A hidden Markov model addressing measurement errors in the response and replicated covariates for continuous nondecreasing processes

Article *in* Biostatistics · February 2019

DOI: 10.1093/biostatistics/kxz004

CITATIONS
4

READS
179

4 authors:

Lizbeth Naranjo
Universidad Nacional Autónoma de México
35 PUBLICATIONS **376** CITATIONS
SEE PROFILE

Carlos J. Pérez
Universidad de Extremadura
125 PUBLICATIONS **1,870** CITATIONS
SEE PROFILE

Ruth Fuentes-Garcia
Universidad Nacional Autónoma de México
27 PUBLICATIONS **244** CITATIONS
SEE PROFILE

Jacinto Martín
Universidad de Extremadura
92 PUBLICATIONS **1,929** CITATIONS
SEE PROFILE

A hidden Markov model addressing measurement errors in the response and replicated covariates for continuous nondecreasing processes

LIZBETH NARANJO*

Departamento de Matemáticas, Facultad de Ciencias, Universidad Nacional Autónoma de México, Circuito Exterior s/n, Ciudad Universitaria, Del. Coyoacán, C.P. 04510 Ciudad de México, Mexico

lizbethna@ciencias.unam.mx

CARLOS J. PÉREZ

Departamento de Matemáticas, Facultad de Veterinaria, Universidad de Extremadura, Avda. de la Universidad s/n, C.P. 10003 Cáceres, Spain

RUTH FUENTES-GARCÍA

Departamento de Matemáticas, Facultad de Ciencias, Universidad Nacional Autónoma de México, Circuito Exterior s/n, Ciudad Universitaria, Del. Coyoacán, C.P. 04510 Ciudad de México, Mexico

JACINTO MARTÍN

Departamento de Matemáticas, Facultad de Ciencias, Universidad de Extremadura, Avda. de Elvas s/n, C.P. 06006 Badajoz, Spain

SUMMARY

Motivated by a study tracking the progression of Parkinson's disease (PD) based on features extracted from voice recordings, an inhomogeneous hidden Markov model with continuous state-space is proposed. The approach addresses the measurement error in the response, the within-subject variability of the replicated covariates and presumed nondecreasing response. A Bayesian framework is described and an efficient Markov chain Monte Carlo method is developed. The model performance is evaluated through a simulation-based example and the analysis of a PD tracking progression dataset is presented. Although the approach was motivated by a PD tracking progression problem, it can be applied to any monotonic nondecreasing process whose continuous response variable is subject to measurement errors and where replicated covariates play a key role.

Keywords: Bayesian methods; Measurement error; Nondecreasing process; Parkinson's disease; Replicated measurements; Voice features.

1. INTRODUCTION

Parkinson's disease (PD) is a neurodegenerative disorder with symptoms that persist and worsen over time. According to the USA Parkinson's Disease Foundation and the European Parkinson's Disease Association,

nearly one million people in the USA and 1.2 million in Europe are living with this medical condition. Its cause is unknown, and at present there is no cure. There are treatment options, such as medication and surgery, to manage its symptoms. Early detection and subsequent progression tracking play a fundamental role for the quality of life of the patient.

Voice and speech are also affected by PD. This is because voice production depends on respiratory and articulatory functions which are disrupted in people with PD. Non-dopaminergic changes can also affect cognition and mood, which may have an important impact on communication. Vocal impairment is recognized as one of the earliest signs of PD (Duffy, 2005). Even in the early stages of PD, there are subtle abnormalities in speech that might not be perceptible to listeners, but can be evaluated objectively by performing acoustic analyses of recorded speech, i.e., the voices of the subjects are recorded to extract some specific characteristics of the speech signals, which are used to detect the disease or track its progression.

Voice recordings have become a potential non-invasive low-cost biomarker for PD. In the recent literature, there are many techniques that have considered feature extraction from speech recordings and machine learning to distinguish healthy people from those with PD (Little *and others*, 2009; Sakar *and others*, 2013; Hariharan *and others*, 2014; Novotny *and others*, 2014; Naranjo *and others*, 2016, 2017). However, less attention has been paid to tracking the disease's progression. Some authors have addressed the problem of finding a statistical model to relate speech parameters and Unified Parkinson's Disease Rating Scale (UPDRS) scores (Tsanas *and others*, 2010; Eskidere *and others*, 2012; Castelli *and others*, 2014; Arias-Vergara *and others*, 2016). UPDRS is the most commonly used scale to follow the longitudinal course of PD (Ramaker *and others*, 2002). In general, recruiting a large number of people with PD is difficult, and even more so to engage them for tracking over a long period of time. This has resulted in their being a relatively small number of acoustic databases. In addition, these databases have been specifically designed for the purpose of diagnosis rather than for tracking the progression of the disease. Gender and the variety of medications also complicate the collection of the data.

The database most commonly used to develop methods for tracking PD progression with acoustic features is based on one multicenter study, which tested the feasibility of a computer-based At-Home Testing Device (AHTD) (Goetz *and others*, 2009). The AHTD database contains acoustic features extracted from replicated voice recordings and is publicly available. Although this voice recording database has been carefully collected and the acoustic features have been extracted appropriately, regression models have been applied without matching the right statistical model to the experimental design (Tsanas *and others*, 2010; Eskidere *and others*, 2012). It has become necessary to develop methods employing a statistical model that is appropriate to the experimental design so as to provide a more accurate tracking of the disease's progression. Such a new approach should address simultaneously three aspects: (i) a time-dependent process, which is assumed to be continuous and monotonic nondecreasing; (ii) measurement errors in the response; and (iii) replications in the covariates. The need to address these three aspects in the model will be justified in the next three paragraphs.

PD, as a neurodegenerative disease, is characterized by a worsening of its status over time. The natural progression of the disease is caused by an irreversible and continual death of neurons, which is reflected in disease status measures such as UPDRS scores. The natural history of PD has been widely studied. For instance, a linear progression trend is one biologically plausible hypothesis (see, e.g., Chan and Holford, 2001 and references therein). Other studies have suggested that the disease may progress more rapidly initially and the rate of deterioration slows in more advanced stages of the disease (see, e.g., Jankovic and Kapadia, 2001). Consensus is achieved in considering PD as a nondecreasing process. Other medical diseases, such as caries, are also characterized by nondecreasing processes (García-Zattera *and others*, 2012).

UPDRS enables the quantification of the type, number, and severity of extrapyramidal signs. Since this quantification is partially based on subjective criteria, disagreement among raters in the interpretation of these criteria introduces measurement errors in these scores (Richards *and others*, 1994). This may

result in an imprecise clinical characterization of the patient and in the loss of reliability when UPDRS values are analyzed. Therefore, scoring UPDRS is subjective and the clinician's assessment of a patient's score may be subject to error. Post *and others* (2005) suggested that the amount by which raters disagree should be quantified before starting longitudinal studies of disease progression or clinical trials. For example, Satten and Longini (1996) proposed a hidden Markov model subject to measurement error to describe the progression of longitudinal observations of CD4, a marker of the progression of the human immunodeficiency virus (HIV).

Several voice recordings were obtained for each patient, in different sessions. The acoustic features have been extracted from them. For each subject, the acoustic features should be identical at a concrete session, but imperfections in technology and their own biological variability result in differences in the replicated features (see, e.g., Naranjo *and others*, 2016). Naranjo *and others* (2017) addressed both the within-subject variability and progression over time.

Motivated by PD progression tracking based on the AHTD database, an hidden Markov model with continuous state-space is proposed in this article. This model addresses measurement errors in the response and replications in the covariates, as well as employing a statistical treatment appropriate for nondecreasing processes. A fully Bayesian framework will be presented, and an efficient Markov chain Monte Carlo (MCMC) method has been derived to solve the computational problem. The approach has been applied to simulated data in order to evaluate its performance and also to the AHTD database. As far as the authors' knowledge extends, an approach addressing these three issues simultaneously (nondecreasing process, measurement error, and replications) has not yet been proposed.

The outline of this article is as follows. Section 2 presents the motivation for this approach. Section 3 presents the proposed approach and how it handles nondecreasing time processes, measurement errors in the response, replicated measurements in the explanatory variables, and conditional independence assumptions. In Section 4, the Bayesian analysis is presented. Section 5 is devoted to an experimental study with the AHTD database and the significance of the results. Next, conclusions are presented in Section 6. Lastly, a supplementary material available at *Biostatistics* online document presents the derivation of the full conditional distributions, a sensitivity analysis of the AHTD application, and a simulation-based experiment.

2. Motivating problem

The AHTD study was presented by Goetz *and others* (2009). Originally, 52 early-stage subjects, having idiopathic PD diagnosed within the previous 5 years of the trial's onset, were recruited from six medical centers in the USA. Patients could not be on symptomatic therapies for PD. Inclusion criteria considered that by best estimates and follow-ups the patients could remain untreated during the study period of 6 months. Subjects who showed clinical decline needed dopaminergic medication and were excluded from the study. The subjects were physically assessed and assigned UPDRS scores at baseline, 3 months, and 6 months later into the trial. During the 6 months the trial lasted, six phonations of sustained /a/ were recorded weekly for each patient. This means that the patients produced six vowel phonations for some seconds in each session, trying to keep the pitch and loudness as constant as possible.

Tsanas *and others* (2010) considered the information and voice recordings for the 42 out of the 52 patients who remained in the trial and for which at least 20 valid study sessions were performed. The voice recordings from the AHTD study are not publicly available, but the features extracted and analyzed by Tsanas *and others* (2010) for these 42 patients (28 men and 14 women) are available online in the AHTD database at the UCI Machine Learning Repository (https://archive.ics.uci.edu/ml/datasets/Parkinsons+Telemonitoring). Specifically, the database contains the name, age, gender, time interval from baseline recruitment date, motor UPDRS and total UPDRS (in both cases, real and interpolated values), and 16 biomedical voice measures, i.e., five measures of

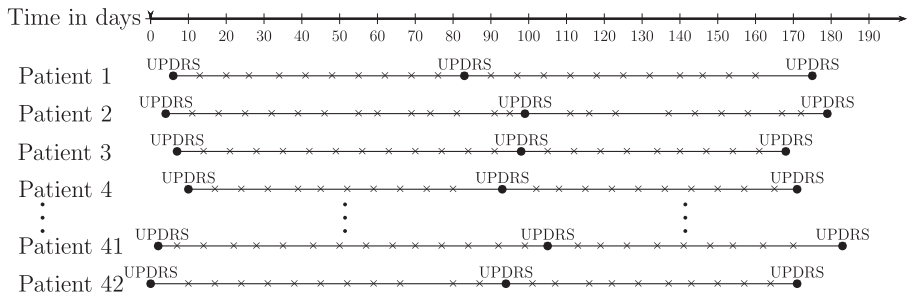

Fig. 1. Experimental design in a time scale based on days. Black points represent replicated voice recordings and UPDRS measures. Crosses represent only replicated voice recordings.

variation in fundamental frequency, six measures of variation in amplitude, two measures of the ratio of the noise to the tonal components in the voice, one nonlinear dynamical complexity measure, one signal fractal scaling exponent, and one nonlinear measure of fundamental frequency variation. This provided acoustic features with temporal and within-subject dependencies. The experimental design is summarized in Figure 1.

In the analysis, we performed for the 42 subjects, only real UPDRS scores (at 0, 3, and 6 months) and six acoustic features extracted from their corresponding voice recording sessions at each time point are used from the AHTD database, avoiding the use of interpolated UPDRS scores. A detailed explanation of the selection of the acoustic features is presented in Section 5. Note that the AHTD database considers both motor and total UPDRS scores. 'Motor UPDRS' refers to Part III of the questionnaire applied to each subject, and is the section that assesses the motor skills. 'Total UPDRS' refers to the whole questionnaire and is the core outcome measure of many studies. Higher scores indicate more advanced stages of the disease. Figure 2 shows the profiles of the subjects for motor and total UPDRS scores at baseline, 3, and 6 months.

Since the patients were unmedicated, it is to be expected that both UPDRS scores are nondecreasing, since PD symptoms do not improve without medication. However, it can be observed that many profiles show decreasing patterns in the response. For example, patient number 17 had motor UPDRS observed scores of 19.1, 33.9, and 20.4 at baseline, 3 months, and 6 months, respectively. This means that in 3 months, an increase of 77.5% was observed, in contrast with other similar early-stage studies, such as Shoulson and Parkinson Study Group (1992), which found a rate of progression for one complete year of 51.1% for unmedicated patients. Even more, in the next 3 months, a decrease of 60.2% was found, which is not realistic under no medication. Post *and others* (2005) found an inter-rater variability for motor UPDRS of less than five points (out of a total of 68 points). There are some profiles with motor scores outside of the range attributable to a reasonable inter-rater variability. The same happens for the total UPDRS. We believe that some unrealistic fluctuations may be explained by a poor UPDRS assessment.

This problem has been the motivation for deriving a statistical model that appropriately matches the experimental design, addressing three aspects that have not been considered simultaneously, as far as the authors' knowledge extends, i.e., (i) assuming the process to be monotonic nondecreasing in time, since under no medication PD is not expected to improve; (ii) including mechanisms to correct the possible errors when measuring UPDRS scores; and (iii) addressing the within-subject variability resulting from replicated acoustic features (not the statistical concept of repeated measurements) originating from several voice recordings obtained at the same time for each subject.

Only gender and acoustic features will be used as predictors in this article, since the goal is that the model can be implemented in the future in an expert system. An expert system should be able to remotely obtain voice recordings from Smartphones, and process them by extracting automatically the gender and

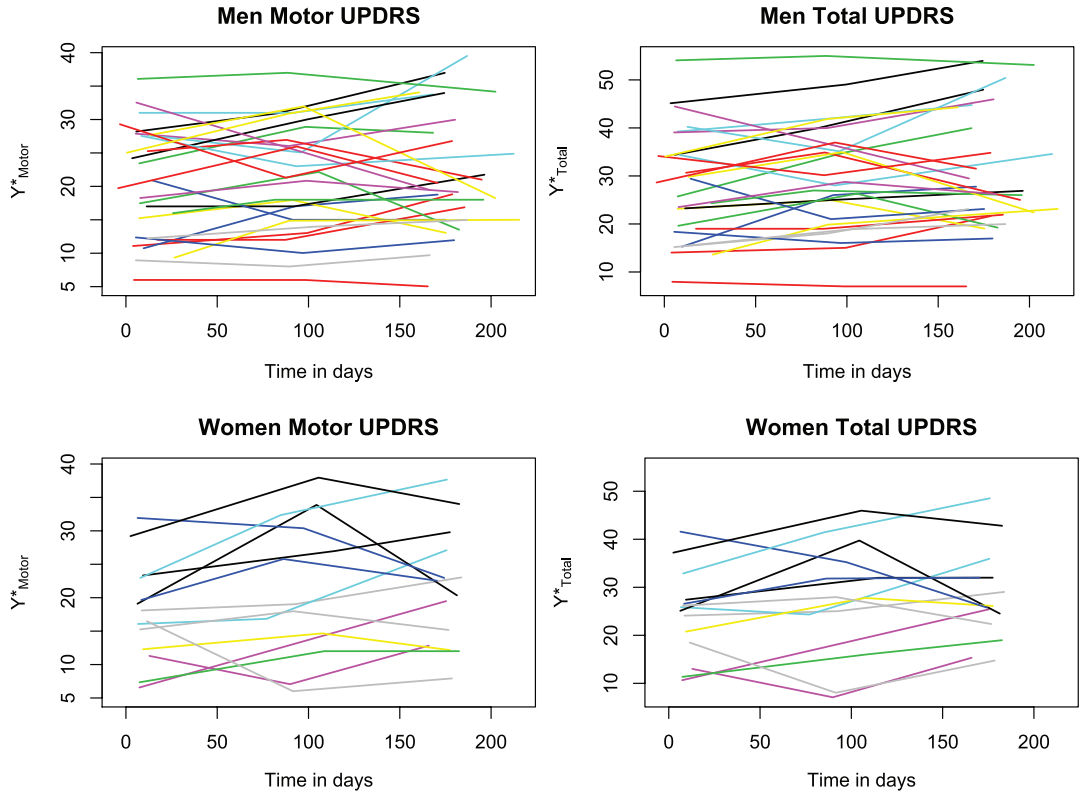

Fig. 2. Profiles of the AHTD database.

acoustic features. However, other variables as the time since diagnosis, could be included if they were available. An accurate system would help neurologists to carry out a more continuous follow-up.

3. Approach

The approach considers a time-dependent process, which is continuous and monotonic nondecreasing. It addresses the measurement errors in the response and the replications in the explanatory variables, and it takes into account several conditional independence assumptions. This approach seeks to appropriately match the experimental design and to provide accurate estimations to assist neurologists in tracking the progression of the disease in their patients.

3.1. *A continuous monotonic nondecreasing process*

Suppose that the UPDRS score is obtained for the ith subject, $i = 1, \ldots, N$, at time points $\{t_{ik} : k = 1, \ldots, K\}$. Without loss of generality, the subjects are assumed to have the same number of time points, which will be denoted by K. Let Y_{ik} be the true continuous response for the ith subject at time t_{ik}, and put $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{iK})'$. They denote the UPDRS score process, which is assumed to be monotonically nondecreasing, i.e., $Y_{i1} \leq Y_{i2} \leq \cdots \leq Y_{i,K-1} \leq Y_{iK}$, for all $i = 1, \ldots, N$. This assumption is consistent with the fact that the PD worsens with time if there is no medication treatment.

Note that Y_{ik} is not observed. Instead, Y_{ik}^* is recorded and is subject to measurement error. Hence, Y_{ik} is treated as a latent variable. The unknown response vector $\boldsymbol{Y}_i$ is related to a set of covariates $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$. Let $\boldsymbol{x}_{ik}$ be an L-dimensional vector of time-varying covariates associated with the ith subject at time point t_{ik}, measured with error, and suppose $\boldsymbol{x}_i = (\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{iK})'$ represents the vector of covariates for the ith subject. These covariates will be the acoustic features. In addition, let $\boldsymbol{z}_{ik}$ be an M-dimensional vector, possibly the time-varying covariates associated with the ith subject at time point t_{ik}, and let $\boldsymbol{z}_i = (\boldsymbol{z}_{i1}, \ldots, \boldsymbol{z}_{iK})'$ represent the vector of observed covariates for the ith subject. These covariates could be the gender and age.

Let η_{ik} be the linear predictor for the ith subject at time t_{ik} that consists of linear combinations of the covariates $\boldsymbol{x}_{ik}$ and $\boldsymbol{z}_{ik}$,

$$\eta_{ik} = \alpha + \boldsymbol{x}_{ik}'\boldsymbol{\beta} + \boldsymbol{z}_{ik}'\boldsymbol{\gamma}, \tag{3.1}$$

where $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are, respectively, the L-dimensional and M-dimensional vectors of coefficients for the covariates $\boldsymbol{x}_{ik}$ and $\boldsymbol{z}_{ik}$, and α is the intercept parameter. Note that different intercept parameters for each time (α_k) or different slope parameters depending on the time ($\boldsymbol{\gamma}_{t_{ik}}$ or $\boldsymbol{\gamma}_k$) could be incorporated with the $\boldsymbol{z}_{ik}$ covariates. Also, in Equation (3.1), the parameters $\boldsymbol{\beta}_i$ and $\boldsymbol{\gamma}_i$ could be defined in an hierarchical structure as well as taking into account random effects by means of α_i.

Assume that for the first time point, $k = 1$, the true unobserved continuous response Y_{i1} is related to the linear predictor η_{i1} through a normal distribution,

$$Y_{i1} \sim \text{Normal}\left(\eta_{i1}, \sigma^2\right). \tag{3.2}$$

For the next time points, $k = 2, \ldots, K$, due to the monotonic nondecreasing continuous nature of the process ($Y_{i,k-1} \leq Y_{ik}$), the true unobserved continuous response Y_{ik} is related to the linear predictor η_{ik} through a truncated normal distribution, namely

$$Y_{ik}\big|Y_{i,k-1} = y_{i,k-1} \sim \text{Normal}\left(\eta_{ik}, \sigma^2\right) \text{I}\left[Y_{ik} \geq y_{i,k-1}\right], \tag{3.3}$$

where $\text{I}[\cdot]$ denotes the indicator function, i.e., $\text{I}[A] = 1$ if A is true, and $\text{I}[A] = 0$ otherwise. In this way, the first-order Markov chain property for continuous processes is assumed and the truncation allows the nondecreasing restriction to be satisfied at the same time as maintaining model conjugacy.

In addition, Equations (3.2) and (3.3) could also be employed for heteroskedastic models, i.e., by using different variances σ_k^2 for each time point or σ_i^2 for each subject. However, in the AHTD database, homoskedasticity has been assumed, because 6 months is not enough time to properly notice non-constant variances.

3.2. *Measurement errors in the response*

The true response variables Y_{ik} are latent variables assumed to be prone to measurement errors. Let Y_{ik}^* be the non error-free continuous response for subject i at time t_{ik}, and set $\boldsymbol{Y}_i^* = (Y_{i1}^*, \ldots, Y_{iK}^*)'$. They denote the observed UPDRS scores, which may have been measured with errors, and therefore, could have both nondecreasing or decreasing patterns. For each continuous latent variable, its measurement error is defined as follows, where Y_{ik}^* is assumed to be conditionally dependent on Y_{ik} and normally distributed,

$$Y_{ik}^*|Y_{ik} = y_{ik} \sim \text{Normal}\left(y_{ik}, \tau^2\right). \tag{3.4}$$

Equation (3.4) is the usual expression for defining the classical additive measurement error model (Carroll *and others*, 2006; Buonaccorsi, 2010), i.e., $Y_{ik}^* = Y_{ik} + \varepsilon_{ik}$, where $\varepsilon_{ik} \sim \text{Normal}(0, \tau^2)$ and ε_{ik} is independent of Y_{ik}.

Note that in the AHTD database there is no information about the neurologists who examined the patients. However, if there were detailed information about who examined each patient i at each time k, then a different variance parameter $\tau^2_{\xi_{ik}}$ could be used to estimate the degree of error made by each examiner ξ_{ik}.

3.3. *Replications in the explanatory variables*

Assume that the covariates $\boldsymbol{z}_i$ are exactly known, but that the $\boldsymbol{x}_i$ have been measured with J replicates. More specifically, $\boldsymbol{X}^*_{ikj} = (X^*_{ikj1}, \ldots, X^*_{ikjL})'$ is the jth replication of the unknown vector of covariates $\boldsymbol{X}_{ik} = (X_{ik1}, \ldots, X_{ikL})'$ and a linear relationship (additive measurement error model (Carroll *and others*, 2006; Buonaccorsi, 2010)), is assumed. One must assume a distribution for the latent variables $\boldsymbol{X}_{ik}$. In this case, multivariate normal distributions are used, i.e., $\boldsymbol{X}_{ik} \sim \text{Normal}_L(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Now, instead of the covariates $\boldsymbol{X}_{ik}$, their replicates $\boldsymbol{X}^*_{ikj}$ are recorded, and

$$\begin{aligned} \boldsymbol{X}^*_{ikj} &= \boldsymbol{X}_{ik} + \boldsymbol{\delta}_{ikj}, \\ \boldsymbol{\delta}_{ikj} &\sim \text{Normal}_L(\boldsymbol{0}, \boldsymbol{\Lambda}), \end{aligned} \tag{3.5}$$

for $j = 1, \ldots, J$, where $\boldsymbol{\Lambda}$ is an $L \times L$ matrix of variances and covariances, which allows for the replicates between covariates to be dependent or independent, by using a non-diagonal or a diagonal matrix $\boldsymbol{\Lambda}$, respectively. In the case of the AHTD database, the assumption of multivariate normal distributions takes into account the correlations between the errors in the acoustic features. The error vector $\boldsymbol{\delta}_{ikj}$ is independent of $\boldsymbol{X}_{ik}$, implying that $\boldsymbol{X}^*_{ikj}$ is a surrogate for $\boldsymbol{X}_{ik}$. Moreover, the errors are independent for each ith subject, at each kth time point t_{ik}, and for each jth replication.

Note that the covariates $\boldsymbol{X}_{ik}$ denote the acoustic features of the ith subject at time point t_{ik}, but they are not exactly known. Instead, the covariates $\boldsymbol{X}^*_{ikj}$ are the observed acoustic features, but they are not identical for each subject in a concrete time. The relation defined in (3.5) allows taking into account the within-subject variability.

Independence-based methods should not be used when the data have been obtained by replicating voice recordings from the same subjects. This fact artificially increases the sample size and provides inaccurate estimations. Some authors have proposed to aggregate related data before learning, by using some different functions as the mean, minimum, maximum, or a linear trend prediction. However, simplifying all the replicated measures from each subject into one single measure (for each voice feature) leads to a loss of information. The within-subject variability is removed when aggregating data and statistical methods should address this instead of removing it.

3.4. *Conditional independence assumptions*

The full approach, which comprises Equations (3.2), (3.3), (3.4), and (3.5), is an inhomogeneous hidden Markov model with continuous state-space. Figure 3 displays the probabilistic graphical model showing the dependencies among the variables in the proposed model, with the usual convention of graphical models where square boxes represent observed variables and ovals represent latent variables. The direction of the arrows indicates conditional dependence.

The process is characterized by the following conditional independence assumptions:

(A.1) $\perp\!\!\!\perp_{1 \le i \le N} \boldsymbol{Y}^*_i \mid \boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_N, \tau^2$, i.e., the observed continuous response vectors for each subject are independent given the true unobserved continuous responses and the variance of the measurement error.

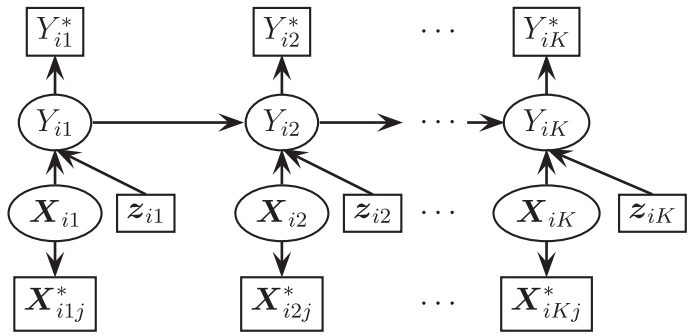

Fig. 3. Graphical representation of the proposed model. Square boxes represent observed variables and ovales represent latent variables. The direction of the arrows indicates conditional dependence.

(A.2) $\boldsymbol{Y}^*_i \perp\!\!\!\perp \boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{i-1}, \boldsymbol{Y}_{i+1}, \ldots, \boldsymbol{Y}_N, \mid \boldsymbol{Y}_i, \tau^2, \forall i$, i.e., the distribution of the observed continuous response vector for a subject only depends on his/her true unobserved continuous response vector and variance of the measurement error.

(A.3) $\perp\!\!\!\perp_{1\le k\le K} Y^*_{ik} \mid \boldsymbol{Y}_i, \tau^2, \forall (i,k)$, i.e., the observed continuous responses for a subject at one point of time is independent of that for the same subject at any other point of time, given the subject's true unobserved continuous response vector and the measurement error variance.

(A.4) $Y^*_{ik} \perp\!\!\!\perp Y_{i1}, \ldots, Y_{i,k-1}, Y_{i,k+1}, \ldots, Y_{iK} \mid Y_{ik}, \tau^2$, i.e., the distribution of the observed continuous response in the kth examination only depends on the subject's true unobserved continuous response at the same examination and variance of the measurement error.

The conditional independence assumptions (A.1)–(A.4) lead to distributions, which are free of other model parameters. In particular, they are independent on parameters σ^2, α, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\Lambda}$, and also they are independent on latent variables $\boldsymbol{X}$ and observed variables $\boldsymbol{X}^*$ and $\boldsymbol{z}$. Note that assumptions (A.1)–(A.4) are a natural extension of the usual conditional independence assumptions for the hidden Markov model (Cappé *and others*, 2005). These assumptions are analogous to those defined by García-Zattera *and others* (2012), but with a different approach and in a different context. Moreover, here the independence assumptions are defined for the latent continuous variables by using both the latent continuous response variable Y_{ik} and the observed continuous response variable subject to measurement error Y^*_{ik}.

4. BAYESIAN ANALYSIS

This section presents the prior distributions, which together with the model specified in the previous section, enable the use of a MCMC method to sample from the posterior distribution of interest (Gilks *and others*, 1996).

4.1. *The prior distributions*

A prior distribution is defined. Some components of the prior are conditionally conjugate. The coefficients in the linear predictor are specified to be normal: $\alpha \sim \text{Normal}_1(a, A)$, $\boldsymbol{\beta} \sim \text{Normal}_L(\boldsymbol{b}, \boldsymbol{B})$, and $\boldsymbol{\gamma} \sim \text{Normal}_M(\boldsymbol{c}, \boldsymbol{C})$.

The prior distribution for the variance of the response variable is specified as inverse Gamma: $\sigma^2 \sim \text{InverseGamma}(s_\sigma, r_\sigma)$. The prior distribution for the variance of the measurement error is also specified as inverse Gamma: $\tau^2 \sim \text{InverseGamma}(s_\tau, r_\tau)$.

Finally, the prior distribution for the covariance matrix of the replicated covariates is specified as inverse Wishart: $\boldsymbol{\Lambda} \sim \text{InverseWishart}(\boldsymbol{V}, \nu)$.

Note that all these distributions allow obtaining the full conditional posterior distributions easily. Moreover, if initial information is available, prior informative distributions can be elicited (O'Hagan *and others*, 2006).

4.2. *Exploring the posterior distribution*

Under the conditional independence assumptions defined in subsection 3.4, the likelihood function considering the observed and latent variables is given by

$$\begin{aligned}
&\mathcal{L}\left(\boldsymbol{Y}, \boldsymbol{X}, \alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma^2, \tau^2, \boldsymbol{\Lambda} \mid \boldsymbol{Y}^*, \boldsymbol{X}^*, \boldsymbol{z}\right) \\
&= \prod_{i=1}^{N} \left\{ \left[\prod_{k=1}^{K} \mathrm{P}\left(Y_{ik}^* | Y_{ik}, \tau^2\right) \right] \mathrm{P}\left(Y_{i1} | \boldsymbol{X}_{i1}, \boldsymbol{z}_{i1}, \alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma^2\right) \right. \\
&\left. \times \left[\prod_{k=2}^{K} \mathrm{P}\left(Y_{ik} | Y_{i,k-1}, \boldsymbol{X}_{ik}, \boldsymbol{z}_{ik}, \alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma^2\right) \right] \left[\prod_{k=1}^{K} \left\{ \prod_{j=1}^{J} \mathrm{P}\left(\boldsymbol{X}_{ikj}^* | \boldsymbol{X}_{ik}, \boldsymbol{\Lambda}\right) \right\} \mathrm{P}\left(\boldsymbol{X}_{ik}\right) \right] \right\}.
\end{aligned} \tag{4.6}$$

The joint posterior distribution of the unobservable latent variables $\boldsymbol{Y}$ and $\boldsymbol{X}$, and the parameters α, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, σ^2, τ^2 and $\boldsymbol{\Lambda}$ is obtained by using the likelihood function (4.6) and the prior distributions defined in Section 4.1, and it is given by

$$\begin{aligned}
&\pi\left(\boldsymbol{Y}, \boldsymbol{X}, \alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma^2, \tau^2, \boldsymbol{\Lambda} \mid \boldsymbol{Y}^*, \boldsymbol{X}^*, \boldsymbol{z}\right) \\
&\propto \mathcal{L}\left(\boldsymbol{Y}, \boldsymbol{X}, \alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma^2, \tau^2, \boldsymbol{\Lambda} \mid \boldsymbol{Y}^*, \boldsymbol{X}^*, \boldsymbol{z}\right) \pi\left(\alpha\right) \pi\left(\boldsymbol{\beta}\right) \pi\left(\boldsymbol{\gamma}\right) \pi\left(\sigma^2\right) \pi\left(\tau^2\right) \pi\left(\boldsymbol{\Lambda}\right).
\end{aligned} \tag{4.7}$$

This approach uses the relationship between the covariates and the latent variables jointly with the prior distributions to achieve a posterior inference of the latent variables and model parameters.

A graphical representation of the proposed approach is shown in Figure S1 in supplementary material available at *Biostatistics* online. This graphical representation is based on doodle objects of WinBUGS (Lunn *and others*, 2000) and indicates that the model can also be implemented in general Bayesian software, such as OpenBUGS, JAGS, or STAN.

The joint posterior density is not directly tractable for computing. Hence, a Gibbs sampling algorithm with Metropolis-Hastings steps has been developed. The supplementary material available at *Biostatistics* online document presents the full conditional distributions for the proposed approach. The algorithm has been developed and implemented in the R language (https://cran.r-project.org/) and JAGS (http://mcmc-jags.sourceforge.net/). The source code and instructions that allow to replicate the analysis for a simulation-based dataset can be found in the GitHub repository through the link https://github.com/lizbethna/HMMprogressionUPDRS.

5. Application to the AHTD database

Some acoustic features in the AHTD database are highly correlated since they describe similar measurements. Therefore, they provide redundant information to the model. For example, there are five jitters, measuring the variation in fundamental frequency, but with slight differences in the formulation

(Baken and Orlikoff, 2000). They are pairwise highly correlated. In the signal processing field, it is usual to obtain similar features and use all of them, leading in many cases to multicollinearity problems. Naranjo *and others* (2017) used one acoustic feature from each group in a PD detection context. We will also consider this reduction in our analysis, by using Jitter in Percentage, Local Shimmer, Harmonic-to-Noise Ratio (HNR), Recurrence Period Density Entropy (RPDE), Detrended Fluctuation Analysis (DFA), and Pitch Period Entropy (PPE).

The following specifications will be used to estimate both the clinician's motor UPDRS scores (motor UPDRS) and the clinician's total UPDRS scores (total UPDRS) based on acoustic features and gender as covariables. Three scenarios have been considered based on the physiological differences in the vocal apparatus of men and women: (i) the acoustic features and the gender are considered as predictor variables for all subjects; (ii) only men are considered with their acoustic features as predictor variables; and (iii) only women are considered with their acoustic features as predictor variables. In each scenario, the acoustic features are individually standardized to have mean 0 and standard deviation 1.

The following prior distributions have been used: $\alpha \sim \text{Normal}(0, 10000)$, $\boldsymbol{\beta} \sim \text{Normal}_L(\mathbf{0}, \text{diag}(10000))$ with $L = 6$, $\gamma \sim \text{Normal}(0, 10000)$, $\sigma^2 \sim \text{InverseGamma}(0.001, 0.001)$, $\tau^2 \sim \text{InverseGamma}(0.001, 0.001)$, $\boldsymbol{X}_{ik} \sim \text{Normal}_L(\mathbf{0}, \text{diag}(1))$ and $\boldsymbol{\Lambda} \sim \text{InverseWishart}(\text{diag}(1), L)$.

Gelman (2006) reported that using the inverse gamma as prior distributions for the variance parameters could provide sensitive results in hierarchical models. A sensitivity analysis is presented in the supplementary material available at *Biostatistics* online document to study the robustness of the prior distribution of σ^2 and τ^2. Different prior distributions are used and the results show that the choice of prior distribution does not affect the conclusions.

A total of 30 000 iterations with a burn-in of 10 000 and a thinning period of 10 generated values was used, yielding a sample of length 2000. With these specifications, the chain generated by using the MCMC sampling algorithm appears to have converged. The convergence analysis was performed by using the `BOA` package (Smith, 2007). The chains require a long burn-in period and the previous specifications are enough to provide evidence of convergence for all parameters.

The proposed approach provides information about the relative contribution of the voice feature variables to the progression process through the regression parameters $\boldsymbol{\beta}$, the differences between gender through the regression parameter γ, and about the variability of the process through the standard deviation parameter σ. Moreover, an estimate of the degree of the error made in the UPDRS score is quantified through the standard deviation parameter τ. Table 1 presents the posterior estimates of the regression and standard deviation parameters of the proposed model, including their corresponding means, standard deviations, and 2.5% and 97.5% quantiles (95% intervals). The observed and estimated scores for the 42 subjects are presented in Table S1 in supplementary material available at *Biostatistics* online.

With respect to motor UPDRS, Schüpbach *and others* (2009) found a progression rate of 39.34% in 6 months, starting from a baseline of 13.42 ± 5.93 points with unmedicated early-stage patients. The estimated progression rate in 6 months obtained with the proposed approach is 31.58%, which is closer to 39.34% than the observed progression rate, which is 10.56%. Shoulson and Parkinson Study Group (1992) found a rate of progression for motor UPDRS scores of 51.07% in one year, also with unmedicated patients starting from a baseline of 16.8 ± 8.8. In spite of the fact that the progression rate is measured in one year, it can be observed that the estimated scores are proportionally closer than the observed ones, indicating that the error correction may be in the right direction. With respect to the total UPDRS, Shoulson and Parkinson Study Group (1992) found a rate of progression in 1 year of 51.61% starting from a baseline of 25.4 ± 11.6, which is also closer to the estimated progression rate than to the observed one. This supports the idea that the proposed approach may also be adequately correcting some profiles of total UPDRS.

Figure S2 in supplementary material available at *Biostatistics* online shows the profiles based on the AHTD data. Figure 4 shows the observed and estimated trajectories of motor UPDRS and total UPDRS for four subjects. The dots represent the observed responses y^*_{ik}, and the lines represent the mean and

Table 1. *AHTD database: means, standard deviations (SD), and 95% intervals of the posterior estimates of the regression and SD parameters*

	Motor UPDRS								
	All			Men			Women		
	Mean	SD	95% interval	Mean	SD	95% interval	Mean	SD	95% interval
α	17.206	1.162	(14.868, 19.440)	17.002	1.244	(14.449, 19.349)	17.781	2.128	(13.554, 21.899)
β_1	−7.634	2.807	(−13.126, −2.304)	−2.220	3.974	(−10.061, 5.666)	−7.503	5.466	(−16.927, 4.439)
β_2	0.424	2.166	(−3.877, 4.576)	7.344	3.975	(−0.422, 15.181)	−1.149	4.138	(−10.011, 4.920)
β_3	−1.286	2.634	(−6.600, 3.731)	−0.336	3.806	(−7.761, 7.240)	1.272	3.996	(−5.849, 9.340)
β_4	1.560	1.817	(−2.037, 5.075)	−2.014	2.493	(−6.991, 2.818)	6.170	3.004	(0.649, 11.396)
β_5	−3.313	0.960	(−5.199, −1.396)	−4.212	1.295	(−6.774, −1.709)	−3.558	1.595	(−6.608, −0.478)
β_6	8.088	2.847	(2.500, 13.641)	5.725	3.962	(−2.255, 13.496)	9.771	4.040	(1.705, 17.391)
γ	−2.301	1.944	(−6.134, 1.502)	—	—	—	—	—	—
σ	4.738	0.717	(3.249, 6.092)	5.148	0.776	(3.614, 6.699)	1.359	1.325	(0.157, 4.724)
τ	4.701	0.523	(3.808, 5.852)	4.513	0.622	(3.475, 5.912)	5.142	0.834	(3.782, 7.056)
	Total UPDRS								
	All			Men			Women		
	Mean	SD	95% interval	Mean	SD	95% interval	Mean	SD	95% interval
α	23.145	1.452	(20.186, 25.897)	22.706	1.657	(19.283, 25.754)	26.288	2.418	(21.019, 30.261)
β_1	−6.935	3.689	(−14.162, 0.349)	−2.219	5.647	(−13.204, 9.090)	−0.140	7.271	(−12.114, 16.778)
β_2	−1.117	2.918	(−6.839, 4.646)	7.787	5.564	(−2.931, 18.918)	−6.649	5.391	(−19.274, 2.714)
β_3	−3.006	3.409	(−9.673, 3.677)	−1.237	5.233	(−11.545, 9.200)	−2.858	3.927	(−9.811, 4.310)
β_4	1.604	2.419	(−3.116, 6.320)	−1.918	3.457	(−8.839,4.779)	5.668	3.077	(0.151, 11.881)
β_5	−4.056	1.256	(−6.542, −1.616)	−5.910	1.784	(−9.450, −2.433)	−2.289	1.506	(−5.296, 0.478)
β_6	8.251	3.657	(1.081, 15.373)	7.217	5.443	(−3.646, 17.716)	5.755	4.004	(−0.965, 14.130)
γ	−3.375	2.554	(−8.349, 1.591)	—	—	—	—	—	—
σ	6.898	0.741	(5.472, 8.406)	7.654	0.918	(5.958, 9.556)	1.408	1.494	(0.062, 5.183)
τ	5.211	0.568	(4.228, 6.449)	4.846	0.669	(3.722, 6.329)	6.230	0.921	(4.634, 8.244)

the 2.5% and 97.5% quantiles of the estimated true responses with nondecreasing patterns $\widehat{y_{ik}}$. Patients 2 and 8 provide nondecreasing observed motor and total UPDRS scores, whereas patients 17 and 21 exhibit patterns that contain measurement errors given the values of the differences between consecutive scores. The corrections performed seem to be in agreement with the dynamic progression of unmedicated early-stage PD patients.

In the supplementary material available at *Biostatistics* online document, a simulation study based on the AHTD database is presented. It shows how the proposed model properly recovers the real parameters in a simulation-based context.

6. Conclusion

An inhomogeneous hidden Markov model with continuous state-space has been proposed to address measurement errors in the response, any within-subject variability of replicated covariates, and monotonic nondecreasing processes. The proposed Bayesian framework has been implemented by deriving an efficient MCMC sampling algorithm, which has been tested on a simulation-based experiment and applied to track the progression of PD based on the At-Home Testing Device (AHTD) database (Goetz *and others*, 2009) of acoustic features extracted from voice recordings. Results are encouraging and could potentially

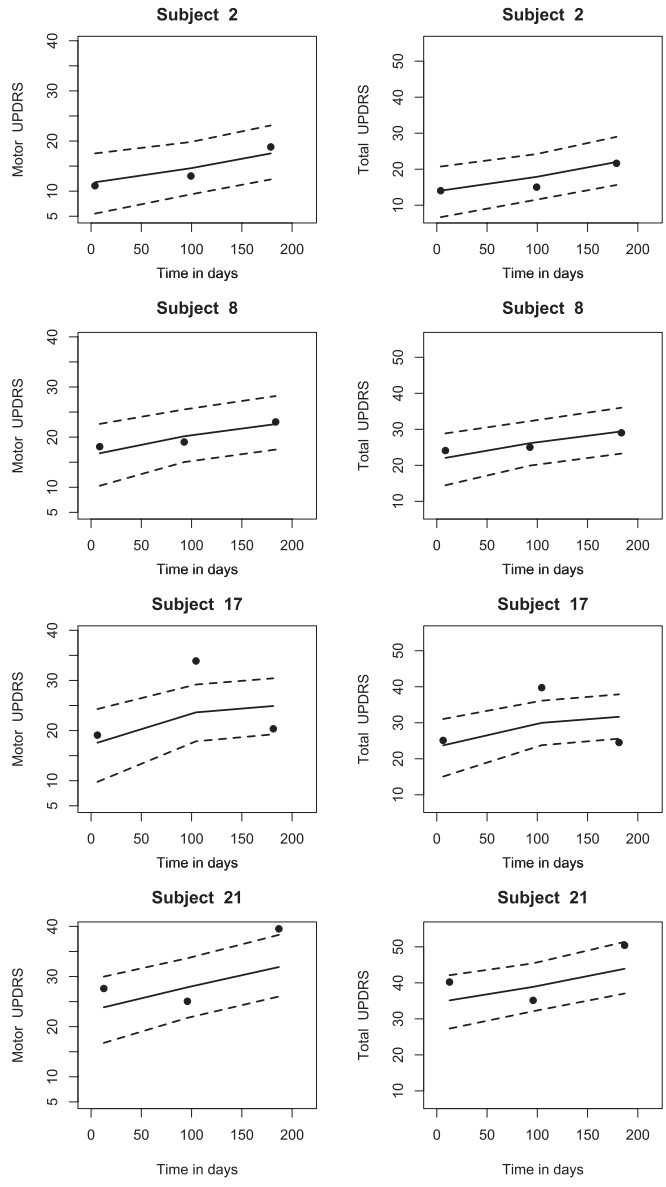

Fig. 4. AHTD database: analysis of some profiles of motor and total UPDRS.

improve the tracking of the progression of PD. Although the approach has been motivated by a PD context, it is applicable to any monotonic nondecreasing process whose continuous response variable is subject to measurement errors and where replications are considered. This is the case for several neurodegenerative diseases that are characterized by a progressive decline and have no cure. Two particular cases are multiple sclerosis of the primary progressive type and Huntington's disease.

Changes in the proposed model could allow the inclusion of several examiners involved in the scoring of the responses for a given subject across time, very useful in real situations. Moreover, different

specifications of the regression parameters or of the variance parameters could also be modified to allow heterogeneity models. In addition, the number of replications could be different for each individual: a small variation in the model specification could be made to take this into account. Besides, different prior distributions could also be considered, although it seems that the proposed model is robust to different specifications of the prior distributions of the variance parameters. In addition, the proposed model can be modified to handle monotonic nonincreasing responses.

Only gender and acoustic features have been used as predictors, since the goal is that in the future, the proposed approach can be implemented in a real-time expert system through Android or iOs-based smartphones. Since the approach pools information across individuals, UPDRS estimates showing the current state of each patient can be performed. This would lead to a warning system that would consider a certain threshold or a fast UPDRS increase to encourage a visit to the neurologist. Such an intelligent telediagnosis system would be very valuable for patients, medical staff, and administration, since it would be based on acoustic biomarkers that are objective, low cost and noninvasive. The database could be increased and more accurate results would be obtained. Therefore, a well calibrated system would facilitate a fast and frequent remote tracking of the progression of the disease. This would help neurologists, although human expertise cannot be completely put aside.

The AHTD database has many limitations, including the small number of women participating as well as the fact that there is no information about how many examiners participated and their degree of expertise. It would be interesting to conduct a new experiment with a greater sample size, using several examiners in a controlled way. This would enable the analysis of the model performance on a different database, assessing the accuracy of the approach.

The AHTD database has many limitations, but it is the only publicly available database of these characteristics. Some them are the reduced duration of the experiment, the fact that there is no information about how many examiners participated and their degree of expertise and the small number of women participating. Although, there is a great difficulty to recruit patients for this kind of experiments, it would be interesting to conduct a new experiment with a greater sample size of both sexes with several examiners in a controlled way. This would enable the analysis of the model performance on a different database, assessing several aspects, including accuracy of the approach.

Supplementary material

Supplementary material is available at http://biostatistics.oxfordjournals.org.

Acknowledgments

We would like to thank Athanasios Tsanas for providing information about the dataset. We are also grateful to three anonymous reviewers who provided insightful comments and suggestions that have greatly improved both the readability and the content of this article. *Conflict of Interest*: None declared.

Funding

This research has been supported by UNAM-DGAPA-PAPIIT, Mexico (Project IA106416), Ministerio de Economía, Industria y Competitividad, Spain (Projects MTM2014-56949-C3-3-R and MTM2017-86875-C3-2-R), Junta de Extremadura, Spain (Projects IB16054 and GR18108), and the European Union (European Regional Development Funds).

REFERENCES

ARIAS-VERGARA, T., VASQUEZ-CORREA, J. C., OROZCO-ARROYAVE, J. R., VARGAS-BONILLA, J.F. AND NÖTH, E. (2016). Parkinson's disease progression assessment from speech using GMM-UBM. *17th Annual Conference of the International Speech Communication Association (INTERSPEECH 2016): Understanding Speech Processing in Humans and Machines.*, 1933–1937. International Speech Communication Association (ISCA).

BAKEN, R. J. AND ORLIKOFF, R. F. (2000). *Clinical Measurement of Speech and Voice*, 2nd edition. San Diego, CA: Singular Thomson Learning.

BUONACCORSI, J. P. (2010). *Measurement Error: Models, Methods and Applications*. Boca Raton, FL: Chapman and Hall/CRC.

CAPPÉ, O., MOULINES, E. AND RYDEN, T. (2005). *Inference in Hidden Markov Models*. Berlin: Springer-Verlag.

CARROLL, R. J., RUPPERT, D., STEFANSKI, L. A. AND CRAINICEANU, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*, 2nd edition. Boca Raton, FL: Chapman and Hall/CRC.

CASTELLI, M., VANNESCHI, L. AND SILVA, S. (2014). Prediction of the unified Parkinson's disease rating scale assessment using a genetic programming system with geometric semantic genetic operators. *Expert Systems with Applications* **41**, 4608–4616.

CHAN, P. L. S. AND HOLFORD, N. H. G. (2001). Drug treatment effects on disease progression. *Annual Review of Pharmacology and Toxicology* **41**, 625–659.

DUFFY, J. R. (2005). *Motor Speech Disorders: Substrates, Differential Diagnosis, and Management*. Amsterdam: Elsevier.

ESKIDERE, Ö., ERTAÇ, F. AND HANILÇI, C. (2012). A comparison of regression methods for remote tracking of Parkinson's disease progression. *Expert Systems with Applications* **39**, 5523–5528.

GARCÍA-ZATTERA, M. J., JARA, A., LESAFFRE, E. AND MARSHALL, G. (2012). Modeling of multivariate monotone disease processes in presence of misclassification. *Journal of the American Statistical Association* **107**, 976–989.

GELMAN, A. (2006). Prior distributions for variance parameters in hierarchical models (Comment on Article by Browne and Draper). *Bayesian Analysis* **1**, 515–534.

GILKS, W. R., RICHARDSON, S. AND SPIEGELHALTER, D. J. (1996). *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.

GOETZ, C. G., STEBBINS, G. T., WOLFF, D., DELEEUW, W., BRONTE-STEWART, H., ELBLE, R., HALLETT, M., NUTT, J., RAMIG, L., SANGER, T. *and others*. (2009). Testing objective measures of motor impairment in early Parkinson's disease: feasibility study of an at-home testing device. *Movement Disorders* **24**, 551–556.

HARIHARAN, M., POLAT, K. AND SINDHU, R. (2014). A new hybrid intelligent system for accurate detection of Parkinson's disease. *Computer Methods and Programs in Biomedicine* **113**, 904–913.

JANKOVIC, J. AND KAPADIA, A. S. (2001). Functional decline in Parkinson's disease. *Archives of Neurology* **58**, 1611–1615.

LITTLE, M. A., MCSHARRY, P. E., HUNTER, E. J., SPIELMAN, J. AND RAMIG, L. O. (2009). Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE Transactions on Biomedical Engineering* **56**, 1015–1022.

LUNN, D. J., THOMAS, A., BEST, N. AND SPIEGELHALTER, D. (2000). WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing* **10**, 325–337.

NARANJO, L., PÉREZ, C. J., CAMPOS-ROCA, Y. AND MARTÍN, J. (2016). Addressing voice recording replications for Parkinson's disease detection. *Expert Systems With Applications* **46**, 286–292.

NARANJO, L., PÉREZ, C. J. AND MARTÍN, J. (2017*a*). Addressing voice recording replications for tracking Parkinson's disease progression. *Medical and Biological Engineering and Computing* **55**, 365–373.

NARANJO, L., PÉREZ, C. J., MARTÍN, J. AND CAMPOS-ROCA, Y. (2017*b*). A two-stage variable selection and classification approach for Parkinson's disease detection by using voice recording replications. *Computer Methods and Programs in Biomedicine* **142**, 147–156.

NOVOTNY, M., RUSZ, J., CMEJLA, R. AND RUZICKA, E. (2014). Automatic evaluation of articulatory disorders in Parkinson's disease. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **22**, 1366–1378.

O'HAGAN, A., BUCK, C. E., DANESHKHAH, A., EISER, J. R., GARTHWAITE, P. H., JENKINSON, D. J., OAKLEY, J. E. AND RAKOW, T. (2006). *Uncertain Judgements: Eliciting Experts' Probabilities*. Chichester, UK: Wiley.

POST, B., MERKUS, M. P., DE BIE, R. M. A., DE HAAN, R. J. AND SPEELMAN, J. D. (2005). Unified Parkinson's disease rating scale motor examination: are ratings of nurses, residents in neurology, and movement disorders specialists interchangeable? *Movement Disorders* **20**, 1577–1584.

RAMAKER, C., MARINUS, J., STIGGELBOUT, A. M. AND VAN HILTEN, B. J. (2002). Systematic evaluation of rating scales for impairment and disability in Parkinson's disease. *Movement Disorders* **17**, 867–876.

RICHARDS, M., MARDER, K., COTE, L. AND MAYEUX, R. (1994). Interrater reliability of the unified Parkinson's disease rating scale motor examination. *Movement Disorders* **9**, 89–91.

SAKAR, B. E., ISENKUL, M. E., SAKAR, C. O., SERTBAS, A., GURGEN, F., DELIL, S., APAYDIN, H. AND KURSUN, O. (2013). Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. *IEEE Journal of Biomedical and Health Informatics* **17**, 828–834.

SATTEN, G. A. AND LONGINI, I. M. J r. (1996). Markov chains with measurement error: estimating the 'strue' course of a marker of the progression of human immunodeficiency virus disease. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **45**, 275–309.

SCHÜPBACH, W. M. M., CORVOL, J. C, CZERNECKI, V., DJEBARA, M. B., GOLMARD, J. L., AGID, Y. AND HARTMANN, A. (2009). Segmental progression of early untreated Parkinson's disease: a novel approach to clinical rating. *Journal of Neurology, Neurosurgery & Psychiatry* **81**, 20–25.

SHOULSON, I. AND Parkinson Study Group. (1992). An interim report of the effect of selegiline (l-deprenyl) on the progression of disability in early Parkinson's disease. *European Neurology* **32**, 46–53.

SMITH, B. J. (2007). BOA: An R package for MCMC output convergence assessment and posterior inference. *Journal of Statistical Software* **21**, 1–37.

TSANAS, A., LITTLE, M. A., MCSHARRY, P. E. AND RAMIG, L. O. (2010). Accurate telemonitoring of Parkinson's disease progression by non-invasive speech tests. *IEEE Transactions Biomedical Engineering* **57**, 884–893.

[*Received March 18, 2018; revised January 25, 2019; accepted for publication January 29, 2019*]