POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Bayesian Scalar-on-Function Models for Functional Data with Wearable Applications

Bayesian Statistics project

Master of Science in Mathematical Engineering

February 2024

Students:

Gabriele Bruni - gabriele1.bruni@mail.polimi.it
Stefania Colombo - stefania4.colombo@mail.polimi.it
Silvia D'Amicantonio - silvia.damicantonio@mail.polimi.it
Chiara Facco - chiara.facco@studenti.unimi.it
Luca Meanti - luca.meanti@mail.polimi.it
Michele Sparviero - michele.sparviero@mail.polimi.it

Tutors: Riccardo Corradin - Yordan Raykov

Academic Year: 2023-24

Contents

1 | Introduction

Parkinson's disease (PD) is the second most common neurodegenerative disorder, surpassed only by Alzheimer's disease. It is characterized by a decline in health-related quality of life and an escalation of both psychological and physical challenges over time.
The symptoms of Parkinson's disease span both nonmotor and motor domains. The former symptoms include psychological problems, cognitive impairment, and sleep disturbances. On the motor side, individuals may experience slowness of movement, rigidity, and tremors [1].
Over the past three decades, the number of patients diagnosed with PD has more than doubled, to over seven million today. Moreover, despite the potential contribution of an aging population, it's noteworthy that one in three patients is under the age of 65, with occurrences reported even in individuals in their twenties and thirties.

While our scientific knowledge of the physical symptoms of PD is well-established, there is a growing need for research to shift focus towards understanding the broader impact of the disease and the significant challenges faced by individuals with the diagnosis [2].
In line with this commitment, we develop our project. In Section 2 we provide a comprehensive overview of the gathered data and the preprocessing techniques applied. Then, in Section 3, we present our Bayesian models, starting with a brief recap on the meaning and construction of a Scalar-on-Function regression model. Next, we move to the core aspects of our analysis, the Unconstrained and Constrained models, both generalized to a binary response regression, given the categorical nature of our response. Moreover, we add an extension to Mixed Models, taking as random effect the Sex and Medication group variables. Subsequently, we systematically discuss the outcomes of our analysis in Section 4, emphasizing noteworthy results and critically assessing the implications of our findings. In the concluding segment, Section 5, we synthesize the key takeaways from our analysis and draw overarching conclusions.

2 | Data and preprocessing

2.1. The Personalized Parkinson Project

The dataset we employ is derived from the Personalized Parkinson Project, a study carried out at the Radboud University Medical Center in the Netherlands and conducted on 650 individuals affected by Parkinson's disease (under some inclusion criteria such as age over 18 and disease duration below 5 years).

To be more precise, the study started on October 1^{st}, 2017 with an estimated completion on December 31^{st}, 2023 and the latest update dating back to April 2023.

Participants are required to reach the study site for three tests (an initial baseline visit and 2 annual follow-up visists), but, more importantly, to continuously wear an investigational device called Verily Study Watch.
Outcomes include:

- Clinical assessment (involving both motor and neuro-psychological tests);
- Collection of biospecimens (i.e. stool, whole blood, and cerebrospinal fluid);
- Magnetic resonance imaging (both structural and functional);
- ECG recordings (both 12-lead and Holter);
- collection of physiological and environmental data in daily life (enabled through the Verily Study Watch);

2.2. Dataset

The first piece of data we work with is made of a structure with 160 patients associated to the following variables:

Variable	Brief description
ID	Patient identification number
GroupSeverity	Intensity of tremor on daily activities (0,1,2,3,4)
Age	Age of the patient
Gender	Gender of the patient (1: Male; 2: Female)
PayedJob	Whether the patient is paid for the study (1) or not (0)
DailyActivities	Level of daily activities (1,2,3,4,5,6,7,8)
FistSympYear	Year of onset of the first symptom
WearingOff	Decreasing effectiveness of Levodopa medication
Medication	Group of medication
TremorProbability	Z-scored retrained classifier for the tremor

2.3. Preprocessing data

Before applying any analyses, we preprocess our dataset to handle its complexity. Initially, we experiment with two smoothing methods:

1. the Kalman Filter;
2. the Savitzky–Golay (Savgol) Filter.

The former relies on the state space representation, using noise parameters for smoothing. In contrast, the latter acts as a convolutional filter, similar to a specifically weighted moving average.
We primarily use Kalman Filtering for smoothing, resorting to Savgol Filtering only for select tests involving over-smoothed data.

After the smoothing stage, we perform a registration procedure to align the functional data. Formally, our covariate dataset is represented as:

$$X = \{c_{i,j,t}\} \qquad i \in \{1, ..., 159\},\ j \in \{1, ..., n_i\},\ t \in \{1, ..., 288\},$$

where i represents the patient level, j the day of observation, and t the time of observation within the day.
Using the square-root velocity framework, we align the data and extract the Karcher mean, representing the "average curve" to which other curves are aligned. Specifically, we perform registration on the n_i curves to ensure alignment at the patient level.

3 | Methodologies

In the upcoming subsections, we show how the models were constructed. The main references we used to build the latter come from the paper "An informative prior distribution on functions with application to functional regression" [3].

3.1. Review on Scalar on Function Regression

First of all, as mentioned, we need to handle variables with both scalar and functional values and to do so we rely on the Scalar on Function Regression Model, as presented in "Methods for Scalar on Function Regression" [4].
In this context, we address regression tasks where a set of functional covariates is employed to explain the scalar response.
Let:

- $\{x_i(t)\}_{i\in I},\ x_i : T \to \mathbb{R}$ for all $i \in I$ functional covariates;
- $\{y_i\}_{i\in I},\ y_i \in \mathbb{R}$ for all $i \in I$ scalar response;

Then the model can be expressed as:

$$Y_i = \mu + \int_T x_i(t)\beta_t\, dt + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2) \quad \forall\, i \in I.$$

where, Y_i represents the finite and scalar response, $\int_T x_i(t)\beta_t\, dt$ denotes the integral term involving the functional predictor $x_i(t) \in L^1(T)$ with separable and continuous functional coefficient $\beta_t \in T$. The error term ε_i follows a normal distribution $\mathcal{N}(0, \sigma^2)$, μ is a constant belonging to $\mathbb{R}$ and $T \subset \mathbb{R}^d$ is a product of compact intervals of $\mathbb{R}$.

3.2. Unconstrained Model

Initially we assume μ to be zero and σ^2 to be known, building the following Bayesian Model:

$$\begin{aligned} Y_i \,|\, \beta &\sim \mathcal{N}\left(\int_T x_i(t)\beta_t\, dt,\ \sigma^2\right) \quad \forall\, i \in I \\ \beta &\sim GP(0, K) \end{aligned}$$

with $\sigma^2 \in \mathbb{R}^+$ and K a continuous covariance function with a Gaussian kernel. We refer to this model as *Unconstrained*, as it does not assume that the trajectories of β follow specific behaviours.
For this first and straightforward model, with the likelihood and prior as described the posterior distribution of the β exists in closed form:

$$\beta\,|Y \sim GP(m, K^*),$$

$$m(t) = Lx(t)' \left(\Sigma + \sigma^2 \mathbb{I}_n\right)^{-1} Y$$
$$K^*(s,t) = K(s,t) - Lx(s)' \left(\Sigma + \sigma^2 \mathbb{I}_n\right)^{-1} Lx(t)$$

where

- $Lx(t)' = (Lx_1(t), \ldots, Lx_n(t))$ is the transpose of the column vector $Lx(t)$ with entries $Lx_i(t) := \int_T K(t,s)\, x_i(s)\, ds$, $i = 1, \ldots, n$;
- Y is the column vector with entries Y_i;
- $\Sigma \in \mathbb{R}^{n \times n}$ with entries $\Sigma_{i,j} = R(x_i, x_j) := \int_T \int_T K(s,t)\, x_i(s)\, x_j(t)\, ds\, dt$, $i, j = 1, \ldots, n$;
- $\mathbb{I}_n \in \mathbb{R}^{n \times n}$ denotes the identity matrix.

Furthermore, Y marginally is a multivariate normal $\mathcal{N}_n(0, \Sigma + \sigma^2 \mathbb{I}_n)$

Hence we have already all the information and there is no need of using MCMC methods.
However, when looking at our data, it is easly noticeable that for each curve of the functional response there are intervals of time in which it assumes null value. To incorporate this prior information in our model we need to introduce some *constraints*: let $T_0 \subset T$ be the union of such intervals, we constrain the support of the functional coefficient β_t to be zero in T_0, namely $\beta(t) = \beta_t \approx 0 \; \forall\, t \in T_0$.

3.3. Constrained Model

3.3.1. From unconstrained to constrained

Starting from the *Unconstrained model*, with μ and σ^2 known:

$$Y_i | \, \beta \sim \mathcal{N}\left(\mu + \int_T x_i(t) \beta_t \, dt, \, \sigma^2\right)$$
$$\beta \sim GP(0, K),$$

we now incorporate the prior information that the functional coefficients β are equal to zero on a given set $T_0 \subset T$.
We can build a *Constrained* Gaussian Process β^0 from the *Unconstrained* one, β, since, with probability one, the latter can be decomposed into the sum of two independent Gaussian Random Processes β^0 and β^1 such that β^0 is zero on T_0. Nevertheless β^0 and β^1 share the same smoothness properties with β (Theorem A.1).
It can be shown that the distribution of β^0 can be viewed as the distribution of β conditioned by $\beta_\tau = \beta(\tau) = 0 \; \forall \tau \in T_0$. In the following we refer to τ as the time instants in T_0, while with t we refer to instants in T in general.
A more detailed explanation can be found in the appendix.
Our Bayesian model becomes:

$$
\begin{aligned}
Y_i \,|\, \beta^0,\, \mu,\, \sigma^2 \;&\sim \mathcal{N}\left(\mu + \int_T x_i(t)\beta^0(t)\,dt, \sigma^2\right) \\
\beta^0 |\, \sigma^2, T_0 &\sim GP(0, \sigma^2 K^0) \\
p(\mu, \sigma^2, T_0) &\propto \frac{1}{\sigma^2} p(T_0)
\end{aligned}
$$

Note that we assume (μ, σ^2) to be independent of T_0 and that it has a noninformative prior distribution $p(\mu, \sigma^2) \propto \frac{1}{\sigma^2}$. In this case the posterior distribution of β^0 is:

$$
\begin{aligned}
\beta^0 \,|\mu, \sigma^2, T_0 &\sim \mathrm{GP}(m, K^*), \\
m(t) &= L^0 x(t)' \, M^{-1} \, (Y - \mu \mathbf{1}_n) \\
K^*(s,t) &= \sigma^2 [K^0(s,t) - L^0 x(s)' \, M^{-1} L^0 x(t)]
\end{aligned}
$$

While the conditional distribution of Y gven (μ, σ^2, T_0) is

$$
Y_i \,|\, \mu, \sigma^2, T_0 \sim \mathcal{N}_n \left(\mu \mathbf{1}_n,\, \sigma^2 M\right)
$$

where

- $L^0 x(t)' = (L^0 x_1(t), \ldots, L^0 x_n(t))$ is the transpose of the column vector $Lx(t)$ with entries $L^0 x_i(t) = \int_T K_0(t,s)\, x_i(s)\, ds$, $i = 1, \ldots, n$;
- Σ^0 is the nxn matrix with entries $\Sigma^0_{ij} = \int_T \int_T K^0(s,t) x_i(s) x_j(t) ds ds$
- $M = \Sigma^0 + \mathbb{I}_n$
- $\mathbf{1}_n = (1, ..., 1)'$

With some calculations we can also show that

$$
p(Y|T_0) \propto |M|^{-\frac{1}{2}} S_{11}^{-\frac{1}{2}} b^{-\frac{n-1}{2}},
$$

and that

$$
\mu, \sigma^2 | Y, T_0 \sim NIG(\frac{S_{1Y}}{S11}, \frac{1}{S11}, \frac{n-1}{2}, b)
$$

where

- $S_{ab} = a^T \Sigma^{-1} b \;\forall\, a, b \in \mathbb{R}^n$
- $b = \frac{1}{2}\big(S_{YY} - \frac{S_{1Y}^2}{S_{11}}\big)$

To simulate from the posterior of (μ, σ^2, T_0) we first sample from the posterior of T_0 using a MH algorhitm.

But, in order to do so, we need to set a prior on T_0.

3.3.2. An estimate of T_0

Let's introduce a couple of concepts:

Array T_0

We build a discrete grid of g time instants to represent T.
As a consequence the most practical choice for the structure of T_0 is a vector of entries $\{0,1\}^g$ such that in position i^{th} there is a 1 if the i^{th} element of the grid T belongs to T_0.

Runs of T_0

We refer to $r(T_0)$ as the number of runs of T_0 and of $T \setminus T_0$.
A run is a sequence of consecutive elements of T namely an interval of the discrete grid of times.
For example consider $g = 10$ and $T_0 = [0,0,0,0,1,1,1,0,0,1]$ then $r(T_0) = 4$ because we have two sequences of zeros and two sequences of ones.

We choose to adopt the following prior:

$$p(T_0) := e^{-\alpha r(T_0)}, \quad \alpha \in \mathbb{R}^+$$

Note that the number of runs $r(T_0)$ is determined by the number of intervals of T_0. This prior assigns less probability to candidate of T_0 with large $r(T_0)$. Moreover, to use MH to estimate posterior distributions, indeed we need an initial guess for T_0 and in particular for $r(T_0)$.
Let $\mathcal{S}_k = \{T_0 \in T, r(T_0) = k\}$ and c_k its cardinality, we have:

$$\mathbb{P}(r(T_0) = k) = \sum_{T_0 \in \mathcal{S}_k} p(T_0) = \frac{c_k e^{-\alpha k}}{\sum_{i=1}^{g} c_i e^{-\alpha i}} \tag{3.1}$$

As inital value of $r(T_0)$ we use the expected number of runs, namely we fix α and g and take the k that maximize (3.1).
Alpha is a parameter used to tune the prior probability.

3.3.3. Random Walk Metropolis-Hastings

Having both a prior and a likelihood, we can easily recover the posterior kernel, which is used to implement a MH algorithm. We proceed as follows:

- Set g and α and compute the expected number of runs k;
- Initialize the chain with $T_0^{(0)}$, a random combination of g zeros and ones with exactly k runs;

- While $i < 5.000$:
  - Generate T_0^{new} a candidate for $T_0^{(i+1)}$, by randomly choosing if removin/adding a 1 in a run's end or if splitting a run of ones by setting to 0 a random internal point;
  - Accept the new candidate with rate:

$$\text{rate} = \frac{p(Y|T_0^{new})p(T_0^{new})}{p(Y|T_0^{(i)})p(T_0^{(i)})}$$

3.3.4. Metropolis-Hastings customized version

Starting from our results of vanilla MH applied to the constrained model using the synthetic data, we implement a new version of the algorithm with the aim of improving the estimation of T_0.

To do such a thing, we decide to introduce an adaptive threshold for the *acceptance rate*. It consists of the update of the previous value of the threshold γ using past information from the chain, in order to make the chain itself explore regions of T_0 where the posterior takes more mass.

The update is made every step looking back at the last 100 iteration. In particular, the value of the threshold is updated in the following way:

- at each step i we save the information about acceptance/rejection of the proposal point $T_0^{(i)}$;
- we look back at the previous 100 steps and we count the number of candidates accepted A, then we update the threshold γ:

$$\gamma^{(new)} = 1.0 - 0.02 \cdot \frac{(100 - A)}{100}$$

We accept our candidate point $T_0^{(i)}$ if the *acceptance rate* is strictly greater than γ. The rational behind this alternative method is strongly related to our knowledge about the posterior kernel. By accepting candidates for which the acceptance rate is higher than $\gamma \in [0.98, 1]$, we are forcing the chain to move forward region of space where the posterior probability assigns more mass. The latter feature of this alternative method strongly supports a loss of ergodicity in the constructed chain. The adaptivity of the threshold is useful to avoid the well-known problem of a "stuck" chain as the output of the method.

3.4. Mixed Effect Models

Our next step is the introduction of group-specific parameters inside the previous model. These parameters are called *random effects* and in this section we present

models with different random effects. This development of our research has to intended as a first step in introducing hierarchical structures in our framework. Moreover, the huge computational cost is a substantial limit in our case.

3.4.1. Random Effects - Random Intercept

We define the following model

$$Y_{ij} = \int_T x_i(t)\beta(t)\,dt + z_{ij}\theta + \theta_0 + \varepsilon_i,$$

where we consider a random effect specific for the group $j \in \{0,\ 1\}$ of the observation $i \in \{1, ..., 159\}$. In detail we have:

$$\begin{aligned} \beta \,|\, \mu, \sigma^2, T_0 &\sim GP(0, \sigma^2 K^0) \\ \theta \,|\, \mu_s, \sigma_s^2 &\sim \mathcal{N}(\mu_s, \sigma_s^2) \\ \theta_0 \,|\, \mu_0, \sigma_0^2 &\sim \mathcal{N}(\mu_0, \sigma_0^2) \\ \varepsilon_i &\sim \mathcal{N}(0, \sigma^2) \end{aligned}$$

3.4.2. Random Effects - Random Slope

Let the model be

$$Y_{ij} = \int_T x_i(t)\beta(t)dt = \int_T x_{ij}(t)\theta(t)dt + \int_t x_{ij}(t)\gamma_j(t)dt + \varepsilon_i$$

$$\begin{aligned} \theta|\mu, \sigma^2, T_0 &\sim GP(0, \sigma^2 K^0) \\ \gamma_j &\sim GP(0, \sigma_0^2 K^0) \qquad j \in \{0, 1\} \\ \varepsilon_i &\sim \mathcal{N}(0, \sigma^2) \end{aligned}$$

The covariates are:

- functional variable $x_i(t) \quad i = 1 : 159\ , \quad t \in T$;
- variable $\lambda_i = \mathbf{1}_{\text{lev doses of pat i} \geq 5} \quad i = 1 : 159$;

3.5. Logistic Regression Model

Since our models involve modeling a categorical response, on top of them we apply a *logit transformation* to predict the probability of belonging to a specific group. Formally speaking, since $Y \in \{0, 1\}$, if we obtain fitted values $\hat{y}$ from the regression task, we proceed with

$$F(\hat{y}) = \frac{1}{1 + e^{-\hat{y}}}.$$

Despite the fact that the response can assume three different values (i.e. {1,2,3}), we merge the last two classes in order to perform binary classification. The choice of merging those groups instead of the first ones is motivated by the small number of observations belonging to class 3.

Finally, we classify Y as 1 if the logit function of the fitted values is greater than 0.5, 0 otherwise.

4 | Results

4.1. Synthetic Data Applications

We first test all the proposed methods on a synthetic dataset. The main advantages of doing so are the following:

- we can reduce the dataset complexity and dimensionality to test many alternative choices for hyperparameters and priors. In fact, one of the main issues in our research is the computational cost of many procedures. To give an example, the estimation of T_0 via MH requires around 25 seconds for every iteration, in our real-world applications;
- we can proceed with a supervised approach. In a synthetic environment we actually know how accurate is our outcome, by construction.

4.1.1. Unconstrained Model

We simulate a covariate set from a centered gaussian process with covariance K, where $K_{i,j} = e^{-(T_i - T_j)^2}$, and $T_i = -3 + \Delta i$, $i \in \{0, ..., 99\}$, $\Delta = 6/100$. We use $\sigma = 1$ and $\beta(t) = \sin(t\pi/4)$ with $t \in T$. Starting from there, we can easily reproduce the scalar response and check the results.

Our results are quite accurate even by changing σ, the structure of K or the actual functional parameter to generate the actual scalar response β. A graphical view of the outcome is shown in the following graphs (figure 1).

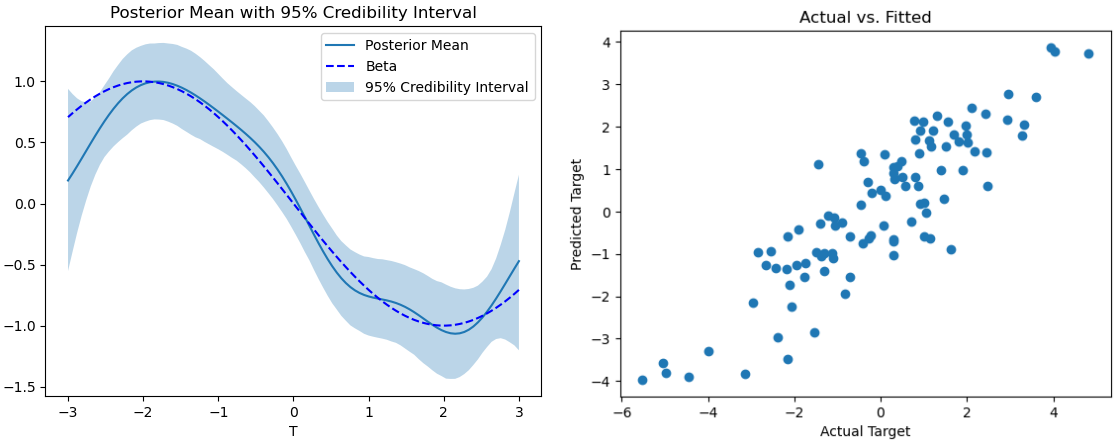

Figure 1: Unconstrained model results using synthetic data: on the left it is shown the posterior mean of the estimated β parameter with the corresponding credible interval. On the right the comparison between the real targets and the predicted ones.

4.1.2. The Estimation of T0

With the same dataset we assume $T_0 = [-1, 0.5] \subset T = [-3, 3]$. After a projection of β, thanks to the RKHS properties, we obtain the actual constrained functional parameter β_0. We construct a similar supervised environment and generate the functional response.

We proceed in the estimation of T_0 with both the proposed methods, highlighting the efficiency of the customized MH by testing it with only 500 iterations and a burn-in of 100. On the other hand, the vanilla MH is run with 2000 iterations and burn-in 1000. Once we get the chain, we use the mode as an estimate of T_0.

We would like to underline that the method proposed in our references is strongly initialization-dependent, while the custom one avoids this issue by construction.
The results are shown below (figure 2). It's important to note that all the curves depicted have a maximum value of 1. To enhance clarity in the graphical representation, we multiply the estimation of T_0 for the Vanilla MH by a factor of 1.1, and the estimation for our customized MH by a factor of 1.2.

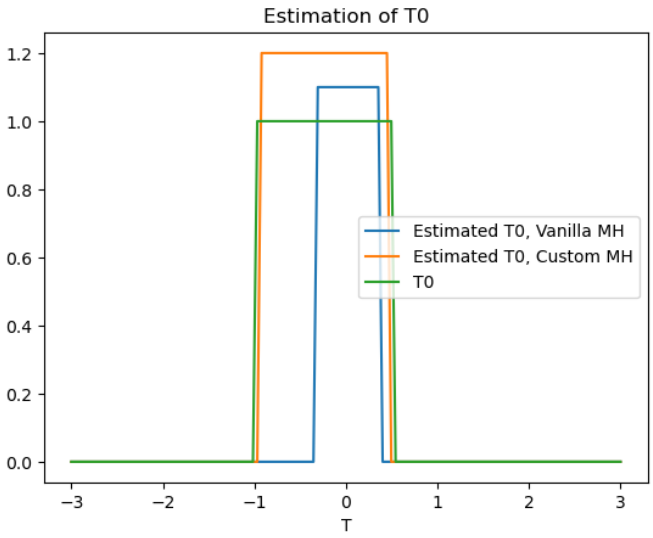

Figure 2: Different estimations of T_0 obtained with the two Metropolis-Hastings procedures illustrated, using synthetic generated data. In green it is shown the real T_0.

4.1.3. Constrained Model

If T_0 is given (or estimated), we can use it to calculate K_0 and to proceed in the assessment of the constrained model.
We use the same data as in the latter subsection and we evaluate analytically the posterior distribution of the functional parameter.

We want to underline that, in general, the prior (and hence the posterior) covariance matrix is not invertible with a Gaussian kernel as the one we adopt. To avoid a null

determinant, we add a diagonal matrix d with $d_{i,i} = 0.0001 \; \forall i \in \{0, ...g\}$.
In the following graph (figure 3) we plot both the estimated parameter and the actual parameter without the constraints, to highlight the impact of them in the interval T_0.

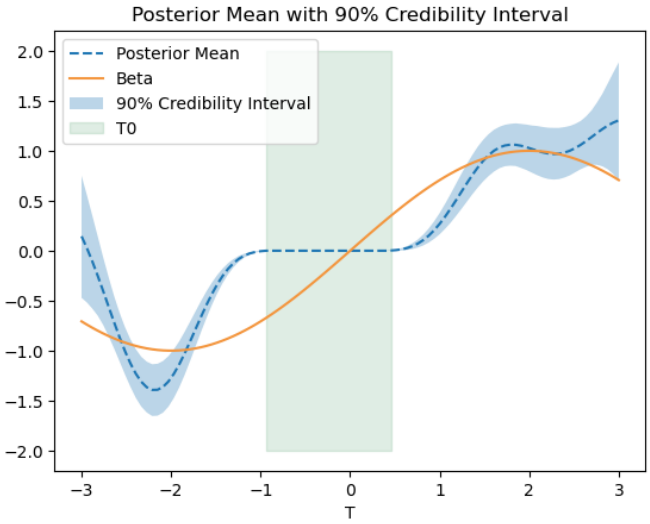

Figure 3: Constrained model results using synthetic generated data: posterior mean of the estimated β parameter with the corresponding credible interval. In green it is highlighted T_0. Finally in orange it is shown the true parameter.

4.2. Real Data Applications

From now on, we consider the original collection of data, but in its smooth and aligned format (refer to section 2.3 for more details).
In particular, our main focus is on severity prediction using the tremor probability as the functional covariate. The models proposed are the same as shown in the latter section, while we make two main hyperparameter choices, imposing $\sigma = 1$ and the prior kernel' scale $l = 5$.

Note that, in any application, we use the logit transform on top of the model to enforce binary classification. Concerning the Unconstrained and Constrained Model, we estimate analytically the posterior of the functional parameter, while for our Mixed Effect extension, we rely on Stan since we have no explicit distribution.

4.2.1. Unconstrained Model

The Unconstrained Model has to be intended as a benchmark for further developments. The following graphs (figure 4) show both the posterior mean with suitable credibility intervals to assess uncertainty and the accuracy of the model. As it is evident from the second plot, the model often overestimates the severity, resulting

in an overall 62.26% of correctly classified observations.

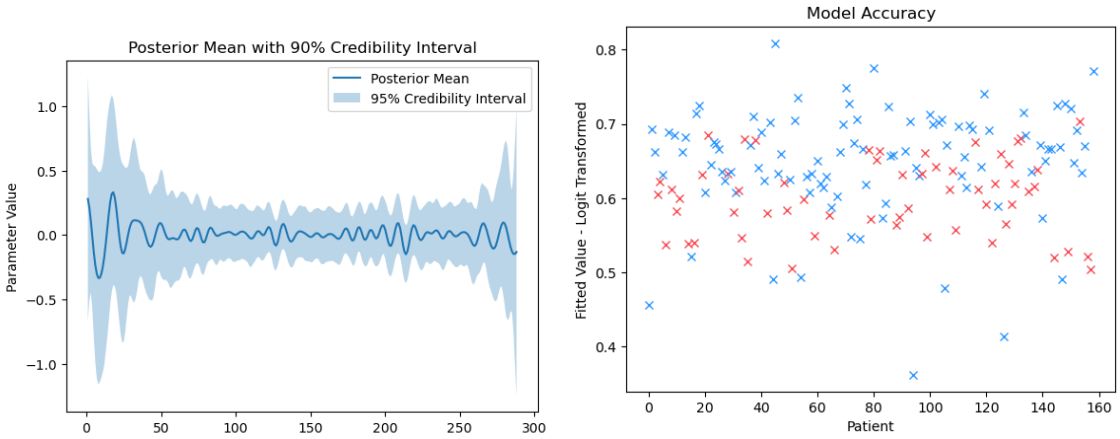

Figure 4: Unconstrained model results using real data: on the left it is shown the posterior mean of the estimated β parameter with the corresponding credible interval. On the right the fitted values are reported for the severity target. In blue the correct classified ones, in red the misclassified ones.

4.2.2. Estimation of T0

We proceed in the estimation of T_0, both with a regular RWMH and our customized proposal (5). In both cases the outcome is a scatter vector, supporting our guess of a strong noisy component misleading the posterior parameter estimation.

We construct many chains to study the behavior of T_0 along them and we conclude that our best choice is the one obtained with $\alpha = 1$, in terms of convergence and stationarity.
On the other hand, we highlight that we have no evidence about the ergodicity of the chain constructed by means of our customized MH method.

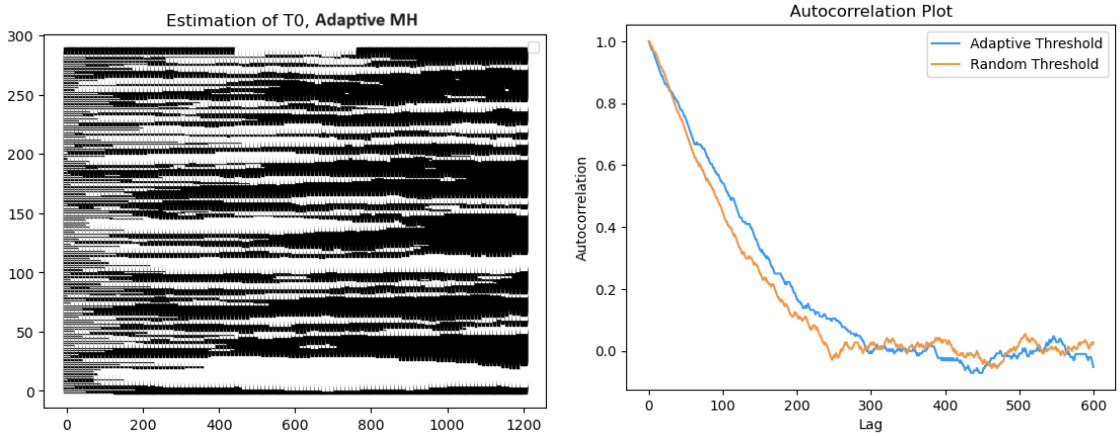

Figure 5: On the left: evolution of the estimated interval at different steps of the chain. The vertical axis represents the T interval, where the portion belonging to T_0 is highlighted in black. The horizontal axis represent the interation number. On the right: comparison between the autocorrelation plot of the two proposed Metropolis-Hastings methods.

4.2.3. Constrained Model

Right after the estimation of T_0, we proceed by applying our theoretical framework to assess the Constrained Model. It is evident from the graph below (figure 6) that our constraints are perfectly adopted by the posterior mean in terms of shape. In addition, the results encourage the usage of such a method, thanks to substantial improvements in accuracy (70.44%).

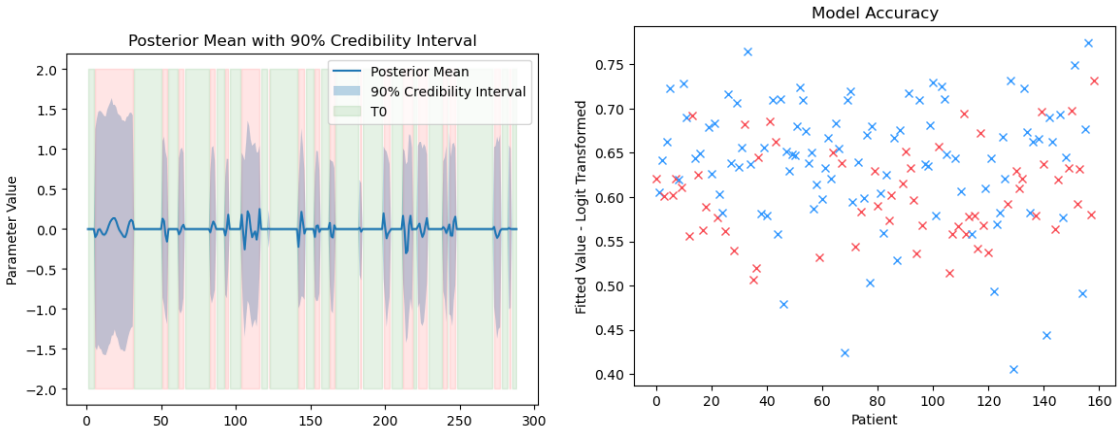

Figure 6: Constrained model results using real data: on the left it is shown the posterior mean of the estimated β parameter with the corresponding credible interval. In green it is highlighted the estimated T_0. On the right fitted values are reported for the severity target. In blue the correct classified ones, in red the misclassified ones.

It is worth noticing that the Constrained model performs 5% better than the Un-

constrained one, in a 3-fold cross-validation environment.

4.2.4. Mixed Effect Model with Random Intercept

We firstly interpret *gender* as a random intercept effect. Here, we aim to assess whether men and women have different severity scores, keeping fixed the tremor probability.
The chain, constructed through 3000 iterations after a burn-in of 3000 ones, is convergent and reliable. Our posterior parameter has a Gaussian Process shape and the outcome supports the idea of differences between the two groups. In particular, we find out that, on average, keeping fixed the tremor probability, women have around 0.07 more in the severity score. See figure 7.

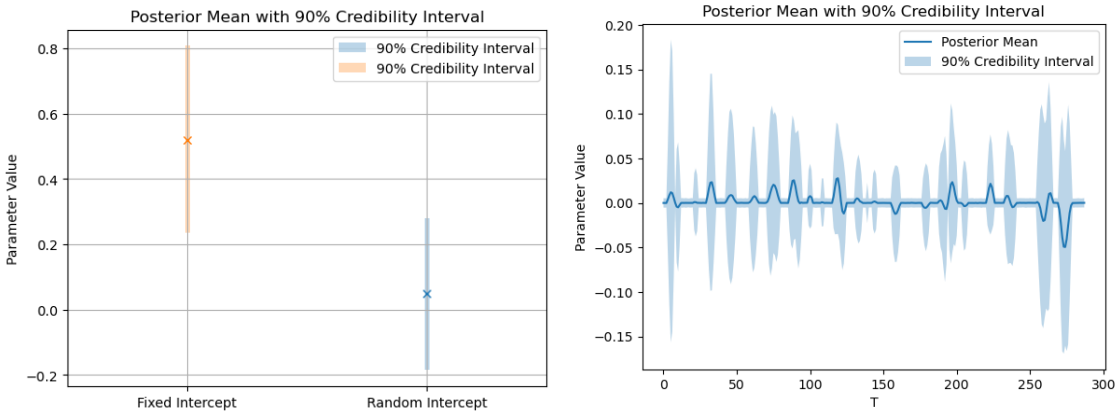

Figure 7: Random intercept model results using real data: on the left it is shown both the fixed and random intercept with the corresponding credible intervals. On the right it is presented the posterior mean of the estimated β parameter with the corresponding credible interval.

4.2.5. Mixed Effect Model with Random Slope

Levodopa is a popular treatment against Parkinson's symptoms. The fact that we have variability among the *number of doses* taken by patients encourages us to develop a hierarchical structure in this direction. Moreover, by modifying the level of tremor, the treatment can lead to misleading estimation of the functional parameter.

With this in mind, we construct a random slope extension of the means of the hierarchical structures induced by the number of Levodopa doses. This has to be considered as a proposed extension for the Unconstrained Model.

The outcome is not strongly reliable, since the huge computational costs make unfeasible to develop a convergent chain. This happens since the estimation of two functional parameters consists of the estimation of 288×2 distributions, resulting from the evaluation of 320 integrals at every step of the chain (We use 300 iterations after a burn-in of 300 ones).

However, this hierarchical extension of the model overperforms the constrained one, and this allows us to consider this direction as promising. See figure 8.

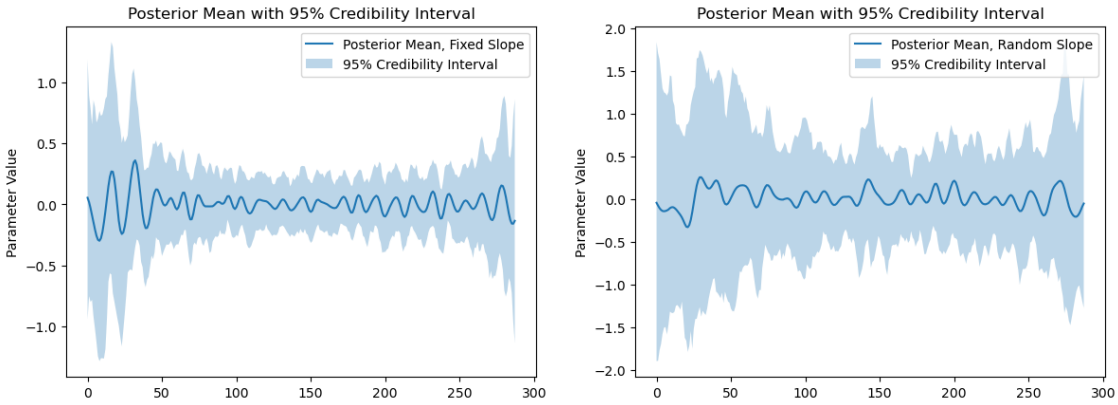

Figure 8: Random slope model results using real data: the plots show the posterior mean of β with the corresponding credible interval in two different cases. On the left the one obtained with the fixed slope model, on the right the one with random slope model.

5 | Conclusion

The following table highlights the results we obtained.

Models	Accuracy	LPML
Unconstrained	62.26%	-206.61
Constrained	70.44%	-223.37
Random Intecept	67.3%	-167.08
Random Slope	72.06%	-188.82

There are many possible extensions of the reported models. Just to name few of them:

- Multiple scalar-on-function regression;
- Incorporation more complex hierarchical structures;
- Use Metropolis-within-Gibbs to update the estimate of T_0 while estimating the Mixed Effect model;
- Further enhancement of the method for the estimation of T_0.

Overall, the results presented are satisfactory in terms of accuracy and information obtained. However, the computational costs pose many limits to our research.

A | Appendix A

A.1. From Unconstrained to Constrained

Let's first revise the concept of *Reproducing kernel Hilbert space*:

RKHS

Let X be an arbitrary set, H a Hilbert space of real-valued functions on X. H is called a **reproducing kernel Hilbert space** (RKHS) endowed with an inner product $\langle \cdot, \cdot \rangle_H$ if there exists a function $K : X \times X \to \mathbb{R}$ with the following properties:

- $\forall\, x \in X \;\; K(x, \cdot) \in H$
- *reproducity property*: $\langle f(\cdot), K(x, \cdot) \rangle = f(x) \quad \forall f \in H$

K is called **reproducing kernel** and it is symmetric and positive definite.

Denote H the RKHS with kernel K. It is known that H can be identified with an Hilbert space $\mathcal{H}$ of real-valued functions on T such that the map:

$$\begin{aligned} \Theta \colon \mathcal{H} &\to H \\ g(t) &\mapsto \Theta(g)(t) := \mathbb{E}[g\beta_t] \end{aligned}$$

is a linear isometry.
Now we define

$$H^0 := \{f \in H \colon f\,|_{T_0} = 0\}.$$

The idea is to find β_t^0 as the projection in $\mathcal{H}^0$ of $\beta_t \in \mathcal{H}$. Thanks to the isometry Θ, the following commutative diagram is well-defined:

$$\begin{array}{ccc} \mathcal{H} & \xrightarrow{\Theta} & H \\ \downarrow{\scriptstyle \mathcal{P}} & & \downarrow{\scriptstyle P} \\ \mathcal{H}^0 & \xrightarrow{\Theta} & H^0 \end{array}$$

where P is the orthogonal projection onto H^0, $\mathcal{H}^0 = \Theta^{-1}(H^0)$ is a Hilbert subspace of $\mathcal{H}$ and $\mathcal{P} = \Theta^{-1} P\, \Theta$ is the orthogonal projection onto $\mathcal{H}^0$.
Now consider $\mathcal{H}^1 := (\mathcal{H}^0)^\perp$ and $\mathcal{Q}$ the orthogonal projection onto the latter space. Then we can decompose the random process β_t:

$$\beta_t = \mathcal{P}\beta_t + \mathcal{Q}\,\beta_t = \beta_t^0 + \beta_t^1 \quad a.s.$$

with $\beta_t^0 \in \mathcal{H}^0$, $\beta_t^1 \in \mathcal{H}^1$.

Moreover, the null set on which the decomposition fails is independent of t:

Theorem A.1. *The following results hold:*

- β_t^0 *and* β_t^1 *are independent;*
- *with probability one the processes* β, β^0 *and* β^1 *are continuous;*
- $\beta_t = \beta_t^0 + \beta_t^1 \quad \forall\, t \in T$*;*
- $\beta^0 \sim GP(0, K^0)$ *with* $K^0(s,t) = PK_s(t)$ *and* $\beta^1 \sim GP(0, K^1)$ *with* $K^0(s,t) = QK_s(t)$*, where* $K_s(t)$=$K(s,t)$*;*
- $\beta_t^0 = 0 \quad \forall\, t \in T_0 \subset T$.

Moreover, as a consequence of the following proposition we know that if $T \in \mathbb{R}$ and if the trajectory of β has $m-1$ absolutely continuous derivatives and a square integrable m^{th} derivative, then the trajectory of β^0 shares the same smoothness property

Proposition A.1. *Let* $\tilde{K}$ *be a continuous positive definite kernel on* $T \times T$ *such that* $H \in \tilde{H}$ *where* $\tilde{H}$ *denotes the RKHS with kernel* $\tilde{K}$*. If the trajectory of* β *belongs to* $\tilde{H}$ *with probability one, then the trajectory of* β^0 *(and* β^1 *) belongs to* $\tilde{H}$ *with probability one as well.*

Finally, for how we define H^0 and as a consequence of the reproducity property:

$$\forall f \in H^0, \forall\, \tau \in T^0 \quad f(\tau) = 0 \iff \langle f(\tau), K_\tau(s) \rangle = 0$$

allowing to deduce:

$$H^1 = \overline{span\{K_\tau \colon \tau \in T_0\}}^H.$$

The latter equation is essential to compute projections in the afomentioned subspaces, hence to estimate β_t^0.

A.1.1. Mathematical projection

Since we showed that H^1 is the closure in H of the subspace spanned by $K_\tau \; \forall \tau \in T^0$:

1 for some real-valued vector $a = (a_1, \ldots, a_m)'$:

$$\forall\, g \in H^1 \; g(t) \approx \sum_{j=1}^{m} a_j K_{\tau_j}(t), \quad \{\tau_1, \ldots, \tau_m\} \subset T_0$$

hence $\forall\, f \in H$:

$$Pf(t) \approx f(t) - \sum_{j=1}^{m} a_j K_{\tau_j}(t) \tag{A.1}$$

2 knowing that $Pf(\tau_j) = 0 \; \forall\, \tau_j \in T_0$, we can solve the following linear system to find the values of the vector $\hat{a} = (\hat{a}_1, \ldots, \hat{a}_m)'$:

$$\begin{pmatrix} f(\tau_1) \\ f(\tau_2) \\ \vdots \\ f(\tau_m) \end{pmatrix} = \begin{pmatrix} K(\tau_1,\tau_1) & K(\tau_1,\tau_2) & \cdots & K(\tau_1,\tau_m) \\ K(\tau_2,\tau_1) & K(\tau_2,\tau_2) & \cdots & K(\tau_2,\tau_m) \\ \vdots & \vdots & \ddots & \vdots \\ K(\tau_m,\tau_1) & K(\tau_m,\tau_2) & \cdots & K(\tau_m,\tau_m) \end{pmatrix} \begin{pmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \vdots \\ \hat{a}_m \end{pmatrix}$$

Note that such a solution exists and is unique when K is positive definite;

3 finally we find the projection of f onto H^0 by evaluating (A.1)

From a computational point of view, consider the finite grids of discrete values of T_0 and T. Let's define:

- $f_t \in \mathbb{R}^g$ the column vector of values $[f_t]_i = f(t_i), \quad t = (t_1, ...t_g) \in T$;
- $f_\tau \in \mathbb{R}^m$ the column vector of values $[f_\tau]_i = f(\tau_i), \quad \tau = (\tau_1, ...\tau_m) \in T_0$;
- $K_{tt} \in \mathbb{R}^{g \times g}$ with entries $[K_{tt}]_{ij} = K(t_i, t_j), \quad t = (t_1, ...t_g) \in T$;
- $K_{\tau\tau} \in \mathbb{R}^{m \times m}$ with entries $[K_{\tau\tau}]_{ij} = K(\tau_i, \tau_j), \quad \tau = (\tau_1, ...\tau_m) \in T_0$;
- $K_{t\tau} \in \mathbb{R}^{g \times m}$ with entries $[K_{t\tau}]_{ij} = K(t_i, \tau_j)$;
- $a = (a_1, \ldots, a_m)'$.

Then, the projection of f onto H^0 can be computed as:

$$f_t^0 = f_t - K_{t\tau} K_{\tau\tau}^{-1} f_\tau$$

And the projected kernel is the $g \times g$ matrix K_{tt}^0 with entries $[K_{tt}^0]_{ij} = K^0(t_i, t_j) \, \forall \, i, j = 1, \ldots, g$ and $(t_1, ...t_g) \in T$ approximated as:

$$K_{tt}^0 \approx K_{tt} - K_{t\tau} K_{\tau\tau}^{-1} K_{t\tau}'$$

A.1.2. Numerical approximation

Let X be the $n \times g$ matrix with entries $[X]_{ij} = x_i(t_j)$, we compute:

- Integral operator L with kernel K^0

$$L^0 x(t)' \approx \delta K_{tt}^0 X'$$

- Covariance matrix with kernel K^0

$$\Sigma^0 \approx \delta^2 X K_{tt}^0 X'$$

where $\delta = t_{i+1} - t_i$ depends on the discretization step chosen.

Bibliography

[1] C. A. Davie, "A review of parkinson's disease," *British Medical Bulletin*, 2008. https://doi.org/10.1093/bmb/ldn013.

[2] B. Den Oudsten, G. Van Heck, and J. De Vries, "Quality of life and related concepts in parkinson's disease: a systematic review," *Movement Disorders*, 2007. https://doi.org/10.1002/mds.21567.

[3] C. Abraham, "An informative prior distribution on functions with application to functional regression," *Statistica Neerlandica*, 2023. https://doi.org/10.1111/stan.12322.

[4] P. Reiss, J. Goldsmith, L. H. Shang, and R. T. Ogden, "Methods for scalar on function regression," *International Statistical Review*, 2017. https://www.jstor.org/stable/44840886.

[5] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. the MIT Press, 2006. www.GaussianProcess.org/gpml.

[6] J. Shi and B. Wang, *Curve prediction and clustering with mixtures of Gaussian process functional regression models*. Springer Science, 2008. https://link.springer.com/article/10.1007/s11222-008-9055-1.