



**HAL**  
open science

# Classification in postural maintenance based on stochastic process modeling

Christophe Denis

► **To cite this version:**

Christophe Denis. Classification in postural maintenance based on stochastic process modeling. 2011. hal-00653316v1

**HAL Id: hal-00653316**

**<https://hal.science/hal-00653316v1>**

Preprint submitted on 19 Dec 2011 (v1), last revised 10 Sep 2012 (v2)

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

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

# Classification in postural maintenance based on stochastic process modeling

Christophe Denis  
Laboratoire MAP5, UMR CNRS 8145,  
Université Paris Descartes, Sorbonne Paris Cité, Paris, France  
christophe.denis@parisdescartes.fr

## Abstract

This article contributes to the study of *postural style*, focusing on the issue of classifying subjects in terms of how they maintain posture. Here, we specifically tackle the statistical problem of classifying subjects sampled from a two-class population. Each subject (enrolled in a cohort of 54 participants) goes through two experimental protocols designed to evaluate potential deficits in maintaining posture. Measures of foot pressure are obtained at discrete times throughout the protocols. The classification procedure is a two-step procedure. In the first step, the data are modeled by parametric diffusion processes. The parameters can change at some unknown change points. Both parameters and change points are estimated from the data. In the second step, we use the  $V$ -fold cross-validation principle and the super-learning methodology to build two classifiers based on the parameters and change points estimators. We achieve a satisfactory 91% rate of correct classification.

*Keywords:* Change point estimation; Classification; Cross-validation; Postural maintenance; Stochastic process modeling

## 1 Introduction

This article contributes to the study of postural maintenance. Posture is fundamental for physical activity. A deficit in postural maintenance often results in falling, which is particularly hazardous in elderly people. The main objective of the research in postural maintenance is to adapt protocols for functional rehabilitation for people who display deficits in maintaining posture. In this paper, we focus on the issue of how to classify subjects in terms of postural maintenance.

A cohort of 54 subjects has been followed at the Center for the Study of Sensorimotor Functioning (CESEM) of the University Paris Descartes (Chambaz and Denis, 2011). Some of the subjects did not exhibit deficit in postural maintenance while the others were

hemiplegic subjects. Each subject completed four experimental protocols designed to evaluate his/her potential deficits in maintaining posture. Each protocol is divided into three phases: a first phase of 15 seconds with no postural perturbation, a second phase of 35 seconds with postural perturbation followed by a last phase of 20 seconds without postural perturbation. During each protocol, measurements of the center-of-pressure of each foot are performed at discrete times, resulting in time-indexed trajectories. Our objective is to classify subjects into two classes, normal and hemiplegic, from these trajectories.

The structure of a generic trajectory is complex: due to the perturbation phase of the protocols, the behavior of the patient can change abruptly. Consequently, the characteristics of the measured trajectories are time-inhomogeneous. This makes statistical inference and classification difficult. The classification problem is linked to functional data classification problems. There is a sizeable literature on methods for classifying functional data. Methods such as linear discriminant analysis (James and Hastie, 2001), classification based on principal component analysis (Hall *et al.*, 2001) or functional data version of the nearest neighbour classification rule (Biau *et al.*, 2005) have been used to classify functional data. For a general introduction to functional data analysis, see Ramsay and Silverman (2005). Our work relies on a previous study carried out by Chambaz and Denis (2011) where they specifically tackle the statistical problem of classifying subjects from this very dataset. Our aim is to improve on their procedure by taking into account the dynamic of the data.

In order to bypass the problem of the complexity of the trajectories, Chambaz and Denis (2011) propose to base their classification procedure on summary measures of the trajectories. Each protocol is summarized by a vector of small dimension involving the averages of observations computed over the last/first five seconds before/after the beginning/end of the second phase of the protocol. Their classification procedure, based on these summary measures, is a two-step procedure. In the first step they rank the protocols by decreasing order of relevance. In the second step, they derive four classifiers that involve the best (i.e., more informative), the two best, the three best and then all four protocols. This two-step procedure relies on the targeted maximum likelihood learning (see van der Laan and Rubin, 2006) and the super-learning methods (see van der Laan *et al.*, 2007). The authors obtain a 74% rate of correct classification with the best protocol and a 81% rate with the two best protocols.

The aim of the present work is to improve on the rate of correct classification provided by Chambaz and Denis (2011). Because, going through several protocols is unpleasant and sometimes painful, we decided to use only the two best protocols as identified in Chambaz and Denis (2011). The summary measures proposed in Chambaz and Denis (2011) do not take into account either the temporal dynamic of the protocols or the specifics of each patient during the protocols (e.g., each patient has his/her own reaction time to the perturbations). The choice of a 5-second threshold around the beginning/end of the perturbation phase is arbitrarily chosen. This is why we propose to build new summary measures based on a stochastic process to model the temporal aspect of the trajectories and the individuals' behavior changes during the three phases of the protocols. Therefore, we assume that the trajectories are observations of a stochastic process with unknown parameters that may change at some unknown change points. The process parameters depend on the change

points. So, we first estimate the sequence of change points, then the parameters of the process on each interval. The new summary measures rely on the estimators of both the change points and the parameters. As in Chambaz and Denis (2011), our classification procedure, based on the new summary measures, involves the construction of two classifiers based on one of the two best protocols and on the two best protocols, respectively. The construction of these classifiers involves the  $V$ -fold cross-validation principle and the superlearning methodology. The evaluation of the performances of the classification procedure on the real dataset requires the leave-one-out-rule. We improve on the rate of correct classification to 91%.

The article is organized as follows. In Section 2, we describe the dataset and the data modeling. The estimation of the model parameters is presented in Section 3, together with a simulation study. The classification procedure is formally described in Section 4. In Section 5 we report the results obtained by applying the latter classification procedure to the real dataset. In Section 6 we draw the conclusions of our research and discuss our approach.

## 2 Data and modeling

The dataset, collected at the Center for the study of Sensorimotor Functioning (CESEM) of the University Paris Descartes, is described in Section 2.1. We motivate and introduce the data modeling in Section 2.2.

### 2.1 Original dataset

Let us briefly recall the protocols of the study. For a more detailed description, see Chambaz and Denis (2011). The dataset is composed of a cohort of  $n = 54$  subjects. Among the 54 subjects, 22 are hemiplegic (due to cerebrovascular accident). After an initial medical evaluation, the 32 other subjects are identified as *normal patients*. For each subject some covariates are collected (age, gender, laterality, height and weight).

Each subject completes four protocols that evaluate potential deficits in maintaining posture. A protocol is divided into three phases: a first phase of 15 seconds with no postural perturbation, a second phase of 35 seconds with visual and/or muscular perturbation and a last phase of 20 seconds without perturbation. Therefore, the patient's behavior can change near (in time) the beginning and the end of the perturbation phase (around 15 and 50 seconds).

The ranking of the protocols provided by Chambaz and Denis (2011) shows that only two protocols are really significant in terms of postural informativeness. Since going through several protocols can be quite painful, we use only the trajectories from these two protocols, the specifics of which are described in Table 1.

For each protocol, the center-of-pressure of each foot is recorded at equispaced discrete times. Thus each protocol results in a trajectory  $X_{1:N}$  where  $X_i = (L_i, R_i)$  is the observation at time  $t_i = i\delta$  for  $i = 1, \dots, N$ ,  $N = 2800$  and a time-step  $\delta = 0.025$  seconds. For

$i = 1, \dots, N$ ,  $L_i = (L_i^1, L_i^2) \in \mathbb{R}^2$  ( $R_i = (R_i^1, R_i^2)$ , respectively) gives the position of the center-of-pressure of the left foot (right foot, respectively) on the force-platform at time  $i\delta$ .

| protocol | 1st phase (0→15s) | 2nd phase (15→50s)                  | 3rd phase (50→70s) |
|----------|-------------------|-------------------------------------|--------------------|
| 1        |                   | muscular stimulation                |                    |
| 2        | no perturbation   | eyes closed<br>muscular stimulation | no perturbation    |

Table 1: Specifics of the two protocols considered in this study. A protocol is divided into three phases: a first phase without postural perturbation is followed by a second phase with perturbation, which is itself followed by a last phase without perturbation. Different kind of perturbations are considered. Protocol 1 perturbs the processing of proprioceptive information by the brain (muscular stimulation). Protocol 2 relies on perturbing both the processing of visual data (eyes closed) and proprioceptive information by the brain (muscular stimulation).

We focus as in (as in Chambaz and Denis, 2011) on the one-dimensional trajectory  $C_{1:N}$  which is derived from  $X_{1:N}$  by:

$$C_i = \|B_i - b\|,$$

where  $(B_i)_{i \in \{1, \dots, N\}} = (\frac{1}{2}(L_i + R_i))_{i \in \{1, \dots, N\}}$  and  $b$  is defined as the componentwise median value of  $(B_i)$  during the first phase of the protocol. The process  $C_{1:N}$  provides a relevant description of the sway of the body during the course of the protocol. Figure 1 displays the trajectories  $C_{1:N}$  corresponding to the two different protocols completed by a same hemiplegic subject. Figure 1 confirms the intuition that the patient's changes of behavior do not occur exactly at 15 and 50 seconds. As an illustration, it is easy to retrieve the beginning and ending times of the second phase of protocol 1 from the left-hand plot in Figure 1, but not for protocol 2 from the right-hand plot of the same figure.

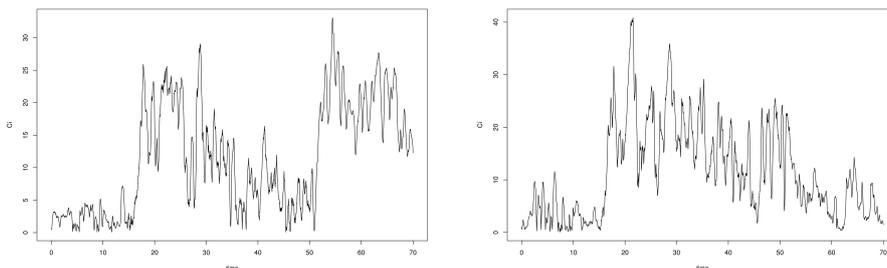


Figure 1: Two trajectories  $i\delta \mapsto C_i$  that correspond to the two different protocols undergone by a same hemiplegic subject (protocol 1 on the left, protocol 2 on the right).

## 2.2 Data modeling

The trajectory  $C_{1:N}$  is considered as the observation at discrete times of a stochastic process  $(C(t))_{t \in [T_0, T]}$  defined by a stochastic differential equation (SDE):

$$\begin{cases} dC(t) = b(C(t), \phi)dt + a(C(t), \sigma)dW(t) \\ C(T_0) = C_0, \end{cases} \quad (1)$$

where  $(W(t))$  is a standard Wiener process,  $b$  is a drift function and  $a$  a volatility function,  $T_0 = \delta$  and  $T = 70$  seconds. Functions  $a$  and  $b$  are known, but they depend on unknown parameters  $\phi$  and  $\sigma$ .

The perturbation phase of the protocols involves that the parameters of  $C_{1:N}$  can change, at least at two change points around 15 and 50 seconds, which corresponds to the beginning and the end of the perturbation phase. Since each patient has a specific behavior during the protocols, we assume that the change points are unknown and that they differ among subjects. A visual investigation of several trajectories  $C_{1:N}$  for different patients shows that a third change point may occur during the perturbation phase. Therefore, we also investigate the case where there are three change points. In the following we denote by  $T_0 = \delta < T_1 = \tau_1\delta < \dots < T_{K-1} = \tau_{K-1}\delta < T_K = T = 70$  the sequence of unknown change points with  $\tau_i \in \mathbb{N}$ , and where  $K$  is either 3 or 4, which corresponds to two or three intermediate change points.

We allow the parameters  $\phi$  and  $\sigma$  defined in (1) to change on each interval  $[T_{k-1}, T_k]$  for  $k = 1, \dots, K$  defined by the change points  $(T_k)$ . Hence, for each  $k = 1, \dots, K$ , we assume that the stochastic process  $C(t)$  on  $[T_{k-1}, T_k]$  is defined by Equation (1) with drift and volatility functions that depend on the unknown parameters  $\phi_k$  and  $\sigma_k$ .

There remains to specify the drift function  $b$  and the volatility function  $a$ . Figure 1 suggests that the variance of  $C_{1:N}$  is not constant over time. In addition the process  $C_{1:N}$  is positive. Therefore we propose to model  $C(t)$  by the classical Cox-Ingersoll-Ross (CIR) process (Kloeden and Platen, 1992). More formally, for all  $k = 1, \dots, K$ ,  $(C(t))_{t \in [T_{k-1}, T_k]}$  is defined by the following SDE:

$$\begin{cases} dC(t) = \lambda_k(\mu_k - C(t))dt + \sigma_k\sqrt{C(t)}dW(t) \\ C(T_{k-1}) = C_{\tau_{k-1}}, \end{cases} \quad (2)$$

where the initial condition  $C(T_{k-1})$  is the observation  $C_{\tau_{k-1}}$  at time  $T_{k-1}$ ,  $\lambda_k$ ,  $\mu_k$  and  $\sigma_k$  are positive constants,  $\lambda_k$  is the time-scale of the process,  $\mu_k$  is the mean of the asymptotic distribution and  $\sigma_k$  represents the standard deviation. If  $\frac{2\mu_k\lambda_k}{\sigma_k^2} \geq 1$  the process stays positive, and admits a stationary distribution. This distribution is a Gamma distribution with shape parameter  $\frac{2\mu_k\lambda_k}{\sigma_k^2}$  and scale parameter  $\frac{\sigma_k^2}{2\lambda_k}$ .

The change points  $T_1, \dots, T_{K-1}$ , and the vector of parameters  $\theta_k = (\lambda_k, \mu_k, \sigma_k)$  are unknown. In a first step we estimate the sequence of change points, then on each interval  $[\hat{T}_{k-1}, \hat{T}_k]$ ,  $k = 1, \dots, K$  we estimate the parameter  $\theta_k$ .

### 3 Estimation of the model parameters

This section is devoted to the first step of our classification procedure: the estimation of the sequence of change points defined in Section 2.2 and the parameters of the diffusion process  $(C(t))_{t \in [T_{k-1}, T_k]}$ . We first define the estimators and recall their properties. Then we give the results of this estimation on the dataset and illustrate the performance of the estimation method with a simulation study.

#### 3.1 Change-point estimation

The methodology used for the estimation of the sequence  $\tau = (\tau_k)_{1 \leq k \leq K-1}$  relies on the statistical theory of change point estimation in piecewise-constant models. In this framework, the distribution parameters of the process are assumed to change at some instants. For instance, the changes can affect the mean and the variance of the distribution. Between two change points, the parameters of the process are assumed to remain constant. Different methods have been proposed for the estimation of the change points, such as sequential methods (Basseville and Nikiforov, 1993), local methods (Gijbels *et al.*, 1999) or the minimization of penalized contrast (Lavielle, 2005). In this paper, we adopt the contrast proposed by Lavielle (2005), but with a known number of change points. It is a global approach, where all the change points are simultaneously detected by minimizing a contrast  $\mathcal{J}(\tau, C_{1:N})$ . This contrast function, based on a Gaussian pseudo log-likelihood, is proposed to detect changes that affect both the mean and the variance. The contrast function is defined as:

$$\mathcal{J}(\tau, C_{1:N}) = \frac{1}{N} \sum_{k=1}^K (\tau_k - \tau_{k-1} + 1) \log(\gamma_{\tau_{k-1}:\tau_k}^2),$$

where

$$\gamma_{\tau_{k-1}:\tau_k}^2 = \frac{1}{\tau_k - \tau_{k-1} + 1} \sum_{i=\tau_{k-1}}^{\tau_k} (C_i - \bar{C}_{\tau_{k-1}:\tau_k})^2$$

and  $\bar{C}_{\tau_{k-1}:\tau_k}$  is the empirical mean of  $(C_{\tau_{k-1}}, \dots, C_{\tau_k})$ . Thus, the estimated sequence of change points denoted by  $\hat{\tau}_N = (\hat{\tau}_{N,1}, \dots, \hat{\tau}_{N,K-1})$  is defined by:

$$\hat{\tau}_N = \arg \min_{\tau} \mathcal{J}(\tau, C_{1:N}).$$

Let us now recall some theoretical results in the framework of piecewise-constant models, when the true number of change points is known (see Lavielle, 1999; Lavielle and Ludeña, 2000). The sequence  $\hat{\tau}_N$  has the following property: Under general conditions, for any  $k \in \{1, \dots, K-1\}$ ,

$$P(|\hat{\tau}_{N,k} - \tau_k| > \eta) \rightarrow 0, \text{ when } \eta \rightarrow \infty \text{ and } N \rightarrow \infty.$$

For notation simplicity, we denote by  $\hat{\tau} = (\hat{\tau}_1, \dots, \hat{\tau}_{K-1})$  the obtained estimator and by  $\hat{T} = (\hat{T}_1, \dots, \hat{T}_{K-1}) = \delta \hat{\tau}$  the instants of change points.

### 3.2 Estimation of the parameters of the diffusion processes

We describe the method used for the estimation of the parameters of the processes  $(C(t))_{t \in [\hat{T}_{k-1}, \hat{T}_k]}$  defined in Equation (2) for  $k = 1, \dots, K$ . On each interval  $[\hat{T}_{k-1}, \hat{T}_k]$ , the estimation of the parameters relies on the minimization of a contrast function proposed by Kessler (1997), and based on the approximate discrete-time Euler-Marumaya's scheme.

Let  $(\tilde{C}_i)_{i \in \{\hat{\tau}_{k-1}, \dots, \hat{\tau}_k\}}$  denote the discrete-time approximation of  $(C(t))_{t \in [\hat{T}_{k-1}, \hat{T}_k]}$ , obtained by the Euler-Marumaya scheme with step size  $\delta$ . For  $i \in \{\hat{\tau}_{k-1}, \dots, \hat{\tau}_k\}$ , we have:

$$\tilde{C}_{i+1} = (1 - \delta\lambda_k)\tilde{C}_i + \delta\lambda_k\mu_k + \sigma_k\sqrt{\delta}\sqrt{\tilde{C}_i}\eta_{i+1}, \quad (3)$$

where  $(\eta_i)_{i \in \{\hat{\tau}_{k-1}+1, \dots, \hat{\tau}_k\}}$  is a sequence of independent random variables with common law  $\mathcal{N}(0, 1)$ . We use the following parameterization  $\theta_k = (\theta_{1,k}, \theta_{2,k}, \theta_{3,k})$ , with  $\theta_{1,k} = (1 - \delta\lambda_k)$ ,  $\theta_{2,k} = \delta\lambda_k\mu_k$  and  $\theta_{3,k} = \sigma_k\sqrt{\delta}$ . The contrast function  $\mathcal{L}_k(\theta_k)$  based on the log-likelihood of  $(\tilde{C}_i)_{i \in \{\hat{\tau}_{k-1}, \hat{\tau}_k\}}$ , is defined by:

$$\mathcal{L}_k(\theta_k) = \sum_{i=\hat{\tau}_{k-1}}^{\hat{\tau}_k-1} \frac{(C_{i+1} - \theta_{1,k}C_i - \theta_{2,k})^2}{C_i\theta_{3,k}^2} + (\hat{\tau}_k - \hat{\tau}_{k-1} + 1) \log(\theta_{3,k}^2).$$

The estimator  $\hat{\theta}_k$  is defined by:

$$\hat{\theta}_k = \arg \min_{\theta_k \in \mathbb{R}_+^3} \mathcal{L}_k(\theta_k). \quad (4)$$

Thus, we obtain explicit estimators:

$$\begin{aligned} \hat{\theta}_{1,k} &= \frac{(\hat{\tau}_k - \hat{\tau}_{k-1} + 1) \sum \frac{C_{i+1}}{C_i} - \sum C_{i+1} \sum \frac{1}{C_i}}{(\hat{\tau}_k - \hat{\tau}_{k-1} + 1)^2 - \sum C_i \sum \frac{1}{C_i}}, \\ \hat{\theta}_{2,k} &= \frac{\sum C_{i+1} - \hat{\theta}_{1,k} \sum C_i}{(\hat{\tau}_k - \hat{\tau}_{k-1} + 1)}, \\ \hat{\theta}_{3,k} &= \sqrt{\frac{\sum \frac{(C_{i+1} - \hat{\theta}_{1,k}C_i - \hat{\theta}_{2,k})^2}{C_i}}{\hat{\tau}_k - \hat{\tau}_{k-1} + 1}}. \end{aligned}$$

Under general assumptions, see Theorem 1 in Kessler (1997), if  $(C(t))_{[T_{k-1}, T_k]}$  is stationary and when  $N \rightarrow +\infty$ ,  $\delta \rightarrow 0$  and  $N\delta \rightarrow +\infty$ , the estimators  $\hat{\theta}_k$ ,  $k = 1, \dots, K - 1$ , defined by (4) are consistent:

$$\hat{\theta}_k \xrightarrow{P_{\theta_k}} \theta_k.$$

### 3.3 Estimation results

For each patient, for each protocol and for  $K = 3$  or  $4$  intervals, we estimate the sequence of change points  $\hat{T}_1, \dots, \hat{T}_{K-1}$  and the parameters  $\hat{\theta}_1, \dots, \hat{\theta}_K$  from the dataset. Results of the

change point estimation are summarized in Table 2 and illustrated in Figure 2. Results of the parameter estimation are summarized in Table 3. In order to discuss the difference in the estimated change points and parameters between the normal and hemiplegic subjects, we give the results separately for each group.

| Normal subjects |              |               |              |               |              |
|-----------------|--------------|---------------|--------------|---------------|--------------|
| protocol        | K = 3        |               | K = 4        |               |              |
| 1               | 18.28 (4.37) | 50.66 (10.53) | 16.62 (5.02) | 31.90 (13.75) | 52.97 (6.85) |
| 2               | 19.75 (7.83) | 52.62 (4.87)  | 15.84 (4.32) | 31.59 (12.47) | 52.68 (4.80) |

| Hemiplegic subjects |              |               |              |               |              |
|---------------------|--------------|---------------|--------------|---------------|--------------|
| protocol            | K = 3        |               | K = 4        |               |              |
| 1                   | 15.27 (3.13) | 45.27 (13.78) | 14.27 (3.89) | 31.18 (11.13) | 51.86 (8.84) |
| 2                   | 16.95 (5.39) | 48.13 (11.60) | 13.95 (7.15) | 30.45 (14.38) | 55.36 (3.35) |

Table 2: Estimation of the sequence  $(T_1, \dots, T_{K-1})$ , for  $K = 3$  and 4 intervals. For each patient and each protocol, we compute the estimators  $\hat{T}_k$  of  $T_k$ ,  $k = 1, \dots, K - 1$ . We provide the empirical means and standard deviations (between parentheses) of the estimators  $\hat{T}_k$  over each group (normal and hemiplegic subjects).

Table 2 shows that whatever the value of  $K$ , the estimated change points over the groups of normal and hemiplegic subjects are not significantly different. Two change points are estimated around the beginning (15 seconds) and the end (50 seconds) of the second phase of the protocol (the perturbation phase). This is consistent with the specifics of the protocols. When  $K = 3$  the change points occur slightly earlier in the hemiplegic group than in the normal group, though not significantly. For each protocol and  $K = 4$ , the third change point is estimated during the perturbation phase (around 30 seconds). This is consistent with previous investigations. As an illustration, we see in Figure 2 that, for each protocol, two change points are estimated around 15 and 50 seconds and are the same for  $K = 3$  or 4 intervals. The third change point is estimated for each protocol during the perturbation phase between 30 and 40 seconds. The third change point seems to be more relevant for protocol 1 than for protocol 2 for this particular subject.

The results provided in Table 3 show that for each protocol and each group, the values of the estimated empirical means of  $\theta_{1,k}$  are close. Consequently, the parameters  $\theta_{1,k}$ ,  $k = 1, \dots, K - 1$  may not be relevant for the classification step. The estimated empirical means of  $\theta_{2,k}$  are significantly different from one protocol to the other and one group to the other. The estimated empirical means of  $\theta_{3,k}$  change over the two groups. We expect that the parameters  $\theta_{2,k}$  and  $\theta_{3,k}$ ,  $k = 1, \dots, K - 1$  can provide relevant information to determine whether the patient is hemiplegic or not.

### 3.4 Simulation study

A simulation study is carried out to evaluate the performance of the estimation procedure described in Section 3.1 and Section 3.2. The data are simulated with two deterministic

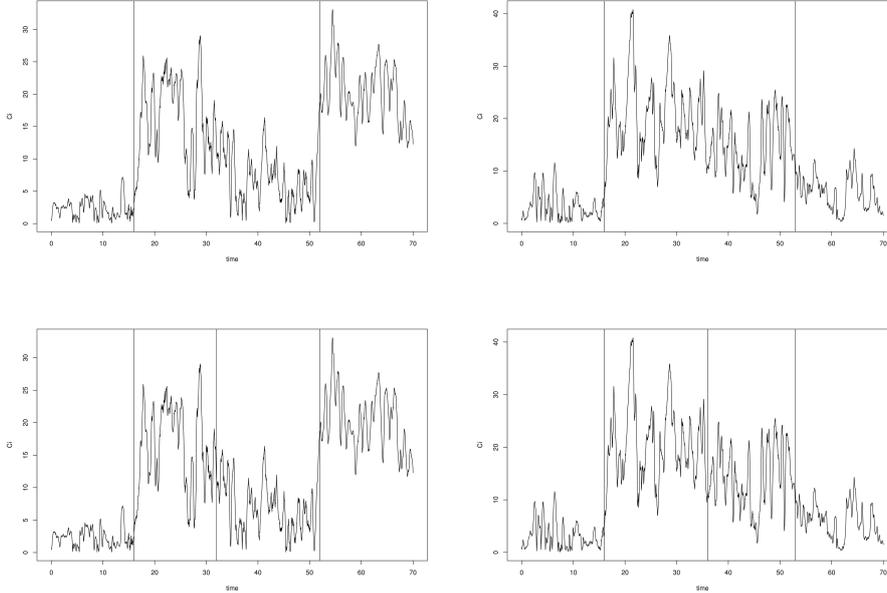


Figure 2: Estimated change points for trajectories of protocol 1 (left column) and protocol 2 (right column). The two trajectories are from the same hemiplegic subject. The top line corresponds to two estimated change points ( $K = 3$ ), the bottom line corresponds to three estimated change points ( $K = 4$ ).

change points  $T_1, T_2$  at respectively 15 and 50 seconds, using the mean parameters obtained in Table 3 for the normal subjects and protocol 1. We simulate 100 datasets using a Euler scheme with step size 0.0025. Then observations at time  $(i\delta)$  with  $\delta = 0.025$  are sub-sampled to build a simulated trajectory. The estimation procedure of the change points and parameters is applied to each simulated trajectory. The means and standard deviations of the change points and parameters estimated from the 100 trajectories are presented in Table 4.

The means of the estimators of the change points are close to the true values. Moreover, the standard deviations are small. The means of the parameter estimators are very satisfactory, although the estimation of  $\theta_{2,2}$  is slightly biased. This is probably due to the fact that the time interval is small.

To conclude this section, we provide in Figure 3 an illustration of the data modeling. The simulated trajectory based on the estimators  $\hat{T}_1, \hat{T}_2$  and  $(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)$  is relatively close to the true trajectory.

| protocol | param.           | Normal subjects |             |             | Hemiplegic subjects |             |             |
|----------|------------------|-----------------|-------------|-------------|---------------------|-------------|-------------|
|          |                  | $k = 1$         | $k = 2$     | $k = 3$     | $k = 1$             | $k = 2$     | $k = 3$     |
| 1        | $\theta_{1,k}$   | 0.98 (0.01)     | 0.99 (0.01) | 0.98 (0.02) | 0.96 (0.02)         | 0.98 (0.01) | 0.98 (0.02) |
|          | $\theta_{2,k}$   | 0.06 (0.04)     | 0.14 (0.12) | 0.11 (0.08) | 0.13 (0.09)         | 0.27 (0.30) | 0.16 (0.13) |
|          | $\theta_{3,k}^2$ | 0.04 (0.03)     | 0.03 (0.04) | 0.03 (0.04) | 0.08 (0.05)         | 0.05 (0.04) | 0.05 (0.05) |
| 2        | $\theta_{1,k}$   | 0.98 (0.02)     | 0.99 (0.01) | 0.98 (0.01) | 0.96 (0.02)         | 0.98 (0.02) | 0.97 (0.02) |
|          | $\theta_{2,k}$   | 0.09 (0.08)     | 0.22 (0.26) | 0.15 (0.14) | 0.15 (0.10)         | 0.37 (0.36) | 0.36 (0.62) |
|          | $\theta_{3,k}^2$ | 0.06 (0.06)     | 0.05 (0.04) | 0.05 (0.06) | 0.1 (0.06)          | 0.1 (0.10)  | 0.09 (0.08) |

Table 3: Estimation of the parameters  $(\theta_1, \theta_2, \theta_3)$  of the CIR stochastic process. For each patient, each protocol and  $K = 3$ , we estimate two change points for the trajectory  $C_{1:N}$ . Then, we compute the estimators  $(\hat{\theta}_{1,k}, \hat{\theta}_{2,k}, \hat{\theta}_{3,k})$ ,  $k = 1, 2, 3$ . We provide the empirical means and standard deviations (between parentheses) of the estimators  $\hat{\theta}_{i,k}$  over each group (normal and hemiplegic subjects).

| Change points estimation |         |               |         |               |         |               |
|--------------------------|---------|---------------|---------|---------------|---------|---------------|
|                          |         | $T_1$         |         | $T_2$         |         |               |
|                          | True    | Estimated     | True    | Estimated     |         |               |
|                          | 15      | 16.65 (2.41)  | 50      | 51.28 (2.42)  |         |               |
| Parameter estimation     |         |               |         |               |         |               |
|                          | $k = 1$ |               | $k = 2$ |               | $k = 3$ |               |
| parameter                | True    | Estimated     | True    | Estimated     | True    | Estimated     |
| $\theta_{1,k}$           | 0.98    | 0.976 (0.010) | 0.99    | 0.986 (0.005) | 0.98    | 0.975 (0.009) |
| $\theta_{2,k}$           | 0.06    | 0.076 (0.030) | 0.14    | 0.189 (0.071) | 0.11    | 0.135 (0.043) |
| $\theta_{3,k}$           | 0.04    | 0.038 (0.002) | 0.03    | 0.029 (0.001) | 0.03    | 0.029 (0.001) |

Table 4: Simulation study. Empirical means and standard deviations of change points and parameters estimators computed from 100 simulated trajectories with two change points.

## 4 Classification procedure

This section is devoted to our classification procedure. We present the definition of the finite-dimensional summary measure in Section 4.1. We formally introduce the statistical framework that we consider in Section 4.2. Finally, we describe our classification procedure in Section 4.3. Our classification procedure involves the cross-validation principle that is described in Section 4.4.

### 4.1 Summary measures

The finite-dimensional summary measure relies on the estimated sequence of change points  $\hat{T}_1, \dots, \hat{T}_{K-1}$ , and on the estimators  $\hat{\theta}_1, \dots, \hat{\theta}_K$  defined in Section 3.2. Under the assumption that, on each interval defined by the estimated change points of the protocols, the process  $(C(t))_{[\hat{T}_{k-1}, \hat{T}_k]}$  defined by (2) reaches the stationary regime, a relevant information from the process  $(C(t))_{[\hat{T}_{k-1}, \hat{T}_k]}$  is the mean of the stationary distribution with  $\frac{\hat{\theta}_{2,k}}{1-\hat{\theta}_{1,k}} = \hat{\mu}_k$

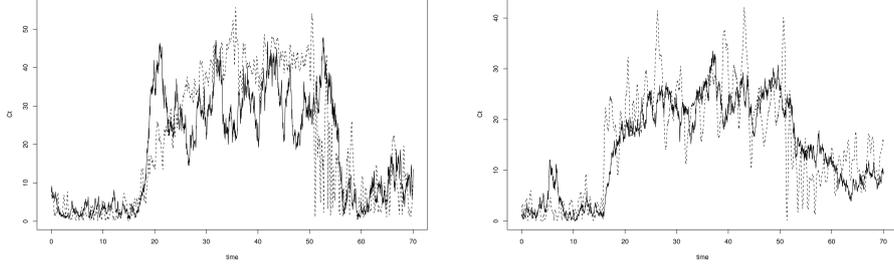


Figure 3: Illustration of the data modeling, for protocol 2: true trajectory (solid line) and simulated trajectory (dotted line). The left plot correspond to a normal patient, the right plot to a hemiplegic patient. Two change points are estimated and parameters  $\theta_k$  are estimated. For each  $k = 1, 2, 3$  and according to Equation (3) with parameter  $\hat{\theta}_k$ , we simulate  $\tilde{C}_{\hat{\tau}_{k-1}:\hat{\tau}_k}$ . The simulated trajectory is then  $\tilde{C}_{1:N} = (\tilde{C}_{1:\hat{\tau}_1}, \tilde{C}_{\hat{\tau}_1:\hat{\tau}_2}, \tilde{C}_{\hat{\tau}_2:N})$ .

as estimator. Furthermore, estimation results given in Section 3.3 have shown that the parameters  $\theta_{2,k}$  and  $\theta_{3,k}$  may provide helpful information for classification. Finally, we decide to focus on the finite-dimensional summary measure  $Y_K$  of  $C_{1:N}$ , which is defined as

$$Y_K = (\hat{\mu}_1, \dots, \hat{\mu}_K, \hat{\theta}_{3,1}, \dots, \hat{\theta}_{3,K}, \hat{T}_1, \dots, \hat{T}_{K-1}). \quad (5)$$

In the sequel, we denote by  $Y_K^j$  the summary measure defined by (5) and associated to protocol  $j$ ,  $j = 1, 2$ .

## 4.2 Classification framework

We want to construct classifiers to determine if the patient is normal or hemiplegic. For a given patient, those classifiers are based both on the finite-dimensional summary measure defined by (5) and on the patient covariates.

For a given patient, we denote by  $O$  the observed data structure:

$$O = (W, Y_3^1, Y_4^1, Y_3^2, Y_4^2, A),$$

where  $W$  is the vector of baseline covariates, for each  $j \in \{1, 2\}$  and  $K \in \{3, 4\}$ , the vector  $Y_K^j \in \mathbb{R}^{3K-1}$  is the summary measure associated to the  $j$ th protocol with  $(K - 1)$  change points. The variable  $A \in \{0, 1\}$  indicates the subject's class (with convention  $A = 1$  for hemiplegic subjects and  $A = 0$  for normal subjects). We assume that the distribution of  $O$  is  $P_0 \in \mathcal{M}$ , where  $\mathcal{M}$  is the set of all possible distributions of  $O$ .

We denote by  $\mathcal{S}$  the set of all possible classifiers which are functions of the vector of baseline covariates  $W$  and the summary measure  $(Y_3^1, Y_4^1, Y_3^2, Y_4^2)$ . Set  $Z = (W, Y_3^1, Y_4^1, Y_3^2, Y_4^2)$ . We recall that, in the classification framework, the risk  $R^{(P_0)}$  associated to a classifier  $S \in \mathcal{S}$  is defined as:

$$R^{(P_0)}(S) = E_{P_0} [\mathbf{1}\{S(Z) \neq A\}].$$

We denote by  $\mathcal{R}^*$  the minimum of these risks:

$$\mathcal{R}^* = \min_{S \in \mathcal{S}} R^{(P_0)}(S).$$

This minimum is achieved at  $S^{(*)}$ , the Bayes classifier, characterized by:

$$S^{(*)}(Z) = \mathbf{1} \left\{ \eta(Z) \geq \frac{1}{2} \right\},$$

where  $\eta(Z) = P_0(A = 1|Z)$ . For  $K = 3, 4$  and  $j = 1, 2$ , let us define the classifier  $S_K^j$  based only on  $W$  and the summary measure  $Y_K^j$  (i.e.  $S_K^j(Z) = S_K^j(W, Y_K^j)$ ):

$$S_K^j(W, Y_K^j) = \mathbf{1} \left\{ \eta_K^j(W, Y_K^j) \geq \frac{1}{2} \right\},$$

where  $\eta_K^j(W, Y_K^j) = P_0(A = 1|W, Y_K^j)$ . For  $(K_1, K_2) \in \{3, 4\}^2$ , let us define the classifier  $S_{K_1, K_2}$  based only on  $W$  and the summary measures  $Y_{K_1}^1$  and  $Y_{K_2}^2$  (i.e.  $S_{K_1, K_2}(Z) = S_{K_1, K_2}(W, Y_{K_1}^1, Y_{K_2}^2)$ ):

$$S_{K_1, K_2}(W, Y_{K_1}^1, Y_{K_2}^2) = \mathbf{1} \left\{ \eta_{K_1, K_2}(W, Y_{K_1}^1, Y_{K_2}^2) \geq \frac{1}{2} \right\},$$

where  $\eta_{K_1, K_2}(W, Y_{K_1}^1, Y_{K_2}^2) = P_0(A = 1|W, Y_{K_1}^1, Y_{K_2}^2)$ . Let us define  $\phi_1$  and  $\phi_2$  as:

$$\begin{aligned} \phi_1 &\in \arg \min_{S_K^j} R^{(P_0)}(S_K^j), \\ \phi_2 &\in \arg \min_{S_{K_1, K_2}} R^{(P_0)}(S_{K_1, K_2}). \end{aligned}$$

The classifier  $\phi_1$  represents the best classifier among those which are functions of  $W$  and a summary measure  $Y_K^j$  with  $K \in \{3, 4\}$  and  $j \in \{1, 2\}$ . This corresponds to the best classifier among all the classifiers based on one protocol only. The classifier  $\phi_2$  represents the best classifier among those which are functions of  $W$  and the summary measures  $Y_{K_1}^1, Y_{K_2}^2$  with  $(K_1, K_2) \in \{3, 4\}^2$ . This corresponds to the best classifier among all the classifiers based on the two protocols. Since  $\phi_1$  and  $\phi_2$  depend on the distribution  $P_0$  which is unknown, we must construct estimators  $\hat{\phi}_1$  and  $\hat{\phi}_2$  of  $\phi_1$  and  $\phi_2$ . First, for each  $K = 3, 4$  and  $j = 1, 2$ , we construct an estimator  $\hat{S}_K^j$  of  $S_K^j$ . Then for each  $K_1 = 3, 4$  and  $K_2 = 3, 4$ , we construct an estimator  $\hat{S}_{K_1, K_2}$  of  $S_{K_1, K_2}$ . We denote by  $\mathcal{H}_1 = \{\hat{S}_3^1, \hat{S}_4^1, \hat{S}_3^2, \hat{S}_4^2\}$  the set of candidate estimators of  $\phi_1$  and by  $\mathcal{H}_2 = \{\hat{S}_{3,3}, \hat{S}_{3,4}, \hat{S}_{4,3}, \hat{S}_{4,4}\}$  the set of candidate estimators of  $\phi_2$ . In a second step, we select  $\hat{\phi}_1$  (respectively  $\hat{\phi}_2$ ) among the set  $\mathcal{H}_1$  (respectively  $\mathcal{H}_2$ ) based on the  $V$ -fold cross-validation principle.

### 4.3 Construction of the classifiers

In this section, we describe the construction of the set  $\mathcal{H}_1$ . Except for the change of notations the procedure is the same for the construction of  $\mathcal{H}_2$ .

Let  $D_n = (O_{(1)}, \dots, O_{(n)})$  be a learning dataset of size  $n$ . For each  $K = 3, 4$  and  $j = 1, 2$ , the construction of  $\hat{S}_K^j$  relies on the plug-in-rule (Tsybakov, 2004). Based on  $D_n$ , we construct an estimator  $\hat{\eta}_K^j(D_n, \cdot)$  of  $\eta_K^j$ , using the super-learning methodology (van der Laan *et al.*, 2007). For a new observation  $(W, Y_K^j)$  independent of  $D_n$ , the plug-in classifier  $\hat{S}_K^j$  is then defined as:

$$\hat{S}_K^j(D_n; W, Y_K^j) = \mathbf{1} \left\{ \hat{\eta}_K^j(D_n; W, Y_K^j) \geq \frac{1}{2} \right\}.$$

We adopt the following classification rule: We decide to classify a new subject with information  $(W, Y_K^j)$  into the group of hemiplegic subjects if  $\hat{S}_K^j(D_n; W, Y_K^j) = 1$  and into the group of normal subjects otherwise. The plug-in classifiers can converge with fast rates under assumptions on the regression function  $\eta_K^j$  and the so-called margin condition (Audibert and Tsybakov, 2007).

#### 4.4 Cross-validation procedure

In this section, we briefly describe the  $V$ -fold cross-validation procedure for the construction of the classifier  $\hat{\phi}_1$ . Except for the change of notations the procedure is the same for the construction of  $\hat{\phi}_2$ . Several studies (practical and theoretical) have been conducted on the cross-validation principle (see Geisser, 1975; Dudoit and van der Laan, 2005, for example).

Let  $V \geq 2$  be an integer. We first consider a regular partition  $(B_v)_{1 \leq v \leq V}$  of the indices  $\{1, \dots, n\}$  such that:

$$\text{Card}(B_v) \in \{ \lfloor n/V \rfloor, \lfloor n/V \rfloor + 1 \}.$$

For each  $v \in \{1, \dots, n\}$ , we denote  $D_n^{(v)}$  (respectively  $D_n^{(-v)}$ ) the data set  $(O_{(i)})_{i \in B_v}$  (respectively  $(O_{(i)})_{i \notin B_v}$ ) and we define:

$$\begin{aligned} P_n^{(v)} &= \frac{1}{\text{Card}(B_v)} \sum_{i \in B_v} \delta_{O_{(i)}} \\ P_n^{(-v)} &= \frac{1}{n - \text{Card}(B_v)} \sum_{i \notin B_v} \delta_{O_{(i)}}, \end{aligned}$$

where  $P_n^{(v)}$  is the empirical distribution of  $D_n^{(v)}$  and  $P_n^{(-v)}$  is the empirical distribution of  $D_n^{(-v)}$ . Let  $\hat{S}_K^j \in \mathcal{H}_1$ . We denote by  $\hat{S}_K^{j(-v)} = \hat{S}_K^j(D_n^{(-v)}; \cdot)$  the estimator based on  $D_n^{(-v)}$  of  $S_K^j$ . We define its empirical cross-validated risk estimator  $\hat{R}$  and its oracle counterpart  $\tilde{R}$  by:

$$\begin{aligned} \hat{R}(\hat{S}_K^j) &= \frac{1}{V} \sum_{v=1}^V R^{(P_n^{(v)})}(\hat{S}_K^{j(-v)}), \\ \tilde{R}(\hat{S}_K^j) &= \frac{1}{V} \sum_{v=1}^V R^{(P_0)}(\hat{S}_K^{j(-v)}), \end{aligned}$$

with

$$R^{(P_n^{(v)})}(\hat{S}_K^{j(-v)}) = \frac{1}{\text{Card}(B_v)} \sum_{i \in B_v} \mathbf{1}\{\hat{S}_K^{j(-v)}(W_i, Y_{K,i}^j) \neq A_i\}.$$

For each  $v \in \{1, \dots, V\}$ , conditionally on  $D_n^{(-v)}$ ,  $R^{(P_n^{(v)})}(\hat{S}_K^{j(-v)})$  is the empirical estimator of the true risk  $R^{(P_0)}(\hat{S}_K^{j(-v)})$  based on  $D_n^{(v)}$ .

The  $V$ -cross-validation principle consists in selecting among  $\mathcal{H}_1$  the classifier which minimizes the cross-validation risk. We define the cross-validated selector  $(\hat{K}, \hat{j})$  as:

$$(\hat{K}, \hat{j}) \in \arg \min_{(K,j)} \hat{R}(\hat{S}_K^j).$$

Its oracle counterpart is defined as:

$$(\tilde{K}, \tilde{j}) \in \arg \min_{(K,j)} \tilde{R}(\hat{S}_K^j).$$

For a new observation  $(W, Y_K^j)$  independent of  $D_n$ , the cross-validated classifier  $\hat{\phi}_1$  is defined as:

$$\hat{\phi}_1(D_n)(Z) = \mathbf{1}\left\{\hat{S}_{\hat{K}}^{\hat{j}}(D_n; Z) \geq \frac{1}{2}\right\}.$$

A natural way to evaluate the performances of the cross-validated classifier is to compare the risk difference between  $\tilde{R}(\hat{\phi}_1(D_n))$  and  $\tilde{R}(\hat{S}_{\tilde{K}}^{\tilde{j}}(D_n))$ . Dudoit and van der Laan (2005) provide the following asymptotic results:

$$0 \leq E \left[ \tilde{R}(\hat{\phi}_1(D_n)) - \mathcal{R}^* \right] \leq E \left[ \tilde{R}(\hat{S}_{\tilde{K}}^{\tilde{j}}(D_n)) - \mathcal{R}^* \right] + O \left( \frac{\log(\text{Card}(\mathcal{H}_1))}{\sqrt{np_n}} \right),$$

where  $p_n$  is the proportion of observations among  $n$  which are in the validation sets. Note that in the  $V$ -fold cross-validation framework  $p_n \simeq \frac{1}{V}$ .

## 5 Classification results

In this section, we present the results of our classification procedure described in Section 4 applied to the real dataset. We evaluate the performance of our classification on the real dataset in Section 5.1. Finally, we give an extension of the classification procedure in Section 5.2.

### 5.1 Performance of the classification procedure

The evaluation of our classification procedure applied to the real data relies on the leave-one-out rule.

Formally, let  $D_n = (O_{(1)}, \dots, O_{(n)})$  be the real dataset ( $n = 54$ ). For each  $i \in \{1, \dots, n\}$  we denote  $D_n \setminus \{O_{(i)}\}$  by  $D_{(-i)}$  and  $\{O_{(i)}\} \setminus \{A_{(i)}\}$  by  $Z_{(i)}$ . For each  $j = 1, 2$  and  $i = 1, \dots, n$ , we first compute  $\hat{\phi}_j^{(-i)}$  the estimator of  $\phi_j$  based on  $D_{(-i)}$  and evaluate its risk  $R_{(i)}^j$  on  $O_{(i)}$ :

$$R_{(i)}^j = \mathbf{1}_{\{\hat{\phi}_j^{(-i)}(Z_{(i)}) \neq A_{(i)}\}}.$$

Then, we compute the performance  $\text{Perf}_j$ :

$$\text{Perf}_j = \frac{1}{n} \sum_{i=1}^n \mathcal{R}_{(i)}^j,$$

which indicates the leave-one-out error of  $\hat{\phi}_j$  on the real dataset. For the  $V$ -fold cross-validation step we use  $V = 10$ . For the specifics of the super-learner used here we refer to the appendix in Chambaz and Denis (2011). The library of estimation procedure used here is the random forest procedure (Breiman, 2001) for different values of tuning parameters. The performance results are provided in Table 5.

| protocol        | $j = 1$ | $j = 2$ |
|-----------------|---------|---------|
| $\text{Perf}_j$ | 0.83    | 0.80    |

Table 5: Leave one out performances  $\text{Perf}_j$  of our classification procedure applied to the real dataset.  $\text{Perf}_1$  corresponds to the classifier  $\hat{\phi}_1$  (one protocol) and  $\text{Perf}_2$  corresponds to the classifier  $\hat{\phi}_2$  (two protocols).

The best performance is achieved by the classifier  $\hat{\phi}_1$  which involves only one protocol. The proportion of subjects correctly classified equals to 83%. So, we improve on the performance rates of correct classification compared to Chambaz and Denis (2011), which are respectively 74% with one protocol and 80% with two protocols. These results confirm that the new summary measures are more adapted to take into account the temporal dynamics of the protocols.

For each  $i = 1, \dots, n$ , it is always protocol 1 that is selected by the cross-validation for the construction of  $\hat{\phi}_1^{(-i)}$ . Conversely, the ranking of the protocols proposed by Chambaz and Denis (2011) shows that protocol 2 is more informative than protocol 1. It is not surprising since the information provided by a protocol for the classification depends on the definition of the summary measures used. Thus, a new definition of summary measures can involve a new ranking of the protocols.

Moreover, for each  $j = 1, 2$  and  $i = 1, \dots, n$ , it is generally  $K = 3$  (two change points) which is selected by the cross-validation for the construction of  $\hat{\phi}_1^{(-i)}$ . Therefore, only the two change points involved in the perturbation phase of the protocols seem to provide helpful information for the classification.

In order to evaluate the relevance of the change point estimation for the classification step, we compute the performance of the classification procedure with two fixed change points  $T_1 = 15$  and  $T_2 = 50$  seconds. We obtain a rate of correct classification equal to 66% with one protocol and 74% with two protocols. Thus, the information provided by the estimated change points is very relevant.

## 5.2 Extension of the summary measure

In this section, we propose an extension of our classification procedure by enriching the vector of summary measures defined in Section 4.1 with the summary measure used in Chambaz and Denis (2011). Let  $U^j$  denote the vector of these summary measures associated to

the protocol  $j$  defined by Chambaz and Denis (2011). The construction of  $U^j$  also relies on  $C_{1:N}$ , and is defined as follows:

$$U^j = (\bar{C}_1^+ - \bar{C}_1^-, \bar{C}_2^- - \bar{C}_1^+, \bar{C}_2^+ - \bar{C}_2^-)$$

where

$$\begin{aligned} \bar{C}_1^- &= \frac{\delta}{5} \sum_{i \in [10/\delta, 15/\delta[} C_i, & \bar{C}_1^+ &= \frac{\delta}{5} \sum_{i \in [15/\delta, 20/\delta]} C_i, \\ \bar{C}_2^- &= \frac{\delta}{5} \sum_{i \in [45/\delta, 50/\delta[} C_i, & \bar{C}_2^+ &= \frac{\delta}{5} \sum_{i \in [50/\delta, 55/\delta]} C_i. \end{aligned}$$

In this new classification procedure, for  $K = 3, 4$  and  $j = 1, 2$ , the summary measure  $Z_K^j$  writes as:

$$Z_K^j = (Y_K^j, U^j).$$

Therefore, for a given patient the observed data structure  $O$  writes as:

$$O = (W, Z_3^1, Z_4^1, Z_3^1, Z_4^1, A)$$

We then apply the classification procedure described in Section 4. The results are shown in Table 6. The best performance is achieved with the classifier that involves summary

| protocol  | $j = 1$ | $j = 2$ |
|-----------|---------|---------|
| Perf $_j$ | 0.83    | 0.91    |

Table 6: Leave-one-out performances Perf $_j$  of our classification procedure applied with the extended summary measures. Perf $_1$  corresponds to the classifier  $\hat{\phi}_1$  (one protocol) and Perf $_2$  corresponds to the classifier  $\hat{\phi}_2$  (two protocols).

measures from the two protocols. The proportion of subjects correctly classified equals to 91%, which is really much better than the performances obtained so far. Thus, the summary measures used by Chambaz and Denis (2011) seem to provide additional information for classification.

## 6 Discussion

In this article, we tackle the statistical problem of classifying subjects from a dataset which consists of complex trajectories. We model the trajectories by a stochastic process with unknown parameters that change at some unknown change points. Then summary measures which rely on estimators of the change points and the parameters of the process are proposed. The classification procedure based only on the summary measures involves the  $V$ -fold cross validation principle. We improve on previous results (Chambaz and Denis, 2011), showing that it is possible to obtain a 91% rate of correct classification (evaluated by the leave-one-out rule).

We show that data modeling, especially as regards change point estimation, is crucial for the classification step. The stochastic process modeling with change point estimation makes it possible to take into account the temporal dynamics of the data and the specifics of the protocols. Moreover, we show that the CIR process may be well adapted for modeling trajectories  $C_{1:N}$ . In previous investigations, we studied other classes of stochastic processes, as the Ornstein-Uhlenbeck process and geometric Brownian motion, but the results of the modeling were not relevant.

In this paper, we focus on summary measures based on trajectories  $C_{1:N}$  pertaining to distances. A natural way to enrich these summary measures would be to add information pertaining to patients' orientation on the force platform during the protocols. Thus, summary measures based on a 2-D stochastic process (as the Ornstein-Uhlenbeck process) modeling of  $B_{1:N}$  could provide helpful information for classification. We will consider this approach in future work.

## Acknowledgements

The author thanks warmly his supervisor A.Samson for her helpful suggestions throughout this work.

## References

- Audibert, J.-Y. and Tsybakov, A. B. (2007). Fast learning rates for plug-in classifiers. *Ann. Statist.* **35**, 608–633. ISSN 0090-5364.
- Basseville, M. and Nikiforov, I. V. (1993). *Detection of abrupt changes: theory and application*. Prentice Hall Information and System Sciences Series. Prentice Hall Inc., Englewood Cliffs, NJ. ISBN 0-13-126780-9.
- Biau, G., Bunea, F. and Wegkamp, M. (2005). Functional classification in Hilbert spaces. *IEEE Trans. Inform. Theory* **51**, 2163–2172.
- Breiman, L. (2001). Random forests. *Machine Learning* **45**, 5–32.
- Chambaz, A. and Denis, C. (2011). Classification in postural style. *preprint MAP5 2011-09*.
- Dudoit, S. and van der Laan, M. J. (2005). Asymptotics of cross-validated risk estimation in estimator selection and performance assessment. *Stat. Methodol.* **2**, 131–154. ISSN 1572-3127.
- Geisser, S. (1975). The predictive sample reuse method with applications. *J. Amer. Statist. Assoc* **70**, 320–328.
- Gijbels, I., Hall, P. and Kneip, A. (1999). On the estimation of jump points in smooth curves. *Ann. Inst. Statist. Math.* **51**, 231–251. ISSN 0020-3157.

- Hall, P., Poskitt, D. and Presnell, B. (2001). A functional data-analytic approach to signal discrimination. *Technometrics* **43**.
- James, G. and Hastie, T. (2001). Functional linear discriminant analysis for irregularly sampled curves. *J. Roy. Statistic. Soc. Ser. B* **63**, 533–550.
- Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.* **24**, 211–229. ISSN 0303-6898.
- Kloeden, P. and Platen, E. (1992). *Numerical Solution of Stochastic Differential Equation*, vol. 23 of *Stochastic modelling and applied probability*. Springer.
- Lavielle, M. (1999). Detection of multiple changes in a sequence of dependent variables. *Stochastic Process. Appl.* **83**, 79–102. ISSN 0304-4149.
- Lavielle, M. (2005). Using penalized contrasts for the change-point problem. *Signal Processing* **85**, n. 8, 1501–1510.
- Lavielle, M. and Ludeña, C. (2000). The multiple change-points problem for the spectral distribution. *Bernoulli* **6**, 845–869. ISSN 1350-7265.
- Ramsay, J. and Silverman, B. (2005). *Functional Data Analysis*. Springer, New-York.
- Tsybakov, A. B. (2004). *Introduction à l'estimation non-paramétrique*, vol. 41 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer-Verlag, Berlin.
- van der Laan, M. J., Polley, E. C. and Hubbard, A. E. (2007). Super learner. *Stat. Appl. Genet. Mol. Biol.* **6**, Art. 25, 23 pp. (electronic). ISSN 1544-6115.
- van der Laan, M. J. and Rubin, D. (2006). Targeted maximum likelihood learning. *Int. J. Biostat.* **2**, Art. 11, 40. ISSN 1557-4679.