This is the accepted version of the article:

Lizbeth Naranjo, Carlos J. Pérez, Yolanda Campos-Roca. Monitoring Parkinson's disease progression based on recorded speech with missing ordinal responses and replicated covariates, Computers in Biology and Medicine, Volume 134, 104503, 2021. https://doi.org/10.1016/j.compbiomed.2021.104503

Monitoring Parkinson's disease progression based on recorded speech with missing ordinal responses and replicated covariates

Lizbeth Naranjo[a], Carlos J. Pérez[b], Yolanda Campos-Roca[c]

[a]*Departamento de Matemáticas, Facultad de Ciencias, Universidad Nacional Autónoma de México, 04510 Ciudad de México, Mexico*
[b]*Departamento de Matemáticas, Facultad de Veterinaria, Universidad de Extremadura, 10003 Cáceres, Spain*
[c]*Departamento de Tecnologías de los Computadores y de las Comunicaciones, Escuela Politécnica, Universidad de Extremadura, 10003 Cáceres, Spain*

Abstract

Monitoring Parkinson's Disease (PD) progression is an important task to improve the life quality of the affected people. This task can be performed by extracting features from voice recordings and applying specifically designed statistical models, leading to systems that improve the ability of monitoring the progression of PD in an objective, remote, non-invasive, fast, and economically sustainable way. An experiment has been conducted with 36 subjects to study the progression of the PD over 4 years by using the Hoehn and Yahr (HY) scale and features extracted from the phonation of the vowel /a/. The collected dataset had many missing data, which should be addressed jointly with the non-decreasing nature of the disease and the within-subject variability due to the use of replicated features. In order to handle these issues, a Hidden Markov model for longitudinal data was designed and implemented by using a data augmentation scheme based on different latent variables. Markov chain Monte Carlo methods were used to generate from the posterior distribution. The proposed approach has been tested on simulated data, providing good accuracy rates in the context of a multiclass problem. It also has been applied to the real data obtained from the conducted experiment, providing imputed and predicted HY stages compatible with the progression of PD. The conducted experiment and the proposed approach contribute to fill a gap in the scientific literature on experiments and methodologies for tracking PD progression based on acoustic features

and the HY scale. This would help to derive an expert system that can be integrated into the protocols of neurology units in hospital centers.

Keywords: Hidden Markov model, Missing data, Non-decreasing process, Ordinal response, Parkinson's disease, Replicated measurements

---

1. Introduction

Parkinson's Disease (PD) is a long-term neurodegenerative disorder that mainly affects the motor system with symptoms including tremor, stiffness, instability, lack of coordination, or difficulty with walking. Non-motor symptoms such as cognitive and behavioral problems are also relevant. Besides, voice production is affected by motor and cognitive problems. These symptoms begin gradually and get worse over the time.

According to the Parkinson's Disease Foundation, it is estimated that PD currently affects 7 to 10 million people worldwide, being the most relevant neurodegenerative disorder after Alzheimer's disease, but with a faster growth. The Global Burden of Disease study projects to reach 13 million people affected by 2040 [1]. This gives an idea of the magnitude of the problem and justifies the great effort in research and medical care to improve the life quality of people suffering from this up-to-now incurable disorder.

The early diagnosis of PD is key to improve the life quality of people who suffer from it. The diagnosis of this disorder is not evident and takes time (between 1 and 3 years). Besides, it requires the intervention of specialized neurologists. Tracking PD progression is also very important. Receiving continuous monitoring of the progression of the disease is especially interesting, since the symptoms fluctuate significantly throughout the day and, in general conditions, the neurologist can only assess the patient's situation at the specific time of day in which the physical consultation is carried out (once a year in many public health systems). The dose of medication and its administration can be customized according to the evolution of the patient's symptoms.

In recent years, Computer-Aided Diagnosis (CAD) systems have been built to aid in the detection and monitoring of many diseases [2], and, in particular, those detectable by voice [3]. Since voice production is affected by PD, CAD systems, based on features extracted from voice recordings, can be used for these tasks. Moro-Velazquez et al. [4] present a recent review of the advances in PD detection and assessment using voice. Many

approaches have been derived for PD detection, according to different experiments involving different phonation tasks (/a/ sustained phonation [5]; /ka/ syllable phonation [6]; and phonations of isolated words, rapid repetition of the syllables, sentences, and read texts [7]), feature extraction procedures (Jitter, Shimmer, MFCCs, HNR, RPDE, DFA, Entropies...[4]), or classification approaches (k-nearest neighbor, random forests, gradient boosting...[8]). However, the number of studies for tracking PD progression is much more limited, as there is an underlying difficulty in conducting long-term studies. Furthermore, regression-based approaches for longitudinal studies, addressing the difficulties related to the particular experimental designs, are more difficult to derive and apply than classification models for cross-sectional studies. Regression analysis is a set of statistical techniques used to evaluate the relationship among variables, i.e., it tries to determine if one or more independent variables can explain the variability produced in a dependent variable (response variable) [9]. Regression models for numerical and ordinal response are especially interesting in this longitudinal context.

PD can be monitored by using the Unified Parkinson's Disease Rating Scale (UPDRS) [10]. UPDRS is a numerical scale used to measure the course of PD. It enables the quantification of the type, number, and severity of extrapyramidal signs. This quantification is partially based on subjective criteria, therefore disagreement among raters in the interpretation of these criteria may happen [11]. UPRDS contains 45 questions for rating, divided into four parts: (I) Mental state (16 points); (II) Daily life activities (52 points); (III) Motor aspects (68 points); and (IV) Complication of treatments (23 points). Some studies have found approximations to UPDRS scores from regression models applied to features extracted from voice recordings both from linear and nonlinear regression techniques. Tsanas et al. [12] used classical least square regression and regression trees, whereas Hemmerling and Wojcik-Pedziwiatr [13] used multiple linear regression, random forest regression and support vector machine regression for the same purpose. However, most of the approaches carried out in the scientific literature have been developed on the basis of a single multicenter study conducted in the United States with a total of 52 subjects with PD during a short period of 6 months [14]. The voice recordings in this study are not publicly available, but the features, which were extracted for a subgroup of 42 patients, can be down-

loaded from the UCI Machine Learning Repository[1]. Based on this dataset, some approaches have been developed and applied. Besides Tsanas et al. [12], Eskidere et al. [15] used this dataset for model performance assessment with regression models based on support vector machine, least square support vector machines, multilayer perceptron neural network, and general neural network regression methods. Naranjo et al. [16] developed a binary regression model that addressed voice recording replications. Nilashi et al. [17] developed an approach based on ensembles of deep belief network and self-organizing map.

The application of UPDRS scale (or MDS-UPDRS [18]) requires the presence of the patient in a hospital center, as well as extensive physical examinations by qualified medical personnel. Besides, the inter-rater agreement is not as good as desirable [19]. The Hoehn and Yahr (HY) scale is simpler and used very often to assess the level of disability produced by PD [20]. Originally, it contained 5 stages on an ordinal scale, but later was modified by including two more stages to help in describing the intermediate course of the disease [21]. Skodda et al. [22] conducted a longitudinal study to assess the progression and speech impairment in the course of PD based on the UPDRS motor and HY scales, although they simplified the HY scale to only 3 stages. They focused on analyzing the correlation between perceptual speech scores (articulation, fluency, prosody...) and basic linear speech parameters, and they made a comparison between PD and control groups. They did not try to make predictions based on the acoustic features.

There is a lack of longitudinal studies to predict HY stages based on acoustic features. This kind of prediction can be performed based on a regression model for ordinal response. This has motivated conducting an experiment with 36 people suffering from PD over 4 years, with stages ranging from 1 to 4 (no subject reached stage 5). Carrying out this experiment led to some challenges that required the construction of a specific approach. The first challenge was the existence of non-response. The experiment was conducted in the headquarters of the Regional Association for Parkinson's Disease of Extremadura (Spain), and the participating subjects were volunteers, which were invited to participate every year. Some of them were not available during some periods of time (illness, travels...) or even permanently. Therefore, missing data were obtained in the following rates: 11% after one year, 22%

[1] `https://archive.ics.uci.edu/ml/datasets/Parkinsons+Telemonitoring`

after two years, and 39% after three years. In spite of that, a great amount of informative data was collected and it was needed to build appropriate statistical methodology to address it. Besides, PD is a neurodegenerative disorder, so the people that suffer from it get progressively worse over the time. This situation also requires to address the problem as a non-decreasing process. Many diseases have been modelled taking this fact into account such as caries experience [23], aortic aneurysm [24], or even PD [25].

Addressing the issues of non-response and non-decreasing process should be completed with managing the within-subject variability produced by the use of replicated voice recordings. There exists a relevant variability between features extracted from two or more voice recordings of the same subject at a particular time, so using only one utterance per subject may provide different results depending on the voice recording that has been selected. Imperfections in technology and the very biological variability result in values that are similar (but not identical) for voice recordings from a particular subject, rather than for recordings from different individuals. Some authors have ignored this issue and have treated the data as if they were independent when there exists a clear dependent nature (see, e.g., the treatment given by Tsanas et al. [12] or Nilashi et al. [17]). However, this within-subject variability can be properly addressed [25]. The experiment involved in this article considers three replications per subject at each time point.

In this article, a Hidden Markov Model (HMM) has been developed and applied to address the lack of response in the HY ordinal scale for tracking PD progression, non-decreasing process, and within-subject variability produced by the voice recording replications. Due to the difficulty of addressing these issues, latent variables have been introduced in the model, which has allowed to solve the model by using Markov chain Monte Carlo (MCMC) methods. The resulting approach has been applied to the data collected from the previously described long-term experiment, which has been conducted specifically for this task. Therefore, the proposed approach is part of a CAD system that contributes to improve the ability of monitoring the progression of PD in an objective, remote, non-invasive, fast, and economically sustainable way. This is in line with the development and use of tools, technologies and digital solutions for health and care, as an essential pillar of modern medicine. Although the approach has been derived for monitoring the progression of PD, it is applicable to other problems that share these characteristics.

The outline of the rest of the article is as follows. Section 2 presents the

data collection through the description of the participants, speech recordings, recording devices, and feature extraction procedures. Section 3 describes the HMM, including the details about dealing with the missing ordinal response, non-decreasing process, and replicated covariates. Also, the posterior distribution is explored. Section 4 presents the results obtained from a simulation-based case and from real data obtained with the conducted experiment. Next, conclusions are shown in Section 5. Finally, imputed and predicted stages are presented in an Appendix.

2. Data collection

2.1. Participants

A total of 36 subjects having PD have been involved in this study, being 12 women (33.3%) and 24 men (66.7%). The mean (standard deviation) age was 69 (7.93) years at the beginning of the study. All had a definitive diagnosis by their neurologists and were medicated with levodopa. They attended different activities at the headquarters of the Regional Association for Parkinson's Disease of Extremadura in Cáceres and Mérida (Spain) during the years 2016 to 2019.

The modified HY scale was used to measure how Parkinson's symptoms progress and the level of disability. It has the following stages:

- Stage 0 - No signs of disease.
- Stage 1 - Symptoms on one side only (unilateral).
- Stage 1.5 - Symptoms unilateral and also involving the neck and spine.
- Stage 2 - Symptoms on both sides but no impairment of balance.
- Stage 2.5 - Mild symptoms on both sides, with recovery when the 'pull' test is applied (the doctor stands behind the person and asks her/him to maintain the balance when pulled backwards).
- Stage 3 - Balance impairment, mild to moderate disease, physically independent.
- Stage 4 - Severe disability, but still able to walk or stand unassisted.
- Stage 5 - Needing a wheelchair or bedridden unless assisted.

Table 1 shows the gender, age (at baseline), and HY stages of the subjects involved in this study. All the subjects were volunteers and signed an informed consent to participate in this study.

2.2. Voice recordings

The task was the sustained phonation of /a/ vowel. This phonation was performed as constantly as possible at comfortable pitch and loudness, and was kept for 5 seconds on one breath. The task was repeated three times per subject in each recording session.

The phonations were recorded using a portable computer with an external sound card (TASCAM US322) and a headband microphone (AKG 520) featuring a cardiod pattern. The sampling rate was 44.1 KHz and the resolution was 16 bits/sample. Version 2.0.5 of Audacity software was used for the voice recordings.

The research protocol was approved by the Bioethical Committee from the University of Extremadura (Spain).

2.3. Feature extraction

A reduced number of acoustic features (concretely, four) have been considered to describe different aspects on voice impairment in PD. Also, a gender feature has been taken into account. The reasons for the selection of these variables are explained in the following paragraphs.

Hypokinetic dysarthria is a common speech disorder experienced by people with PD [26]. It is caused by neurologic damage affecting one or several of the motor components of speech. Therefore, it may involve impaired functioning of any or all processes of voice production, including respiration, phonation, articulation, resonance, and prosody. Sustained vowel phonation is a simple vocal task that allows for considering some of these aspects.

Cepstral peak prominence (CPP) has been reported to be a reliable feature to characterize PD [27]. Classical perturbation and noise features (such as jitter, shimmer or harmonic-to-noise ratio) have been proposed to detect phonatory impairment in PD [4]. However, voices with severe disorders may compromise the reliability of these traditional measures. In comparison to jitter or shimmer, CPP has the advantage that it does not require the identification of individual cycles. High CPP values correspond to a well-defined harmonic structure, whereas disturbed periodicity can decrease the amplitude of the cepstral peak.

Table 1: Gender, age (at baseline), and HY stages of the involved subjects. NA represents missing data.

Subject	**Gender**	**Age**	**Stages by year**			
			2016	**2017**	**2018**	**2019**
1	M	60	3.0	3.0	3.0	3.0
2	M	67	3.0	NA	4.0	NA
3	M	46	1.0	1.5	2.0	2.0
4	M	69	2.0	2.0	2.0	2.0
5	M	69	1.0	1.5	2.0	2.0
6	F	65	2.0	2.0	2.5	2.5
7	F	70	2.5	3.0	NA	3.0
8	M	70	2.0	NA	2.0	NA
9	M	83	4.0	4.0	4.0	NA
10	M	69	3.0	3.0	NA	NA
11	M	75	1.0	1.5	1.5	2.0
12	F	82	3.0	3.0	3.0	NA
13	F	74	3.0	3.0	3.0	4.0
14	M	72	3.0	3.0	NA	NA
15	M	80	1.0	1.0	3.0	3.0
16	F	75	3.0	3.0	3.0	3.0
17	M	69	2.5	3.0	NA	3.0
18	M	66	2.5	3.0	3.0	3.0
19	F	73	2.5	2.5	NA	NA
20	M	77	3.0	3.0	NA	NA
21	M	53	1.0	1.5	2.0	2.0
22	F	73	3.0	3.0	NA	NA
23	M	60	2.0	2.0	3.0	NA
24	F	72	3.0	4.0	4.0	NA
25	F	58	4.0	4.0	4.0	4.0
26	M	79	2.0	3.0	3.0	3.0
27	M	72	1.0	2.0	2.0	2.0
28	M	69	2.0	2.0	2.5	2.5
29	M	71	3.0	3.0	4.0	NA
30	F	55	1.0	1.5	2.0	3.0
31	M	74	2.0	2.0	2.0	2.0
32	M	71	3.0	3.0	3.0	3.0
33	F	65	3.0	NA	4.0	NA
34	F	69	2.0	2.5	2.5	2.5
35	M	62	1.0	NA	NA	3.0
36	M	70	1.5	2.5	2.5	NA

Mel Frequency Cepstral Coefficients (MFCCs) are also relevant features. They allow for detecting differences in the resonant characteristics of the vocal tract. They have been previously proposed as a robust feature for PD diagnosis [28] and also for estimation of PD severity [13]. According to Bang et al. [29], patients with PD present an asymmetric centralization of tongue position during the phonation of sustained vowels, which produces a remarkable decrease in the vowel space area in comparison to healthy controls. Vowel space area depends on the two first formant frequencies, therefore features that capture basic properties of the spectral envelope (such as the MFCCs) are a suitable choice to describe these phenomena. In order to avoid multicolinearity problems, only the 5th coefficient is used, since it showed the highest predictive capability in previous experiments of automatic PD detection.

Furthermore, some studies demonstrate that vocal fold vibration is highly influenced by nonlinearities in tissue and air movement. In the case of pathological voice from PD patients, some physiological conditions such as incomplete vocal fold closure emphasize this nonlinear behavior [30]. These phenomena often challenge traditional methods based on the Fourier transform. Two features based on nonlinear analysis have been selected to capture these nonlinear aspects of speech: Recurrence Period Density Entropy (RPDE) and correlation dimension (D2) [31]. The former is based on the notion of nonlinear recurrence, which can be considered as a generalization of periodicity. D2 gives an idea of the complexity of a voice signal. Signals with a well-defined harmonic structure show a simple oscillation pattern which leads to a low D2 value. However, in the case of dysphonic voices, the oscillation patterns are more difficult to predict, the value of D2 increases and this value is associated to complex dynamics.

Regarding gender aspects, it is worth noting that there are more men than women diagnosed with PD, and the disease affects men and women differently. Although our experiment is not based on a simple random sampling, the proportion is 2:1 favoring the men, which reflects the proportion at the headquarters of the Regional Association for Parkinson's Disease of Extremadura, where the patients were recruited. Besides, it has been previously shown that gender may have an influence on which features have more potential to provide a good performance in the discrimination between PD participants and controls [32]. Therefore, gender has been included as a feature.

After feature extraction, a spreadsheet with 36 rows (36 subjects) and 49 columns (gender + 4 acoustic features $\times$ 3 replications $\times$ 4 years) constitute

the dataset to feed the proposed model.

3. Approach

In this section, the HMM-based approach is proposed. It is composed of several parts. Firstly, the ordinal regression model is formulated. Then, a data augmentation framework is considered to address the monotone non-decreasing process. Next, it is explained how the replications in the explanatory variables are integrated in the model, and how the missing data are addressed. Finally, the prior distribution is presented and the posterior distribution is derived. Using MCMC methods to solve the problem is proposed.

3.1. Ordinal response model

Let Y_{it} be the ordinal response for the subject i at time t, for $i = 1, \ldots, N$ and $t = 1, \ldots, T$, where Y_{it} takes one of K categories, having covariates $\boldsymbol{x}_{it}$ and $\boldsymbol{z}_{it}$. Let

$$p_{itk} \quad = \quad \mathbb{P}[Y_{it} = k | \boldsymbol{x}_{it}, \boldsymbol{z}_{it}]$$

be the probability that the ith subject at time t is classified in the kth category, for $k = 1, \ldots, K$. The covariates $\boldsymbol{x}_{it}$ and $\boldsymbol{z}_{it}$ are used in the linear predictor η_{it}, where

$$\eta_{it} \quad = \quad \boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{z}_{it}'\boldsymbol{\gamma}, \tag{1}$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$, with $\boldsymbol{\beta}$ being an L-dimensional coefficient vector for $\boldsymbol{x}_{it}$, and $\boldsymbol{\gamma}$ being an M-dimensional coefficient vector for $\boldsymbol{z}_{it}$.

For ordered response categories, the ordinal model can be defined by cutpoints $\kappa_0, \kappa_1, \ldots, \kappa_K$. These cutpoints and linear predictor make that the probabilities p_{itk} are defined as:

$$p_{itk} \quad = \quad \Psi\left(\kappa_k - \eta_{it}\right) - \Psi\left(\kappa_{k-1} - \eta_{it}\right),$$

where $\Psi(\cdot)$ is a cumulative distribution function (see Albert and Chib [33]). The cumulative distribution function of a Gaussian distribution will be considered. The vectors of unknown cutpoints are defined as: $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_{K-1})'$, where κ_1 is set to 1 to avoid parameter identifiability problems. Besides, by convention $\kappa_0 = -\infty$ and $\kappa_K = +\infty$.

3.2. A monotone non-decreasing process

Based on the data augmentation framework for the ordinal regression model proposed by Albert and Chib [33] and Cowles [34], let W_{it} be the continuous response for the subject i at time t. They are related to Y_{it} by:

$$Y_{it} = k \quad \text{if} \quad \kappa_{k-1} < W_{it} \leq \kappa_k, \text{ for } k = 1, \ldots, K. \tag{2}$$

Assume that $\{W_{it}\}$ is a monotone non-decreasing continuous process, i.e.,

$$W_{i1} \leq W_{i2} \leq \cdots \leq W_{iT},$$

and then,

$$\begin{aligned} W_{i1} &\sim \mathrm{N}\left(\eta_{i1}, 1\right), \quad t = 1, \\ W_{it} \big| W_{i,t-1} = w_{i,t-1} &\sim \mathrm{N}\left(\eta_{it}, 1\right) \mathrm{I}\left[W_{it} \geq w_{i,t-1}\right], \quad t = 2, \ldots, T. \end{aligned} \tag{3}$$

Note that the first-order Markov chain property has been assumed.

3.3. Replications in the explanatory variables

Assume that the covariates $\boldsymbol{z}_i$ are exactly known, but the covariates $\boldsymbol{x}_i$ have been measured with J replicates. Now, instead of the covariates $\boldsymbol{X}_{it}$, their replicates $\boldsymbol{X}^*_{itj}$ are recorded, and

$$\begin{aligned} \boldsymbol{X}^*_{itj} &= \boldsymbol{X}_{it} + \boldsymbol{\delta}_{itj}, \\ \boldsymbol{\delta}_{itj} &\sim \mathrm{Normal}_L\left(\mathbf{0}, \boldsymbol{\Lambda}\right), \end{aligned}$$

for $j = 1, \ldots, J$, where the covariates $\boldsymbol{X}_{it}$ represent the acoustic features, and they are latent variables having distributions $\boldsymbol{X}_{it} \sim \mathrm{Normal}_L(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. This assumption is related to the classic additive error model [35].

3.4. Missing response data

Missing data are presented in the response variables. Therefore, if a response Y_{it} is missing, then the notation Y^{miss}_{it} will be used. For a missing data point Y^{miss}_{it} its corresponding latent variable W_{it} is computed from (3). The missing data Y^{miss}_{it} could be imputed by using (2), i.e., by randomly simulating the response variable $Y^{imputed}_{it}$ by:

$$Y^{imputed}_{it} = k \quad \text{if} \quad \kappa_{k-1} < W_{it} \leq \kappa_k, \text{ for } k = 1, \ldots, K.$$

If no covariates $\boldsymbol{X}^*_{itj}$ are available, it is possible to simulate W_{it} from (3) by using $\eta_{it} \sim \mathrm{N}\,(0, 1000)$.

The full model is an inhomogeneous HMM that comprises the equations defined along this section. Figure 1 displays the probabilistic graphical model representation, showing the dependencies among the variables. The proposed model follows conditional independence assumptions, similar to those defined by Naranjo et al. [36].

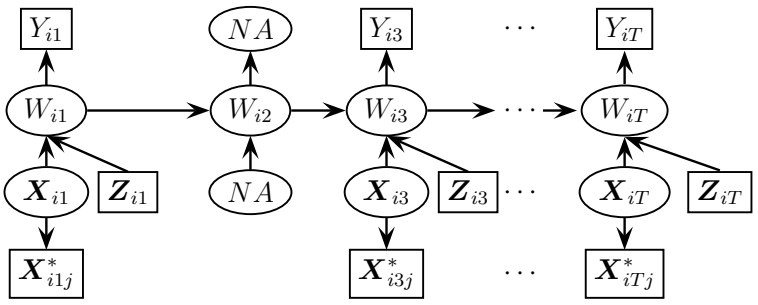

Figure 1: Graphical representation of an example of the proposed HMM with missing data at time $t = 2$. Square boxes represent observed variables, ovals represent latent variables, and the directions of the arrows indicate conditional dependence.

3.5. Exploring the posterior distribution

The prior distribution has been chosen in the following way, having selected some components to have conditionally conjugate distributions. For the regression coefficients in the linear predictor, normal distributions are considered, i.e., $\boldsymbol{\beta} \sim \mathrm{N}_L(\boldsymbol{b}, \boldsymbol{B})$ and $\boldsymbol{\gamma} \sim \mathrm{N}_M(\boldsymbol{c}, \boldsymbol{C})$. For the cutpoints $\boldsymbol{\kappa}$, flat prior distributions are used, i.e., $\pi(\boldsymbol{\kappa}) \propto 1$. For the covariance matrix of the replicated covariates, the prior distribution is specified as an inverse Wishart, $\boldsymbol{\Lambda} \sim \mathrm{InverseWishart}(\boldsymbol{V}, \nu)$.

The likelihood function of the proposed model has the following form:

$$
\begin{aligned}
&\mathcal{L}\left(\boldsymbol{W}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\kappa}, \boldsymbol{\Lambda} \mid \boldsymbol{Y}, \boldsymbol{Y}^{miss}, \boldsymbol{X}, \boldsymbol{z}\right) \qquad (4)\\
&= \prod_{i=1}^{N} \Bigg\{ \mathrm{P}\left(Y_{i1} | W_{i1}, \boldsymbol{\kappa}\right) \mathrm{P}\left(Y_{i1}^{miss} | W_{i1}, \boldsymbol{\kappa}\right) \mathrm{P}\left(W_{i1} | \boldsymbol{X}_i, \boldsymbol{z}_{i1}, \boldsymbol{\beta}, \boldsymbol{\gamma}\right)\\
&\times \left[\prod_{t=2}^{T} \mathrm{P}\left(Y_{it} | W_{it}, \boldsymbol{\kappa}\right) \mathrm{P}\left(Y_{it}^{miss} | W_{it}, \boldsymbol{\kappa}\right) \mathrm{P}\left(W_{it} | W_{i,t-1}, \boldsymbol{X}_i, \boldsymbol{z}_{it}, \boldsymbol{\beta}, \boldsymbol{\gamma}\right) \right]\\
&\times \left[\prod_{t=1}^{T} \left\{ \prod_{j=1}^{J} \mathrm{P}\left(\boldsymbol{X}_{itj}^{*} | \boldsymbol{X}_{it}, \boldsymbol{\Lambda}\right) \right\} \mathrm{P}\left(\boldsymbol{X}_{it}\right) \right] \Bigg\}.
\end{aligned}
$$

Therefore, from the likelihood (4) and the prior distributions, the joint posterior distribution is given by:

$$
\begin{aligned}
&\pi\left(\boldsymbol{W}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\kappa}, \boldsymbol{\Lambda} \mid \boldsymbol{Y}, \boldsymbol{Y}^{miss}, \boldsymbol{X}, \boldsymbol{z}\right)\\
&\quad \propto \mathcal{L}\left(\boldsymbol{W}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\kappa}, \boldsymbol{\Lambda} \mid \boldsymbol{Y}, \boldsymbol{Y}^{miss}, \boldsymbol{X}, \boldsymbol{z}\right) \pi\left(\boldsymbol{\beta}\right) \pi\left(\boldsymbol{\gamma}\right) \pi\left(\boldsymbol{\kappa}\right) \pi\left(\boldsymbol{\Lambda}\right).
\end{aligned}
$$

Note that to make posterior inferences, the relationship between the covariates and the latent variables jointly with the prior distributions are used. Figure 2 shows a graphical representation of the proposed model. This is based on doodle objects of WinBUGS [37].

The posterior distribution (5) is not directly tractable due to the complexity of the model. Applying MCMC methods is required [38].

4. Results

Firstly, a simulation-based case is presented to show the model performance. Then, the approach is applied to the real data coming from the conducted experiment and the results are displayed as well as discussed.

4.1. Simulation-based case

A simulation-based procedure to generate the datasets for testing the approach is described. A total of 100 datasets are simulated. Each dataset has $N = 75$ subjects, $L = 4$ covariates, each one measured $J = 3$ times (replications) at each one of the $T = 4$ time points, and $K = 4$ categories. The covariates $\boldsymbol{x}_{it}$ are simulated from uniform distributions, i.e. $x_{itl} \sim \mathrm{U}(0, 1)$, for $i = 1, \ldots, N$, $t = 1, \ldots, T$, $l = 1, \ldots, L$. Then, their replications are simulated

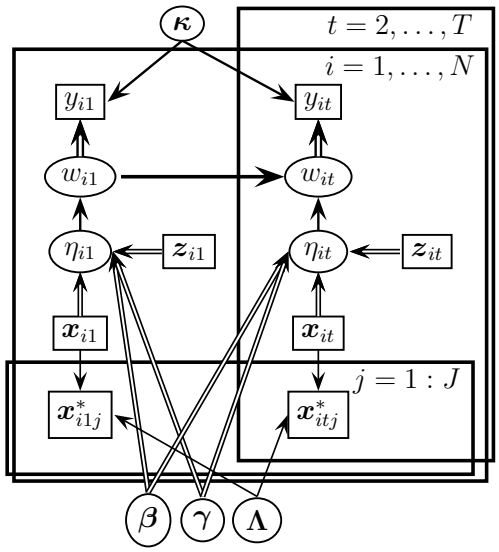

Figure 2: Flowchart for the proposed model.

from the normal distributions, $x_{itlj} \sim \text{Normal}(0, \lambda_l)$, where the variances are $\boldsymbol{\lambda} = (0.2^2, 0.25^2, 0.3^2, 0.35^2)$. The covariates $\boldsymbol{z}_i = (z_{i1}, z_{i2})$ are simulated from $z_{i1} \sim \text{Bernoulli}(0.5)$ and $z_{i1} \sim \text{Poisson}(5)$, considering they could refer to the gender and another variable that is not subject to replications. Linear predictors in Equation (1), η_{i1} and η_{it}, are computed by using the mean of the replicated covariates and regression parameters $\boldsymbol{\gamma} = (0.4, 0.2)'$ and $\boldsymbol{\beta} = (1.0, 0.5, -1.5, -2.0)'$.

The latent variables W_{it}, that follow the monotone non-decreasing continuous process, are simulated by using Equation (3). Then, the ordinal response variables Y_{it} are generated by using Equation (2), where the cutpoints are $\boldsymbol{\kappa} = (0, 1.5, 3.0)'$.

The following prior distributions have been used: (i) for the regression parameters, $\beta_l \sim \text{N}(0, 10000)$ for $l = 1, \ldots, L$, $\gamma_1 \sim \text{N}(0, 10000)$, $\gamma_2 \sim \text{N}(0, 10000)$; (ii) for the cutpoints, $\kappa_k \sim \text{N}(0, 1)$ where $\kappa_k < \kappa_{k+1}$, for $k = 2, \ldots, K-1$; (iii) for the covariates, $X^*_{itjl} \sim \text{N}(x_{itjl}, \lambda_l)$, $X_{itjl} \sim \text{N}(0, 1)$, $\lambda_l \sim \text{InvGamma}(0.001, 0.001)$, where $i = 1, \ldots, N$, $t = 1, \ldots, T$, $j = 1, \ldots, J$.

The proposed approach has been implemented in JAGS[2] through the

[2] `http://mcmc-jags.sourceforge.net/`

R[3] platform. Source code and instructions to replicate the simulated data and estimate the parameters can be downloaded from the GitHub repository through the link https://github.com/lizbethna/HMMprogressionOrdinalMissing.git.

Table 2 summarizes the means and standard deviations of the model parameters based on the 100 generated datasets. The estimated means of the posterior distributions of the model parameters associated to each time point can be compared to the true values from which the datasets were generated. It can be observed how the model parameters are reasonable well recovered, reporting small biases. Note that the model parameters can be estimated without using external information about the missing data.

Table 2: Means and standard deviations (SD) of the posterior estimates of the model parameters based on the 100 generated datasets.

Variable	**Parameter**	**True**	**Mean**	**SD**
Z_1	γ_1	0.4	0.4384	0.2063
Z_2	γ_2	0.2	0.2174	0.0461
X_1	β_1	1.0	1.1234	0.3440
X_2	β_2	0.5	0.5044	0.3034
X_3	β_3	-1.5	-1.6135	0.3306
X_4	β_5	-2.0	-2.2783	0.3746
Cutpoints	κ_1	0	0	—
	κ_2	1.5	1.6082	0.1628
	κ_3	3.0	3.1791	0.2273
Variances	λ_1	0.0400	0.0390	0.0001
	λ_2	0.0625	0.0617	0.0001
	λ_3	0.0900	0.0906	0.0001
	λ_4	0.1225	0.1125	0.0001

Data were simulated for a future time t_5 in relation with the 100 datasets. Since data were simulated, the responses are exactly known and the model can be applied to assess the predictive capability. Table 3 shows the confusion matrix of the observed data versus the estimated predictions for a future time t_5. Results show a global accuracy rate of 78.29%. This accuracy rate is very

[3] `https://cran.r-project.org/`

close to 80%, which is a relevant threshold given the fact that there are 4 classes and the decision boundaries are more complex than in the binary case.

Table 3: Relative frequencies between observed and predicted stages in time t_5.

Time t_5	**Predicted**			
Observed	Stage 1	Stage 2	Stage 3	Stage 4
Stage 1	0.0000	0.0016	0.0000	0.0000
Stage 2	0.0000	0.0678	0.0324	0.0000
Stage 3	0.0000	0.0433	0.4138	0.0233
Stage 4	0.0000	0.0012	0.1150	0.3013

In order to support the previous results in terms of accurate predictions, Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) for ordinal regression are calculated by following the proposal by Baccianella et al. [39]. Firstly, the cumulative probabilities are calculated for the observed values P_{it} and the predictions $\widehat{P_{it}}$, i.e.:

$$\begin{aligned} P_{it} &= \mathbb{P}\left[Y_{it} \leq y_{it} | \boldsymbol{x}_{it}, \boldsymbol{z}_{it}, \boldsymbol{\beta}, \boldsymbol{\gamma}\right], \\ \widehat{P_{it}} &= \mathbb{P}\left[Y_{it} \leq \widehat{y_{it}} | \boldsymbol{x}_{it}, \boldsymbol{z}_{it}, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}\right]. \end{aligned}$$

Then, MAE and RMSE are calculated by:

$$\begin{aligned} \text{MAE} &= \frac{1}{N \cdot T} \sum_{i=1}^{N} \sum_{t=1}^{T} \left| P_{it} - \widehat{P_{it}} \right|, \\ \text{RMSE} &= \sqrt{\frac{1}{N \cdot T} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(P_{it} - \widehat{P_{it}} \right)^2}. \end{aligned}$$

Finally, for time t_5 both criteria were calculated and the means and standard deviations based on the 100 datasets are presented in Table 4. It can be observed that both mean values are small as well as the standard deviations. This also supports the idea that the prediction accuracy is good.

As a summary, the recovering of model parameters and the values obtained for accuracy rate, MAE and RMSE are indicative of the good performance of this multiclass classification approach.

Table 4: Means and standard deviations of MAE and RMSE based on 100 datasets for time t_5.

Criterion	Mean	Standard deviation
MAE	0.035449	0.005178
RMSE	0.080724	0.012229

4.2. Real data-based case

The proposed model is applied to the experiment described in Section 2. A total of 36 PD patients with missing responses and covariates were used to impute responses and predictions during the time that the experiment lasted. Conducting this experiment has been difficult due to the amount of missing data. When the recruitment took place (2016), there were 36 respondents; the next year only 32 (89%) were available; the next year, this number was reduced to 28 (78%), and the last year 22 (61%) finished the experiment. In 2020, the experiment was not performed due to the coronavirus pandemic. This virus especially affects to the elder people, and the access to the headquarters of the Regional Association for Parkinson's Disease of Extremadura (Spain) was limited. The great amount of valuable data collected is used to impute data for 2017-2019 and to make predictions for 2020.

Relevant acoustic features were extracted from the collected voice recordings. These features had been previously proven to work properly for PD detection and monitoring progression. They are CPP, D2, RPDE, and MFCC5. Including more features did not improve the predictive capacity and may introduce multicollinearity problems.

The following prior distributions have been used: (i) for the regression parameters, $\beta_l \sim \mathrm{N}_L(0, 10000)$, for $l = 1, \ldots, L$ with $L = 4$ covariates, $\gamma \sim \mathrm{N}(0, 10000)$; (ii) for the cutpoints for all times $\kappa_k \sim \mathrm{N}(0, 1)$ where $\kappa_k < \kappa_{k+1}$, for $k = 2, \ldots, K-1$, with $K = 6$ categories; (iii) for the replicated covariates, $X^*_{itjl} \sim \mathrm{N}(x_{itjl}, \lambda_l)$, $X_{itjl} \sim \mathrm{N}(0, 1)$, $\lambda_l \sim \mathrm{InvGamma}(0.001, 0.001)$, where $i = 1, \ldots, N$, $t = 1, \ldots, T$, $j = 1, \ldots, J$, with $N = 36$ subjects, $T = 4$ time points, and $J = 3$ replications.

A total of 10,000 iterations with a burn-in of 1,000 and a thinning period of 5 generated values was used, yielding a sample of length 2,000. With these specifications, the chain generated by using the MCMC sampling algorithm appears to have converged.

Table 5 presents the posterior estimates of the regression parameters of

the proposed model, summarizing their corresponding posterior means, standard deviations, and 2.5% and 97.5% quantiles (95% intervals).

Table 5: Means, standard deviations (SD), and 95% intervals of the posterior estimates of the model parameters.

Variable	**Parameter**	**Mean**	**SD**	**95% interval**
Gender	γ^F	1.3425	0.2993	(0.7493, 1.9276)
CPP	β^F_{CPP}	-0.7630	0.2747	(-1.3220, -0.2415)
D2	β^F_{D2}	0.0386	0.2485	(-0.4570, 0.5248)
RPDE	β^F_{RPDE}	0.1933	0.3452	(-0.4806, 0.8641)
MMFCC5	β^F_{MMFCC5}	-0.4294	0.1981	(-0.8194, -0.0494)
Cutpoints	κ_1	0	—	—
	κ_2	0.3102	0.0982	(0.1478, 0.5306)
	κ_3	1.2043	0.1460	(0.9350, 1.5053)
	κ_4	1.6312	0.1657	(1.3196, 1.9600)
	κ_5	3.3099	0.2556	(2.8064, 3.8023)
Variances	λ_1	0.2197	0.0139	(0.1933, 0.2490)
	λ_2	0.3372	0.0215	(0.2981, 0.3824)
	λ_3	0.3042	0.0199	(0.2672, 0.3452)
	λ_4	0.2261	0.0145	(0.1995, 0.2573)

The observed HY stages, imputed HY stages for years 2017 to 2019, and predicted HY stages for year 2020 are presented in Table 6.

Years from 2017 to 2019 contain a certain amount of missing data, but in 2020 the experiment did not continue and the proposed approach has provided a full prediction based on the previous observed data (features and HY stages). The impact of COVID-19 on older people and, particularly, on people suffering from PD has had devastating consequences in terms of deaths and worsening of life quality. The headquarters of the Regional Association for Parkinson's Disease of Extremadura (Spain) were partially closed in 2020, and the activities were reduced. In order to analyze the model performance, we compare the HY stage predictions obtained with the proposed approach to the actual HY stage in 2020. From the 36 subjects that began the experiment, 9 subjects were not available due to their decease, 13 subjects abandoned the activities (7 in a definitive way and 6 with the intend of going back later), and only 14 continued the activities in 2020 and had a medical evaluation with the HY stage. A total of 10 out 14 predictions performed by the proposed

Table 6: Gender, age (at baseline), and HY stages of the involved subjects. Imputed (2017-2019) and predicted stages (2020) are presented in bold.

Subject	**Gender**	**Age**	**Stages by year**				
			2016	**2017**	**2018**	**2019**	**2020**
1	M	60	3.0	3.0	3.0	3.0	**3.0**
2	M	67	3.0	**3.0**	4.0	**4.0**	**4.0**
3	M	46	1.0	1.5	2.0	2.0	**2.5**
4	M	69	2.0	2.0	2.0	2.0	**2.5**
5	M	69	1.0	1.5	2.0	2.0	**2.5**
6	F	65	2.0	2.0	2.5	2.5	**3.0**
7	F	70	2.5	3.0	**3.0**	3.0	**3.0**
8	M	70	2.0	**2.0**	2.0	**2.5**	**3.0**
9	M	83	4.0	4.0	4.0	**4.0**	**4.0**
10	M	69	3.0	3.0	**3.0**	**3.0**	**4.0**
11	M	75	1.0	1.5	1.5	2.0	**2.5**
12	F	82	3.0	3.0	3.0	**3.0**	**4.0**
13	F	74	3.0	3.0	3.0	4.0	**4.0**
14	M	72	3.0	3.0	**3.0**	**3.0**	**4.0**
15	M	80	1.0	1.0	3.0	3.0	**3.0**
16	F	75	3.0	3.0	3.0	3.0	**3.0**
17	M	69	2.5	3.0	**3.0**	3.0	**3.0**
18	M	66	2.5	3.0	3.0	3.0	**3.0**
19	F	73	2.5	2.5	**3.0**	**3.0**	**3.0**
20	M	77	3.0	3.0	**3.0**	**3.0**	**4.0**
21	M	53	1.0	1.5	2.0	2.0	**2.5**
22	F	73	3.0	3.0	**3.0**	**3.0**	**4.0**
23	M	60	2.0	2.0	3.0	**3.0**	**3.0**
24	F	72	3.0	4.0	4.0	**4.0**	**4.0**
25	F	58	4.0	4.0	4.0	4.0	**4.0**
26	M	79	2.0	3.0	3.0	3.0	**3.0**
27	M	72	1.0	2.0	2.0	2.0	**2.5**
28	M	69	2.0	2.0	2.5	2.5	**3.0**
29	M	71	3.0	3.0	4.0	**4.0**	**4.0**
30	F	55	1.0	1.5	2.0	3.0	**3.0**
31	M	74	2.0	2.0	2.0	2.0	**2.5**
32	M	71	3.0	3.0	3.0	3.0	**3.0**
33	F	65	3.0	**3.0**	4.0	**4.0**	**4.0**
34	F	69	2.0	2.5	2.5	2.5	**3.0**
35	M	62	1.0	**2.0**	**3.0**	3.0	**3.0**
36	M	70	1.5	2.5	2.5	**3.0**	**3.0**

approach were accurate (71.43%), whereas the other 4 only differed in one stage. Specifically, two subjects were predicted to have stage 2.5, whereas the true stage was 3, one subject was predicted to have stage 3 and had 2.5, and finally one subject was predicted as having stage 3 and had stage 4. The results are reasonably good for a multiclass ordered problem, given the great amount of missing data and the small size of the sample.

Figure 3 shows the observed stages and estimated trajectories for six subjects with the median imputed and predicted responses. The shades represent the 2.5% and 97.5% quantiles of the estimated trajectories. Subjects 1, 21, and 35 got accurate predictions, whereas subjects 8, 10, and 36 abandoned the activities and were not assessed for HY stages in 2020.

These six subjects represent several profiles that are found in the experiment. The same plots are presented for the rest of the subjects involved in the experiment in tables A.4, A.5, and A.6 of Appendix A. Subject 1 is a typical case of PD subject, where the symptoms are stabilized through the years, and all the observations are available. The prediction suggested that the patient would keep in stage 3 in t_5 and it was right. Subject 8 only provides two observed stages (t_1 and t_3), then stages for t_3 and t_4 are imputed by the model. An HY increase is imputed for t_4, and another increase is projected for t_5. In these two last years, the quantile bands suggest that the imputation and prediction are performed with more uncertainty. The next subject to analyze is number 10. This patient only has data for the first two years. He was estimated to keep in stage 3 during two years more and was predicted to be in stage 4 in t_5. The slope of the trajectory suggests that the symptoms may worsen in the last year. Subject 21 is a typical case where the disease advances faster than in the previous patient. The prediction for t_5 is accurate and captures the essence of this increase with a larger uncertainty provided by the quantile bands. Now, the subject 35 represents patients with a very fast advance of PD, since he goes from stage 1 to stage 3 in three years. The imputations for time t_2 and t_3 capture this increase with a large uncertainty since 4 stages are involved in this increase. The prediction for t_5 is accurate for this subject. Finally, subject 36 has data for the first three years and the approach imputes the stage for time t_4 and predicts the one for time t_5. An increase to stage 3 is imputed, which is maintained for the last year.

The very few studies addressing the problem of tracking PD progression based on acoustic features are based on the UPDRS scale, as commented in the introduction section. Most of them are based on short-term studies.

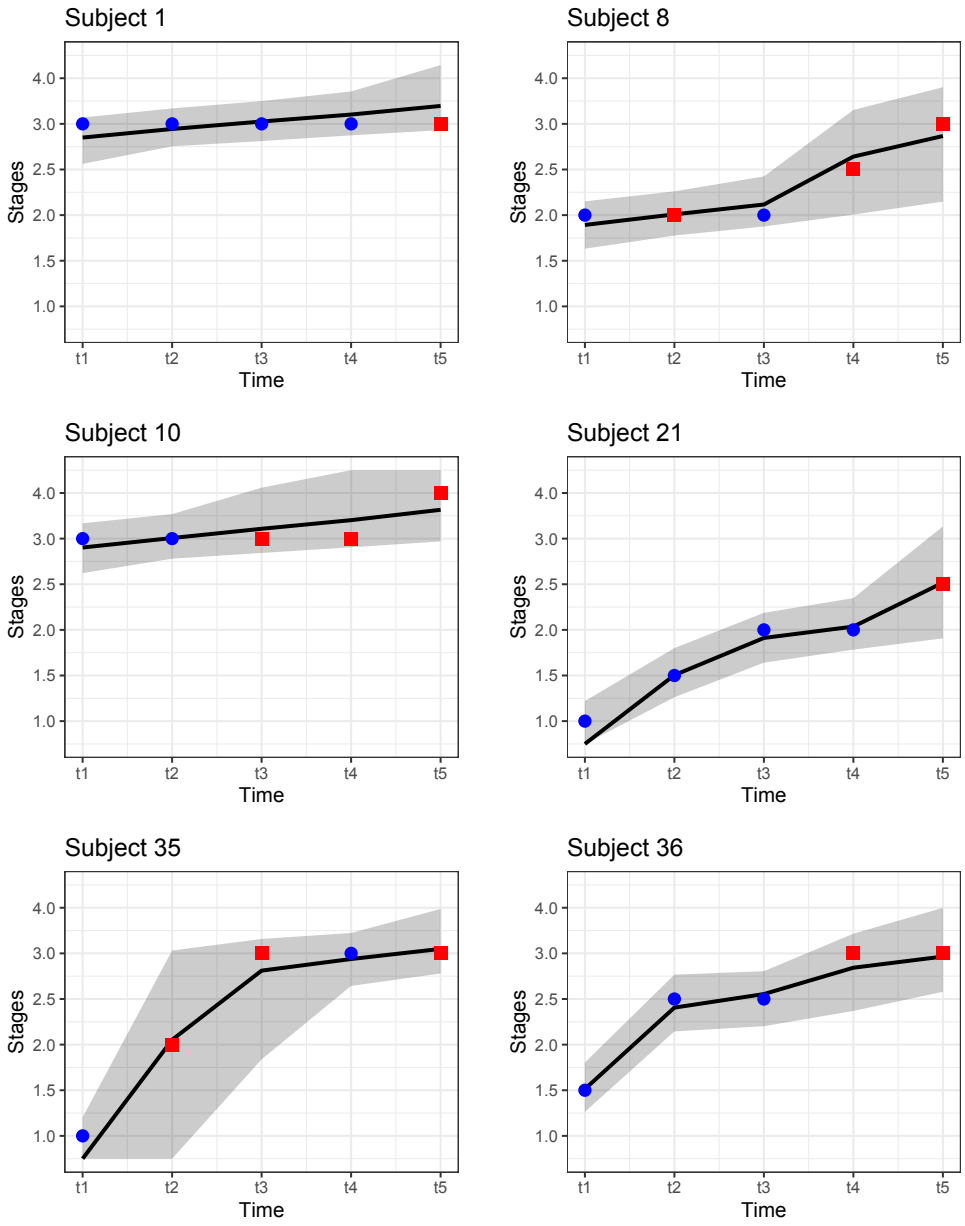

Figure 3: Observed stages (blue circle) and estimated trajectories for six typical subjects with imputed and predicted data (red square).

Arias-Vergara et al. [40] showed the difficulty of conducting long-term experiments, since they began with 62 subjects at the first year and finished with only 7 in the fourth year (also considering the UPDRS scale). This confirms the need of developing and using methodologies addressing missing data to make imputations and predictions. To the best of the author's knowledge, the conducted experiment and the proposed approach are the only ones in the scientific literature considering HY scales and acoustic features for long-term. Pérez et al. [41] conducted a pilot experiment to track PD severity based on the HY scale with 32 patients in two years, which was clearly insufficient to make any conclusion.

5. Conclusion

Tracking the progression of PD can be addressed with the use of acoustic features in an objective, remote, non-invasive, fast, and economically sustainable way. Some approaches have been used for the UPDRS scale, but there is a lack of studies and approaches for the HY scale. The conducted experiment provided challenging data that must be addressed with a specific approach that is able to manage missing data, the non-decreasing nature of the disease, and replicated covariates obtained from repeated voice recordings.

An HMM for the longitudinal data has been derived to manage these three issues. Latent variables have been included in the approach, leading to a data augmentation scheme, which has been solved by using MCMC methods. The proposed approach has been tested on simulated data, providing good results, and also has been applied to the data coming from the experiment. The imputed data and predictions for future year are compatible with the progression of the disease and provide profiles that usually appear under this disease. The conducted experiment and the proposed approach fill a gap in the scientific literature on experiments and methodologies for tracking PD progression based on acoustic features and the HY scale.

In spite of the recruitment difficulties, it would be very interesting to conduct long-term studies with many more participants and even more years to which the proposed approach can be applied. Learning from past data is a key aspect in this context. Also, it would be interesting to apply the proposed approach to features extracted from other recording protocols different from the /a/ sustained phonation, such as continuous speech and/or those based on diadochokinetic tests. This would help to derive an expert system that can be integrated into the protocols of neurology units in hospital centers.

Acknowledgments

This research has been supported by *Agencia Estatal de Investigación*, Spain (Project MTM2017-86875-C3-2-R), *Junta de Extremadura*, Spain (Projects IB16054, GR18108, and GR18055), and the European Union (European Regional Development Funds). Lizbeth Naranjo has also been partially supported by *UNAM-DGAPA-PAPIIT*, Mexico (Project IN118720).

All procedures performed in this study involving human participants were in accordance with the ethical standards of the Bioethical Committee of the University of Extremadura (Spain) and with the 1964 Helsinki Declaration and its later amendments.

No competing interests are declared.

Appendix A. Observed, imputed and predicted stages

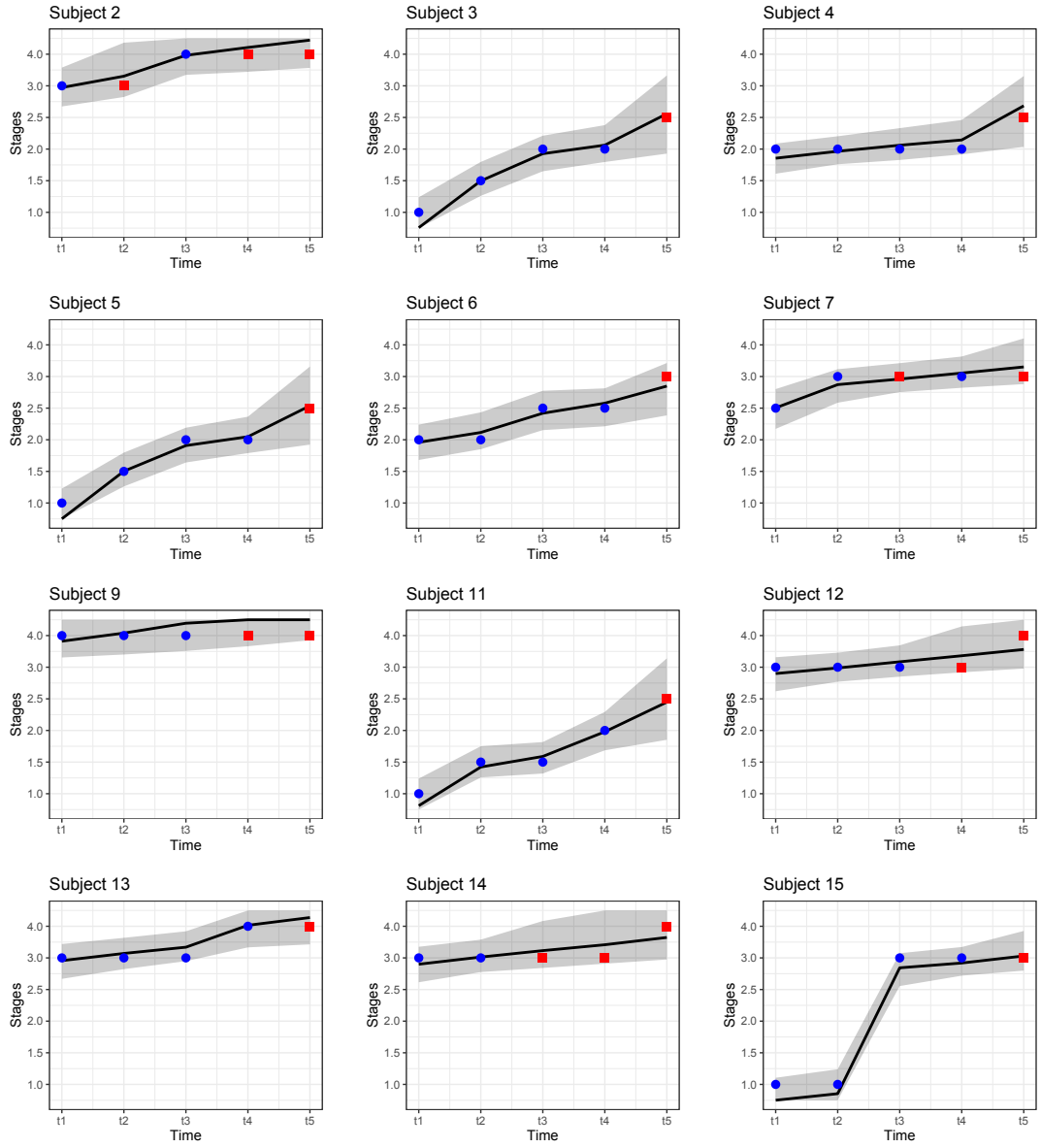

Figure A.4: Observed stages (blue circle) and estimated trajectories for twelve subjects with imputed and predicted data (red square).

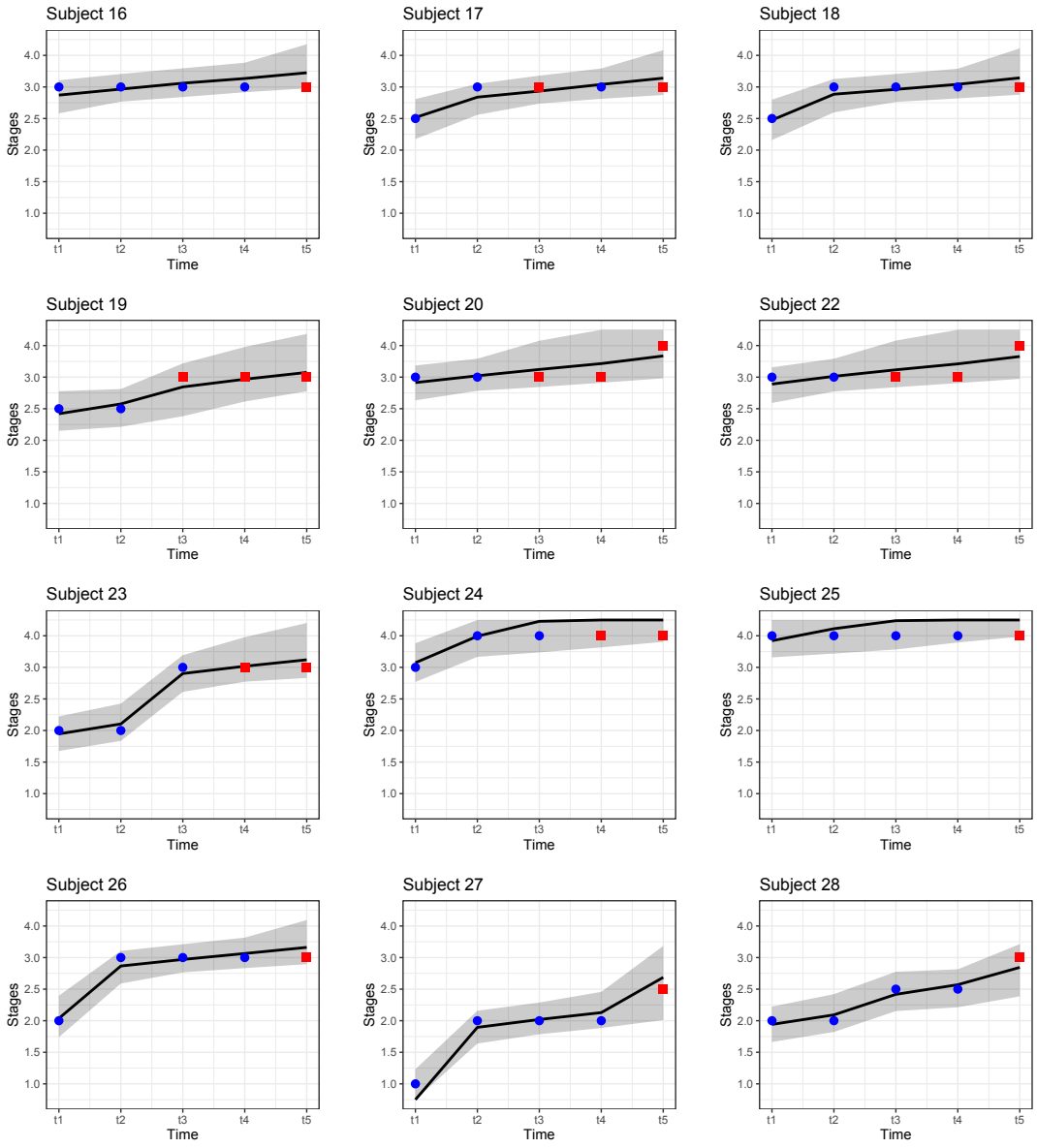

Figure A.5: Observed stages (blue circle) and estimated trajectories for twelve subjects with imputed and predicted data (red square).

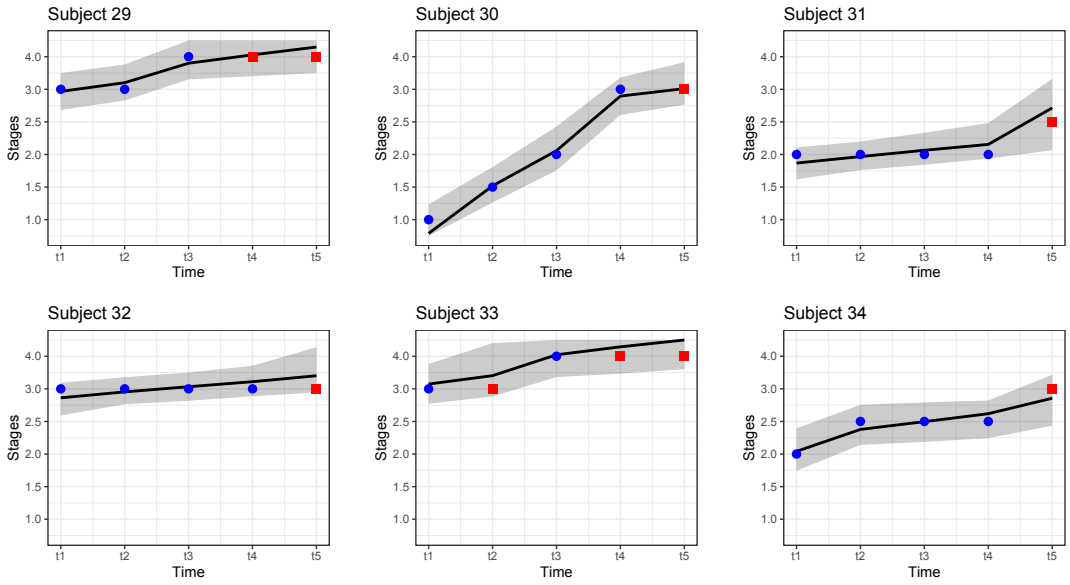

Figure A.6: Observed stages (blue circle) and estimated trajectories for six subjects with imputed and predicted data (red square).

References

[1] E. R. Dorsey, B. R. Bloem, The Parkinson pandemic - A call to action, JAMA Neurology 75 (1) (2018) 9–10.

[2] J. Yanase, E. Triantaphyllou, A systematic survey of computer-aided diagnosis in medicine: Past and present developments, Expert Systems with Applications 138 (2019) 112821.

[3] S. Hegde, S. Shetty, S. Rai, T. Dodderi, A survey on machine learning approaches for automatic detection of voice disorders, Journal of Voice 33 (6) (2019) 947.e11–947.e33.

[4] L. Moro-Velazquez, J. A. Gomez-Garcia, J. D. Arias-Londoño, N. Dehak, J. I. Godino-Llorente, Advances in Parkinson's disease detection and assessment using voice and speech: A review of the articulatory and phonatory aspects, Biomedical Signal Processing and Control 66 (2021) 102418.

[5] M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, L. O. Ramig, Suitability of dysphonia measurements for telemonitoring of Parkinson's disease, IEEE Transactions on Biomedical Engineering 56 (4) (2009) 1015–1022.

[6] D. Montaña, Y. Campos-Roca, C. J. Pérez, A diadochokinesis-based expert system considering articulatory features of plosive consonants for early detection of Parkinson's disease, Computer Methods and Programs in Biomedicine 154 (2018) 89–97.

[7] J. R. Orozco-Arroyave, F. Hönig, J. D. Arias-Londoño, J. Vargas-Bonilla, K. Daqrouq, S. Skodda, J. Rusz, E. Nöth, Automatic detection of Parkinson's disease in running speech spoken in three different languages, Journal of the Acoustic Society of America 139 (1) (2016) 481–500.

[8] D. Jain, A. K. Mishra, S. K. Das, Machine learning based automatic prediction of Parkinson's disease using speech features, in: P. Bansal, M. Tushir, V. E. Balas, R. Srivastava (Eds.), Proceedings of International Conference on Artificial Intelligence and Applications, Springer, 2021, pp. 351–362.

[9] S. Chatterjee, A. S. Hadi, Regression Analysis by Example, 4th Edition, Wiley Series in Probability and Statistics, Wiley, 2006.

[10] S. Fahn, R. L. Elton, Recent developments in Parkinson's disease, Vol. 2, Macmillan Health Care information, Florham Park, NJ, 1987, Ch. Unified Parkinson's Disease Rating Scale (UPDRS), pp. 153–164.

[11] B. Post, M. P. Merkus, R. M. A. de Bie, R. J. de Haan, J. D. Speelman, Unified Parkinson's disease rating scale motor examination: Are ratings of nurses, residents in neurology, and movement disorders specialists interchangeable?, Movement Disorders 20 (12) (2005) 1577–1584.

[12] A. Tsanas, M. A. Little, P. E. McSharry, L. O. Ramig, Accurate telemonitoring of Parkinson's disease progression by non-invasive speech tests, IEEE Transactions Biomedical Engineering 57 (4) (2010) 884–893.

[13] D. Hemmerling, M. Wojcik-Pedziwiatr, Prediction and estimation of Parkinson's disease severity based on voice signal, Journal of Voice(In press) (2020).

[14] C. G. Goetz, G. T. Stebbins, D. Wolff, W. DeLeeuw, H. Bronte-Stewart, R. Elble, M. Hallett, J. Nutt, L. Ramig, T. Sanger, A. D. Wu, P. H. Kraus, L. M. Blasucci, E. A. Shamim, K. D. Sethi, J. Spielman, K. Kubota, A. S. Grove, E. Dishman, C. B. Taylor, Testing objective measures of motor impairment in early Parkinson's disease: Feasibility study of an at-home testing device, Movement Disorders 24 (4) (2009) 551–556.

[15] O. Eskidere, F. Ertaç, C. Hanilçi, A comparison of regression methods for remote tracking of Parkinson's disease progression, Expert Systems with Applications 39 (5) (2012) 5523–5528.

[16] L. Naranjo, C. J. Pérez, J. Martín, Addressing voice recording replications for tracking Parkinson's disease progression, Medical and Biological Engineering and Computing 55 (3) (2017) 365–373.

[17] M. Nilashi, H. Ahmadi, A. Sheikhtaheri, R. Naemi, R. Alotaibi, A. Abdulsalam Alarood, A. Munshi, T. A. Rashid, J. Zhao, Remote tracking of Parkinson's disease progression using ensembles of deep belief network and self-organizing map, Expert Systems with Applications 159 (2020) 113562.

[18] C. Goetz, S. Fahn, P. Martinez-Martin, W. Poewe, C. Sampaio, G. Stebbins, M. Stern, B. Tilley, R. Dodel, B. Dubois, R. Holloway, J. Jankovic, J. Kulisevsky, A. Lang, A. Lees, S. Leurgans, P. LeWitt, D. Nyenhuis, C. Olanow, O. Rascol, A. Schrag, J. Teresi, J. Van Hilten, N. LaPelle, Movement disorder society-sponsored revision of the unified Parkinson's disease rating scale (MDS-UPDRS): process, format, and clinimetric testing, Movement Disorders 22 (2007) 41–47.

[19] E. Smith, J. Bertoni, D. Bhatti, D. Torres-Russotto, Is the UPDRS a reliable tool for detecting the worse side in Parkinson's disease?, Neurology 94 (15 Supplement) (2020).

[20] M. M. Hoehn, M. D. Yahr, Parkinsonism: onset, progression and mortality, Neurology 17 (5) (1967) 427–442.

[21] C. G. Goetz, W. Poewe, O. Rascol, C. Sampaio, G. T. Stebbins, C. Counsell, N. Giladi, R. G. Holloway, C. G. Moore, G. K. Wenning, M. D. Yahr, L. Seidl, Movement disorder society task force report on the Hoehn and Yahr staging scale: Status and recommendations, Movement Disorders 19 (9) (2004) 1020–1028.

[22] S. Skodda, W. Grönheit, N. Mancinelli, U. Schlegel, Progression of voice and speech impairment in the course of Parkinson's disease: A longitudinal study, Parkinson's Disease Article ID 389195 (2013).

[23] M. J. García-Zattera, A. Jara, E. Lesaffre, G. Marshall, Modeling of multivariate monotone disease processes in presence of misclassification, Journal of the American Statistical Association 107 (499) (2012) 976–989.

[24] L. Naranjo, L. J. R. Esparza, C. J. Pérez, A hidden Markov model to address measurement errors in ordinal response scale and non-decreasing process, Mathematics 8 (4) (2020) 622.

[25] L. Naranjo, C. J. Pérez, R. Fuentes-García, J. Martín, A hidden Markov model addressing measurement errors in the response and replicated covariates for continuous nondecreasing processes, Biostatistics 21 (4) (2020) 743–757.

[26] J. R. Duffy, Motor Speech Disorders: Substrates, Differential Diagnosis, and Management, Elsevier, 2012.

[27] S. Jannetts, A. Lowit, Cepstral analysis of hypokinetic and ataxic voices: Correlations with perceptual and other acoustic measures, Journal of Voice 28 (6) (2014) 673–680.

[28] C. O. Sakar, G. Serbes, A. Gunduz, H. C. Tunc, H. Nizam, B. E. Sakar, M. Tutuncu, T. Aydin, M. E. Isenkul, H. Apaydin, A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable Q-factor wavelet transform, Applied Soft Computing 74 (2019) 255–263.

[29] Y. Bang, K. Min, Y. H. Sohn, S. Cho, Acoustic characteristics of vowel sounds in patients with Parkinson's disease, NeuroRehabilitation 32 (2013) 649–654.

[30] J. R. Orozco-Arroyave, J. D. Arias-Londoño, J. F. Vargas-Bonilla, E. Nöth, Analysis of speech from people with Parkinson's disease through nonlinear dynamics, in: T. Drugman, T. Dutoit (Eds.), Advances in Nonlinear Speech Processing, Vol. LNAI 7911 of Lecture Notes in Artificial Intelligence, Springer-Verlag, 2013, pp. 112–119.

[31] A. Tsanas, M. A. Little, C. Fox, L. O. Ramig, Objective automatic assessment of rehabilitative speech treatment in Parkinson's disease, IEEE Transactions on Neural Systems and Rehabilitation Engineering 22 (1) (2013) 181–190.

[32] M. Picillo, A. Nicoletti, V. Fetoni, B. Garavaglia, M. T. Pellecchia, The relevance of gender in Parkinson's disease: a review, Journal of Neurology 264 (2017) 1583–1607.

[33] J. Albert, S. Chib, Bayesian analysis of binary and polychotomous response data, Journal of the American Statistical Association 88 (422) (1993) 669–679.

[34] M. K. Cowles, Accelerating Monte Carlo Markov chain convergence for cumulative-link generalized linear models, Statistics and Computing 6 (1996) 101–111.

[35] J. P. Buonaccorsi, Measurement Error: Models, Methods and Applications, Chapman and Hall/CRC, Boca Raton, FL, 2010.

[36] L. Naranjo, E. Lesaffre, C. J. Pérez, A mixed hidden Markov model for multivariate monotone disease processes in the presence of measurement errors, Statistical Modelling(In press) (2020).

[37] D. J. Lunn, A. Thomas, N. Best, D. Spiegelhalter, WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility, Statistics and Computing 10 (4) (2000) 325–337.

[38] W. R. Gilks, S. Richardson, D. J. Spiegelhalter, Markov Chain Monte Carlo in Practice, Chapman and Hall, London, 1996.

[39] S. Baccianella, A. Esuli, F. Sebastiani, Evaluation measures for ordinal regression, in: Ninth International Conference on Intelligent Systems Design and Applications, 2009, pp. 283–287.

[40] T. Arias-Vergara, J. C. Vasquez-Correa, J. R. Orozco-Arroyave, J. Vargas-Bonilla, E. Nöth, Parkinson's disease progression assessment from speech using GMM-UBM, in: INTERSPEECH 2016, 2016, pp. 1933–1937.

[41] C. J. Pérez, Y. Campos-Roca, L. Naranjo, J. Martín, Diagnosis and tracking of Parkinson's disease by using automatically extracted acoustic features, Journal of Alzheimer's Disease and Parkinsonism 6 (5) (2016) 260.