regression methods include linear regression, nonlinear regression with feedforward neural networks and ensemble learning with linear and nonlinear regression. In addition, several feature extraction methods discussed in chapter 4 are considered.

Chapter 5

Neural Networks for Initial Contact Prediction

Contents

5.1	Nonlinear regression with Neural Networks	69
5.2	Predicting postoperative knee flexion and pelvic tilt at initial contact of children with cerebral palsy	72

5.1 Nonlinear regression with Neural Networks

Artificial neural networks (NN) are a mathematical tool inspired by the human brain function in order to allow machines to learn like humans do (McCulloch & Pitts, 1943). NN are based on connectionist approaches and represent one of the regression methods derived from artificial intelligence. Rosenblatt (1958) proposed an algorithm for pattern recognition called the *perceptron*, but at the time it was too computationally costly to optimize such model. With the advances in computational power and the creation and formalization of the backpropagation algorithm (Werbos, 1974; Rumelhart, Hinton & Williams, 1986), it became possible to quickly train multilayer NN. Since then, scientific interest in NN has considerably grown, making them one of the most widely used machine learning technique in many different applications.

In this section we will mainly focus on the multilayer perceptron or feedforward neural networks (FNN), and how they will be employed to find a nonlinear regression of postoperative gait with respect to the preoperative gait and surgery.

A multilayer perceptron is a feedforward network composed of an input layer, an output layer and one or more hidden layer, as shown in Figure 5.1. There are exclusively connections between neurons belonging to adjacent layers. The input layer units or neurons correspond to the input variables, and the rest of the neurons denote nonlinear functions of the previous layer unit states. The output units have frequently linear activation functions (Dreyfus et al., 2008).

For a neuron i , its state value is given by

$$s_i = f \left(\sum_{j=1}^m w_{ij} s_j \right) \quad (5.1)$$

where f is the activation function, j and m are units and the number of units in the previous layer respectively, a w_{ij} is the weight of the connection of neurons j and i . Activation functions

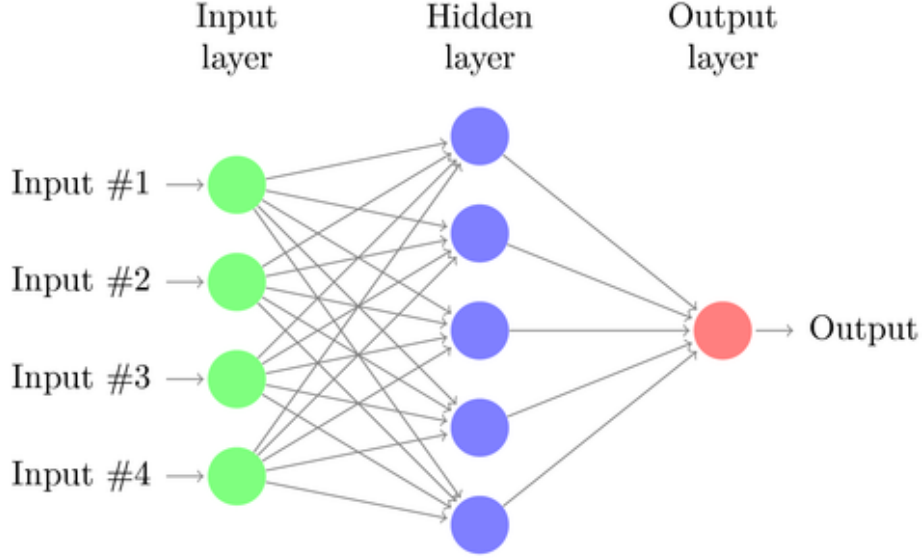


Figure 5.1 – Multilayer Perceptron. Green circles correspond to input variables, blue circles correspond to hidden neurons and the red circle represents the output. Black arrows represent the connections between units. Each connection is associated to a weight w .

are usually bounded and linear in a specific subspace. The most typical activation functions are sigmoid, such hyperbolic tangent. Then Equation 5.1 becomes:

$$s_i = \tanh \left(\sum_{j=1}^m w_{ij} s_j \right) = \frac{e^{\sum_{j=1}^m w_{ij} s_j} - e^{-\sum_{j=1}^m w_{ij} s_j}}{e^{\sum_{j=1}^m w_{ij} s_j} + e^{-\sum_{j=1}^m w_{ij} s_j}} \quad (5.2)$$

The training of a neural network consists in adjusting the set of weights $W = \{w_{ij}\}$ to minimize the approximation error, defined for example by Equation 4.26. Since there is no analytical solution, the weights are initialized (i.e. randomly) and then iterative optimization methods are used, such as gradient descent techniques, to update the weights:

$$W(\tau + 1) = W(\tau) - \mu \nabla J(W(\tau)) \quad (5.3)$$

Where τ is the current iteration, ∇J is the gradient of the error function and μ modulates the descent and is called the *learning rate*.

The gradient of the error function is usually evaluated by applying the backpropagation algorithm (Duda et al., 2000, sec.6.3), but there are other methods for computing the gradient (Dreyfus et al., 2008). Backpropagation consists in iteratively propagate the error from the output through the hidden layers in order to find $\frac{\partial J}{\partial w}$ for every weight w in the feedforward network. If w_{ki} is a weight that goes to the linear output layer, then

$$\frac{\partial J}{\partial w_{ki}} = \frac{\partial J}{\partial \sum_{z=1}^m w_{kz} s_z} \frac{\partial \sum_{z=1}^m w_{kz} s_z}{\partial w_{ki}} = (\hat{y}_k - y_k) s_k \quad (5.4)$$

Where m is the number of hidden units in the last (or unique) hidden layer.

For a weight w_{ij} that goes from the input towards a hidden unit with activation function f , we have:

$$\frac{\partial J}{\partial w_{ij}} = f' \left(\sum_{z=1}^{N_{in}} w_{iz} x_z \right) \left[\sum_{k=1}^{N_{out}} w_{ki} (\hat{y}_k - y_k) \right] x_j \quad (5.5)$$

The learning algorithm we used in the experiments of this work is the Levenberg-Marquardt algorithm (LM) (Levenberg, 1944; Marquardt, 1963). In LM, weights are updated with the equation

$$W(\tau + 1) = W(\tau) - (H - \lambda \text{diag}(H))^{-1} \nabla J(W(\tau)) \quad (5.6)$$

where H is the Hessian matrix at τ , whose elements are $H_{\alpha\beta} = \frac{\partial^2 J}{\partial w_\alpha \partial w_\beta}$, that can also be computed by backpropagation.

In particular, in this work NN training have been performed by LM algorithm with Bayesian regularization (MacKay, 1991, 1992). Bayesian regularization consists in minimizing a linear combination of errors and weights, as well as modifying this linear combination so that the trained network has optimal generalization capacity.

The learning process is stopped when the learning error converges to a stable value. To avoid over-fitting, a subset of the training set, called the *validation set* is tested at each iteration. The validation set is not considered for weight adjustment. Iterations can then be stopped, for example, when validation error arrives to a minimum and starts to increase. This method is called *early stopping* (Bishop, 2006, pp.259-261). Additionally there can be a *test set*, composed of other examples outside the training and validation sets that are used to evaluate the generalization capacity of the network.

Multilayer perceptrons are considered parsimonious universal approximators due to their capability to approximate numerous functions (Bishop, 2006, ch.5) and it has been demonstrated that one hidden layer suffices to generate arbitrary functions (Duda et al., 2000, ch.6). In practice, the most useful network architecture is with one hidden layer with sigmoid activation function and a unique linear output (Dreyfus et al., 2008). However, there is no method to optimally *a priori* define the number of hidden neurons. The optimal number of hidden units is usually found by training networks with different number of hidden units, and the evaluating the validation or test error. On the other hand, the use of (bounded) sigmoid functions imposes the normalization (centered and divided by the standard deviation) of the data beforehand.

Feedforward neural networks have been used for predicting some postoperative gait variables in CP (Niiler, Richards, Miller, Sun & Castagno, 1999; Hersh, Sun, Richards & Miller, 1997). To test the gait outcome prediction potential of feedforward neural networks, we have conducted an experiment where the gait signals are reduced to the initial contact instant. In particular, we analyzed the prediction of two gait parameters at initial contact using neural networks: knee flexion and pelvic tilt. These two parameters are particularly related in hamstrings muscle function (DeLuca et al., 1998). This experiment, which also uses the variable selection technique described in 4.2, was the object of a journal article that is presented in the next section. FNN will again be considered in section 7.2, in combination with linear regression and ensemble learning.

Predicting postoperative knee flexion and pelvic tilt at initial contact of cerebral palsy children

Omar A. Galarraga^{1,2}, Vincent Vigneron², Bernadette Dorizzi³, Néjib Khouri^{1,4} and Éric Desailly¹

¹ UNAM, Pôle Recherche & Innovation, Fondation Ellen Poidatz, 1 rue Ellen Poidatz, Saint-Fargeau-Ponthierry, France

² IBISC EA 4526, Université d'Évry Val d'Essonne, 40 rue du Pelvoux, Courcouronnes, France

³ Institut Mines-Télécom, Télécom SudParis, UMR 5157 SAMOVAR, 9 rue Charles Fourier, Évry, France

⁴ Hôpital universitaire Necker-Enfants malades, 149 rue de Sèvres, Paris, France

Received 1 March 2015 – Accepted 21 October 2015

Abstract. Increased knee flexion at initial contact (KFIC) is a common gait deviation in Cerebral Palsy. Hamstring lengthening surgery (HL) aims to decrease KFIC, but may affect pelvic tilt (PTIC). In this work, we design a decision-making tool that simulates the effect of orthopedic surgery, with or without HL, on KFIC and PTIC. The postoperative parameters are estimated given preoperative gait, physical examination and surgery. Nonlinear regressions are performed by feedforward neural networks. On test, the mean root-mean squared error for KFIC and PTIC are 9° and 5° respectively. Sixty-three percent and 90% of the lower limbs are estimated within a 10° error range for KFIC and PTIC respectively. The simulator is able to give good estimations independently of the surgery type.

Key words: Prediction, clinical gait analysis, cerebral palsy, nonlinear regression, artificial neural networks

Résumé. Prédiction de la flexion postopératoire du genou et de l'antéversion du bassin au contact initial d'enfants atteints de paralysie cérébrale.

La paralysie cérébrale entraîne des troubles de la marche pouvant être caractérisés par une flexion du genou au contact initial (FGCI) excessive. La chirurgie d'allongement d'ischio-jambiers (AIJ) diminue la FGCI mais peut affecter l'antéversion du bassin (ABCI). L'objectif de ce travail est de simuler l'effet de l'AIJ, dans un contexte de chirurgie multi-sites, sur la FGCI et sur l'ABCI. Les paramètres postopératoires sont estimés en fonction de la cinématique et de l'examen clinique pré-opératoires ainsi que la chirurgie. Des régressions non-linéaires sont faites par réseaux de neurones. L'erreur moyenne en test est 9° et 5° pour la FGCI et l'ABCI respectivement. L'erreur de prédiction est indépendante du type de chirurgie et en général reste dans la variabilité des paramètres cinématiques.

Mots clés : Prédiction, analyse quantifiée de la marche, paralysie cérébrale, régression non-linéaire, réseaux de neurones artificiels

1 Introduction

Cerebral palsy (CP) refers to a group of neurological disorders, caused by brain damaged during development, that affect human movement, balance and posture. CP frequently entails muscle and bone deformities and usually manifests by gait troubles. A typical gait trouble in CP is crouch gait (Gage, Schwartz, Koop, & Novacheck, 2009), which is characterized by an excessive knee flexion at stance phase and terminal swing (Arnold, Liu, Schwartz, Öunpuu, & Delp, 2006).

In order to lessen gait deviations, orthopedic surgery is usually performed on cerebral palsy patients. Multiple bone and soft tissue are modified during a single-event multilevel surgery (SEMLS), which combines several surgical gestures according to functional objectives, techniques and affected body parts, etc. in the same operation. In particular for crouch gait, a common treatment is hamstring lengthening surgery (HL). Its purpose is to decrease knee flexion at late swing and stance phase by increasing hamstrings length surgically (Baumann, Ruetsch, & Schürmann, 1980), especially at

initial contact where hamstrings are generally at their maximum length (Laracca, Stewart, Postans, & Roberts, 2014). This surgery has reportedly given good results (Ma *et al.*, 2006), however, since hamstring muscles go from the pelvis to the knee, it can also affect pelvic tilt during gait (DeLuca, Öunpuu, Davis, & Walsh, 1998). For these reasons, the indication of hamstring lengthening is not always straightforward. Moreover, at this moment there is not reported study or tool, other than sometimes the knowledge and the experience of the surgeon, for predicting the resulting knee flexion and pelvic tilt after a hamstring lengthening surgery.

Physical examination (PE) and clinical gait analysis (CGA) (Gage *et al.*, 2009) are performed on patients to improve diagnosis and assert suitable treatments. Physical examination allows to measure principally passive joint ranges of motion, muscle force and spasticity. Clinical Gait Analysis allows to compute joint angles and angular velocities during gait (motion capture), electrical activity of muscles (by electromyography) and joint forces and moments during walking (force platforms). There are some decision-making tools for hamstring lengthening, based on musculoskeletal simulations (Arnold *et al.*, 2006; Desailly, 2008) or statistical methods (Hicks, Delp, & Schwartz, 2011; Sebsadji *et al.*, 2012). However, these tools are only able to predict qualitative (“good” or “bad”) outcomes. This could be too relative, especially for the patient, and gives no information about variation of gait parameters.

In this paper, the objective is to design a simulator of the effect of SEMLS, on both knee flexion and pelvic tilt at initial contact, considering the inclusion or exclusion of HL. This simulator could be used as a decision-making tool for whether including or not a hamstring lengthening procedure in a treatment plan for cerebral palsy patients.

To achieve the objective, supervised learning techniques are applied on data of cerebral palsy children that underwent orthopedic surgery. Postoperative knee flexion and pelvic tilt at initial contact are estimated knowing preoperative gait and physical examination and given a type of surgery. Initial contact is the instant when the foot touches the ground during walking. Nonlinear regression is performed between the postoperative and preoperative parameters using multi-layer feedforward neural networks (Duda, Hart, & Stork, 2001).

2 Materials and method

The database is composed of clinical gait analyses and physical examinations of $N_{pat} = 99$ cerebral palsy children that have undergone orthopedic surgery. To simplify the problem, both sides lower limbs were separated, which gives $N = 191$ lower limbs (7 limbs are not valid due to missing data whether on the Clinical Gait Analysis or the physical examination). Sixty percent of limbs had hamstring lengthening as part of their single-event multilevel surgery. The possible concurrent surgeries are

Table 1. Distribution of patients’ ages at preoperative and postoperative CGA / PE, and surgery.

	\bar{X} (years)	σ (years)
Preoperative CGA / PE	11.80	3.30
Surgery	12.60	3.24
Postoperative CGA / PE	14.76	3.32

Table 2. Considered preoperative parameters.

Label	Parameter
PTIC	Pelvis tilt at initial contact
POIC	Pelvis obliquity at initial contact
PRIC	Pelvis rotation at initial contact
HFIC	Hip flexion at initial contact
HAIC	Hip abduction at initial contact
HRIC	Hip rotation at initial contact
KFIC	Knee flexion at initial contact
AFIC	Ankle dorsiflexion at initial contact
FFIC	Foot-ground flexion at initial contact
HF _{cont} IC	Contralateral hip flexion at initial contact
HA _{cont} IC	Contralateral hip abduction at initial contact
HR _{cont} IC	Contralateral hip rotation at initial contact
KF _{cont} IC	Contralateral knee flexion at initial contact
AF _{cont} IC	Contralateral ankle flexion at initial contact
FF _{cont} IC	Contralateral foot flexion at initial contact
PoplA	Popliteal angle

rectus femoris transfer, hip flexor release, femoral derotational osteotomy, patella lowering, triceps lengthening and foot bony surgery. Forty-five percent of patients are at Gross Motor Function Classification System (GMFCS) level I, 39% at level II and 16% at level III. Male subjects represent 61% of the population and 39% are female. The mean ages and standard deviation of the patients at the three different stages considered (preoperative, surgery and postoperative) can be seen in table 1.

From each Clinical Gait Analysis, fifteen gait angles were considered to describe each limb, considering both ipsilateral and contralateral joints configuration at ipsilateral initial contact. With the ipsilateral popliteal angle (Thompson, Baker, Cosgrove, Saunders, & Taylor, 2001) taken from the physical examination, the preoperative vectors were composed of 16 parameters (table 2). We collected $n = 714$ preoperative gait cycles ($n \ll 191$ because several gait cycles were recorded per Clinical Gait Analysis). The input table is therefore composed of 714 rows, where each row is a vector of 16 parameters.

The target output consists of the mean postoperative knee flexion and pelvic tilt at initial contact of the limbs. The target output matrix was then transformed as in figure 1 to become a matrix of two columns (mean postoperative knee flexion and pelvic tilt at initial contact) and $n = 714$ rows.

Before computing regression functions between the target outputs (postoperative parameters) and the input (preoperative parameters), an input variable selection is performed. The utilized variable selection method

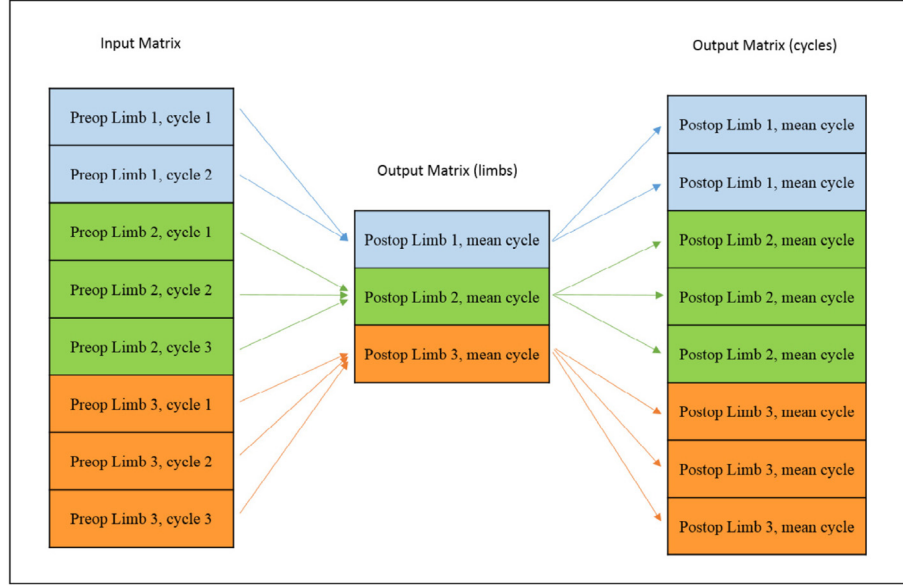


Fig. 1. Composition of input and target output matrices. There are more gait cycles than lower limbs, therefore there are more rows in the input matrix than in the target output target at first. Each input row is associated to the row in the target output matrix that corresponds to the same lower limb.

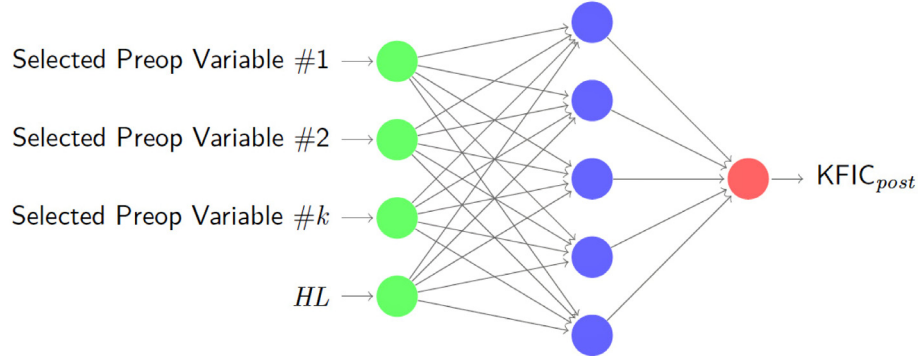


Fig. 2. Architecture of one-hidden-layer perceptron. From left to right: input, hidden and output layer. In this case, the output is the postoperative KFIC. The architecture for the postoperative PTIC is exactly the same. The lines correspond to the weighted connections between neurons.

is the probe technique described by Guyon and Elisseeff (2003). This technique consists on adding a random variable (probe) to the input candidate variables that is completely unrelated to the output target. Then all candidate variables with the probe are ranked. The candidate variables that are better ranked than the probe are selected, and the other candidates are rejected. The ranking technique is a Gram-Schmidt orthogonalization, which consists on iteratively compute the correlation between the candidate variables and the target output and then projecting all variables (input and output) into the orthogonal space of the candidate with the highest correlation to the target output.

Once we had selected the input variables for both output parameters, nonlinear regression was performed using multi-layer perceptron with one hidden layer (Duda *et al.*,

2001). Therefore, the artificial neural network consisted of three layers: an input layer, a hidden layer and an output layer (Fig. 2). The input was composed of the selected variables and a binary parameter that corresponds to the hamstring lengthening surgery, as shown in figure 2. $HL = 1$ if the single event multilevel surgery included hamstring lengthening and $HL = 0$ if not. The output layer corresponds to the postoperative parameter (KFIC or PTIC).

In a multilayer perceptron, each neuron or unit is connected to all the units in the previous and the following layer, but has no connections with the neurons in the same layer. The weights (connections) are optimized in order to minimize the error between the target output and the nonlinear function of the network as in equation (1).

$$J = (g(x, w) - y)^2 \quad (1)$$

Table 3. Test error for KFIC and PTIC separately and together, and according to the type of surgery. HL = 1 if surgery includes hamstring lengthening and HL = 0 if not. Comparison to the prediction error of the postoperative mean, median and preoperative data (no change). “No change” gives the same preoperative KFIC and PTIC as postoperative prediction for each patient. “Mean” gives always the mean postoperative parameters for all the patients (15.24° for PTIC and 26.61° for KFIC) as prediction. “Median” always predicts the median of the postoperative parameters for all patients (14.91° for PTIC and 26.12° for KFIC).

	Method	Limbs	Knee flexion (KFIC)	Pelvic tilt (PTIC)
RMSE (°) $\bar{X} + \sigma$	No change	All	15 ± 11	8 ± 5
		$HL = 1$	20 ± 11	8 ± 4
		$HL = 0$	10 ± 8	7 ± 5
	Mean	All	12 ± 7	8 ± 5
		$HL = 1$	11 ± 6	8 ± 5
		$HL = 0$	12 ± 7	7 ± 4
	Median	All	12 ± 7	8 ± 5
		$HL = 1$	11 ± 6	8 ± 5
		$HL = 0$	12 ± 7	7 ± 4
	NN	All	9 ± 6	5 ± 4
		$HL = 0$	9 ± 5	5 ± 5
$HL = 0$		9 ± 5	5 ± 5	
RMSE ≤ 10°	NN	63%	58%	90%

where g is the nonlinear function of the network, x is the input, w corresponds to the weights of the connections and y is the target output.

The weights were optimized iteratively using Levenberg-Marquardt and backpropagation algorithms (Bishop, 2006). At each iteration or epoch, the training error decreases. To avoid overfitting, some input samples were used for validation. The validation set was not used for updating the weights, but to test the network at each epoch. When the validation error stopped decreasing, the training was stopped. This stop criteria is called early stopping (Bishop, 2006). In our experiments, we utilized 20% of training samples for validation, choosing them randomly. The number of hidden units was varied from $m = 1$ to 15. The optimal number of hidden units was then selected as the one that gave the lowest mean test error over all the patients.

In order to have a prediction error measure for all patients, the proposed method was tested for each patient separately. When testing a patient p , all the associated gait cycles to both limbs of patient p were removed from the training set. Then a perceptron was trained with the new training set (divided into training and validation sets) and the resulted network was afterwards tested with the gait cycles of patient p . The prediction error was the difference between the real postoperative parameters and the estimation given by the neural network. This prediction error was compared to three prediction pseudo-methods: “Mean”, “median” and “no change” predictions. “Mean” is a pseudo-predictor that gives always the same prediction for all patients, and this predictions corresponds to the mean postoperative PTIC and KFIC in the database. “Median” works similarly to “mean”, but its predictions correspond to median postoperative KFIC and PTIC. On the other, “no change” is a pseudo-

predictor that indicates the preoperative KFIC and PTIC as postoperative prediction, which means that the surgery would have no effect on these parameters. A statistical test, specifically a Student’s t -test was performed in order to measure significant difference between the neural network estimator and the pseudo-predictors.

3 Results

The selected preoperative variables for postoperative KFIC ordered by importance are: KFIC, FF_{cont}IC, PoplA, HR_{cont}IC, AFIC, HA_{cont}IC and PTIC (Tab. 2). The selected inputs for postoperative PTIC are: PTIC, PRIC, HFIC, HAIC, HRIC, FFIC, HA_{cont}IC, HR_{cont}IC, KF_{cont}IC, FF_{cont}IC and PoplA.

Best results were obtained for $m = 10$ hidden units for both postoperative knee flexion and pelvic tilt at initial contact. Table 3 shows the mean and standard deviation test error over all limbs for both postoperative knee flexion and pelvic tilt at initial contact, according to the type of surgery (with or without hamstring lengthening) and the percentages of limbs under 10° of error for both angles considered separately and together. Comparison to prediction of the postoperative mean and median, as well as preoperative parameters (no change) are also provided in table 3. We consider the root-mean-squared error (RMSE) between the estimation and the real postoperative value.

The distribution of the tests errors of both knee flexion and pelvic tilt at initial contact is shown in figure 3. In both cases, the RMSE vary from 0° to 26° for pelvic tilt at initial contact, and from 2° to 30° for knee flexion at initial contact.

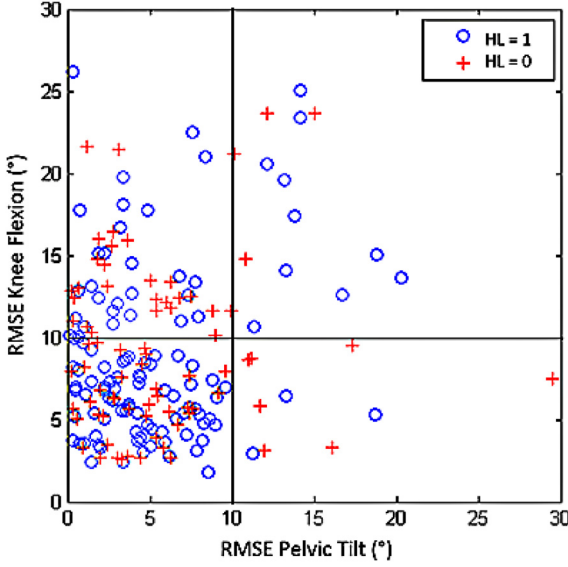


Fig. 3. Distribution of the test RMSE. RMSE on KFIC with respect to RMSE on PTIC. Circles correspond to HL = 1 (surgery with hamstring lengthening) and crosses correspond to HL = 0 (surgery without hamstring lengthening). The circles correspond to those lower limbs that had hamstrings lengthening and the crosses correspond to those who had another kind of surgery. Points left to the vertical line are lower limbs estimated with less than 10° error on pelvic tilt at initial contact and points under the horizontal line are lower limbs estimated with less than 10° error on knee flexion at initial contact. Points in the low left squared represent limbs estimated within a 10° error range in both pelvic tilt and knee flexion at initial contact.

Statistical significance tests between the proposed method and other considered predictors give 0, 0.03 and 0.04 for “no change”, “mean” and “median” respectively. This means that there is significant difference between the performance of the proposed method and the other predictors.

4 Discussion

For the first time, both postoperative knee flexion and pelvic tilt at initial contact are predicted given the preoperative gait, physical examination and type of surgery.

The proposed simulator performs always better than the postoperative mean parameters over all patients, as well as the postoperative median and “no change” prediction (where the prediction is the same preoperative parameter). In addition, statistical tests show that there are significant differences between the proposed simulator and the pseudo-predictors (mean, median and no change). However, the global performance is not too far from the performance of the mean or median.

The error rates are almost the same for those limbs that had hamstring lengthening and those who had an-

other kind of surgery, which means that the simulator is able to predict independently of the type of surgery.

Since surgical treatment implies various risks and potential complications, a threshold for the difference between estimated and measured postoperative knee flexion has to be defined in order to determine whether or not an estimation is acceptable. First of all, we need to consider the intrasubject gait variability, which is higher in CP children than in normal children (Steinwender *et al.*, 2000). Variability of knee flexion at initial contact and mean pelvic tilt in cerebral palsy patients has been measured via test-retest repeatability (Klejman, Andrysek, Dupuis, & Wright, 2010). In particular, Klejman *et al.* (2010) compute the minimum detectable change (MDC) for these two parameters. The MDC is the minimal variation of the parameter for which we can be sure that is a real change and not part of the variability of the parameter. The MDC for knee flexion at initial contact and mean pelvic tilt are 8.1° and 9.3° respectively. If we add the uncertainties linked to surgery and rehabilitation, we consider an acceptable prediction an estimation with less than 10° of error. The proposed system predicts 58% of limbs within this range of error in both knee flexion and pelvic tilt at initial contact. In addition, the expected error for both considered postoperative parameter is inferior to the MDC, which means that in general the prediction error is under the variability of the signals.

When the system predicts well the knee flexion at initial contact, the pelvic tilt at initial contact is likely to be well predicted as well (93% of the cases). Moreover, the postoperative knee flexion at initial contact is more complicated to estimate than the postoperative pelvic tilt at initial contact.

Around a third of the patients are not well estimated. Regarding the medical application, it is important to apply surgery only if a good result can be asserted. Conversely, it is crucial to avoid a surgery plan if a bad outcome is most likely. For this reason, it would be interesting to detect a priori patients that will be badly estimated. With this strategy, patients more likely to be badly estimated, could be rejected by the simulator. For example, if a new patient is too far from the training patients in the input space, no estimation will be given. For this purpose, we propose the computation of confidence intervals that can help to reject or accept a patient or a prediction. For instance, if the confidence interval for a given patient is too wide, the prediction will be considered as unprecise.

The variable selection applied allows to decrease the dimension of the problem from 16 to 7 for the knee flexion at initial contact, which is less than the half of the original candidates (56% of dimensionality reduction). For the pelvic tilt, the dimensionality reduction is 31%, hence lower than for the knee flexion. This also decreases complexity of the nonlinear regression. Moreover, with the Gram-Schmidt orthogonalization, we maximize the correlation of the input with the target output and, at the same time, we reduce redundancy between the entry variables. In addition, this variable selection method constitutes the

input for regression taking into account the statistical relations between the target output and the candidates and not arbitrarily as in Hicks *et al.* (2011). This allows to find relations between some parameters that are not traditionally considered by clinicians. For example, we have found several important relations between contralateral angles, such as contralateral hip abduction and rotation and contralateral foot-ground angle, which are not considered in previous works as Sebsadji *et al.* (2012) and Hicks *et al.* (2011).

The potential use of the proposed system should be limited to patients and surgeries that are similar to those in our database. The effect of concurrent surgeries has been considered as random for the purposes of this study, but they also have an effect on gait parameters. Another limitation of the method is that it does not consider some other clinical data as input (GMFCS, topography, spasticity, etc.). Nevertheless, the system can deal with gait symmetry and asymmetry (i.e., hemiplegia and diplegia) because it considers bilateral data for each limb.

Finally, the proposed application could predict good or bad outcomes of orthopedic surgery, as Desailly (2008), Arnold *et al.* (2006), Hicks *et al.* (2011) and Sebsadji *et al.* (2012), after a clinical interpretation of the estimated postoperative parameters. Summing the estimated resulting gait parameters, this allows clinicians and patients to have a clearer idea *a priori* of the probable treatment outcome. However, the system output should be considered with caution given the delicate medical application, because a significant part of the patients are still not well predicted. Several perspectives of development of the prediction system are in process.

Acknowledgements. This work is part of a project funded by Fondation Ellen Poidatz, Fondation Bettencourt Schueller and Ile de France region. The authors would like to thank the UNAM medical and technical team of Fondation Ellen Poidatz for useful discussions and for recording all data used in this work.

Bibliography

- Arnold, A.S., Liu, M.Q., Schwartz, M.H., Öunpuu, S., & Delp, S.L. (2006). The role of estimating muscle-tendon lengths and velocities of the hamstrings in the evaluation and treatment of crouch gait. *Gait & Posture*, 23 (3), 273–281.
- Baumann, J.U., Ruetsch, H., & Schürmann, K. (1980). Distal hamstring lengthening in cerebral palsy. *International Orthopaedics*, 3 (4), 305–309.
- Bishop, C. (2006). *Pattern Recognition and Machine Learning*. Information Science and Statistics, Springer.
- DeLuca, P.A., Öunpuu, S., Davis, R.B., & Walsh, J.H. (1998). Effect of hamstring and psoas lengthening on pelvic tilt in patients with spastic diplegic cerebral palsy. *Journal of Pediatric Orthopaedics*, 18 (6), 712–719.
- Desailly, E. (2008). *Analyse biomécanique 3D de la marche de l'enfant déficient moteur. Modélisation segmentaire et modélisation musculo-squelettique*. PhD thesis, Université de Poitiers.
- Duda, R., Hart, P., & Stork, D. (2001). *Pattern Classification*. Wiley-Interscience, 2nd edition.
- Gage, J., Schwartz, M., Koop, S., & Novacheck, T. (2009). *The identification and treatment of gait problems in cerebral palsy*. Mac Keith Press, 2nd edition.
- Guyon, I. & Elisseeff, A. (2003). An introduction to variable and feature selection. *The Journal of Machine Learning Research*, 3, 1157–1182.
- Klejman, S., Andrysek, J., Dupuis, A., & Wright, V. (2010). Test-Retest reliability of discrete gait parameters in children with cerebral palsy. *Archives of Physical Medicine and Rehabilitation*, 91 (5), 781–787.
- Hicks, J.L., Delp, S.L., & Schwartz, M.H. (2011). Can biomechanical variables predict improvement in crouch gait?. *Gait & Posture*, 34 (2), 197–201.
- Laracca, E., Stewart, C., Postans, N., Roberts, A. (2014). The effects of surgical lengthening of hamstring muscles in children with cerebral palsy – The consequences of pre-operative muscle length measurement. *Gait & Posture*, 39 (3), 847–851.
- Ma, F.Y.P., Selber, P., Natrass, G.R., Harvey, A.R., Wolfe, R., & Graham, H.K. (2006). Lengthening and transfer of hamstrings for a flexion deformity of the knee in children with bilateral cerebral palsy. Technique and preliminary results. *Journal of Bone & Joint Surgery, British Volume*, 88 (2), 248–254.
- Sebsadji, A., Khouri, N., Djemal, K., Yepremian, D., Hareb, F., Hoppenot, P., & Desailly, E. (2012). Description and classification of the effect of hamstrings lengthening in cerebral palsy children multi-site surgery. *Computer Methods in Biomechanics and Biomedical Engineering*, 15 (supl.1), 177–179.
- Steinwender, G., Saraph, V., Scheiber, S., Zwick, E.B., Uitz, C., & Hackl, K. (2000). Intrasubject repeatability of gait analysis data in normal and spastic children. *Clinical Biomechanics*, 15 (2), 134–139.
- Thompson, N.S., Baker, R.J., Cosgrove, A.P., Saunders, J.L., & Taylor, T.C. (2001). Relevance of the popliteal angle to hamstrings length in cerebral palsy crouch gait. *Journal of Pediatric Orthopaedics*, 21 (3), 383–387.

Chapter 6

Predicting Postoperative Gait Using Multiple Linear Regression

Contents

6.1	Multiple linear regression	79
6.2	Predicting Postoperative Gait in Cerebral Palsy	81

6.1 Multiple linear regression

Linear regression consists in approximating the target output y as a linear combination of linear functions, thus a straight line if the dimension of the input x is $\dim(x) = 1$ as:

$$\hat{y} = ax + b \quad (6.1)$$

where \hat{y} is the approximation of y , a is the slope of the straight line and b is the independent term. Hence, the target output can be expressed as

$$y = \hat{y} + \epsilon = ax + b + \epsilon \quad (6.2)$$

where ϵ is the approximation error.

When the input is composed of multiple variables of dimension N_{in} , the approximation becomes a plane ($N_{in} = 2$) or a hyperplane ($N_{in} > 2$), and linear regression is usually called multiple linear regression. In such case, Equation 6.2 becomes:

$$y = a_1x_1 + a_2x_2 + \dots + a_{N_{in}}x_{N_{in}} + b + \epsilon = \begin{pmatrix} a_1 & a_2 & \dots & a_{N_{in}} & b \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{N_{in}} \\ 1 \end{pmatrix} + \epsilon \quad (6.3)$$

If the output is also multivariate with dimension N_{out} over n examples, we can generalize Equation 6.3 as:

$$Y = AX + \epsilon \quad (6.4)$$

With

$$Y = \begin{pmatrix} y_{1,1} & \cdots & y_{1,n} \\ \vdots & \ddots & \vdots \\ y_{N_{out},1} & \cdots & y_{N_{out},n} \end{pmatrix}$$

$$X = \begin{pmatrix} x_{1,1} & \cdots & x_{1,n} \\ \vdots & \ddots & \vdots \\ x_{N_{in},1} & \cdots & x_{N_{in},n} \\ 1 & \cdots & 1 \end{pmatrix}$$

$$A = \begin{pmatrix} a_{1,1} & \cdots & a_{1,N_{in}+1} \\ \vdots & \ddots & \vdots \\ a_{N_{out},1} & \cdots & a_{N_{out},N_{in}+1} \end{pmatrix}$$

From Equation 6.4 and Equation 4.26, the cost function that we need to minimize A is:

$$J = \frac{1}{2} (Y - AX)^2 \quad (6.5)$$

Deriving Equation 6.5 with respect to A , we obtain:

$$\frac{\partial J}{\partial A} = (Y - AX) X^T = YX^T - AXX^T \quad (6.6)$$

Where X^T is the transposed of X . By making Equation 6.6 equal to zero, we get:

$$YX^T = AXX^T \longrightarrow A = YX(XX^T)^{-1} \quad (6.7)$$

Where $X(XX^T)^{-1}$ is the Moore-Penrose pseudoinverse. To find a minimum, the matrix (XX^T) needs to be positive definite. This condition is respected if X has full rank, thus $n > N_{in}$ (each example is considered as independent). This condition obliges us to reduce the dimension of the preoperative vector beforehand, if we want to consider all the kinematic signals.

In statistical learning problems, linear regression is usually the first method that is tested. Then, if results are not satisfactory, nonlinear regression is tested. In the previous chapter we have seen how nonlinear regression can be applied to predict gait variables at a single gait cycle instant. Since we are interested in predicting the whole postoperative kinematic signals, we conducted another experiment to predict complete gait curves using linear regression with parametric confidence intervals. This experiment, where principal component analysis is considered for dimensionality reduction [see section 4.4], was the subject of another journal article that is presented below.



Full length article

Predicting postoperative gait in cerebral palsy



Omar A. Galarraga C.^{a,b}, Vincent Vigneron^b, Bernadette Dorizzi^c, Néjib Khouri^{a,d},
Eric Desailly^{a,*}

^a UNAM, Pôle Recherche & Innovation, Fondation Ellen Poidatz, 1 Rue Ellen Poidatz, Saint-Fargeau-Ponthierry, France

^b IBISC—EA 4526, Université d'Evry Val d'Essonne, 40 Rue du Pelvoux, Courcouronnes, France

^c SAMOVAR—UMR 5157, Télécom SudParis, Institut Mines-Télécom, 9 Rue Charles Fourier, Evry, France

^d Chirurgie Orthopédique Pédiatrique, Hôpital Universitaire Necker-Enfants Malades, 149 Rue de Sèvres, Paris, France

ARTICLE INFO

Article history:

Received 13 January 2016

Received in revised form 6 September 2016

Accepted 6 November 2016

Keywords:

Cerebral palsy

Single-event multilevel surgery

Outcome prediction

Clinical gait analysis

Machine learning

ABSTRACT

In this work, postoperative lower limb kinematics are predicted with respect to preoperative kinematics, physical examination and surgery data. Data of 115 children with cerebral palsy that have undergone single-event multilevel surgery were considered. Preoperative data dimension was reduced utilizing principal component analysis. Then, multiple linear regressions with 80% confidence intervals were performed between postoperative kinematics and bilateral preoperative kinematics, 36 physical examination variables and combinations of 9 different surgical procedures. The mean prediction errors on test vary from 4° (pelvic obliquity and hip adduction) to 10° (hip rotation and foot progression), depending on the kinematic angle. The unilateral mean sizes of the confidence intervals vary from 5° to 15°. Frontal plane angles are predicted with the lowest errors, however the same performance is achieved when considering the postoperative average signals. Sagittal plane angles are better predicted than transverse plane angles, with statistical differences with respect to the average postoperative kinematics for both plane's angles except for ankle dorsiflexion. The mean prediction errors are smaller than the variability of gait parameters in cerebral palsy. The performance of the system is independent of the preoperative state severity of the patient. Even if the system is not yet accurate enough to define a surgery plan, it shows an unbiased estimation of the most likely outcome, which can be useful for both the clinician and the patient. More patients' data are necessary for improving the precision of the model in order to predict the kinematic outcome of a large number of possible surgeries and gait patterns.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Orthopaedic surgery is usually performed in order to lessen gait abnormalities observed in patients with cerebral palsy (CP). Multiple bones and muscles are operated during a Single Event Multilevel Surgery (SEMLS) [1], which combines several procedures in the same surgery.

Clinical Gait Analysis (CGA) is used in combination with physical examination in order to propose a suitable surgery to patients with CP [1]. However, surgical decision making is not yet fully standardized. Different surgical procedures may be proposed to address the same gait deviation and different decision making algorithms may be used by medical teams to define surgical plans. Moreover, once the indication is established it is difficult for the

surgeon and furthermore for the patient to predict the effect of the surgery. Recently, several decision-making tools based on statistical machine learning have been developed for predicting surgery outcome in SEMLS. Reinbolt et al. [2] used linear discriminant analysis for predicting good or bad outcomes of rectus femoris transfer for patients with stiff knee. For predicting good or bad outcomes of hamstring lengthening, Arnold et al. [3] utilized hierarchical log-linear analysis and Sebsadji et al. [4] used support vector machines both combined with musculoskeletal models. Schwartz et al. [5] used random forests for predicting good or bad outcomes of psoas lengthening. All of the above methods give qualitative outcome predictions of improvement or non-improvement, but they do not help the surgeon nor the patient to predict how the latter will walk after surgery.

Some other methods predict some quantitative gait parameters. Hicks et al. [6] used multiple linear regression for predicting post-treatment knee flexion during stance for patients presenting crouch gait and also established good or bad outcomes based on these predictions. Sullivan et al. [7] used regression analysis and

* Corresponding author at: 1 Rue Ellen Poidatz, 77310 Saint Fargeau-Ponthierry, France.

E-mail address: eric.desailly@fondationpoidatz.com (E. Desailly).

Hersh et al. [8] used artificial neural networks to predict knee flexion during gait after rectus femoris transfer. Galarraga et al. [9] utilized artificial neural networks for predicting postoperative knee flexion and pelvic tilt at initial contact with or without hamstring lengthening. All previous works already mentioned are based on one type of abnormal gait patterns in CP (i.e. stiff knee or crouch gait) or on one principal surgical procedure, without considering the effect of other surgical procedures and their combinations. Niiler et al. [10] considered rectus femoris transfer and concurrent surgeries (hamstring lengthening, Achilles lengthening and gastrocnemius lengthening) of 68-patient series (94 lower limbs) and performed linear regressions for predicting postoperative knee range of motion during gait.

Despite these previous works, surgery planning remains difficult and global gait outcome prediction is still incomplete. Moreover, it is difficult to explain postoperative expected outcome to patients and their families, who might struggle to imagine a realistic outcome based on the predicted parameters.

The objective of this study was to use statistical machine learning techniques to develop a system able to predict postoperative kinematic curves of children with CP based on preoperative physical examination and 3-D gait analysis, and a proposed surgery plan.

2. Materials and methods

2.1. Population and data description

This retrospective study analyzed anonymous data of children with CP that have undergone SEMLS within a ten year period between 2004 and 2014. These children have had physical examination and CGA before and at least one year after surgery. Gait analysis was performed pre and postoperatively in the same laboratory. From 2004 to 2007, the acquisition was performed with a SAGA 3RT Biogesta system and, since 2008, with a Vicon system. Lower limb marker placements were identical in all the exams and kinematic data were computed from the acquisition's raw data (marker coordinates) with the same custom software based on a modified Helen Hayes [11,12] model with anatomical markers on the femoral condyles and the medial-malleolus. Fifteen kinematic angles were considered for each patient: pelvic tilt, pelvic

obliquity, pelvic rotation and hip flexion, hip adduction, hip rotation, knee flexion, ankle dorsiflexion and foot progression for both lower limbs. Surgical data were decomposed into combinations of $N_s = 9$ surgery categories: hip bony surgery, hip soft tissue surgery, rectus femoris surgery (transfer or release), hamstring lengthening, patella lowering, distal femoral osteotomy, shank bony surgery, ankle-foot soft tissue surgery and foot bony surgery. The surgical categories have been established depending on their functional objective and joint or segment that is modified. In these categories, some different surgical procedures are grouped in the same class if their functional objective and the affected joints or segments are alike [see Supplementary data for examples]. For each lower limb j , a surgery binary code $S_j = (s_{j,1} \dots s_{j,N_s})^T$ was attributed where $s_{j,i} = \begin{cases} 1 & \text{if gesture } i \text{ was conducted on patient } j \\ 0 & \text{if gesture } i \text{ was not conducted on patient } j \end{cases}$ with $i = 1, \dots, N_s$ and T is the transpose operator.

2.2. Preprocessing

The variables that have been measured during physical examination varied depending on the patient and the clinician that performed the exam. For this reason, we considered 36 parameters that were measured at a minimum rate of 80% in our database. These parameters include information about size and weight; hip, knee and ankle ranges of motion; muscle force; and spasticity (details in Supplementary data).

Fig. 1 shows all the stages of the method. Physical examination missing data were replaced using iterative robust model-based imputation (IRMI) [13]. This technique consists on initializing missing values and then iteratively perform linear regressions of each column with respect to the others. The initialization begins by searching the lower limbs with the nearest physical examination profile considering only the non-missing data with a k -Nearest Neighbor algorithm [14] for $k=5$ and ends by replacing each missing value by the median over the 5 nearest neighbors.

Kinematic data were automatically segmented into gait cycles utilizing the high pass algorithm (HPA) [15]. Then gait cycles were resampled and normalized to 51 points (2% of gait cycle) as in [16] and mean gait cycles were computed for each limb. A right and a left kinematic preoperative gait vectors were composed with the fifteen kinematic signals of both limbs normalized respectively by

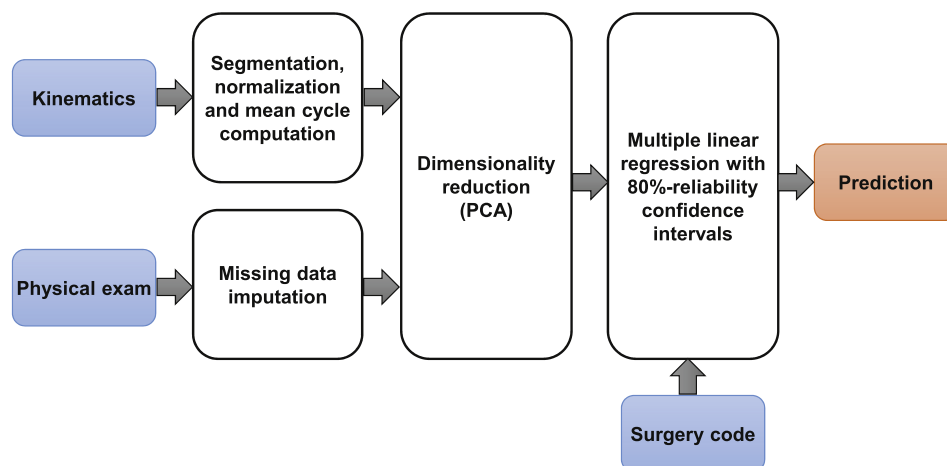


Fig. 1. Method stages from CGA and physical examination data (see Supplementary data) to prediction. Kinematic signals were segmented into gait cycles and normalized to 51 points per angle and cycle. Missing data from physical examination were imputed with the IRMI algorithm. The dimension of the concatenated vectors of preprocessed kinematics and physical examination data (see Supplementary data) was reduced using PCA. Then a multiple linear regression between postoperative kinematics and the low-dimensional preoperative vectors and surgery codes was performed. Confidence intervals with 80% reliability were computed.

the right and the left gait cycle. Postoperative gait vectors were solely described by the kinematic data of the considered limb like in [16]. Preoperative kinematic and imputed physical examination data (see Supplementary data) were gathered and projected into a lower-dimensional space using principal component analysis (PCA) [17]. Different PCA projection dimensions from 1 to 63 were tested for regression, keeping between 25% and 99% of the data inertia.

2.3. Regression model

After preprocessing, a multiple linear regression [18] was performed between the postoperative kinematic data of all lower limbs and the low-dimensional preoperative data plus the surgery data.

The multiple linear regression is represented by the equation $Y = AX + \epsilon$, where

$$Y = \begin{pmatrix} y_{1,1} & \cdots & y_{N_{LIMB},1} \\ \vdots & \ddots & \vdots \\ y_{1,N_{out}} & \cdots & y_{N_{LIMB},N_{out}} \end{pmatrix} \text{ represents the postoperative}$$

kinematic data of all lower limbs,

$N_{out} = 51 \times 9 = 459$ is the total number of output points (nine

51-point kinematic angles for each limb), $X =$

$$\begin{pmatrix} x_{1,1} & \cdots & x_{N_{LIMB},1} \\ \vdots & \ddots & \vdots \\ x_{1,DIM} & \cdots & x_{N_{LIMB},DIM} \\ s_{1,1} & \cdots & s_{N_{LIMB},1} \\ \vdots & \ddots & \vdots \\ s_{1,N_s} & \cdots & s_{N_{LIMB},N_s} \\ 1 & \cdots & 1 \end{pmatrix} \text{ includes low-dimensional preopera-}$$

tive data x_j and surgery codes S_j , $DIM = 1, \dots, 63$ is the dimension

of the PCA projection, $\epsilon = \begin{pmatrix} e_1 \\ \vdots \\ e_{N_{LIMB}} \end{pmatrix}^T$ is the training error vector

and $A = \begin{pmatrix} a_{1,1} & \cdots & a_{1,DIM+N_s+1} \\ \vdots & \ddots & \vdots \\ a_{N_{out},1} & \cdots & a_{N_{out},DIM+N_s+1} \end{pmatrix}$ is the regression matrix to

be estimated and whose parameters were learned using the least-squares method [18] as $A = Y(X^T X)^{-1} X^T$ with $DIM \leq n$.

Parametric confidence intervals for prediction are constructed for each limb and for each point of the gait curves using the expression

$$\hat{y}_0 \pm t_{n-k-1, \alpha/2} \sqrt{1_{N_{out} \times 1} + \hat{\sigma}^2 x_0^T (X \cdot X^T)^{-1} x_0}, \quad \text{where}$$

$$\hat{\sigma}^2 = \begin{pmatrix} \hat{\sigma}_1^2 \\ \vdots \\ \hat{\sigma}_{N_{out}}^2 \end{pmatrix}, \quad \hat{\sigma}_i^2 = \frac{SSE_i}{N_{LIMB}-k-1}, \quad SSE_i \text{ is the sum of quadratic errors}$$

for point i , $k = DIM + N_s$ is the number of input parameters, $x_0 =$

$$\begin{pmatrix} x_{0,1} \\ \vdots \\ x_{0,k} \\ 1 \end{pmatrix} \text{ is a test input vector, } \hat{y}_0 = \begin{pmatrix} \hat{y}_{0,1} \\ \vdots \\ \hat{y}_{0,N_{out}} \end{pmatrix} \text{ is the prediction of } y_0$$

and $t_{n-k-1, \alpha/2} = 1.28$ for $\alpha = 80\%$ confidence level with respect to a

unilateral Student distribution with $n - k - 1$ degrees of freedom [20].

2.4. Performance assessment

For every test limb and gait angle, the prediction performance was evaluated by the root-mean-square error:

$RMSE = \sqrt{\frac{\sum_{i=1}^{N_{points}} (\hat{y}_i - y_i)^2}{N_{points}}}$, where \hat{y}_i and y_i are respectively the prediction and the expected resampled postoperative gait signal for any i among the $N_{points} = 51$ points of a normalized kinematic angle. In addition, the overall RMSE per limb was computed with $N_{points} = N_{out} = 51 \times 9 = 459$, which corresponds to the concatenation of the nine predicted kinematic curves. The overall RMSE per patient (considering both limbs predictions at the same time) was also considered.

The method was tested using a leave-one-out jackknife procedure [14,19]: when a patient k is tested, data of both limbs of patient k were removed from the training set. Then a multiple linear regression was done with data of the other remaining patients and subsequently the data of patient k (both limbs separately) were tested on the trained model. This process was repeated for all the patients in the database.

These prediction errors were compared to errors of two pseudo-predictors: “Mean-P” and “No Change-P”. “Mean-P” is a constant pseudo-predictor equal to the mean postoperative gait cycle over all patients in the database. “No Change-P” is a pseudo-predictor that indicates the preoperative gait cycle for each patient as the postoperative prediction, which means that the surgery would have no effect on these parameters.

To measure the reliability of the parametric confidence intervals, the number of postoperative mean cycle points of test patients that were outside of the prediction band were counted for each gait angle.

The effect of the gait deviations for a single kinematic angle on the prediction performance was studied by comparing the RMSE to the preoperative Gait Variable Score (GVS) [21]. The reference database used for computing GVS is composed of 14 subjects presenting no pathology with an average age of 16 y/o (SD = 8 y/o). These reference data were acquired and processed under the same conditions of the patient's database.

Finally, the prediction performance was compared to the minimal detectable change (MDC) [22] of gait patterns in CP.

3. Results

Data of $N_{PAT} = 115$ children with CP ($N_{LIMB} = 230$ lower limbs) were included. 34 (29.6%) patients have a Gross Motor Function Classification System (GMFCS) level I, 57 (49.6%) a level II and 24 (20.9%) a level III. Male subjects represent 61% of the population and 39% are female. The mean ages of patients are 11.8 y/o (SD = 3.3) and 14.8 y/o (SD = 3.3) for preoperative and postoperative CGA respectively. Mean age at surgery is 12.6 (SD = 3.2) y/o, with postoperative CGA around 18 month after surgery. The percentages of patients that underwent each surgery category are: 33.9% hip bony surgery, 23.5% hip soft tissue surgery, 49.1% rectus femoris surgery, 50.9% hamstring lengthening, 19.6% patella lowering, 10% distal femoral osteotomy, 4.8% shank bony surgery, 50.0% ankle-foot soft tissue surgery and 23.48% foot bony surgery. There are 75 different combinations of surgical procedures and the most frequent combination has been performed over 13 lower limbs in the database (hip bony surgery with rectus femoris surgery, hamstring lengthening and ankle-foot soft tissue surgery). The average gait deviation index (GDI) [16] variation (Postoperative GDI – Preoperative GDI) is 8.5 (SD = 9.3, MAX = 37.6, MIN = –16.2).

Best results were obtained with PCA dimension of $DIM = 9$, which contains 82% of the total variance. After dimensionality reduction and reprojection into the original space, 57.6% of the original kinematic values that were outside a 2-standard-deviation (2-SD) band are preserved. Original values outside this 2-SD band represent 5% of the kinematic data. For physical examination data (see Supplementary data), 2.1% of the outlier values were preserved after reprojection. Originally 4.6% of the physical examination values were outside the 2-SD band. The kinematic angles that lost the most part of their original variance after PCA are hip adduction (bilaterally) with 61% of variance loss, knee flexion from 75% to 85% of the gait cycle with 64% of variance loss and pelvic obliquity with 51% of variance loss. The rest of the kinematic variables preserved more than 50% of their original variance, with an average loss of 18%. All the following results correspond to regressions with the optimal PCA dimension, thus to input vectors of size 18 (PCA projection of size $DIM = 9$ plus surgery code of size $N_s = 9$).

The simulator outputs the predicted kinematic curves for the test limb with their respective 80%-reliability confidence band (Fig. 2).

Mean prediction errors over all patients of the proposed method vary from 3.7° (pelvic obliquity) to 9.9° (foot progression), depending on the kinematic angle (Table 1). Best pseudo-predictors mean performances vary from 3.9° (pelvic obliquity) and 11.0° (foot progression). There are significant differences between the proposed method RMS errors and the pseudo-predictors for all the angles, except for hip adduction, hip rotation and ankle dorsiflexion. For all the kinematic angles, the mean prediction error of the proposed method is smaller or equal to the prediction error of pseudo-predictors. Overall (considering all the kinematic angles at the same time), 65.6% of the limbs are better predicted with the proposed method. Moreover, if both lower limbs of each patient are considered, the proposed method gives better predictions for 71.3% of the patients.

Unilateral mean sizes over all patients of the 80%-reliability confidence intervals vary from 5.1° (pelvic obliquity) and 14.6° (foot progression) [Table 1]. The standard deviations of the intervals' sizes are between 0.1° and 0.3° . The percentages of test points inside the parametric confidence intervals vary from 76.2% (pelvic tilt and knee flexion) to 80.6% (foot progression) [Table 1].

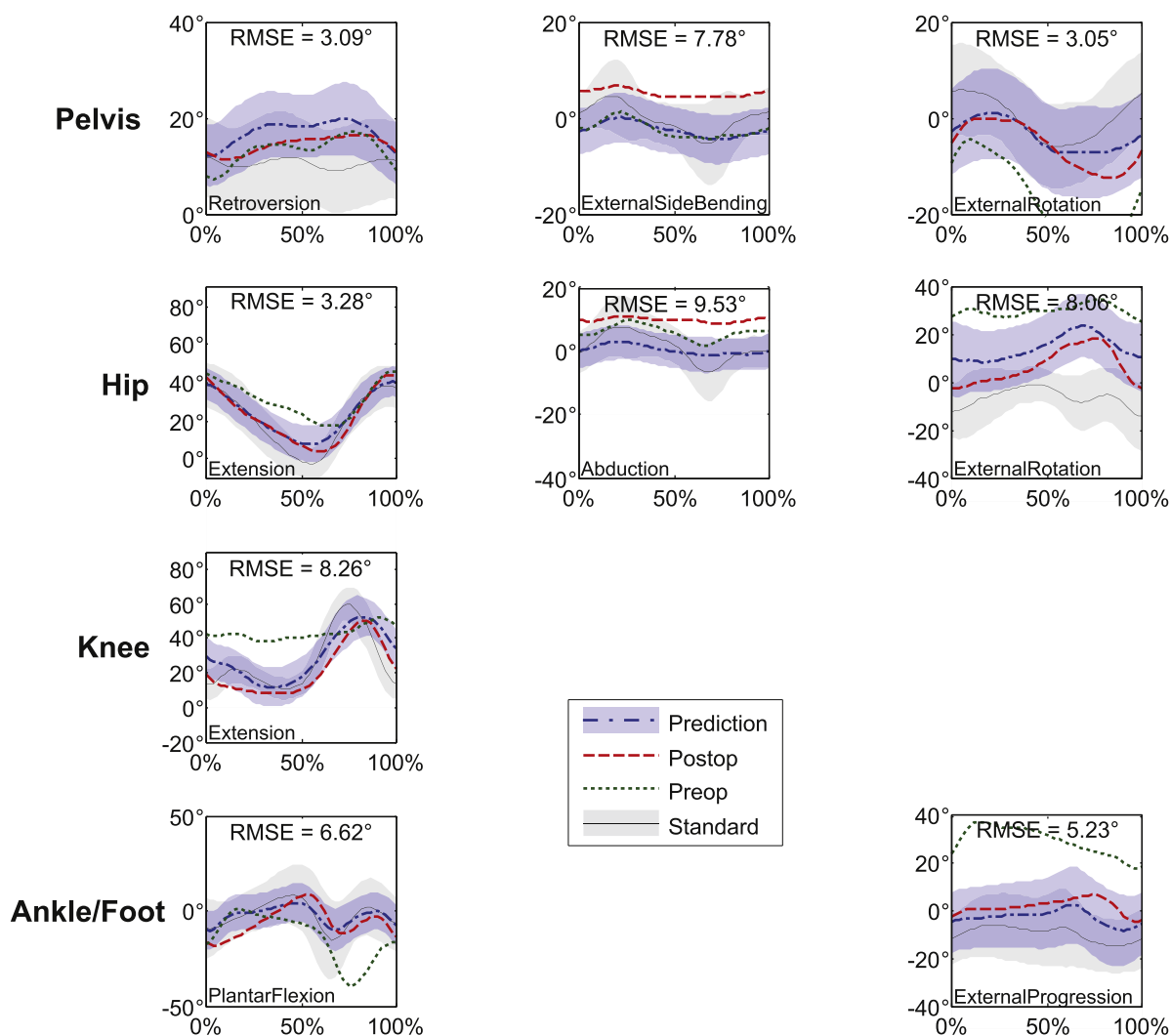


Fig. 2. System output example. Dash-dotted blue lines correspond to the estimation and blue bands represent their confidence intervals. Dashed red lines correspond to measured (real) mean postoperative gait cycle. Dotted green lines represent mean preoperative gait cycle. Solid black lines and grey bands correspond to the standards (non-pathological gait) and their two-standard-deviation bands. These standards come from the reference database described in Section 2.4. In this example the surgery for this limb consists of bony and muscle hip surgery, rectus femoris surgery (RFT), hamstring lengthening (HL) and muscle ankle/foot surgery. The contralateral limb surgery consists on rectus femoris surgery (RFT) and hamstring lengthening (HL). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1

Comparison of the performances of the proposed method and the pseudo-predictors (Mean-P, No change-P) per kinematic angle. Unilateral sizes and percentages of test points within the prediction bands of the proposed method.

		Pelvic Tilt	Pelvic obl.	Pelvic rot.	Hip flex.	Hip add.	Hip rot.	Knee flex.	Ankle dors.	Foot prog.
RMSE (°) Mean (SD)	Predictor:									
	No Change-P	8.9 (5.6)*	5.7 (3.1)*	10.3 (5.7)*	11.0 (5.8)*	6.2 (3.3)*	15.8 (9.5)*	17.2 (10.2)*	14.9 (9.7)*	19.2 (12.3)*
	Mean-P	7.1 (3.5)*	3.9 (1.9)*	7.0 (3.9)*	7.6 (4.0)*	4.2 (2.0)	10.2 (6.2)	10.6 (4.3)*	7.5 (4.0)	11.0 (7.3)*
	Proposed	5.1 (3.2)	3.7 (1.9)	6.6 (3.8)	6.8 (3.7)	4.1 (1.8)	9.7 (5.6)	9.0 (3.9)	7.5 (4.2)	9.9 (7.3)
Confidence intervals	Inside points on test (%)	76.2	79.3	79.9	79.0	76.9	78.3	76.2	78.0	80.6
	Mean size (°)	7.0	5.1	9.3	9.0	5.5	13.5	11.6	10.1	14.6
	SD of size (°)	0.1	0.1	0.2	0.2	0.1	0.3	0.2	0.2	0.3

* Significant difference with respect to errors of the proposed method ($p < 0.05$).

The prediction error is uniformly distributed with respect to the preoperative GVS [21], with a maximal Pearson correlation coefficient of 0.28 (hip adduction) [Fig. 3].

All the mean prediction errors are smaller than their associated MDC and at least 58.3% up to 100% of the limbs are predicted with an error smaller than the MDC depending on the predicted parameter [Table 2].

4. Discussion

For the first time postoperative lower limb kinematics are predicted with respect to preoperative kinematics, preoperative physical examination and a large number of combinations of surgical procedures in a SEMLS context for patients with CP that walk with different gait patterns. Other outcome-predictive

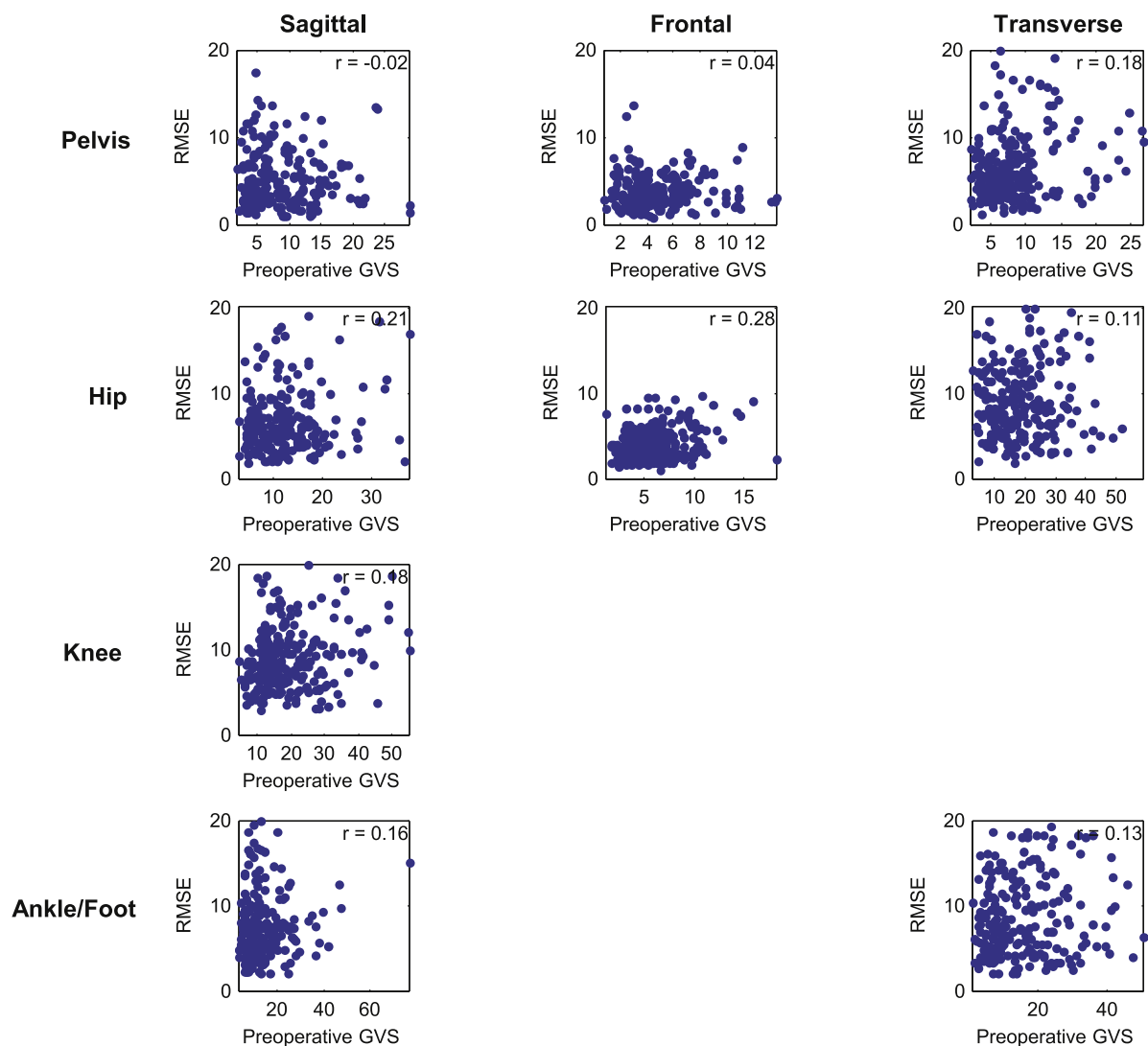


Fig. 3. Prediction Error (RMSE) with respect to preoperative GVS for each kinematic angle.

Table 2

Comparison of the proposed method performance to variability of the signals: RMSE vs. MDC.

Parameter	MDC [22] (°)	Prediction RMSE		% of limbs where RMSE < MDC
		Mean (°)	SD (°)	
Sagittal				
Mean tilt	9.3	4.3	3.5	89.6
Pelvis ROM	3.9	3.2	2.7	67.8
Min hip flexion	9.6	6.3	5.5	80.0
Hip flexion IC	12.0	5.1	4.1	93.9
Hip ROM	12.4	6.4	5.1	87.8
Max knee flexion	11.4	7.2	5.4	78.7
Min knee flexion	8.9	7.4	6.1	65.2
Knee flexion IC	8.1	7.7	5.7	58.3
Knee ROM	12.1	9.6	7.1	70.4
Max ankle dorsi	10.8	6.5	5.2	83.0
Min ankle dorsi	10.3	6.9	5.7	77.0
Ankle dorsi IC	10.2	6.6	5.8	80.0
Ankle ROM	8.5	5.8	5.2	75.7
Frontal				
Min up obliquity	6.6	3.3	2.4	91.3
Max up obliquity	7.7	3.3	2.4	94.8
Max hip adduction	13.1	3.5	2.6	100.0
Min hip adduction IC	12.5	3.7	2.6	99.6
Min hip adduction swing	15.0	3.6	2.7	100.0
Transverse				
Max int rotation	14.4	5.7	4.7	94.4
Min int rotation	10.5	5.5	4.9	85.2
Max hip int rotation	16.1	9.0	6.9	85.2
Min hip int rotation	12.6	9.9	7.6	70.0

methods are usually based on one principal surgical gesture [2,4,5,7,8,10] or give only qualitative predictions [2,4–6]. The method integrates physical examination and CGA data, which represent the usual information utilized for treatment decision making [1] and are considered complementary information because they are uncorrelated [23].

If we consider the original problem with an input vector of 810 elements (765 kinematic elements, plus 36 physical examination variables and the 9-element surgery code), before PCA, and with 228 lower limbs in the training set (114 patients, excluding the test patient), there are infinite solutions of the least-squares optimization for the regression. PCA allows to reduce the size of the input vector in order to find a unique least-squares solution. PCA dimensionality reduction maximizes the variance and eliminates data redundancy [17], however it implies some information loss. The principal variables that were affected by this information loss were physical examination data (see Supplementary data) as well as some kinematic data such as hip adduction (bilaterally), knee flexion from 75% to 85% of the gait cycle and pelvic obliquity. Even with this dimensionality reduction stage, given the heterogeneity of gait patterns in CP and the number of different possible surgical combinations with respect to the number of patients in the database, the regression problem remains very complex. This represents a limitation for the precision of the system. If there were more patients in the training database, the dimensionality reduction stage might be less important or even omitted. In any case, with or without dimensionality reduction, prediction precision would be higher if more patients' data were available. For this reason, it would be interesting to consider data from other gait analysis laboratories to considerably increase the database size, and we would like to encourage data sharing between these laboratories, which is very limited at the present. This would also facilitate the use of other regression methods, especially nonlinear regression, which may also improve the performance.

The prediction errors are generally smaller than the variability of the CP gait parameters measured by the MDC [22]. The MDC is a

measure of intersession variability of these parameters and is an estimation of the minimal amount of change that is needed to exceed measurement error.

In addition, the prediction errors are in average smaller than the naive predictors' errors, especially Mean-P. This means that the proposed model gives richer information than just knowing the average SEMLS outcome [24], which is a tendency to slightly improve gait shown by the GDI [16] variation of the considered series. Nevertheless, the system is not yet accurate enough to find the most suitable surgery or treatment.

On the other hand, the prediction errors and the preoperative Gait Variable Scores (GVS) [21] are uncorrelated (Fig. 3). This means that the system is able to perform in the same way independently of the severity of the preoperative state of patients. However, the usage of the system should be limited to patients and surgeries similar to those composing the training database.

The system does consider contralateral preoperative state (kinematics and physical examination) but does not consider contralateral surgery, which has an effect on the postoperative kinematics. Another limitation of the system is the grouping of similar but not identical surgical procedures in one surgical category. Both the grouping and the omission of the contralateral surgery were necessary to reduce the variability of the treatment, but it also introduces a source of error that cannot be controlled by the system.

Even if further validation and especially external validation should be considered before clinical use, the adjunction of thresholds for preoperative parameters (kinematics and physical examination) could be established (i.e. within ± 2 SD) in order to filter patients having gait parameters that are far from those in the training base. Also surgery plans might be unpredicted if their combinations are too different from those in the training database.

Frontal plane angles (pelvic obliquity and hip adduction) are the angles predicted with the smallest mean errors over all the patients. However, this does not mean that this is always the case for all the patients (see Fig. 2). In addition, the global performances

for these angles are the same as predicting the postoperative mean, with no significance statistical difference for hip adduction.

Sagittal plane angles (pelvic tilt, hip flexion, knee flexion and ankle dorsiflexion) are better predicted than transverse plane angles, but have bigger prediction errors than frontal plane angles. Performances for sagittal plane angles are significantly more accurate than predicting the postoperative mean (Mean-P) except for ankle dorsiflexion.

Highest prediction errors are obtained for transverse plane angles (pelvis rotation, hip rotation and foot progression), which are also the angles with highest variability in children with CP (Table 2) [22]. In addition, confidence intervals for angles in transverse plane are the largest in general.

Computed parametric confidence intervals give prediction intervals with 80% reliability. This reliability has been confirmed for test data (76–81% of tested points per angle). The sizes of these confidence intervals are specific to the patient, surgery, angle and instant of gait cycle. These intervals show an estimation of the prediction uncertainty. For example, if a confidence interval were considered too large, its associated prediction would be considered as inaccurate. Prediction bands give a set of probable solutions that might help both clinicians and patients to better understand and discuss the most likely surgery outcome.

After evaluation of the most likely outcome, a qualitative prediction could also be presented in order to decide if the outcome would be good or bad, as most of the previous works do [2,4–6]. Decision thresholds over the gait variables variation could be established for this purpose, and each medical team could adjust those threshold according to theirs needs.

The proposed system could be used as a decision-making tool for SEMLS that shows the most likely surgery outcome in terms of gait to both clinicians and patients with a confidence level of at least 76%. The system represents an easy way for visualizing the most likely surgery outcome, which allows to facilitate clinician-patient treatment discussion. The output of the system could serve as a motivation for patients in order to proceed with a surgical treatment.

In conclusion, the proposed method is able to give a preview of the most likely surgery outcome for different surgical combinations and gait patterns, given a relatively small data set. The effect of each surgical procedure and preoperative parameters is not well estimated for frontal plane angles and ankle dorsiflexion, because these predictions tend too closely to the postoperative mean kinematics. Nevertheless, the postoperative mean is a good predictor for frontal plane angles, presenting tenuous mean prediction errors and standard deviations. On the other hand, the system is more sensitive to the input variables for both sagittal and transverse plane angles (except for ankle dorsiflexion), but is more precise for sagittal plane angles, especially for pelvic tilt and hip flexion. Finally, although there are still some developments and validations to be done, this work represents an encouraging progress towards general treatment outcome prediction, in order to help clinicians to choose optimal treatment and to help patients to better understand it.

Conflict of interest

None.

Acknowledgements

This study is part of the SIM-PC2 project promoted by the Fondation Ellen Poidatz and supported by Fondation Ellen Poidatz, Fondation Bettencourt Schueller and Région Île de France.

The authors would like to thank UNAM medical and technical team of Fondation Ellen Poidatz for useful discussions and for recording all the data used in this work.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.gaitpost.2016.11.012>.

References

- [1] J.R. Gage, M.H. Schwartz, S.E. Koop, T.F. Novacheck, *The Identification and Treatment of Gait Problems in Cerebral Palsy*, 2nd edition, MacKeith Press, London, 2009.
- [2] J.A. Reinbolt, M.D. Fox, M.H. Schwartz, S.L. Delp, Predicting outcomes of rectus femoris transfer surgery, *Gait Posture* 30 (July (1)) (2009) 100–105.
- [3] A.S. Arnold, M.Q. Liu, M.H. Schwartz, S. Öunpuu, S.L. Delp, The role of estimating muscle-tendon lengths and velocities of the hamstrings in the evaluation and treatment of crouch gait, *Gait Posture* 23 (April (3)) (2006) 273–281.
- [4] A. Sebsadji, N. Khouri, K. Djemal, D. Yepremian, F. Hareb, P. Hoppenot, E. Desailly, Description and classification of the effect of hamstrings lengthening in cerebral palsy children multi-site surgery, *Comp. Methods Biomech. Biomed. Eng.* 15 (September (Suppl. 1)) (2012) 177–179.
- [5] M.H. Schwartz, A. Rozumalski, W. Truong, T.F. Novacheck, Predicting the outcome of intramuscular psoas lengthening in children with cerebral palsy using preoperative gait data and the random forest algorithm, *Gait Posture* 37 (April (4)) (2013) 473–479.
- [6] J.L. Hicks, S.L. Delp, M.H. Schwartz, Can biomechanical variables predict improvement in crouch gait? *Gait Posture* 34 (June (2)) (2011) 197–201.
- [7] K. Sullivan, J. Richards, F. Miller, P. Castagno, N. Lennon, Predicting the outcome of surgery for children with cerebral palsy using pre-operative gait analysis, *Gait Posture* 3 (2) (1995) 92.
- [8] L.A. Hersh, J.Q. Sun, J.G. Richards, F. Miller, The prediction of post-operative gait patterns using neural networks, *Gait Posture* 5 (2) (1997) 151.
- [9] O.A. Galarraga, C.V. Vigneron, B. Dorizzi, N. Khouri, E. Desailly, Predicting postoperative knee flexion and pelvic tilt at initial contact of cerebral palsy children, *Mov. Sport Sci. - Sci Mot.* 93 (2016) 87–92.
- [10] T.A. Nüller, J.G. Richards, F. Miller, Concurrent surgeries are a factor in predicting success of rectus transfer outcomes, *Gait Posture* 26 (June (1)) (2007) 76–81.
- [11] R.B. Davis III, S. Öunpuu, D. Tyburski, J.R. Gage, A gait analysis data collection and reduction technique, *Hum. Mov. Sci.* 10 (October (5)) (1991) 575–587.
- [12] M.P. Kadaba, H.K. Ramakrishnan, M.E. Wootten, et al., Measurement of lower extremity kinematics during level walking, *J. Orthop. Res.* 8 (3) (1990) 383–392.
- [13] M. Templ, A. Kowarik, P. Filzmoser, Iterative stepwise regression imputation using standard and robust methods, *Comput. Stat. Data Anal.* 55 (10) (2011) 2793–2806.
- [14] R.O. Duda, P.E. Hart, D.G. Stork, *Pattern Classification*, 2nd edition, Wiley-Blackwell, New York, 2000.
- [15] E. Desailly, Y. Daniel, P. Sardain, P. Lacouture, Foot contact event detection using kinematic data in cerebral palsy children and normal adults gait, *Gait Posture* 29 (January (1)) (2009) 76–80.
- [16] M.H. Schwartz, A. Rozumalski, The gait deviation index: a new comprehensive index of gait pathology, *Gait Posture* 28 (October (3)) (2008) 351–357.
- [17] I.T. Jolliffe, *Principal Component Analysis*, Springer-Verlag, New York, 2002.
- [18] C.M. Bishop, *Pattern Recognition And Machine Learning*, 2nd ed., Springer-Verlag, New York, 2006.
- [19] R.G. Miller, The jackknife—a review, *Biometrika* 61 (April (1)) (1974) 1–15.
- [20] Stuart, Kendall's Advanced Theory of Statistics: Distribution Theory, 6th edition, Wiley-Blackwell, Chichester, West Sussex, 1994.
- [21] R. Baker, J.L. McGinley, M.H. Schwartz, S. Beynon, A. Rozumalski, H.K. Graham, O. Tirosh, The gait profile score and movement analysis profile, *Gait Posture* 30 (October (3)) (2009) 265–269.
- [22] S. Klejman, J. Andrysek, A. Dupuis, V. Wright, Test-retest reliability of discrete gait parameters in children with cerebral palsy, *Arch. Phys. Med. Rehabil.* 91 (May (5)) (2010) 781–787.
- [23] K. Desloovere, G. Molenaers, H. Feys, C. Huetnaerts, B. Callewaert, P.V. de Walle, Do dynamic and static clinical measurements correlate with gait analysis parameters in children with cerebral palsy? *Gait Posture* 24 (November (3)) (2006) 302–313.
- [24] A. Bonnefoy-Mazure, Y. Sagawa Jr., V. Pomero, P. Lascombes, G. De Coulon, S. Armand, Are clinical parameters sufficient to model gait patterns in patients with cerebral palsy using a multilinear approach? *Comput. Methods Biomech. Biomed. Eng.* 19 (7) (2016) 800–806.

Chapter 7

Ensemble Learning With Linear and Nonlinear Regression

Contents

7.1	Arriving to Ensemble Learning Methods	89
7.2	Predicting Postoperative Gait in Cerebral Palsy with Ensemble Learning	91

7.1 Arriving to Ensemble Learning Methods

In the previous experiment, it is uncertain if the model properly learns the effect of each surgical procedure. Besides, leaving nine inputs (half of the total inputs) uniquely for the surgery code penalizes the learning process and obliges us to further reduce the preoperative vector dimension. In addition, given the statistical small data problem (Bishop, 2012), we need a robust model able to learn from a small set of examples. For these reasons, another experiment was conducted using ensemble learning with linear and nonlinear regression for predicting postoperative gait. This experiment derived into another journal article that is presented in the next section. Ensemble learning and the proposed ensemble method are discussed respectively in sections 2.3 and 2.4 of the article in question.

Ensemble based machine learning consists in utilizing several simplified models, whose outputs are then combined in order to give an optimal solution to the problem. In the experiment of section 6.2 we have a unique output for each input limb. This is analogous to having the opinion of a surgeon with respect to the outcome of the surgery. Keeping this analogy, in ensemble learning we would ask several opinions from different surgeons (or any other experts) and then combine their opinions to give a final opinion. In ensemble learning, the first stage is the ensemble generation, and each ensemble E_i is used to train an expert model M_i . But, what should the expert model learn? What should be the architecture of M_i ? (*i.e. linear or nonlinear*).

We have also seen that in most previous works that surgery outcome prediction models are based for a single surgical procedure, such as rectus femoris transfer, hamstring lengthening or psoas lengthening [see section 2.7]. We have assumed since the beginning that the variation Δ of the kinematic signals of a limb is due to its surgery. For instance, if a limb had surgical procedures A and B , we can assume that its kinematic variation would be given by:

$$\Delta = h(\Delta_A, \Delta_B) \tag{7.1}$$

Where h is any function that relates Δ to Δ_A and Δ_B . Δ_A and Δ_B are the variations introduced respectively by surgical procedures A and B .

One possibility is to construct a model tuned for the combination of surgical procedures A and B that directly predicts Δ . Another possibility is to construct a model tuned for surgical procedure A and another model tuned for B , in order to estimate respectively Δ_A and Δ_B . Then by the relation in Equation 7.1 we can predict Δ .

Furthermore, if we had a model that simulates the effect of a single surgery for each surgical procedure, we could probably combine them to predict postoperative gait for any surgical combination.

To answer the first question, it seems logical that the ensembles E_i should be generated such that each expert model M_i learns the effect of a different single surgical procedure. However, surgeries are generally a combination of surgical procedures. Moreover, it is practically impossible to statistically simulate isolated effects of surgical procedures given our database, where few or null number of limbs per surgical procedure did not have any other surgical procedure at the same time [see section 3.3].

Both questions above will be further discussed in the next section.

On the other hand, the combination rule or function (h in example of Equation 7.1) needs to be optimized. In ensemble methods, there are two principal combination rule types: trainable and non-trainable. Trainable combination rules need a specific subset from the training set to be properly optimized. This reduces the amount of data that can be used to train models M_i . Hence non-trainable combination rules are intuitively more suitable for statistically small data sets. The optimal combination rule cannot “learn” with our database. However, we have fortunately *a priori* information on the effect of each surgical procedure, because we know what kinematic variables are significantly affected by each surgical procedure [see section 3.5]. The explanation of how this information was used for combining models will be detailed in the next section.

The other question is what type of ensemble method should be the optimal for our problem. There are numerous ensemble methods, but the most popular are *bagging* (Breiman, 1996), *boosting* (Schapire, 1990), *stacked generalization* (Wolpert, 1992) and *mixture of experts* (Jacobs, Jordan, Nowlan & Hinton, 1991), and their different variations. These methods differ in the generation of the ensembles and/or the combination rule. For instance, in stacked generalization and mixture of experts, the combination rule is learned from data. Conversely, in bagging and boosting, the combination rule is arbitrary. Since we are confronted to a statistical small data problem, any method with trainable combination rule can be discarded. On the other hand, in bagging ensembles are randomly generated by bootstrapped replicas (with replacement). In boosting, the first ensemble is also randomly generated. With random ensembles, M_i would learn the pre-post gait variation of random surgeries comprised in E_i . For this reason, these two other algorithms seem unsuitable for our problem. The *ad hoc* proposed model will be described in the next section.

The usage of simplified models makes ensemble methods have less tendency to overfit (Polikar, 2006), therefore they seem useful for statistical small data problems. Random Forests (Breiman, 2001), which is a variation of bagging, have been previously used to predict surgery outcome in CP (Schwartz et al., 2013). This ensemble method was used for classification (good and poor outcomes) and not for regression of quantitative gait parameters.

Predicting postoperative gait in cerebral palsy using ensemble learning

Abstract

Cerebral palsy (CP) manifests by gait troubles that are generally improved with orthopedic surgery consisting in a combination of several surgical procedures. Surgery programs are often established following a physical examination and a clinical gait analysis (CGA). In this work, postoperative kinematics signals were predicted with respect to preoperative gait and surgery type, using ensemble learning. Data of 134 operated children with CP that had CGA before and after surgery were considered. Preoperative kinematics and physical examination data were projected into a lower-dimensional space using principal component analysis (PCA). An ensemble method was developed where each surgical procedure (9 in total) was associated to a regression model between postoperative kinematics and preoperative data. The regression models were trained using multiple linear regression and nonlinear regression with feed-forward neural networks. Both the PCA-projection dimension and the regression model were optimized per each surgical procedure and kinematic signal. The different outputs of each optimal model were then fused with a non-trainable combination rule. This rule consists of an average weighted by the statistical significance of the models for each postoperative variable. The average prediction errors varied from 3° to 10° depending on the kinematic signal. The majority of the optimal models were whether linear or with few hidden units. The proposed combination rule allows to estimate the contribution of each surgical procedure. This provides a tool for showing the most likely outcome to both the patient and the clinician, allowing a better comprehension and discussion of both the treatment and its outcome.

1 Introduction

Cerebral Palsy (CP) is a group of neurological disorders, caused by brain damage during development, that affect human movement and balance [8]. CP entails muscle control problems, as well as bone and muscle deformities, which typically manifest by walking troubles. In order to lessen gait abnormalities, orthopedic surgery is usually performed on patients with CP. The purpose of the surgery is to improve gait patterns by modifying lower limb's anatomy depending on the patient's condition. For some patients, the objective of the

surgery would be to obtain standard non pathological gait patterns. In some other cases, the objective would be to decrease pain during walking or to allow the patient to walk around his or her home without using a wheelchair. Multiple bones and muscles are modified during a Single Event Multilevel Surgery (SEMLS) [8], which combines several surgical procedures in the same operation.

In general surgical treatment decision is based on physical examination and Clinical Gait Analysis (CGA). Physical examination principally gives information about joint range of motion, muscle forces and spasticity. CGA combines motion capture techniques, force platforms and electromyography for measuring spatial-temporal parameters, kinematics, kinetics and muscle activity during gait [1]. Kinematic signals refer to joint angles during gait, as well as angular velocities. Kinetics allow to measure joint moments and forces.

Recently, several CGA-based decision-making tools for SEMLS, based on statistical machine learning, have been created for predicting surgery outcome. Niiler et al. [17] utilized neural networks for predicting some kinematic signals after rectus femoris transfer surgery over a 24-patient database. Sullivan et al. [25] used regression analysis to predict knee flexion during gait after rectus femoris transfer for 15 patients. Sebsadji et al. [23] predicted good or bad outcomes of hamstring lengthening surgery using Support Vector Machines (SVM) over 36 patients. Reinbolt et al. [19] utilized linear discriminant analysis (LDA) for predicting good or bad outcomes of rectus femoris transfer for 62 patients. Niiler et al. [16] used linear regression for predicting knee range of motion after rectus femoris transfer and concurrent surgical procedures for 68 patients. Galarraga et al. [10] utilized neural networks for predicting postoperative knee flexion at initial contact for SEMLS with or without hamstring lengthening for 99 patients. Galarraga et al. [9] used multiple linear regression over 115 patients for predicting knee flexion during gait. All the methods above are faced to limitations because they all confront statistically small data sets [3], with few number of training examples with respect to the number of parameters to be estimated. This small data situation is typical in medical applications and increases the probability of overfitted models [3]. To circumvent the small data problem, one should utilize robust methods to avoid overfitting. In this regard, Schwartz et al. [22] used an ensemble method that consisted on random forests for predicting good or bad outcomes of SEMLS with or without psoas lengthening surgery over a 800-patient database.

The aforementioned works predict either qualitative parameters (good or bad outcome) or few gait parameters. This complicates interpretation of the predicted outcome, because patient's perception of good or bad outcome may differ from the clinician's perception. Besides, it is hard to imagine how the patients would walk after surgery with only qualitative or few parameters. Another disadvantage of these methods is that they predict in general the effect of a unique surgical procedure and not the combinations of different surgical procedures, except for [16] and [9].

The objective of this study is to develop an ensemble method able to predict postoperative kinematic curves of patients with CP on a statistically small dataset, based on preoperative physical examination, 3-D gait analysis and a proposed surgery plan. The system should also be able to predict the probable contribution of each surgical procedure composing the

proposed SEMLS.

1.1 Ensemble Learning

Ensemble based machine learning consists in utilizing several simplified submodels M_i , whose outputs are combined in order to compute an optimal predictor or classifier. Let $\mathcal{D} = \{(\mathbf{x}^{(j)}, \mathbf{t}^{(j)})\}_{j=1}^n$ be the input-target training set. The first stage is to generate the ensembles $E_i \subseteq \mathcal{D}, i = 1, \dots, N_S$ where $E_i = \{(\mathbf{x}^{(\sigma_i(j))}, \mathbf{t}^{(\sigma_i(j))})\}_{j=1}^n$ and $\sigma_i(j)$ indicates a partition on \mathcal{D} such that $\sigma_i(j) \in \{1, 2, \dots, n\}$. This means that each model $M_i(\varphi_i; \mathbf{x})$ is trained on the training set E_i and the weight vectors $\{\varphi_i\}_{i=1}^{N_S}$ are optimized by maximizing $P(\varphi_i|M_i; \mathbf{x})$. Since each trained model M_i gives an estimation of the most likely output, the outputs of the models should then be combined in order to give a unique solution.

Ensemble learning methods are known to have less tendency of overfitting [18]. There are numerous ensemble methods, the most popular being bagging [5], boosting [20], stacked generalization [27] and mixture of experts [11], and their different variations. The principal differences between ensemble methods are the generation of the ensembles and the combination rule. For instance, in stacked generalization and mixture of experts, the combination rule is learned from data, and is thus a trainable combination rule. Conversely, in bagging and boosting, the combination rule is arbitrary. Trainable combination rules need a specific subset from the training set to be properly optimized to the detriment of the optimization of first stage models. Hence non-trainable combination rules are intuitively more suitable for statistically small data sets.

2 Materials and Methods

2.1 Data Description

The database is composed of $N_{PAT} = 134$ children with CP that have undergone SEMLS. In total $n = 232$ lower limbs of the patients have been surgically modified, which means that 98 children had bilateral surgery and 36 children had unilateral surgery. All patients have had a clinical gait analysis (CGA) before and after surgery. The average ages of patients are 11.8 y/o (SD=3.3) and 14.8 y/o (SD=3.3) for preoperative and postoperative CGA respectively. The average age at surgery is 12.6 (SD=3.2) y/o, with postoperative CGA around 18 month after surgery. Surgical data were decomposed into combinations of $N_s = 9$ surgery categories. The percentages of limbs that underwent each surgery category vary from 4.5% (bony shank) to 50.0% (hamstring lengthening) [see table 1].

There are 80 different combinations of surgical procedures and the most frequent combination has been performed over 13 lower limbs in the database (hip bony surgery with rectus femoris surgery, hamstring lengthening and ankle-foot soft tissue surgery, and also only hip bony surgery), with an average number of 3 examples per surgery combination. Moreover, three surgical procedures are always associated to other surgeries and, thus, are

Table 1: Considered surgical procedures categories and their frequencies in the database.

Segment/Joint	Category	Number of limbs	
Pelvis/Hip	Bony Hip Surgery	89	33.2%
	Muscle Hip Surgery	62	23.1%
Thigh/Knee	Rectus Femoris Surgery	129	48.1%
	Hamstring Lengthening	134	50.0%
	Patella Lowering	50	18.7%
	Distal Femoral Osteotomy	27	10.1%
Shank	Bony Shank Surgery	12	4.5%
Foot/Ankle	Muscle Ankle/Foot Surgery	131	48.9%
	Bony Foot Surgery	63	23.5%

never isolated in the database. These three categories are: rectus femoris, patella lowering and bony shank.

For each lower limb j , a surgery binary code $\mathbf{s}^{(j)} = (s_1^{(j)}, \dots, s_{N_s}^{(j)})^T$ was attributed where

$$s_i^{(j)} = \begin{cases} 1 & \text{if procedure } i \text{ was conducted on limb } j \\ 0 & \text{if procedure } i \text{ was not conducted on limb } j \end{cases} \quad (1)$$

with $i = 1, \dots, N_s$ and T is the transpose operator.

For each lower limb, gait kinematics were described by fifteen timeseries: pelvic tilt, pelvic obliquity, pelvic rotation and hip flexion, hip adduction, hip rotation, knee flexion, ankle dorsiflexion and foot progression angles for both lower limbs.

In this study we considered data recorded by several clinicians, thus measurements of physical examination may have been reported or not depending on the patient and the clinician that performed the exam. For this reason, we consider only 36 parameters that were measured at a minimum rate of 80% in our database. These parameters include information about size and weight; hip, knee and ankle ranges of motion; muscle force; and spasticity (details in supplementary data).

2.2 Preprocessing

Physical examination missing data were imputed using iterative robust model-based imputation (IRMI) [26]. This technique consists in initializing missing values and then iteratively performing linear regressions of each column with respect to the others. We initialized missing values with k-Nearest Neighbor algorithm for $k = 5$ and then computed the median of the nearest neighbors. In this case, the nearest neighbor of any limb is another limb with the closest physical examination values, considering only variables where values are available for both limbs.

Kinematic data were automatically segmented into gait cycles utilizing the high pass algorithm (HPA) [6]. Then gait cycles were resampled and normalized to 51 points (2% of gait cycle) as in [21] and mean gait cycles were computed for each limb. A right and a left

kinematic preoperative gait vectors were composed with the fifteen kinematic signals of both limbs normalized respectively by the right and the left gait cycle. Postoperative gait vectors were solely described by the kinematic data of the considered limb like in [21]. Preoperative kinematic and imputed physical examination data were gathered and projected into a lower-dimensional space using principal component analysis (PCA) [12]. Different PCA projection dimensions from 1 to 63 were tested, keeping between 25% and 99% of the inertia.

For any limb j , the input vector is $\mathbf{x}^{(j)} = (x_1^{(j)}, \dots, x_D^{(j)})^T$, where D is the dimension of the PCA projection. The target output vector for limb j is $\mathbf{y}^{(j)} = (y_1^{(j)}, \dots, y_{N_{out}}^{(j)})^T$, where $N_{out} = 51 \times 9 = 459$ (9 gait signals of 51 points each).

2.3 Ensemble Method for Postoperative Gait Prediction

In order to predict surgery effect on CP gait for a statistically small data set, an ensemble method with non-trainable combination rule was developed. For each surgical procedure, a regression model was generated between postoperative kinematic signals and lower-dimensional preoperative input vectors [figure 1]. Each model M_i (figure 1) was trained with data from limbs that have had surgical procedure i , independently of the concurrent surgeries. For example, if a limb's surgery consisted of the combination of three surgical procedures, this limb's data would be used for training three different models.

The regression models M_i were one-hidden-layer feedforward neural networks (NN) [7]. The number of hidden units were varied from $m = 0, \dots, 10$ to select optimal models, where $m = 0$ corresponds to a multiple linear regression. Multiple linear regression parameters were estimated with Least-Squares [4]. For $m = 1, \dots, 10$ neural networks were optimized using bayesian regularization [14, 13].

For each kinematic angle and PCA projection dimension, a leave-one-out cross validation [4] was performed on each considered model. The model that gave minimal cross-validation error was considered as the optimal model. Thus, each model's architecture was optimized with respect to the input size vector (PCA projection dimension) [section 2.2] and the number of hidden units.

The outputs of the different models were then combined in order to give a unique prediction. If we considered that each kinematic variable varies due to the effect of each surgical procedure, each model output should be weighted by the probability of that output being caused by the surgical procedure that is associated to it. Thus, for a kinematic variable k , the pre-postoperative kinematic variation Δ_k is considered to be given by:

$$\Delta_k = \Delta_k^{s_1} + \dots + \Delta_k^{s_9} \quad (2)$$

Where $\Delta_k^{s_i}$ is the variation triggered by surgical procedure i .

Since in equation (2), only Δ_k is known for every limb in the training base. If we compare the distributions of $h(\Delta_k/s_i = 1)$ and $h(\Delta_k/s_i = 0)$, thus when surgical procedure i is present and absent, we can estimate the probability of Δ_k being caused by s_i and not the other present surgical procedures. Therefore, for each kinematic variable and surgical

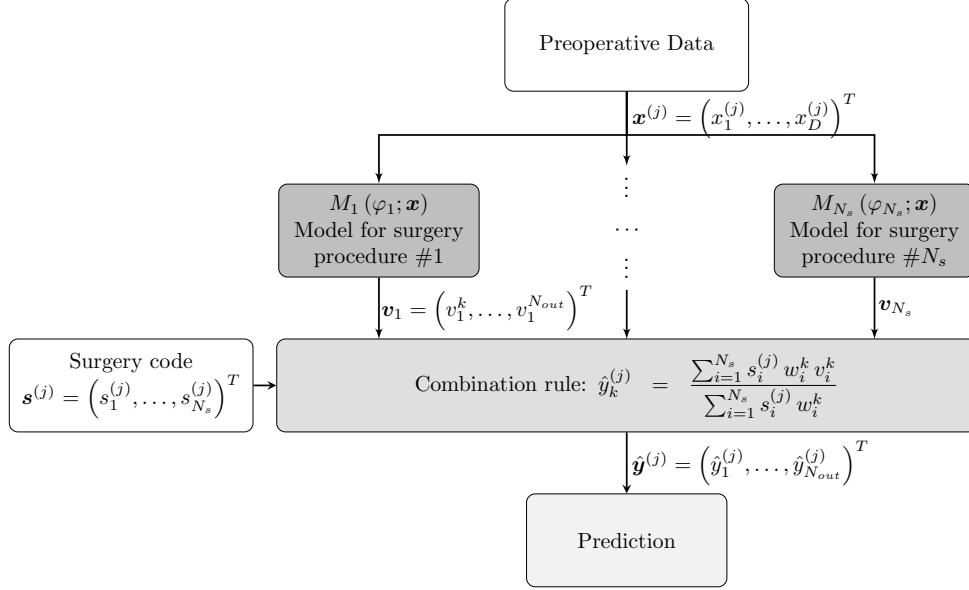


Figure 1: Ensemble method diagram. A model M_i is trained for each surgical procedure. Model output's are then combined by a weighted average of the effects of the considered surgical procedures.

procedure, we performed a two-sample t-test on the kinematic variation of patients that had surgical procedure i and those who did not have that procedure. The p-value of this test is the probability that the kinematic variation was not caused by that particular surgical procedure but by the any other surgical procedure, a mixture of them, noise or other factors that are not considered in this work (*e.g.* rehabilitation).

Each output was weighted depending on the significance of the model (or surgery) on the considered output variable as in equation (3) below:

$$\hat{y}_k^{(j)} = \frac{\sum_{i=1}^{N_s} s_i^{(j)} w_i^k v_i^k}{\sum_{i=1}^{N_s} s_i^{(j)} w_i^k} = \frac{\sum_{i=1}^{N_s} \alpha_i^{j,k} f_i(x^k)}{\sum_{i=1}^{N_s} \alpha_i^{j,k}}, \quad k = 1, \dots, N_{out}. \quad (3)$$

with $s_i^{(j)} \in \{0, 1\}$ [see equation (1)]. Since $s_i^{(j)}$ is binary, equation (3) represents a weighted average of only the outputs of the models associated to the considered surgical procedures. If a limb's surgery is composed of a unique surgical procedure, the prediction will be identical to the output of the model associated to that particular surgical procedure.

The weights w_i^k represent the statistical significance of model i on variable k which is computed as:

$$w_i^k = 1 - p_i^k \quad (4)$$

Where p_i^k is the p-value of the two-group independence t-test [24] when comparing gait

variations of limbs that underwent surgical procedure i against limbs that did not have surgical procedure i .

Summarizing, the proposed method consists in an ensemble method where the ensembles are generated according to some input categories (in this case, surgical procedures categories) and whose outputs are combined by the statistical relevance of each model or ensemble on each output variable. In this case, there were $N_s = 9$ generated ensembles, due to the N_s considered surgery categories.

2.4 Performance assessment

The method was tested using a leave-one-out jackknife procedure [15, 7]: when a patient k was tested, data of both limbs of patient k were removed from the training set. Then the regression was done with data of the other remaining patients and subsequently the data of patient k was tested on the trained model. This process was repeated for all the patients in the database.

Prediction performance was evaluated by the root-mean-square error (RMSE) for every test limb j and gait angle g calculated as:

$$RMSE = \sqrt{\frac{1}{N_{points}} \sum_{i=1}^{N_{points}} (\hat{g}_i - g_i)^2} \quad (5)$$

where \hat{g}_i and g_i are respectively the prediction and the expected resampled postoperative gait signal among the $N_{points} = 51$ points of a time-normalized kinematic angle. These prediction errors were compared to the errors of two naive predictors: “Mean-P” and “NoChange-P” [see table 2]. “Mean-P” is a naive predictor that gives always the same prediction for all limbs, and this prediction corresponds to the mean postoperative gait cycle over all patients in the training database. The median naive predictor was not considered due to similarities on the mean and the median of the considered kinematic data. On the other hand, “NoChange-P” is a naive predictor that indicates the preoperative gait cycle for each limb as the postoperative prediction, which means that the surgery would have no effect on these parameters. Ω is the space of all the limbs in the database and j_{cont} is the contralateral limb associated to limb j (i.e. if limb j is the right limb of patient k , then j_{cont} is the left limb of patient k).

On the other hand, the proposed combination rule was compared to two other typical combination rules: average and median.

3 Results

The weights of each model in the proposed combination rule [equation (3)] vary from 0 to 1 depending on the surgical procedure and the gait variable [figure 2]. Some gait signals are significantly affected by several surgical procedures, *i.e.* pelvic tilt and knee flexion, whilst other signals remained almost unaffected by any surgical procedure, *e.g.* pelvic obliquity.

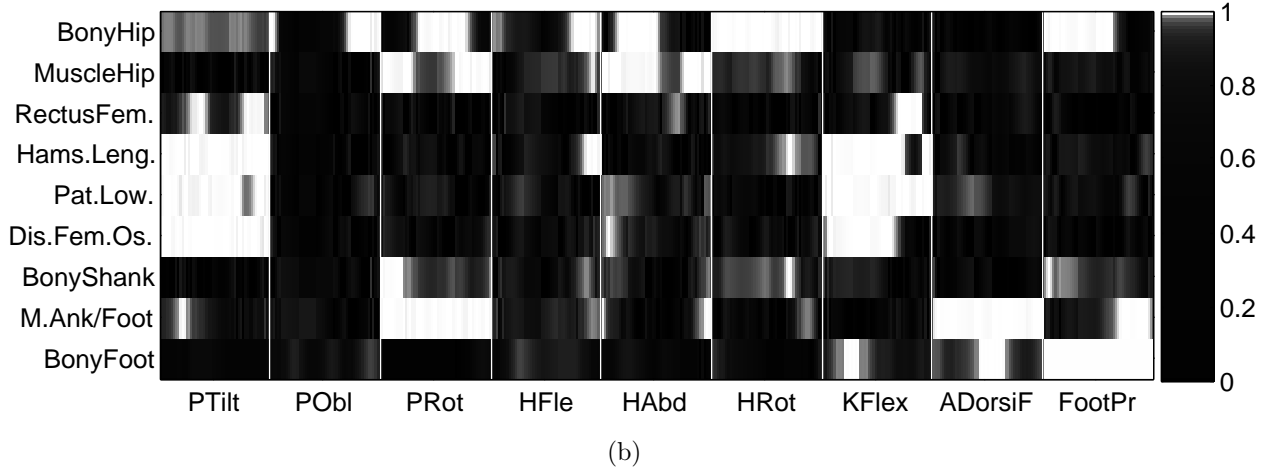
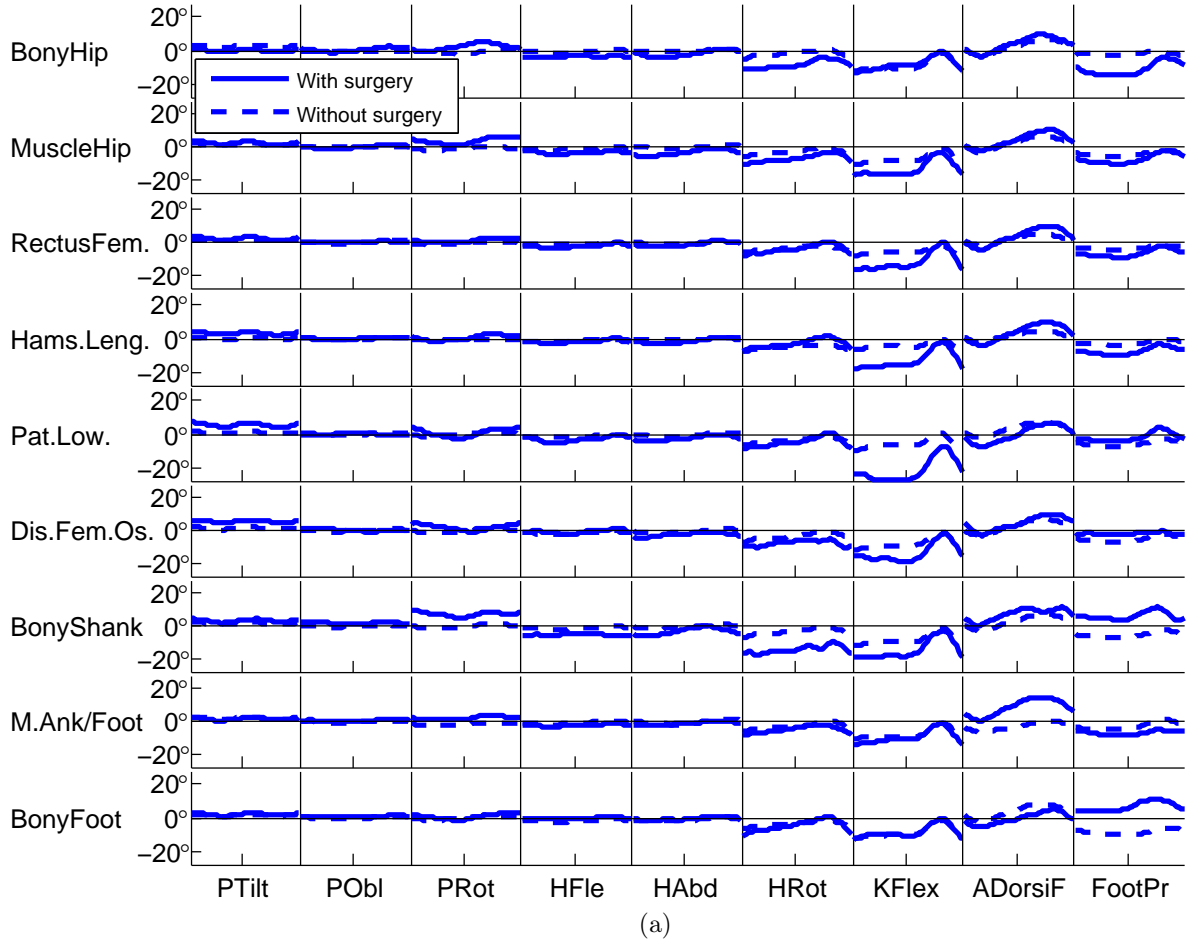


Figure 2: Weights for combination rule. 2a Average variations (Post-Preoperative) of patients that underwent the considered surgical procedure (solid blue) and patients who did not undergo the procedure (dashed blue). 2b Weights of surgical procedures on gait variables. The weights vary from 0 (black) or no significant to 1 (white) maximal significance depending on the p-value of the independence test per surgical procedure over gait variables [see equations (3)- (4)].

Table 2: Naive predictors’ formulae.

Name	Formula
Mean-P	$\hat{Y}^{(j)} = \text{mean} \left(\Omega \setminus \{Y^{(j)}, Y_{cont}^{(j)}\} \right) = \frac{\sum_z Y^z}{n-2}$ <p style="text-align: center;">where $z \in \{1, \dots, n\} \setminus \{j, j_{cont}\}$</p>
NoChange-P	$\hat{Y}^{(j)} = Y_{preop}^{(j)}$ <p style="text-align: center;">where $Y_{preop}^{(j)}$ is the preoperative kinematic vector of limb j</p>

The system outputs the estimated postoperative kinematic signals for the test limb along with their respective preoperative and postoperative (if known) signals [see figure 3]. If the real postoperative curves are known, the prediction RMS errors [equation (5)] per angle are indicated for comparison. The output also includes the standard gait (non-pathological) kinematic curves within their two-standard-deviation band. The estimations of the weighted pre-post variations introduced by each considered surgical procedure are also shown in the output [3].

The validation errors of the optimal models per surgical category vary from 3° (bony foot surgery on pelvic obliquity angle) and 14° (bony shank surgery on foot progression) [table 3] depending on the surgery and the kinematic angle, with a maximal overall error of 8.40° for patella lowering surgery. Optimal PCA projection dimensions vary from 5 to 15, which represent respectively 65.81% and 89.48% of the total variance [table 4]. Most of the optimal regression models are either linear or with one hidden unit, with some others with 4, 5, 9 or 10 hidden units [table 5].

The average prediction errors over all patients of the proposed method vary from 4° (pelvic obliquity and hip adduction) to 10° (hip rotation, knee flexion and foot progression), depending on the kinematic angle [table 6]. Best naive predictors mean performances vary from 4° (pelvic obliquity and hip adduction) and 11° (foot progression) [table 6]. The proposed method give significantly ($p < 0.05$, where p in this case is the p -value of a paired t -test [24]) smaller errors than the naive predictors for all the considered kinematic signals.

Regarding the combination rule, all three tested rules (average, median and the proposed weighted average) give in general equivalent performances [table 7].

4 Discussion

In this work, postoperative lower limb kinematics are predicted with respect to preoperative kinematics, preoperative physical examination and surgery using a non-trainable rule ensemble method. A large number of combinations of surgical procedures in a single-event multilevel surgery (SEMLS) context were considered, for patients with CP that present different gait patterns.

One of the advantages of the proposed ensemble method is its ability to give good estimations for a high number of surgical combinations that may not necessarily be in the training

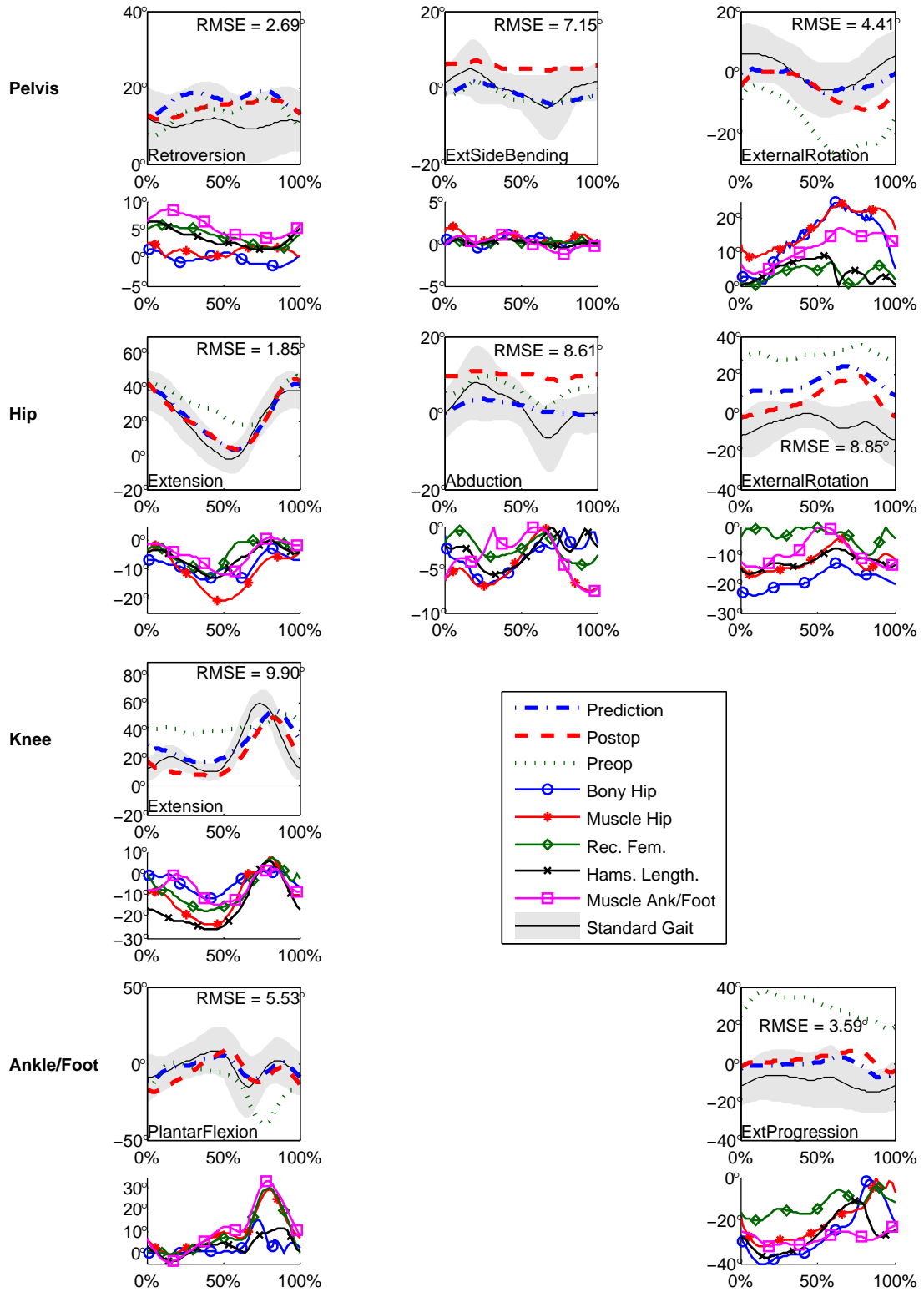


Figure 3: System output example. Predicted postoperative signals (dashed dotted blue), measured postoperative curves (dashed red) and preoperative signals (dotted green), with the estimated variations introduced by each considered surgical procedure.

Table 3: Cross-validation error of constructed models per kinematic angle and surgical category.

	Bony Hip	Muscle Hip	Rectus Femoris	Hams. Length.	Patella Lowering	Distal Fem. Osteotomy	Bony Shank	Muscle Ankle/Foot	Bony Foot
Pel. Tilt	5.41°	7.83°	5.89°	5.62°	7.72°	5.89°	4.17°	5.69°	6.93°
Pel. Obl.	3.55°	3.89°	3.91°	3.95°	4.37°	3.69°	3.56°	3.53°	3.49°
Pel. Rot.	6.91°	6.69°	6.62°	7.01°	8.10°	6.74°	8.89°	7.00°	6.87°
Hip Flex.	6.60°	8.42°	7.00°	6.91°	8.82°	7.02°	7.34°	6.80°	6.85°
Hip Add.	4.00°	4.02°	3.97°	3.95°	4.07°	4.78°	4.78°	4.28°	4.33°
Hip Rot.	9.99°	11.00°	10.13°	9.90°	11.13°	9.84°	7.91°	9.69°	10.41°
Knee Flex.	10.75°	10.51°	10.27°	9.76°	10.66°	9.66°	11.59°	9.89°	10.72°
Ank. DFl.	7.68°	8.58°	7.37°	7.35°	8.28°	5.74°	4.98°	7.81°	8.04°
Foot Prog.	9.38°	10.43°	9.75°	10.98°	12.46°	9.00°	14.14°	9.97°	10.88°
Overall	7.41°	7.93°	7.21°	7.27°	8.40°	6.93°	7.48°	7.18°	7.61°

Table 4: Optimal PCA projection dimensions (P-Dim) per kinematic angle and surgical procedure.

	Bony Hip	Muscle Hip	Rectus Femoris	Hams. Length.	Patella Lowering	Distal Fem. Osteotomy	Bony Shank	Muscle Ankle/Foot	Bony Foot
Pel. Tilt	8	11	7	8	5	13	8	11	11
Pel. Obl.	13	5	11	11	8	8	6	5	5
Pel. Rot.	5	5	10	15	13	6	5	13	5
Hip Flex.	7	11	11	13	5	11	5	12	15
Hip Add.	5	5	11	5	5	5	5	5	5
Hip Rot.	10	7	6	12	13	7	5	10	7
Knee Flex.	13	6	10	5	6	8	8	10	15
Ank. DFl.	12	5	11	5	14	13	8	5	5
Foot Prog.	5	15	6	6	6	7	7	6	5

Table 5: Optimal number of hidden units per kinematic angle and surgical procedure. '0' hidden units corresponds to multiple linear regression.

	Bony Hip	Muscle Hip	Rectus Femoris	Hams. Length.	Patella Lowering	Distal Fem. Osteotomy	Bony Shank	Muscle Ankle/Foot	Bony Foot
Pel. Tilt	0	0	0	0	0	5	0	0	0
Pel. Obl.	0	1	0	0	0	1	10	0	0
Pel. Rot.	1	1	0	0	0	1	1	0	1
Hip Flex.	1	0	0	0	0	0	1	0	0
Hip Add.	1	1	0	1	0	1	9	1	1
Hip Rot.	1	0	0	1	0	0	0	1	1
Knee Flex.	0	0	1	1	1	4	1	0	0
Ank. DFl.	1	0	0	0	0	0	9	0	0
Foot Prog.	1	0	0	0	0	0	9	0	0

Table 6: Prediction performance per predictor and gait signal.

	RMSE ($^{\circ}$) $\mu(\sigma/\sqrt{n})$		
	NoChange-P	Mean-P	Proposed
Pelvic Tilt	6.26 (0.27)*	6.50 (0.26)*	5.55 (0.27)
Pelvic Obl.	4.23 (0.15)*	3.95 (0.12)*	3.69 (0.12)
Pelvic Rot.	7.77 (0.29)*	6.91 (0.23)*	6.53 (0.24)
Hip Flex.	7.32 (0.31)*	7.99 (0.25)*	6.76 (0.24)
Hip Add.	4.87 (0.17)*	4.23 (0.12)*	4.06 (0.12)
Hip Rot.	12.03 (0.50)*	10.06 (0.37)*	9.50 (0.34)
Knee Flex.	13.63 (0.57)*	10.49 (0.34)*	9.71 (0.31)
Ank. DorFl.	13.02 (0.65)*	7.71 (0.27)*	7.51 (0.27)
Foot Prog.	14.65 (0.66)*	10.90 (0.43)*	10.00 (0.43)

*Significant difference with respect to the proposed method ($p < 0.05$)

Table 7: Comparison of performances of combination rules per gait signal.

	RMSE ($^{\circ}$) $\mu(\sigma/\sqrt{n})$		
	Average	Median	Proposed
Pelvic Tilt	5.62 (0.27)	5.65 (0.27)	5.55 (0.27)
Pelvic Obl.	3.68 (0.12)	3.71 (0.11)	3.69 (0.12)
Pelvic Rot.	6.46 (0.24)	6.46 (0.24)	6.53 (0.24)
Hip Flex.	6.75 (4.02)	6.79 (4.05)	6.76 (0.24)
Hip Add.	4.05 (0.12)	4.05 (0.12)	4.06 (0.12)
Hip Rot.	9.49 (0.35)	9.56 (0.35)	9.50 (0.34)
Knee Flex.	9.70 (0.31)	9.72 (0.31)	9.71 (0.31)
Ank. DorFl.	7.41 (0.27)	7.44 (0.27)	7.50 (0.27)
Foot Prog.	9.95 (0.43)	10.03 (0.43)	10.00 (0.43)

database. Other postoperative parameters prediction studies construct a model for every surgical combination and are unable to predict other combinations of the same considered surgical procedures [16]. Nevertheless, the largest validation errors were found in patella lowering (overall) and bony shank (foot progression) surgeries. These are two of the three surgical categories that are never alone in the training database. In addition, bony shank surgery is the category for which there is the fewest number of examples in the database. Even if the proposed method is suitable for small data problems, it still has obvious limitations related to the "lack" of data.

It is interesting to highlight the fact that the majority of the optimal models were whether linear or with few hidden units in the hidden layer of the feedforward neural network. This can be explained by the statistically small data context, where it is difficult to optimize many parameters and therefore less complex models tend to be more suitable to such problems.

On the other hand, certain angles are not significantly affected by certain surgical procedures, which can even simplify the problem. We could, for example, consider only the surgical procedures that have a significant effect over certain angles or variables, and remove the analysis of non-significant surgical procedures. However, this could introduce discontinuities on the gait angles when there are consecutive points and only one is statistically affected by the surgery, which are naturally continuous. This is the reason why we have not used this methodology.

The proposed combination rule gives equivalent performance to the other references rules (mean and median). The advantage of the proposed combination rule is its ability to estimate the significant gait variation introduced by each surgical procedure. This information is useful for surgeons and clinicians in general, who might consider the addition or removal of the considered surgical procedures based on those estimations. The weights of the combination rule are the same for each input vector. Since the statistical effect of surgeries may depend on the preoperative state of the patient, one way of improving the performance of the combination rule would be to compute the statistical influence of surgical procedures according to the input. Thereby, the weights would be the statistical influence of surgical

procedures on an input's neighborhood. However, this approach complicates the problem in two ways: first, the size of the neighborhood has to be established (ideally optimized). Second, there is a risk of finding no limbs with or without certain surgical procedures in a defined neighborhood. In such case, the statistical test cannot be performed and no weight can be computed. To avoid these complications, we have computed the weights considering all the available patients, without taking into account any neighborhood.

Patients that underwent a certain surgical procedure could be seen as a cluster. So when we consider all the surgical procedures, we will get the same number of clusters as surgical procedures. Since a patient can undergo several surgical procedures at the same time, these clusters overlap. In many applications, overlapping clusters are a more natural way to constitute meaningful classes [2]. Moreover, overlapping clusters may allow to find a good compromise between diversity and statistical meaning of clusters on small data sets. In such cases, the proposed ensemble method could be used for any small data set with overlapping clusters: it is only necessary to train a model per each input cluster and then combine their outputs with a weighted average, where the weights are related to the statistical influence of each cluster on the target output.

5 Conclusion

In conclusion, the system provides a tool for showing the most likely surgery outcome to the patient, clinicians and the patient's family. This allows better comprehension and discussion of both the treatment itself and its outcome. This could also motivate the patient to pursue a certain surgical treatment. However, while not externally validated, the usage of the proposed system should be limited to patients that are similar to those in the training base.

After an expert analysis of the probable outcome, it allows to validate or reject a surgical program. It could also allow to modify a surgical program (add or remove certain surgical procedures) after studying the global prediction and the probable contribution of each surgical procedure. This could be used as a decision-making tool for clinicians, especially orthopaedic surgeons.

Acknowledgements

This study is part of the SIM-PC2 project promoted by the Fondation Ellen Poidatz and supported by Fondation Ellen Poidatz, Fondation Bettencourt Schueller and Region Île de France. The authors would like to thank UNAM medical and technical staff of Fondation Ellen Poidatz for useful discussions and for recording all the data used in this work.

References

- [1] Richard W. Baker. Measuring Walking: A Handbook of Clinical Gait Analysis. MacKeith Press, London, 1 edition, 2013.
- [2] Arindam Banerjee, Chase Krumpelman, Joydeep Ghosh, Sugato Basu, and Raymond J. Mooney. Model-based Overlapping Clustering. In Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining, KDD '05, pages 532–537, New York, NY, USA, 2005. ACM.
- [3] C. M. Bishop. Model-based machine learning. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 371(1984), December 2012.
- [4] Christopher M. Bishop. Pattern Recognition And Machine Learning. Springer-Verlag New York Inc., New York, 1st ed. 2006. corr. 2nd printing 2011 edition, 2006.
- [5] Leo Breiman. Bagging predictors. Machine learning, 24(2):123–140, 1996.
- [6] Eric Desailly, Yepremian Daniel, Philippe Sardain, and Patrick Lacouture. Foot contact event detection using kinematic data in cerebral palsy children and normal adults gait. Gait & Posture, 29(1):76–80, January 2009.
- [7] Richard O. Duda, Peter E. Hart, and David G. Stork. Pattern Classification. Wiley-Blackwell, New York, 2nd edition edition, November 2000.
- [8] James R. Gage, Michael H. Schwartz, Steven E. Koop, and Tom F. Novacheck. The Identification and Treatment of Gait Problems in Cerebral Palsy. MacKeith Press, London, 2nd edition edition, 2009.
- [9] O. A. Galarraga C., V. Vigneron, B. Dorizzi, N. Khouri, and E. Desailly. Predicting postoperative knee flexion during gait of cerebral palsy children. Computer Methods in Biomechanics and Biomedical Engineering, pages 1–2, August 2015.
- [10] Omar A. Galarraga C., Vincent Vigneron, Bernadette Dorizzi, Néjib Khouri, and Eric Desailly. Estimation of Postoperative Knee Flexion at Initial Contact of Cerebral Palsy Children using Neural Networks. In Proceedings of the International Conference on Pattern Recognition Applications and Methods, pages 338–342, Lisbon, 2015. SCITEPRESS - Science and and Technology Publications.
- [11] Robert A. Jacobs, Michael I. Jordan, Steven J. Nowlan, and Geoffrey E. Hinton. Adaptive Mixtures of Local Experts. Neural Computation, 3(1):79–87, February 1991.
- [12] I.T. Jolliffe. Principal Component Analysis. Springer Series in Statistics. Springer-Verlag, New York, 2002.
- [13] David J. C. MacKay. Bayesian Interpolation. Neural Computation, 4:415–447, 1991.

- [14] David JC MacKay. The evidence for neural networks. In Maximum Entropy and Bayesian Methods, pages 165–183. Springer, 1992.
- [15] Rupert G. Miller. The jackknife-a review. Biometrika, 61(1):1–15, April 1974.
- [16] Timothy A. Niiler, James G. Richards, and Freeman Miller. Concurrent surgeries are a factor in predicting success of rectus transfer outcomes. Gait & Posture, 26(1):76–81, June 2007.
- [17] Timothy A. Niiler, James G. Richards, Freeman Miller, J. Q. Sun, and Patrick Castagno. Reliability of predictions of post-operative gait in rectus transfer patients using FFT neural networks. Gait & Posture, 9:107, January 1999.
- [18] Robi Polikar. Ensemble based systems in decision making. Circuits and Systems Magazine, IEEE, 6(3):21–45, 2006.
- [19] Jeffrey A. Reinbolt, Melanie D. Fox, Michael H. Schwartz, and Scott L. Delp. Predicting outcomes of rectus femoris transfer surgery. Gait & Posture, 30(1):100–105, July 2009.
- [20] Robert E. Schapire. The strength of weak learnability. Machine learning, 5(2):197–227, 1990.
- [21] Michael H. Schwartz and Adam Rozumalski. The gait deviation index: A new comprehensive index of gait pathology. Gait & Posture, 28(3):351–357, October 2008.
- [22] Michael H. Schwartz, Adam Rozumalski, Walter Truong, and Tom F. Novacheck. Predicting the outcome of intramuscular psoas lengthening in children with cerebral palsy using preoperative gait data and the random forest algorithm. Gait & Posture, 37(4):473–479, April 2013.
- [23] A. Sebsadji, N. Khouri, K. Djemal, D. Yepremian, F. Hareb, P. Hoppenot, and E. Desailly. Description and classification of the effect of hamstrings lengthening in cerebral palsy children multi-site surgery. Computer Methods in Biomechanics and Biomedical Engineering, 15(sup1):177–179, September 2012.
- [24] Stuart. Kendall’s Advanced Theory of Statistics: Distribution Theory. Wiley-Blackwell, Chichester, West Sussex, 6th edition edition, 1994.
- [25] Karen Sullivan, James Richards, Freeman Miller, Patrick Castagno, and Nancy Lennon. Predicting the outcome of surgery for children with cerebral palsy using pre-operative gait analysis. Gait & Posture, 3(2):92, 1995.
- [26] M. Templ, A. Kowarik, and P. Filzmoser. EM-based stepwise regression imputation using standard and robust methods. Research report cs-2010-3, Department of Statistics and Probability Theory, Vienna University of Technology, 2010.
- [27] David H. Wolpert. Stacked Generalization. Neural Networks, 5:241–259, 1992.

Chapter 8

Comparison of the Methods

Contents

8.1	Regression using Fourier gait parameters	109
8.2	Performance comparison	109
8.3	General Discussion	111

The performance of the predictions systems in the previous chapters were evaluated according to gait parameters variability in CP (Klejman, Andrysek, Dupuis & Wright, 2010), gait deviation severity (Baker et al., 2009) or comparing them to the average surgery outcome. In this chapter we will compare these prediction system between them and to another proposed system based on Fourier approximation [subsection 4.1.1], as well as some results from the literature (Niiler et al., 1999; Niiler, 2001).

8.1 Regression using Fourier gait parameters

To compare to the previous methods, we have conducted another experiment for predicting postoperative CP gait based on the trigonometric approximation described in section 4.1.

Kinematic angles were considered separately. For each kinematic angle, preoperative vector consisted of the preoperative Fourier coefficients, plus the 9-bit surgery code. Hence, the dimension of the input was $K_{in} = K_{fourier} + N_s = K_{fourier} + N_s$, where $K_{fourier}$ depends on the order m_{fou} of the series. The order of the series was chosen for each kinematic angle such that the average approximation error was smaller than 5° [see Table 4.3]. The target output consisted of the postoperative Fourier coefficients for the angle in question.

The regressions were performed utilizing feedforward neural networks with one hidden layer and $m = 3$ hidden units. The learning process was done by Bayesian regularization (MacKay, 1991, 1992).

To evaluate the prediction error, the estimated postoperative coefficients were used to reconstruct the kinematic curve using Equation 4.8. Then, the $RMSE$ between the postoperative signals and the predicted reconstructed signals were computed.

Similarly to the previous experiments, test have been performed by the same jackknife leave-one-out procedure (Miller, 1974).

8.2 Performance comparison

In this section we compare the performances of the postoperative gait prediction systems presented in sections 6.2, 7.2 and 8.1. The naive predictor of the mean was again included in the

comparison, as well the performances reported in (Niiler et al., 1999) and (Niiler, 2001). Since the database has been growing since the beginning of the first experiments with each patient that comes for the postoperative CGA, the mentioned experiments were conducted with different numbers of patients. For better comparison, we have conducted the experiments again, with the last version of the database composed of data of 134 patients [see chapter 3]. This is the same database that was used in sections 7.2 and 8.1. Only prediction errors on operated limbs were considered [see Table 8.1].

Table 8.1 – Performance comparison of the prediction methods. Average μ and standard error σ/\sqrt{n} of the prediction RMSE on operated limbs. The first column indicates the kinematic signal. VS+NN stands for Variable Selection and Neural Networks (experiment in section 5.2). FOU corresponds to experiment in section 8.1. PCA+MLR (Principal Component Analysis + Multiple Linear Regression) corresponds to the experiment in section 6.2. PCA+Ens (PCA+Ensemble learning) corresponds to the the experiment in section 7.2. Nii99 and Nii01 correspond respectively to (Niiler et al., 1999) and (Niiler, 2001). Mean-P corresponds to the naive predictor of the postoperative mean.

	RMSE ($^{\circ}$) $\mu(\sigma/\sqrt{n})$						
	Mean-P	VS+NN	FOU	PCA+MLR	PCA+Ens	Nii99	Nii01
PTilt	6.5 (0.3)	5.3 (0.4) ^{† §}	5.9 (0.2)*	5.7 (0.3)*	5.6 (0.3)*	-	-
PObl	4.0 (0.1)	-	4.2 (0.1)*	3.8 (0.1)*	3.7 (0.1)*	-	-
PRot	6.9 (0.2)	-	7.4 (0.2)*	6.7 (0.2)*	6.5 (0.2)*	-	-
HFl	8.0 (0.3)	-	14.3 (0.2)*	7.1 (0.3)*	6.8 (0.2)*	8.1 [§]	6.2 [§]
HAdd	4.2 (0.1)	-	4.4 (0.1)*	4.2 (0.1)	4.1 (0.1)*	-	-
HRot	10.1 (0.4)	-	11.2 (0.4)*	9.9 (0.4)	9.5 (0.3)*	-	-
KFl	10.5 (0.3)	9.3 (0.6) ^{† §}	16.6 (0.3)*	10.2 (0.3)*	9.7 (0.3)*	9.7 [§]	9.2 [§]
ADorF	7.7 (0.3)	-	9.7 (0.3)*	7.9 (0.3)	7.5 (0.3)*	6.7 [§]	-
FPProg	10.9 (0.4)	-	10.8 (0.4)	10.2 (0.4)*	10.0 (0.4)*	-	-

*Significant difference with respect to the naive Mean-P predictor ($p < 0.05$).

[†] Only at initial contact and not the complete signal.

[§] The independence test was not performed.

From Table 8.1, it is worth to comment the following facts:

- All the methods except FOU have prediction errors smaller or equivalent to Mean-P.
- FOU has errors greater than those of Mean-P for all kinematic angles, except for pelvic tilt and foot progression. From these two kinematic angles, only pelvic tilt has significant difference with respect to Mean-P.
- Apart from FOU, only Nii99 at hip flexion and PCA+MLR at ankle dorsiflexion have errors greater than Mean-P.
- There is no significant different between errors of PCA+MLR and Mean-P on hip adduction, hip rotation and ankle dorsiflexion.
- Nii01 gives the best performance for hip flexion, followed by PCA+Ens.
- Nii01 gives the smallest error for knee flexion, followed by VS+NN (at initial contact) and then PCA+Ens.
- VS+NN gives the best performance pelvic tilt (at initial contact), followed by PCA+Ens.
- Nii99 gives the smallest error for ankle dorsiflexion.

- Errors of PCA+Ens are slightly smaller than errors of PCA+MLR for all the kinematic angles.
- PCA+Ens gives the best performances for pelvic obliquity, pelvic rotation, hip adduction, hip rotation, hip rotation and foot progression.
- All the best performance are only slightly better than Mean-P (less than 2° of difference).
- PCA+Ens has significant difference with respect to Mean-P for all the kinematic angles.

8.3 General Discussion

It is complicated to compare the performances of the proposed methods to Nii99 and Nii01 because their results were obtained for a single surgical procedure (rectus femoris transfer), only one gait pattern (stiff knee) and with a different database. However, average prediction RMSE are on the same magnitude, especially for knee flexion, which is the kinematic angle that is affected by rectus femoris transfer [see section 3.5]. In any case, all the proposed methods consider many more surgical procedures and gait patterns.

It is also difficult to compare the performance of VS+NN to the other proposed methods, because VS+NN only predicts gait variables at initial contact and not the complete kinematic curves. Moreover, VS+NN does not consider preoperative kinematic variables at other instant than initial contact. Nevertheless, VS+NN performance is somewhat similar to PCA+MLR and PCA+Ens and slightly better (with differences of less than 1°).

On the other hand, the naive predictor Mean-P performs better than FOU. The FOU system is the proposed method that is methodologically the most similar to Nii99 and Nii01, because they all utilize Fourier series of the gait signals for feature extraction and neural networks for regression. However, in FOU the surgery code is also part of the input, which is not the case in Nii99 and Nii01, where a single type of surgery was considered (rectus femoris transfer). FOU is also the unique proposed method that does not consider physical examination data and contralateral preoperative kinematics.

Of the proposed methods that predicts complete gait cycles, PCA+MLR and PCA+Ens perform overall better than the naive predictor Mean-P. However, PCA+MLR performance is not significantly different from Mean-P for hip adduction, hip rotation and ankle dorsiflexion. Since the surgery is part of the input of a general model, PCA+MLR does not give estimation of the contribution of each surgical procedure. In fact, there is no guarantee that PCA+MLR learns to differentiate the effect of each surgical procedure rather than a global effect of the surgery. However, PCA+MLR gives the linear regression parametric confidence intervals, which represent a valuable information about the prediction variability.

Conversely, all PCA+Ens performances are smaller and significantly different to Mean-P. In PCA+Ens, confidence intervals cannot be directly computed as in PCA+MLR because of the mixture of linear and nonlinear models, and the fusion rule of the models. However, the extra information in PCA+Ens is represented by the estimations of the contributions of each surgical procedure. This information is also valuable to the clinician in order to consider alternative surgical combinations.

Chapter 9

Conclusions and Perspectives

In this thesis, the effect of orthopaedic surgery on cerebral palsy gait is simulated by applying supervised machine learning to biomechanical signals. The simulations of surgery effect are used in order to predict postoperative gait kinematics. Different methods for predicting postoperative gait were considered: multiple linear regression, nonlinear regression with neural networks and ensemble learning. In addition, multiple feature extraction and dimensionality reduction methods were utilized, including curve fitting with B-Splines and Fourier series, a local variable selection technique, and principal component analysis. Regression analyses were performed between postoperative kinematic variables and preoperative kinematics, preoperative physical examination data and the surgical combination that the surgeons performed on the patients. The study comprised five principal stages:

1. Organization of the preoperative and postoperative clinical gait analysis (CGA) database
2. Data conditioning and analysis
3. Dimensionality reduction and feature extraction
4. Regression analyses of postoperative data with respect to preoperative data given the surgery
5. Evaluation and comparison of the tested methods

The utilized database is one of the largest reported preoperative and postoperative clinical gait analyses (CGA) database, with data of 134 operated patients on the last update. The size of this database is larger than the number of considered patients in all cited previous works that predict the effect of surgery [see section 2.7], except for (Schwartz et al., 2013). It is also larger than other pre-postoperative CGA series, such as (Rutz, Donath, Tirosh, Graham & Baker, 2013) that contains data of 121 patients. In addition, a web-like interface was developed to facilitate the input of new data into this database [see for example Appendix 9.2]. For this reason, this database is constantly growing and may be even larger at the moment this document is being written.

The prior data analysis comprised a description of cerebral palsy gait, surgery outcome, as well as statistical evaluations of the effect of each considered surgical procedure and the criteria for selecting surgeries.

All the proposed methods can deal with missing data, by the integration of an imputation stage, prior to dimensionality reduction. The methods have been adapted to deal with statistical small data problems thanks to the construction of robust models. Robustness was obtained by the

usage of simplified models, by combining models (ensemble learning), and by introducing Bayesian regularization into neural network training.

In our model, limb postoperative gait signals are considered as functions of bilateral preoperative gait signals, bilateral preoperative physical examination data and the limb's surgery. We propose that the surgery effect on gait is in fact a combination of individual surgical procedures. The models combine continuous data (gait signals) with discrete and continuous physiological data (physical examination) and with qualitative surgical data. The surgical data was either considered as direct input or to generate ensembles and thereby surgical procedure submodels. However, we believe that quantitative surgical data would improve the performance of the models. For instance, a femoral proximal osteotomy could be modeled as the degree of rotation of the femur or hamstring lengthening could be modeled as a quantified variation of the popliteal angle.

On the other hand, the given predictions are suitable for 3-D animation because they model all the joints of the lower limb. The predicted signals can be continuous and periodical, via the proposed B-Splines or Fourier approximations.

This work represents the first time that the effect of surgery on cerebral palsy gait is simulated to predict numerous postoperative quantitative gait variables for a large number of surgical procedure combinations and a large variety of gait patterns. Previous works focus on qualitatively predict surgery outcome (good or not-good outcomes) or few gait variables (*i.e.* walking speed, knee range of motion, knee flexion cycle, etc.) and are limited to a single surgical procedure (not combinations) or a specific gait pattern (*e.g.* crouch gait or stiff knee).

Apart from some naive prediction systems and regression of Fourier coefficients, three principal postoperative gait prediction methods were described:

- Prediction of knee flexion and pelvic tilt at initial contact using Gram-Schmidt variable selection with probe, and nonlinear regression with multilayer perceptrons.
- Prediction of kinematic signals using principal component analysis (PCA) and multiple linear regression.
- Prediction of kinematic signals using PCA and ensemble learning with linear and nonlinear regression (feedforward neural networks).

The performances of the proposed methods have been compared to the naive prediction systems, to gait parameters variability in cerebral palsy and to gait deviation severity. Prediction errors are in general smaller than gait variables variability and to errors of naive predictor. Moreover, the performance is independent of the preoperative gait deviation severity, meaning that the system is able to make predictions for any type of patient. In addition, the prediction performance is comparable to previous works that predict certain postoperative kinematic signals for a single surgical procedures. Therefore, the proposed methods can be considered as a successful generalization of several previous works.

The principal contributions of this study could be resumed with following points:

- It gives a quantified prediction (estimated postoperative kinematics)
- It simultaneously considers several surgical procedures (and their combinations)
- It considers multiple gait patterns in cerebral palsy
- It considers physical examination data and contralateral kinematics

- The size of the database is larger than in most of the previous works
- It generalizes works that predict a single surgical procedure or for a single gait pattern or estimates less kinematic variables with equivalent performance

The best proposed method is the one that utilizes Principal Component Regression (PCR) and ensemble learning with linear and nonlinear regression by feedforward neural networks. The advantage of this combination with respect to the previous systems that uses PCR is that errors are slightly higher and it estimates the probable contribution of each surgical procedure. On the other hand, the advantage of the multiple linear regression is its simplicity and the presentation of the confidence intervals, which give more information about the prediction variability. Introducing parametric confidence interval computation in the ensemble method is more complex, because it involves the computation of confidence intervals of each submodel and then combining them in an optimal unknown way.

It is worth to mention that when considering linear and nonlinear regression, the optimal models were either linear or nonlinear with small numbers of parameters (small numbers of hidden neurons). This suggests, following the Occam's Razor, that simplified (or not too complex) models are more suitable for statistical small data problems than sophisticated models that learn better but need large amount of data to correctly generalize.

Regarding the representation of gait signals, best results were obtained with global approximation of CGA data by PCA. However, the performances of the experiment that uses local data (discrete CGA data) by variable selection and nonlinear regression are equivalent to the best proposed method. This suggests that a statistical method that introduces time, *e.g.* Elman networks, time-delay networks, temporal Boltzmann machines (Taylor, 2009) or even Kalman filters (Haykin, 2001; Grewal & Andrews, 2001), may give similar or slightly better prediction performances. The use of local data selection for a large number of outputs is complicated, because the variable selection should be adapted to each output. Thus, for each output, the selected input is in general different. This introduces a problem to evaluate the influence of the input on global results.

The most difficult kinematic signals to predict are pelvic rotation, knee flexion and foot progression. The first and third could be explained by the variability and precision of CGA in transverse plane angles (Ramakrishnan & Kadaba, 1991). The second could be explained because many of the considered surgical procedures significantly affects this signal.

This study is of both high scientific and societal interest, given the medical application related to surgical treatment of children with neurological gait troubles. In the first place, the predictive models show the most likely surgery outcome in the form of kinematic curves that clinicians and gait analysts are used to examine. These curves can be used to animate a 3-D avatar to illustrate the probable resulting gait to people that are not used to analyze kinematic signals. This can facilitate the comprehension of the likely outcome for both the clinician and the patient. Without such preview of the most likely outcome, it is easier that the patient misunderstands the qualitative explanation of the outcome by the medical team. This is indeed a major problem nowadays.

The gait outcome prediction given the surgery can help the surgeon to decide between different possible surgical combinations. For example, when surgeons are considering to add a supplementary surgical procedure to an original surgical program, they can test the model with and without the procedure in question and make a final decision by comparing the likely outcomes. The preview of the most likely postoperative gait will also help the medical team to discuss the treatment, because professional opinions on the likely outcome might differ between teammates (*i.e.* two different surgeons in a same team).

Moreover, if all possible surgical combinations are tested and the accuracy of the test is uniformly decent, the system would be able to suggest an optimal surgical combination. However, the current accuracy of the best model makes difficult to guarantee good predictions for all the patients and all the surgical combinations. Hence, if the model is used to find the optimal treatment, the user should interpret the results with caution and confront them to opinions of experienced surgeons.

In addition, if an adequate surgery is tested, the prediction represents an estimation of the functional improvement capacity of the patient, which is in general different. This could be used to motivate the patient to pursue a certain treatment. This psychological motivation would improve patient involvement in the recovery and the rehabilitation, which would in turn positively affect the actual outcome. In particular, if the optimal treatment is known, we could have an estimation of the maximal improvement capacity of the patient.

Societal interest of the proposed systems are resumed in the following points:

- It shows the most likely outcome
- It facilitates patient-clinician communication
- It gives better understanding of treatment outcome to the patient and the patient's entourage
- it may help the surgeon to validate, reject or modify a surgery plan.
- If a higher accuracy is guaranteed, it helps to find the optimal surgery
- It estimates the functional improvement capacity of each patient (if the proposed surgery is adequate)

Globally, this work constitutes a base for predicting postoperative gait with performances limited by the complexity of the problem and the size of the database. Even though the road is still long, this study represents a step towards optimal surgical treatment assessment.

9.1 Limitations

The principal limitation of this study is the small number of samples with respect to the number of parameters to estimate, or the quantity of operated patients with preoperative and postoperative CGA with respect to the number of kinematic variables, physical examination data and surgical procedures that we consider. Despite having used one of the biggest pre-postoperative CGA database, this limitation leads us to a statistical small data context. In such context, the risk of over-fitting is really high, affecting the generalization capacity of the trained models. For this reason, the choice of dimensionality reduction techniques is crucial, as well as the choice of robust regression algorithms.

In practice, this drawback is evident when analyzing the moderate prediction accuracy over test patients. For example, prediction errors are slightly (some degrees) smaller than best naive predictors. This prevents the usage of the proposed systems in seeking the optimal surgical treatment. For example, it is difficult that surgeons change their opinions on the choice of the treatment based on predicted outcome differences of some degrees. In none of the cases this implies that such prediction is irrelevant. It only means that, in some cases, the medical application may be too sensible (risk of surgery) for the model accuracy. When this is the case, the prediction gives nevertheless a good preview of the most likely outcome.

Another limitation is that the proposed models do not consider other CGA data that are commonly used for treatment assessment, such as kinetics or electromyography, among others. Nevertheless, the inclusion of such data would introduce a major missing data problem, because these data are not always available. For example, when the patient uses a technical walking aid (*i.e.* canes or K-Walker), kinetic data cannot be reliably calculated. In addition, the proposed limb-based models do not consider contralateral surgery, which intuitively affects both limbs kinematics. The problem is that it is too difficult to model the effect of compensations between ipsilateral and contralateral limbs, hence the relation between a limb's kinematics and its contralateral is not direct. Given our database, it is practically impossible to find such complex statistical relation.

On the other hand, given the lack of external validation, the usage of the proposed methods should be limited to patients and surgeries that are somewhat similar to those in the considered database.

Since all methods are purely statistical, there are no mechanical constraints on the gait prediction, which may result in physically unfeasible solutions. However, given the intrinsic physical and physiological constraints of the training data, the risk of such solutions might be reduced.

9.2 Perspectives

This thesis presents a basis for gait outcome prediction for orthopaedic surgery in cerebral palsy. Even though the prediction performance is moderate, the results are encouraging for developing this research problem. Since it is a highly complex mathematical problem, in this *big data* era, the first action should be to enlarge databases with preoperative and postoperative CGA data, as well as surgical data. The main issue of constructing and maintaining these databases is the missing of the postoperative exam, because surgeons usually demand the CGA to select the surgical treatment and not always to evaluate its outcome. On the other hand, we would like to encourage the laboratories around the world that have confronted this research question to unify and share their data. However, data sharing would imply to circumvent legal issues about medical and personal data (even if data is anonymized), as well as to find some conventions on the data CGA and surgical data (kinematic model, data representation, preprocessing, etc.).

With big and unified databases, the methods that were used in this report could be enforced by more sophisticated regressors, such as deep learning techniques (LeCun, Bengio & Hinton, 2015) or other time-series models, *e.g.* restricted Boltzmann machines (Taylor, Hinton & Roweis, 2006), that would find much more complex representations of the CGA and surgical data, but are significantly limited in statistical small data problems. Moreover, with more data, trainable combination rule ensemble methods, *i.e.* mixture of experts (Jacobs et al., 1991) or stacked generalization (Wolpert, 1992), could be tested in order to find an optimal model fusion function. Another important point for the future of this subject is the integration of mechanical and statistical models. In this work, as well as the cited related works, models are purely mechanical or statistical. Mechanical models incorporate temporal and physical information to make more realistic simulations, but cannot estimate many parameters in subject-specific models in order to accurately simulate pathological gait. The mixture of both types of models, as planned in the Sim-PC², would conduct to most likely and physically realistic predictions. For example, we could consider a combination between the ensemble learning model presented in this work with the integration of a gait synthesis model as in (Santos, Benamar, Bidaud & Desailly, 2016).

The application of machine learning techniques on clinical gait analysis data is not widely spread in the scientific field. Hopefully this work will increase the interest of the machine learning community in this medical application. Machine learning is more widely studied in other medical

applications, such as electrocardiography (Sameni, 2008; Dubois, Maison-Blanche, Quenet & Dreyfus, 2007) or medical imaging (*i.e.* magnetic resonance or tomography (Kodewitz et al., 2013)), among others. Thereby, hopefully these methods will be improved to crucially help society with this sensible medical application.

References

- Abdel-Aziz, Y. & Karara, H. (1971). Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry. In *Proceedings of the Symposium on Close-Range Photogrammetry*, (pp. 1–18).
- Accardo, P. (2007). *Capute & Accardo's Neurodevelopmental Disabilities in Infancy and Childhood: Neurodevelopmental Diagnosis and Treatment* (3 edition ed.). Baltimore: Brookes Publishing.
- Akaike, H. (1998). Information Theory and an Extension of the Maximum Likelihood Principle. In E. Parzen, K. Tanabe, & G. Kitagawa (Eds.), *Selected Papers of Hirotugu Akaike*, Springer Series in Statistics (pp. 199–213). Springer New York.
- Allano, L., Dorizzi, B., & Garcia-Salicetti, S. (2012). A new protocol for multi-biometric systems' evaluation maintaining the dependencies between biometric scores. *Pattern Recognition*, 45(1), 119–127.
- Allison, P. D. (2009). Missing Data. In R. E. Millsap & A. Maydeu-Olivares (Eds.), *The SAGE Handbook of Quantitative Methods in Psychology* (pp. 72–89). Thousand Oaks, CA: Sage Publications Inc.
- Armand, S., Watelain, E., Mercier, M., Lensel, G., & Lepoutre, F.-X. (2006). Identification and classification of toe-walkers based on ankle kinematics, using a data-mining method. *Gait & Posture*, 23(2), 240–248.
- Arnold, A. S., Liu, M. Q., Schwartz, M. H., Öunpuu, S., & Delp, S. L. (2006). The role of estimating muscle-tendon lengths and velocities of the hamstrings in the evaluation and treatment of crouch gait. *Gait & Posture*, 23(3), 273–281.
- Atkeson, C. G. & Schaal, S. (1995). Memory-based neural networks for robot learning. *Neurocomputing*, 9(3), 243–269.
- Baker, R., McGinley, J. L., Schwartz, M. H., Beynon, S., Rozumalski, A., Graham, H. K., & Tirosh, O. (2009). The gait profile score and movement analysis profile. *Gait & Posture*, 30(3), 265–269.
- Baker, R. W. (2013). *Measuring Walking: A Handbook of Clinical Gait Analysis* (1 ed.). London: MacKeith Press.
- Bishop, C. M. (2006). *Pattern Recognition And Machine Learning* (2nd ed.). New York: Springer-Verlag.
- Bishop, C. M. (2012). Model-based machine learning. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1984).
- Bonnefoy-Mazure, A., Sagawa, Y., Pomero, V., Lascombes, P., De Coulon, G., & Armand, S. (2016). Are clinical parameters sufficient to model gait patterns in patients with cerebral palsy using a multilinear approach? *Computer Methods in Biomechanics and Biomedical Engineering*, 19(7), 800–806.
- Boudaoud, S. (2006). *Analyse de la variabilité de forme des signaux : Application aux signaux électrophysiologiques*. PhD thesis, Université de Nice-Sophia Antipolis.
- Bouisset (1999). *Muscles, posture et mouvement*. Paris: Editions Hermann.

- Boulgouris, N. V., Hatzinakos, D., & Plataniotis, K. N. (2005). Gait recognition: a challenging signal processing technology for biometric identification. *Signal Processing Magazine, IEEE*, 22(6), 78–90.
- Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2), 123–140.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5–32.
- Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and Regression Trees* (1 edition ed.). New York: Chapman and Hall/CRC.
- Burges, C. J. (1998). A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2), 121–167.
- Cantoni, E. & Ronchetti, E. (2001). Robust Inference for Generalized Linear Models. *Journal of the American Statistical Association*, 96(455), 1022–1030.
- Carriero, A., Zavatsky, A., Stebbins, J., Theologis, T., & Shefelbine, S. J. (2009). Determination of gait patterns in children with spastic diplegic cerebral palsy using principal components. *Gait & Posture*, 29(1), 71–75.
- Chen, S., Billings, S. A., & Luo, W. (1989). Orthogonal least squares methods and their application to non-linear system identification. *International Journal of Control*, 50(5), 1873–1896.
- Chow, J. W. & Knudson, D. V. (2011). Use of deterministic models in sports and exercise biomechanics research. *Sports Biomechanics*, 10(3), 219–233.
- Cleland, J., Koppenhaver, S., & Netter, F. H. (2011). *Netter's orthopaedic clinical examination: an evidence-based approach*. Philadelphia, Pa.: Saunders/Elsevier.
- Davis III, R. B., Öunpuu, S., Tyburski, D., & Gage, J. R. (1991). A gait analysis data collection and reduction technique. *Human Movement Science*, 10(5), 575–587.
- De Boor, C. (2001). *A Practical Guide to Splines*, volume 27 of *Applied Mathematical Sciences*. Springer.
- DeLuca, P. A., Ounpuu, S., Davis, R. B., & Walsh, J. H. (1998). Effect of hamstring and psoas lengthening on pelvic tilt in patients with spastic diplegic cerebral palsy. *Journal of Pediatric Orthopedics*, 18(6), 712–718.
- Desailly, E. (2008). *Analyse biomécanique 3D de la marche de l'enfant déficient moteur : Modélisation segmentaire et modélisation musculo-squelettique*. PhD thesis, Université de Poitiers.
- Desailly, E., Daniel, Y., Sardain, P., & Lacouture, P. (2009). Foot contact event detection using kinematic data in cerebral palsy children and normal adults gait. *Gait & Posture*, 29(1), 76–80.
- Desailly, E., Sebsadji, A., Yepremian, D., Djemal, K., Hoppenot, P., & Khouri, N. (2013). Supervised classification of the effect of hamstrings lengthening in cerebral palsy children after single event multilevel surgery. *Gait & Posture*, 38, Supplement 1, S39.
- Desailly, E., Sebsadji, A., Yepremian, D., Hareb, F., & Khouri, N. (2012). Simulation of muscle retraction in cerebral palsy. Validation of a decision support system for surgical lengthening of contractured muscles. *Computer Methods in Biomechanics and Biomedical Engineering*, 15(S1), 263–265.
- Desailly, E., Yepremian, D., Khouri, N., Hareb, F., Lejeune, L., Bouchakour, D., Sardain, P., & Lacouture, P. (2009). Simulation of the muscles maximum extensibility by musculoskeletal modelling. Contribution to the differential diagnoses conducing to the decision of a surgical lengthening in cerebral palsy children. *Gait & Posture*, 30, S136.
- Desloovere, K., Molenaers, G., Feys, H., Huenaerts, C., Callewaert, B., & Walle, P. V. d. (2006). Do dynamic and static clinical measurements correlate with gait analysis parameters in children with cerebral palsy? *Gait & Posture*, 24(3), 302–313.
- Donders, A. R. T., van der Heijden, G. J., Stijnen, T., & Moons, K. G. (2006). Review: A gentle

- introduction to imputation of missing values. *Journal of Clinical Epidemiology*, 59(10), 1087–1091.
- Dorizzi, B., Pellieux, G., Jacquet, F., Czernichow, T., & Munoz, A. (1996). Variable selection using generalized RBF networks: Application to the forecast of the French T-bonds. In *CESA'96 IMACS Multiconference : computational engineering in systems applications*, (pp. 122–127).
- Dreyfus, G., Martinez, J.-M., Samuelides, M., & Collectif (2008). *Apprentissage statistique*. Paris: Eyrolles.
- Dubois, R., Maison-Blanche, P., Quenet, B., & Dreyfus, G. (2007). Automatic ECG wave extraction in long-term recordings using Gaussian mesa function models and nonlinear probability estimators. *Computer Methods and Programs in Biomedicine*, 88(3), 217–233.
- Duda, R. O., Hart, P. E., & Stork, D. G. (2000). *Pattern Classification* (2nd Edition ed.). New York: Wiley-Blackwell.
- El-Yacoubi, M. A., Shaiek, A., & Dorizzi, B. (2011). HMM-based gait modeling and recognition under different walking scenarios. In *2011 International Conference on Multimedia Computing and Systems (ICMCS)*, (pp. 1–5).
- Elliott, B. (2006). Biomechanics and tennis. *British Journal of Sports Medicine*, 40(5), 392–396.
- Farhangfar, A., Kurgan, L., & Dy, J. (2008). Impact of imputation of missing values on classification error for discrete data. *Pattern Recognition*, 41(12), 3692–3705.
- Federolf, P. A., Boyer, K. A., & Andriacchi, T. P. (2013). Application of principal component analysis in clinical gait research: Identification of systematic differences between healthy and medial knee-osteoarthritic gait. *Journal of Biomechanics*, 46(13), 2173–2178.
- Fernández-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems? *The Journal of Machine Learning Research*, 15(1), 3133–3181.
- Gage, J. R., Schwartz, M. H., Koop, S. E., & Novacheck, T. F. (2009). *The Identification and Treatment of Gait Problems in Cerebral Palsy* (2nd Edition ed.). London: MacKeith Press.
- Gage, J. R. & Stout, J. L. (2009). Gait Analysis: Kinematics, Kinetics, Electromyography, Oxygen Consumption and Pedobarography. In J. R. Gage, M. H. Schwartz, S. E. Koop, & T. F. Novacheck (Eds.), *The Identification and Treatment of Gait Problems in Cerebral Palsy* (2nd Edition ed.), number 180-181 in Clinics in Developmental Medicine (pp. 260–284). London: MacKeith Press.
- García-Laencina, P. J., Abreu, P. H., Abreu, M. H., & Afonoso, N. (2015). Missing data imputation on the 5-year survival prediction of breast cancer patients with unknown discrete values. *Computers in Biology and Medicine*, 59, 125–133.
- Goldberg, S. R., Öunpuu, S., Arnold, A. S., Gage, J. R., & Delp, S. L. (2006). Kinematic and kinetic factors that correlate with improved knee flexion following treatment for stiff-knee gait. *Journal of Biomechanics*, 39(4), 689–698.
- Grewal, M. S. & Andrews, A. P. (2001). *Kalman filtering theory and practice using MATLAB*. New York: Wiley.
- Guyon, I. & Elisseeff, A. (2003). An introduction to variable and feature selection. *The Journal of Machine Learning Research*, 3, 1157–1182.
- Haykin, S. S. (2001). *Kalman filtering and neural networks*. New York: Wiley.
- Hedström, L. & Carlberg, E. B. (2015). Missing data in physiotherapists' assessments of children with cerebral palsy. *European Journal of Physiotherapy*, 17(2), 66–73.
- Hemo, Y., Aiona, M. D., Pierce, R. A., Dorociak, R., & Sussman, M. D. (2007). Comparison of rectus femoris transposition with traditional transfer for treatment of stiff knee gait in patients with cerebral palsy. *Journal of Children's Orthopaedics*, 1(1), 37–41.
- Hersh, L. A., Sun, J. Q., Richards, J. G., & Miller, F. (1997). The prediction of post-operative

- gait patterns using neural networks. *Gait & Posture*, 5(2), 151.
- Hicks, J. L., Delp, S. L., & Schwartz, M. H. (2011). Can biomechanical variables predict improvement in crouch gait? *Gait & Posture*, 34(2), 197–201.
- Houmani, N., Garcia-Salicetti, S., & Dorizzi, B. (2012). On measuring forgery quality in online signatures. *Pattern Recognition*, 45(3), 1004–1018.
- Huang, G.-B., Zhu, Q.-Y., & Siew, C.-K. (2006). Extreme learning machine: Theory and applications. *Neurocomputing*, 70(1–3), 489–501.
- Hurvich, C. M. & Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, 76(2), 297–307.
- Jacobs, R. A., Jordan, M. I., Nowlan, S. J., & Hinton, G. E. (1991). Adaptive Mixtures of Local Experts. *Neural Computation*, 3(1), 79–87.
- Jamshidi, A. A., Kirby, M. J., & Broomhead, D. S. (2011). Geometric Manifold Learning. *IEEE Signal Processing Magazine*, 28(2), 69–76.
- Jennett, S. (2008). *Churchill Livingstone's Dictionary of Sport and Exercise Science and Medicine* (First ed.). Edinburgh: Churchill Livingstone.
- Jerez, J. M., Molina, I., García-Laencina, P. J., Alba, E., Ribelles, N., Martín, M., & Franco, L. (2010). Missing data imputation using statistical and machine learning methods in a real breast cancer problem. *Artificial Intelligence in Medicine*, 50(2), 105–115.
- Jiang, X. (2011). Linear Subspace Learning-Based Dimensionality Reduction. *IEEE Signal Processing Magazine*, 28(2), 16–26.
- Jolliffe, I. (2002). *Principal Component Analysis*. Springer Series in Statistics. New York: Springer-Verlag.
- Kadaba, M. P., Ramakrishnan, H. K., Wootten, M. E., & others (1990). Measurement of lower extremity kinematics during level walking. *Journal of orthopaedic research*, 8(3), 383–392.
- Kawata, T. (1972). Fourier Analysis in Probability Theory. In E. Lukacs & Z. W. Birnbaum (Eds.), *Probability and Mathematical Statistics*. Academic Press.
- Kay, R. M., Rethlefsen, S. A., Dennis, S. W., & Skaggs, D. L. (2001). Prediction of postoperative gait velocity in cerebral palsy. *Journal of Pediatric Orthopedics. Part B*, 10(4), 275–278.
- Khoury, N. & Desailly, E. (2013). Rectus femoris transfer in multilevel surgery: Technical details and gait outcome assessment in cerebral palsy patients. *Orthopaedics & Traumatology: Surgery & Research*, 99(3), 333–340.
- Klejman, S., Andrysek, J., Dupuis, A., & Wright, V. (2010). Test-Retest Reliability of Discrete Gait Parameters in Children With Cerebral Palsy. *Archives of Physical Medicine and Rehabilitation*, 91(5), 781–787.
- Kodewitz, A., Lelandais, S., Montagne, C., & Vigneron, V. (2013). Alzheimer's disease early detection from sparse data using brain importance maps. *ELCVIA Electronic Letters on Computer Vision and Image Analysis*, 12(1).
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
- Legendre, P. & Legendre, L. F. J. (1998). *Numerical Ecology*. Elsevier Science.
- Leray, P. L.-P. & Gallinari, P. (2001). De l'utilisation d'OBD pour la sélection de variables dans les perceptrons multicouches. *Revue d'Intelligence Artificielle*, 15, 373–391.
- Levenberg, K. (1944). A Method for the Solution of Certain Non-linear Problems in Least Squares. *Quarterly of Applied Mathematics*, 2(2), 164–168.
- Ma, F. Y. P., Selber, P., Nattrass, G. R., Harvey, A. R., Wolfe, R., & Graham, H. K. (2006). Lengthening and transfer of hamstrings for a flexion deformity of the knee in children with bilateral cerebral palsy. Technique and preliminary results. *Journal of Bone & Joint Surgery, British Volume*, 88(2), 248–254.
- MacKay, D. J. (1992). The evidence for neural networks. In *Maximum Entropy and Bayesian Methods* (pp. 165–183). Springer.

- MacKay, D. J. C. (1991). Bayesian Interpolation. *Neural Computation*, 4, 415–447.
- Marks, M. C. & Chambers, H. G. (2003). Clinical utility of the Duncan-Ely test for rectus femoris dysfunction during the swing phase of gait. *Developmental Medicine & Child Neurology*, 45(11), 763–768.
- Maronna, R. A., Martin, D. R., & Yohai, V. J. (2006). *Robust Statistics: Theory and Methods* (1 edition ed.). Chichester, England: Wiley.
- Marquardt, D. (1963). An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2), 431–441.
- McCulloch, W. S. & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4), 115–133.
- Menache, A. (2000). *Understanding Motion Capture for Computer Animation and Video Games*. Morgan Kaufmann.
- Miller, R. G. (1974). The jackknife-a review. *Biometrika*, 61(1), 1–15.
- Moran, K., Richter, C., & O'Connor, N. E. (2014). Letter to the editor regarding “Application of principal component analysis in clinical gait research” by Federolf and colleagues. *Journal of Biomechanics*, 47(6), 1554–1555.
- Motulsky, H. & Christopoulos, A. (2004). *Fitting Models to Biological Data Using Linear and Nonlinear Regression: A Practical Guide to Curve Fitting* (1 edition ed.). Oxford ; New York: Oxford University Press.
- Niiler, T. A. (2001). *Efficacy of predictions of post-operative gait in rectus transfer patients using neural networks*. PhD thesis, University of Delaware.
- Niiler, T. A., Richards, J. G., & Miller, F. (2007). Concurrent surgeries are a factor in predicting success of rectus transfer outcomes. *Gait & Posture*, 26(1), 76–81.
- Niiler, T. A., Richards, J. G., Miller, F., Sun, J. Q., & Castagno, P. (1999). Reliability of predictions of post-operative gait in rectus transfer patients using FFT neural networks. In *Proceedings of the Gait & Clinical Movement Analysis Society Conference*.
- Novacheck, T. F. (2009). Orthopaedic Treatment of Muscle Contractures. In J. R. Gage, M. H. Schwartz, S. E. Koop, & T. F. Novacheck (Eds.), *The Identification and Treatment of Gait Problems in Cerebral Palsy* (2nd Edition ed.), number 180-181 in Clinics in Developmental Medicine (pp. 445–472). London: MacKeith Press.
- Olsson, M. C., Krüger, M., Meyer, L.-H., Ahnlund, L., Gransberg, L., Linke, W. A., & Larsson, L. (2006). Fibre type-specific increase in passive muscle tension in spinal cord-injured subjects with spasticity. *The Journal of Physiology*, 577(Pt 1), 339–352.
- Oskoui, M., Coutinho, F., Dykeman, J., Jetté, N., & Pringsheim, T. (2013). An update on the prevalence of cerebral palsy: a systematic review and meta-analysis. *Developmental Medicine & Child Neurology*, 55(6), 509–519.
- Palisano, R., Rosenbaum, P., Walter, S., Russell, D., Wood, E., & Galuppi, B. (1997). Development and reliability of a system to classify gross motor function in children with cerebral palsy. *Developmental Medicine and Child Neurology*, 39(4), 214–223.
- Palisano, R. J., Rosenbaum, P., Bartlett, D., & Livingston, M. H. (2008). Content validity of the expanded and revised Gross Motor Function Classification System. *Developmental Medicine and Child Neurology*, 50(10), 744–750.
- Park, J. & Sandberg, I. W. (1993). Approximation and Radial-Basis-Function Networks. *Neural Computation*, 5(2), 305–316.
- Peacock, W. J. (2009). The Pathophysiology of Spasticity. In J. R. Gage, M. H. Schwartz, S. E. Koop, & T. F. Novacheck (Eds.), *The Identification and Treatment of Gait Problems in Cerebral Palsy* (2nd Edition ed.), number 180-181 in Clinics in Developmental Medicine (pp. 89–98). London: MacKeith Press.
- Perry, J. & Burnfield, J. (2010). *Gait Analysis: Normal and Pathological Function* (Second

- Edition edition ed.). Thorofare, NJ: Slack Incorporated.
- Polikar, R. (2006). Ensemble based systems in decision making. *Circuits and Systems Magazine, IEEE*, 6(3), 21–45.
- Raghunathan, T. E., Lepkowski, J. M., Van Hoewyk, J., & Solenberger, P. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey methodology*, 27(1), 85–96.
- Ramakrishnan, H. K. & Kadaba, M. P. (1991). On The Estimation of Joints Kinematics During Gait. *Journal of Biomechanics*, 24(10), 969–977.
- Rani, M. P. & Arumugam, G. (2010). An efficient gait recognition system for human identification using modified ICA. *International journal of computer science and information technology*, 2(1), 55–67.
- Reinbolt, J. A., Fox, M. D., Schwartz, M. H., & Delp, S. L. (2009). Predicting outcomes of rectus femoris transfer surgery. *Gait & Posture*, 30(1), 100–105.
- Robertson, D. G. E., Caldwell, G. E., Hamill, J., Kamen, G., & Whittlesey, S. N. (2013). *Research Methods in Biomechanics* (2nd Revised edition ed.). Champaign, Illinois: Human Kinetics Publishers.
- Rodda, J. M., Graham, H. K., Carson, L., Galea, M. P., & Wolfe, R. (2004). Sagittal gait patterns in spastic diplegia. *Bone & Joint Journal*, 86-B(2), 251–258.
- Rosenbaum, P., Paneth, N., Leviton, A., Goldstein, M., Bax, M., Damiano, D., Dan, B., & Jacobsson, B. (2007). A report: the definition and classification of cerebral palsy April 2006. *Developmental Medicine and Child Neurology. Supplement*, 109, 8–14.
- Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386–408.
- Rousseeuw, P. J. & Van Driessen, K. (2006). Computing LTS Regression for Large Data Sets. *Data Mining and Knowledge Discovery*, 12(1), 29–45.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533–536.
- Rutz, E., Donath, S., Tirosh, O., Graham, H. K., & Baker, R. (2013). Explaining the variability improvements in gait quality as a result of single event multi-level surgery in cerebral palsy. *Gait & Posture*, 38(3), 455–460.
- Sacco, I. C. N., Suda, E. Y., Vigneron, V., & Sartor, C. D. (2015). An 'Importance' Map of Signs and Symptoms to Classify Diabetic Polyneuropathy: An Exploratory Data Analysis. *PLoS ONE*, 10(6).
- Sameni, R. (2008). *Extraction of fetal cardiac signals from an array of maternal abdominal recordings*. PhD thesis, Institut National Polytechnique de Grenoble-INPG.
- Santos, A. P., Benamar, F., Bidaud, P., & Desailly, E. (2016). Dynamic simulation of different multi-contact gait patterns identified in patients affected by cerebral palsy. *Gait & Posture*, 49, 81.
- Schapire, R. E. (1990). The strength of weak learnability. *Machine learning*, 5(2), 197–227.
- Scholtes, V. A. B., Becher, J. G., Beelen, A., & Lankhorst, G. J. (2006). Clinical assessment of spasticity in children with cerebral palsy: a critical review of available instruments. *Developmental Medicine & Child Neurology*, 48(1), 64–73.
- Schutte, L. M., Narayanan, U., Stout, J. L., Selber, P., Gage, J. R., & Schwartz, M. H. (2000). An index for quantifying deviations from normal gait. *Gait & posture*, 11(1), 25–31.
- Schwartz, M. H., Novacheck, T. F., & Trost, J. (2000). A tool for quantifying hip flexor function during gait. *Gait & Posture*, 12(2), 122–127.
- Schwartz, M. H. & Rozumalski, A. (2008). The gait deviation index: A new comprehensive index of gait pathology. *Gait & Posture*, 28(3), 351–357.
- Schwartz, M. H., Rozumalski, A., Truong, W., & Novacheck, T. F. (2013). Predicting the out-

- come of intramuscular psoas lengthening in children with cerebral palsy using preoperative gait data and the random forest algorithm. *Gait & Posture*, 37(4), 473–479.
- Schweizer, K., Cattin, P. C., Brunner, R., Müller, B., Huber, C., & Romkes, J. (2012). Automatic selection of a representative trial from multiple measurements using Principle Component Analysis. *Journal of Biomechanics*, 45(13), 2306–2309.
- Sebsadji, A. (2011). *Evaluation d'un outil d'aide à la décision chirurgicale. Contribution des techniques de classification supervisée*. Master thesis, University of Evry Val d'Essonne.
- Sebsadji, A., Khouri, N., Djemal, K., Yepremian, D., Hareb, F., Hoppenot, P., & Desailly, E. (2012). Description and classification of the effect of hamstrings lengthening in cerebral palsy children multi-site surgery. *Computer Methods in Biomechanics and Biomedical Engineering*, 15(S1), 177–179.
- Shapiro, R. (1978). Direct linear transformation method for three-dimensional cinematography. *Research Quarterly*, 49(2), 197–205.
- Sirovich, L. & Kirby, M. (1987). Low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America A*, 4(3), 519–524.
- Sohrweide, S. (2009). Foot Biomechanics and Pathology. In J. R. Gage, M. H. Schwartz, S. E. Koop, & T. F. Novacheck (Eds.), *The Identification and Treatment of Gait Problems in Cerebral Palsy* (2nd Edition ed.), number 180-181 in Clinics in Developmental Medicine (pp. 205–221). London: MacKeith Press.
- Stühler, E. & Merhof, D. (2012). Principal Component Analysis Applied to SPECT and PET Data of Dementia Patients-A Review. In P. Sanguansat (Ed.), *Principal Component Analysis - Multidisciplinary Applications*. Bibliothek der Universität Konstanz.
- Stoppiglia, H., Dreyfus, G., Dubois, R., & Oussar, Y. (2003). Ranking a random feature for variable and feature selection. *The Journal of Machine Learning Research*, 3, 1399–1414.
- Stout, J. L., Novacheck, T. F., Gage, J. R., & Schwartz, M. H. (2009). Treatment of Crouch Gait. In J. R. Gage, M. H. Schwartz, S. E. Koop, & T. F. Novacheck (Eds.), *The Identification and Treatment of Gait Problems in Cerebral Palsy* (2nd Edition ed.), number 180-181 in Clinics in Developmental Medicine (pp. 555–578). London: MacKeith Press.
- Stuart (1994). *Kendall's Advanced Theory of Statistics: Distribution Theory* (6th Edition ed.). Chichester, West Sussex: Wiley-Blackwell.
- Stulp, F. & Sigaud, O. (2015). Many regression algorithms, one unified model: A review. *Neural Networks*, 69, 60–79.
- Sullivan, K., Richards, J., Miller, F., Castagno, P., & Lennon, N. (1995). Predicting the outcome of surgery for children with cerebral palsy using pre-operative gait analysis. *Gait & Posture*, 3(2), 92.
- Sutherland, D. H. & Davids, J. R. (1993). Common gait abnormalities of the knee in cerebral palsy. *Clinical Orthopaedics and Related Research*, (288), 139–147.
- Taylor, G. W. (2009). *Composable, distributed-state models for high-dimensional time series*. PhD thesis, University of Toronto.
- Taylor, G. W., Hinton, G. E., & Roweis, S. T. (2006). Modeling human motion using binary latent variables. In *Advances in neural information processing systems*, (pp. 1345–1352).
- Templ, M., Kowarik, A., & Filzmoser, P. (2011). Iterative stepwise regression imputation using standard and robust methods. *Computational Statistics & Data Analysis*, 55(10), 2793–2806.
- Tipping, M. E. & Bishop, C. M. (1999). Probabilistic Principal Component Analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(3), 611–622.
- Tosic, I. & Frossard, P. (2011). Dictionary Learning. *IEEE Signal Processing Magazine*, 28(2), 27–38.
- Trost, J. P. (2009). Clinical Assessment. In J. R. Gage, M. H. Schwartz, S. E. Koop, & T. F.

- Novacheck (Eds.), *The Identification and Treatment of Gait Problems in Cerebral Palsy* (2nd Edition ed.), number 180-181 in Clinics in Developmental Medicine (pp. 181–204). London: MacKeith Press.
- Tufféry, S. (2012). *Data mining et statistique décisionnelle* (4e édition ed.). Paris: Editions technip.
- Turk, M. A. & Pentland, A. P. (1991). Face recognition using eigenfaces. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1991. Proceedings CVPR '91*, (pp. 586–591).
- Õunpuu, S., Thomason, P., Harvey, A., & Graham, H. K. (2009). Classification of Cerebral Palsy and Patterns of Gait Pathology. In J. R. Gage, M. H. Schwartz, S. E. Koop, & T. F. Novacheck (Eds.), *The Identification and Treatment of Gait Problems in Cerebral Palsy* (2nd Edition ed.), number 180-181 in Clinics in Developmental Medicine (pp. 147–166). London: MacKeith Press.
- Vignerot, V. (1997). *Méthodes d'apprentissage statistiques et problèmes inverses. Application à la spectrographie*. PhD thesis, University of Evry Val d'Essonne.
- Vignerot, V., Kodewitz, A., Tome, A., Lelandais, S., & Lang, E. (2016). Alzheimer's Disease Brain Areas: The Machine Learning Support for Blind Localization. *Current Alzheimer Research*, 13(5), 498–508.
- Wang, S. (2005). Classification with incomplete survey data: a Hopfield neural network approach. *Computers & Operations Research*, 32(10), 2583–2594.
- Werbos, P. J. (1974). *Beyond regression: new tools for prediction and analysis in the behavioral sciences*. PhD thesis, Harvard University.
- Winters, T. J., Gage, J. R., & Hicks, R. (1987). Gait patterns in spastic hemiplegia in children and young adults. *Journal of Bone and Joint Surgery*, 69(3), 437–441.
- Wolpert, D. H. (1992). Stacked Generalization. *Neural Networks*, 5, 241–259.
- Wren, T. A. L., Gorton III, G. E., Õunpuu, S., & Tucker, C. A. (2011). Efficacy of clinical gait analysis: A systematic review. *Gait & Posture*, 34(2), 149–153.
- YeARGIN-Allsopp, M., Van Naarden Braun, K., Doernberg, N. S., Benedict, R. E., Kirby, R. S., & Durkin, M. S. (2008). Prevalence of cerebral palsy in 8-year-old children in three areas of the United States in 2002: a multisite collaboration. *Pediatrics*, 121(3), 547–554.
- Yohai, V. J. (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics*, 15, 642–665.

List of Publications

Galarraga C., O., Vigneron, V., Dorizzi, B., Khouri, N., and Desailly, E. (2015a). Predicting surgery effect on knee kinematics in cerebral palsy. *Gait & Posture*, 42(S1):S2–S3.

Galarraga C., O. A., Khouri, N., Vigneron, V., Dorizzi, B., and Desailly, E. (2016a). Predicting kinematic outcome of multi-level surgery in cerebral palsy. *Gait & Posture*, 49(S2):1.

Galarraga C., O. A., Vigneron, V., Dorizzi, B., Khouri, N., and Desailly, E. (2015b). Estimation of Postoperative Knee Flexion at Initial Contact of Cerebral Palsy Children using Neural Networks. In *Proceedings of the International Conference on Pattern Recognition Applications and Methods*, pages 338–342. SCITEPRESS - Science and Technology Publications.

Galarraga C., O. A., Vigneron, V., Dorizzi, B., Khouri, N., and Desailly, E. (2015c). Predicting postoperative knee flexion during gait of cerebral palsy children. *Computer Methods in Biomechanics and Biomedical Engineering*, 18(S1):1940–1941.



Galarraga C., O. A., Vigneron, V., Dorizzi, B., Khouri, N., and Desailly, E. (2016b). Predicting postoperative knee flexion and pelvic tilt at initial contact of cerebral palsy children. *Movement & Sport Sciences - Science & Motricité*, 93:87–92.

Galarraga C., O. A., Vigneron, V., Dorizzi, B., Khouri, N., and Desailly, E. (2017a). Predicting postoperative gait in cerebral palsy. *Gait & Posture*, 52:45–51.

Galarraga C., O. A., Vigneron, V., Dorizzi, B., Khouri, N., and Desailly, E. (2017b). Predicting postoperative gait in cerebral palsy using ensemble learning. Submitted.

Appendix A

Physical Examination Data Interface

Nom : GALARRAGA

Prénom : Omar

Date de naissance : 1989-04-28

Sexe : M

Dossier # : 8

Arrêt sur ordre/obs : Oui ☐ Non ☐

Ten. et mar. gen. dressés : Oui ☐ Non ☐

Adresseur :

Examineur 1 :

Examineur 2 :

Date de consultation : 22/04/2014

Taille (cm) :

Poids (Kg) :

Tenue assise : Oui ☐ Non ☐

Pass. Che. Ser. Av D : Oui ☐ Non ☐

Fréq. de chutes (mois) :

Périmètre de marche (m) :

Aide de marche :

Appareillage : Oui ☐ Non ☐

GMFCS :

Tenue debout : Oui ☐ Non ☐

Pass. Che. Ser. Av G : Oui ☐ Non ☐

Déclutés Dorsal

	Droit		Gauche	
	V1	V3	V1	V3
Hanche Flexion (°)	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Hanche Abd Flexion (°)	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Hanche Abd Extension (°)	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Hanche Abd Ext Genou 90° (°)	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Genou Flexion (°)	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Genou Extension (°)	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Angle Poplité (°)	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Angle Poplité bilatéral (°)	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Angle Poplité H 40° (°)	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Lasague (°)	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Tardieu (°)	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Chev. Flex. Dors. Genou 90° (°)	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Chev. Flex. Dors. Genou 0° (°)	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Chev. Fl. D. éversion Gen. 0° (°)	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Chev. Flexion Plantaire (°)	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Ar. Pied Valgus (=) Varus (-) (°)	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Av. Pied Pronation (°)	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Av. Pied Supination (°)	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

Testing abdominaux :

Pont bilatéral :

Pont D :

Pont G :

Oui ☐ Non ☐ Oui ☐ Non ☐ Oui ☐ Non ☐

CMS flicisseurs de hanche :

Testing U :

Laixé du genou :

Testing quadriceps :

CMS quadriceps :

Angle mort du quadriceps (°) :

Hauteur de ronde :

Lasague actif (°) :

Torsion tibiale (°) :

Testing triceps :

Déclutés ventral

	Droit	Gauche
	V1	V1
Hanche Rotation interne (°)	<input type="text"/>	<input type="text"/>
Hanche Rotation externe (°)	<input type="text"/>	<input type="text"/>
Torsion fémorale (°)	<input type="text"/>	<input type="text"/>

Scoliose : Oui ☐ Non ☐

ELY test :

Grand Fessier :

Déclutés latéral

	Droit		Gauche	
	V1	V3	V1	V3
Hanche Extension (GE) (°)	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Hanche Extension (GF) (°)	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

Moyen Fessier :

Maintien cuisse abd hori :

Maintien cuisse abd flex :

Assis

	Droit	Gauche
Rotation Genou (=RE -RI) (°)	<input type="text"/>	<input type="text"/>

Tonus Tronc :

Flicisseurs Hanche :

Releveur confusionnet :

Releveurs :

Inverseurs :

Everseurs :

Debout

	Droit	Gauche
AR Pied en charge (=Val -Vae) (°)	<input type="text"/>	<input type="text"/>
AV Pied en charge (=Val -Vae) (°)	<input type="text"/>	<input type="text"/>
Distension du médiotarse : Oui <input type="radio"/> Non <input type="radio"/>	Oui <input type="radio"/> Non <input type="radio"/>	Oui <input type="radio"/> Non <input type="radio"/>

Cyphose : Oui ☐ Non ☐

Lordose : Oui ☐ Non ☐

Tps de maintien unipodal :

Cloche pied : Oui ☐ Non ☐

Observations :

Figure A.1 – Web-like interface for inputting physical examination data with all the usually performed tests (in French). This interface is part of the developed web-like interface for inputting all the considered data.

Title: Simulation of Surgery Effect on Cerebral Palsy Gait by Supervised Machine Learning

Keywords: Supervised Machine Learning; Regression; Prediction; Clinical Gait Analysis; Cerebral Palsy; Single Event Multi-level Surgery

Abstract: Cerebral Palsy frequently leads to gait troubles. After a physical examination and a Clinical Gait Analysis (CGA), these walking troubles are usually treated by orthopedic surgery, called single event multi-level surgery (SEMLS), in which several surgical corrections are simultaneously done at different levels of the lower limbs. Kinematic improvements obtained by this treatment are sometimes very efficient, but at this moment they remain difficultly predictable. The objective of this thesis is to simulate the effect of surgery on gait parameters, using supervised statistical machine learning. The purpose of the simulator is to show the most likely gait outcome in order to improve decision-making in SEMLS. The database was composed of 134 children with cerebral palsy that have undergone sur-

gery and have had at least one CGA before and after the treatment. Gait signals were preprocessed and physical examination missing data were imputed. Features of the preprocessed data were extracted using different techniques such as curve fitting, variable selection and dimensionality reduction. Then regressions were performed utilizing different methods such as multiple linear regression, feedforward neural networks and ensemble learning. The tested methods and their performances were compared between them and to other methods in the literature. This work represents the first time that the effect of surgery on cerebral palsy gait is quantitatively simulated for a large number of surgical combinations and numerous different gait patterns.

Titre : Simulation de l'effet de la chirurgie sur la marche par apprentissage statistique chez des enfants atteints de paralysie cérébrale

Mots clefs : Apprentissage Supervisé; Régression; Prédiction; Analyse Quantifiée de la Marche; Paralysie Cérébrale; Chirurgie Multi-étagée

Résumé : La paralysie cérébrale (PC) cause souvent d'importants troubles de la marche. Suite à un examen clinique et une analyse quantifiée de la marche (AQM), ces troubles peuvent être traités par une chirurgie orthopédique, dite multi-sites, au cours de laquelle plusieurs corrections chirurgicales sont faites simultanément à différents niveaux des membres inférieurs. Les améliorations cinématiques apportées par ce traitement, bien que parfois très efficaces, demeurent à ce jour difficilement prévisibles. L'objectif de cette thèse est de simuler par apprentissage statistique supervisé l'effet de la chirurgie sur les signaux de marche, notamment les signaux cinématiques. Ce simulateur vise à montrer le résultat probable de la marche postopératoire afin d'aider à la décision chirurgicale. Une base de données constituée de 134 enfants atteints de PC, ayant été opérés et ayant eu au moins une AQM avant et après la chirurgie, a

été exploitée. Les signaux cinématiques ont été prétraités et les données cliniques manquantes ont été imputées. Des caractéristiques des données prétraitées ont été extraites en utilisant différentes méthodes telles que l'approximation des courbes, la sélection de variables et la réduction de dimension par analyse en composantes principales. Ensuite des régressions ont été faites en utilisant différentes méthodes telles que la régression multilinéaire, la régression non linéaire avec des réseaux de neurones et l'apprentissage par ensembles. Les différentes méthodes testées ont été comparées entre elles, ainsi qu'avec d'autres méthodes trouvées dans la littérature. Il s'agit de la première fois que l'effet de la chirurgie sur la marche paralysée cérébrale est simulé de façon quantitative pour des nombreuses combinaisons chirurgicales et des nombreux patterns de marche.