Ubiquity: Technologies for Better Health in Aging Societies
A. Hasman et al. (Eds.)
IOS Press, 2006

Functional data analysis for gait curves study in Parkinson's disease

Alain Duhamel[a,1], Patrick Devos[a] , Jean Louis Bourriez[b], C Preda[a] , L Defebvre[c] and Regis Beuscart[a]
[a] *CERIM (EA2694), Faculté Médecine, Université de Lille2, France*
[b] *Neurologie et pathologie du mouvement, CHRU de Lille, France*
[c] *Neurophysiologie Clinique, CHRU de Lille, France*

Abstract. In Parkinson's disease, precise analysis of gait disorders remains essential for the diagnostic or the evaluation of treatments. During a gait analysis session, a series of successive dynamic gait trials are recorded and data involves a set of continuous curves for each patient. An important aspect of such data is the infinite dimension of the space data belong. Therefore, classical multivariate statistical analysis are inadequate. Recent methods known as functional data analysis allow to deal with this kind of data. In this paper, we present a functional data analysis approach for solving two problems encountered in clinical practice: (1) for a given patient, assessing the reliability of the gait curves corresponding to the different trials (2) performing intra individual curves comparisons for assessing the effect of a therapy. In a first step, each discretized curve was interpolated using cubic B-splines bases in order to ensure the continuous character of data. A cluster analysis was performed on the smoothed curves to assess the reliability and to identify a subset of representative curves for a given patient. Intra individual curves comparisons were carried out in the following way: (1) functional principal component analysis was performed to describe the temporal structure of data and to derive a finite number of reliable principal components. (2) These principal components were used in a linear discriminant analysis to point out the differences between the curves. This procedure was applied to compare the gait curves of 12 parkinsonian patients under 4 therapeutic conditions. This study allowed us to develop objective criteria for measuring the improvements in a subject's gait and comparing the effect of different treatments. The methods presented in this paper could be used in other medical domains when data consist in continuous curves.
Keywords: statistics, gait, Parkinson disease, curves, functional data analysis.

1. Introduction

Degenerative neurological diseases constitute a source of handicap in view of their increased prevalence with age and the aging of the population. Parkinson's disease, which principally leads to a motor deficit, is ranked second (after Alzheimer's disease) in terms of these afflictions. Its prevalence is around 2% in the over-65s and has a growing socio-economic impact [1]. Gait disorders constitute one of the major symptoms of Parkinson's disease and lead to the progressive loss of autonomy [2]. At disease onset, these disorders are limited to mild difficulty in quickening gait, which

[1] Corresponding Author: Alain Duhamel, CERIM, Faculté de Médecine, France , aduhamel@univ-lille2.fr

appears less fluid on one side. Postural instability appears later on and constitutes a true turning point in the progression of the condition. Dopatherapy and treatment with dopaminergic agonists constitutes the benchmark treatment for Parkinson's disease. The efficacy of these drugs on gait disorders has been demonstrated, with improvement in kinematic parameters (increase in speed and stride length). Deep brain stimulation also enables improvement in kinematic parameters, the intensity of which varies according to the target stimulated in the central grey nuclei. In Parkinson's disease, precise analysis of gait disorders remains essential, with several possible orientations: specifying the characteristics for diagnostic purposes, monitoring disease progression and prognosis and evaluating the effect of treatment or deep brain stimulation [3]. In gait analysis performed using a video motion analysis, data consist of continuous curves of measurements made over a gait cycle (e.g. the trace of each lower limb joint and the amplitude in degrees, expressed in different planes during a gait cycle). The current software for gait analysis automatically calculates the kinematic parameters like stride speed or cadence and allows to build a mean curve of the different trials for a subject. However, it does not include adequate statistical methods to deal with curve analysis. For example, is not possible, with the current software, to establish comparisons between different situations for a given patient (on drug versus off drug, for example) or between groups when one is studying a drug treatment or surgical therapy. Since 1998, we have collaborated with the Clinical Neurophysiology Department and the Neurology and Movement Pathology Department of the Lille University Hospital in order to develop statistical tools for application to gait analysis. In a first study, we proposed statistical methods for assessing the reliability of gait curves for a given subject and determining the range of normal values for gait curves in a given population. These methods were applied to a healthy population (60 subjects) [4]. The aim of the present study was to extend these tools for analysing gait curves in a population of patients (Parkinsonians). We focused on two problems : the reliability of gait curves and the intra individual curve comparisons. We used the functional data analysis approach [5]. The methods presented in this paper could be used in other medical domains when data consist in continuous curves.

2. Material and methods

2.1. Data acquisition

Kinematic, spatiotemporal and angular gait parameters are recorded automatically using a VICON® video motion analysis system. Fifteen spherical, retro-reflective markers are placed on the various segments of the pelvis and legs and are illuminated by stroboscopes. The trajectories in all 3 planes are recorded by 6 infrared cameras. For each subject, a series of successive dynamic gait trials were analysed during a given session. Different traces of joint position (amplitude in degrees) such as pelvic tilt, hip flexion/extension or knee flexion/extension were recorded in a discrete way for each subject. Data were expressed as a percent of the gait cycle from 0 to 100 %. To test our statistical tools, we present here the results obtained for the knee flexion/extension. Twelve parkinsonian patients were analysed. A minimum of 8 trials were recorded in four conditions, creating a minimum of 32 curves for each patient : before surgery, when the patient had not received any treatment (Off) and after acute administration of L-dopa alone (On-Dopa) - after surgery, with deep brain stimulation alone (On-Stim) and when the patient received in addition L-dopa (Best-On). In our statistical analysis, each discretized curve was interpolated using cubic B-splines bases in order to ensure the continuous character of data.

2.2. Reliability of gait curves

In clinical practice, the gait curves corresponding to different trials can be scattered and a criteria must be defined in order to decide which curves can be selected as characterising the patient. In an initial study [5] performed on healthy subjects, we demonstrated how the Intraclass Correlation Coefficient (ICC) could be used as a measure of the reproducibility of determining which curves should be considered as representative of a subject's gait. ICC varies between 0 to 1 and can be interpreted as the proportion of variance due to the time-to-time variability in the total variance. In the healthy population, the gait of a subject was considered to be reproducible when a minimum of 4 curves resulted in a ICC value greater than 0.95. For parkinsonian patients, the problem is more complicated. These patients have a very jerky gait, and there can be great variability between the gait curves obtained. In fact, the ICC alone did not (for certain patients) enable reliable selection of representative curves. We then proposed the use of a functional data analysis in addition to the ICC computation. For each subject, the gait curve reliability was assessed by performing the following steps : (1) let k be the number of curves for the subject, compute a k×k distance matrix on the set of the cubic B-splines interpolated curves, using integrals, since the curves are continuous; (2) perform a hierarchical classification using the WARD aggregation criteria on the distance matrix in order to identify a subset of representative curves for the patient; (3) compute the ICC to assess the reliability of the subset of curves.

2.3. Confidence bands for healthy population

The identification of the "normal" values for gait curves from a given population is essential in clinical practice. It means building a confidence band which has the following meaning: the gait curve of a subject randomly drawn from the study population has a 95% probability of falling within the confidence band. The method currently proposed by gait analysis software do not take into account the correlation that exists between the measurements and presuppose a Gaussian model for the distribution of data. Use of this method may thus result in a large distortion between the pre-specified probability and the true coverage probability. We have developed a non-parametric procedure based on the bootstrap method for building reliable confidence bands [5]. The basic idea of the bootstrap is to generate many (pseudo)samples from the original data and to use these samples to compute robust estimator. In the present study, the confidence bands for healthy population were used to assess the gait of parkinsonian patients

2.4. Intra individuals comparisons using functional analysis

The comparison of intra-individual curves is very useful for the clinician. For example, one seeks to assess the effect of L-Dopa treatment or stimulation on a given patient or to compare a patient curve to a standard curve. A first approach consists in representing the gait of a subject by a single curve, which is the mean of the reliable trials and then using the confidence bands of healthy population to visually detect the improvements in the subject's gait as a function of various treatments. However, this approach does not enable one to obtain objective criteria for measuring this improvement and comparing the effect of different treatments. We then propose the use of Functional Principal Component Analysis (FPCA) [6]. A gait curve was modelled using a continuous-time stochastic process, $\mathbf{X}=\{X_t\}_{t\in[0,\ T]}$. Let *n* be the number of representative curves for the patient. Denote by *Y* the categorical variable defined by the different therapeutic conditions (§2.1). The intra individual curves comparisons included the following steps (1) The *n* curves were interpolated using cubic B-splines basis functions. (2) FPCA was performed on the *n* B-splines to describe the temporal structure of data and to derive a finite number of reliable principal components. FPCA is a generalization of the classical principal component analysis of discrete data to continuous stochastic process. Details on this method can be found in *Preda et al.* [6]. (3) The principal components were used in a linear discriminant analysis, with the dependent variable *Y,* to point out the differences between the groups.

3. Results

Twelve parkinsonian patients were analyzed using the tools previously described. Because the different steps of the analysis are identical for each patient, we present here the results obtained for one patient. In figure 1a, the curves of this patient are graphically superimposed in order to visually assess reliability. We observe that two curves seem to be different from the others. Using cluster analysis (1b), we detect with no doubt that these curves (C7 and C8) can be considered as outliers. The ICC value computed on the 7 remaining curves was 0.98 and consequently, this subset of 7 curves was selected to represent the gait of the subject in the Best-On condition.

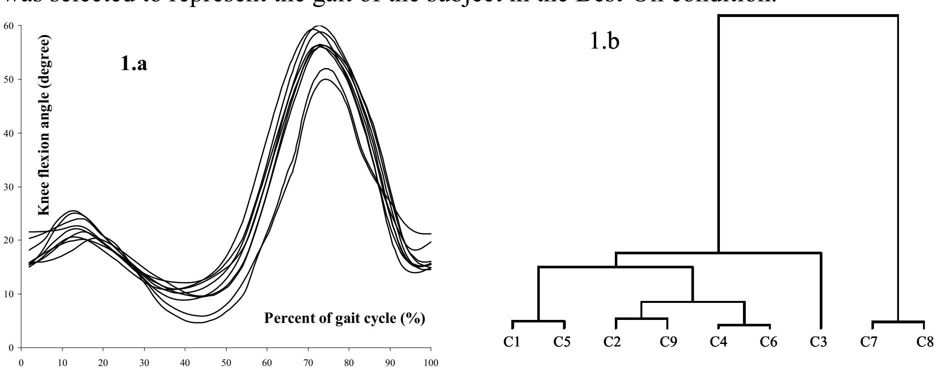

Figure 1 : Gait curves reliability of one patient in the Best-On condition

The use of confidence bands for the healthy population to compare the gait curves is illustrated in figure 2. Each patient's curve represents the mean of the reliable trials in the corresponding treatment condition (Off, On-Dopa, On-Stim and Best-On). Here, one can note that the confidence band enables us to exclude the curve in the condition Off (without treatment) from normality. However, this includes the three other curves : On-Dopa, On-Stim and Best-On. The confidence band does not, therefore, enable us to conclude as to a potential difference in efficacy between these treatments. We then used a functional data analysis approach. This procedure allows to perform a discriminant analysis of curves. In this analysis, the patient is represented by the set of the reliable curves trials in each therapeutic condition. Figure 3 shows the result obtained for the previous patient.

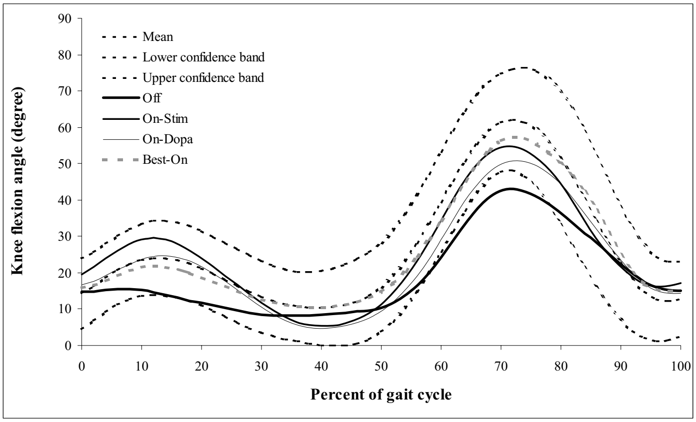

Figure 2 : intra-individual curves comparison using confidence bands

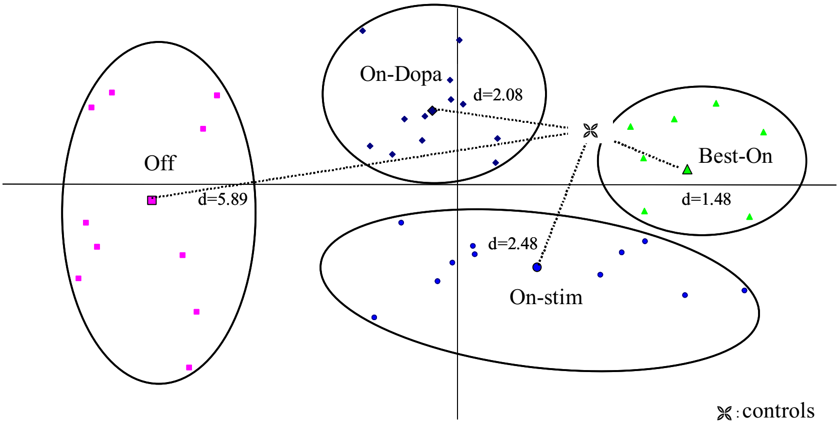

Figure 3 : intra-individual curves comparison using functional data analysis

In this plot, each point represents the projection of a curve in the first discriminant plane. The ellipses indicate the four conditions. For each condition, the mean curve of the corresponding trials is also projected (the central point of the ellipse). Here, one observes that when the parkinsonian patient is not receiving treatment, the gait curves are furthest away from the mean curve for controls (a point represents a curve, that is to say a gait cycle). The group of curves closest to the mean for controls corresponds to the trials recorded in the "Best-On" condition, that is to say with a combination of L-dopa and stimulation. A statistical test (Wilks' Lambda) enables us to show that the centres of gravity for the 4 conditions are different (p<0.0001). Calculation of the distances (using Mahalanobis metric) between the groups' respective centres of gravity and the mean for the controls reveals a progressive improvement in gait curves in the a parkinsonian patient - the "On-Stim" (with stimulation, no L-Dopa) and "On-Dopa" (no stimulation, with L-Dopa) conditions appear to be intermediate.

4. Discussion - conclusion

We presented statistical methods for analysing the gait curves of Parkinsonian patients. These methods make it possible to solve two problems encountered in clinical practice: the selection of reliable curves to represent the gait of a given patient and the comparison of intra-individuals curves. The gait curves are continuous processes. Classical multivariate analysis on the discrete data measurements are generally inefficient to analyse this kind of data because the number of variables (equal to the number of measurements) is greater than the number of individuals (overfit) and also because the continuous-time processes of gait curves are intrinsically smooth. Functional Principal Component Analysis (FCPA) allows to describe the temporal structure of the data and to resume the functional data by means of a finite number of principal components. These principal components do not depend on time and can be used in classical multivariate analysis such as cluster or discriminant analysis. Original contributions to the FPCA methods were developed by our team and used in this study [6]. The different algorithms used to perform FPCA were developed with R. Graphical representations of results provide the physician with support for the interpretation. Different improvements are considered :

- FFT selection for reliability : Fourier transform analysis of the curves could enable determination of those with abnormally high or low harmonics and thus their removal from the calculation.
- Measurements of influence for reliability : calculation of the influence of each curve on calculation of the mean curve (based on the principle of Cook's distance in regression). Curves with too great an influence are deleted.
- Use of the functional data analysis approach for the inter-population comparison of mean curves on different groups of patients. We will experiment the PLS (Partial Least Square) procedure.

We intend to continue this research by analysing other gait parameters such as pelvic tilt or hip flexion/extension. The methods described in this paper could be used in other medical domains when data consist in continuous curves, for example to analyse the cortical rhythm activity from the electroencephalogram.

References

[1] De Lau LM, Schipper CM, Hofman A, Koudstaal PJ, Breteler MM. Prevalence and incidence of Parkinson's disease in Europe, Eur Neuropsychopharmacol. 2005;15(4):473-90.

[2] Allan LM, Ballard CG, Burn DJ, Kenny RA. Prevalence and severity of gait disorders in Alzheimer's and non-Alzheimer's dementias, J Am Geriatr Soc. 2005; 53(10):1681-7.

[3] Defebvre L, Krystkowiak P, Blatt JL, Duhamel A, Bourriez JL, Perina M, Blond S, Guieu JD, Destée A. Influence of pallidal stimulation and L-Dopa on gait and preparatory postural adjustments in Parkinson's disease, Mov. Disord. 2002; 17 : 76-83

[4] Ramsay JO, Silverman BW. Applied functional data analysis : methods and case studies. Springer; 2002.

[5] Duhamel A, Bourriez JL, Devos P, Krystkowiak P, Destee A, Derambure P, Defebvre L. Statistical tools for clinical gait analysis, Gait Posture 2004; 20(2): 204-12.

[6] Preda C, Saporta G. PLS regression on a stochastic process. Computational Statistics and Date Analysis 2005; 48(1) : 149-158.