ORIGINAL PAPER

A Bayesian approach to model individual differences and to partition individuals: case studies in growth and learning curves

Maura Mezzetti[1] (iD) · **Daniele Borzelli**[2] · **Andrea d'Avella**[2,3]

Abstract

The first objective of the paper is to implement a two stage Bayesian hierarchical nonlinear model for growth and learning curves, particular cases of longitudinal data with an underlying nonlinear time dependence. The aim is to model simultaneously individual trajectories over time, each with specific and potentially different characteristics, and a time-dependent behavior shared among individuals, including eventual effect of covariates. At the first stage inter-individual differences are taken into account, while, at the second stage, we search for an average model. The second objective is to partition individuals into homogeneous groups, when inter individual parameters present high level of heterogeneity. A new multivariate partitioning approach is proposed to cluster individuals according to the posterior distributions of the parameters describing the individual time-dependent behaviour. To assess the proposed methods, we present simulated data and two applications to real data, one related to growth curve modeling in agriculture and one related to learning curves for motor skills. Furthermore a comparison with finite mixture analysis is shown.

Keywords Individual differences · Bayesian mixed model · Motor control · Growth curve · Cluster analysis

1 Introduction

Behavioral science faces a dilemma: how to model individual differences without treating every individual as entirely unique? Models that impose uniformity on the population may miss significant, theoretically meaningful differences between individuals. But modeling every individual completely independently is

---

D. Borzelli, A. d'Avella, contributed equally to this study.

Extended author information available on the last page of the article

incompatible with scientific generalization and looses all the information about what they have in common (Samuels 2012). Recent literature has highlighted the risks of modeling aggregated data without considering possible individual differences. Estes (1956) was a pioneer in using formal modeling approaches to build an understanding of the fundamental properties of human cognition and stated that the form of the individual functions or the distribution of individual parameters rarely can correctly be inferred from aggregated data, i.e. obtained by average the data at each point and look for a curve on grouped data. On the other hand, Cohen et al. (2008) demonstrated conditions in which group analysis outperforms individual analysis, for example due to the difficulties or impossibility to collect much data from an individual for each condition.

Longitudinal studies, where data consist of measures repeated over time from the same unit, e.g., an individual (Villarroel et al. 2009), emphasize changes within individuals rather than differences between individuals, which is the focus of cross-sectional studies. Growth and learning curve models are particular cases of longitudinal data. Growth curve data consist of repeated measurements of a growth process over time among a population of individuals. Flexibility of growth curve modelling allows to simultaneously analyze both within-subject effects and between-subject effects.

Similarly, a learning curve is defined as a mathematical description of subject's performance in repetitive tasks. Despite many studies on learning analyze performances averaged across participants (Newell and Rosenbloom 1993), inter-individual differences in the ability to learn new tasks have long been observed (Chapman 1919). Thus, learning curves show an inherent hierarchy related to the individual performance (Rice and Leyland 1996). The confounding nature of averaging becomes serious when the model of acquisition process is approximated by a group learning curve, rather than the curve typical of the individuals (Gallistel et al. 2004). In contrast, measures of subject-specific, or within-subject change, allow to understand how people differ from each other when changing over time. It is, thus, possible to model individual trajectories over time, and to compare these trajectories across individuals and groups (Oravecz and Muth 2018).

Models proposed for longitudinal data include multilevel models, mixed models, latent growth models (Duncan et al. 2006), and random coefficient models (Raudenbush 2001). When the function that depends on parameters of interest is a non-linear function, nonlinear mixed-effects models have been introduced. However, they are considerably more difficult and computationally intensive to fit because, due to nonlinearity, there are no close-form solutions for the temporal parameters (Davidian and Giltinan 1995, 2003). To model longitudinal data, a Bayesian approach is attractive, as it has some unique advantages while its major limitation due to its high computational complexity is being progressively overcome by the dramatic increase recently seen in computing power (Zhang et al. 2007; Oravecz and Muth 2018; Ghosh et al. 1997). The primary benefit of using Bayesian approaches lies in their ability to generate the full posterior distribution of the parameters, without additional constraints or assumptions on prior distributions, as in the frequentist method, and on variables, which are allowed to depart from the standard normal paradigm (Davidian and Giltinan 1995). Fong et al. (2010) review

different approaches to Bayesian mixed model. In Cafarelli et al. (2019) a Bayesian nonlinear hierarchical model for growth curve in marine species has been proposed. Additionally, nonparametric regression with splines has high flexibility because the data is expected to find its regression curve estimation form without being influenced by the researcher's subjectivity (Welham 2009; Becher et al. 2009; Green and Silverman 2019; Lestari et al. 2012).

Here, we aim at estimating individual learning or growth curves with nonlinear time dependence and characterizing inter-individual differences. To address the first objective, we propose a method based on a two stage Bayesian hierarchical model. In the first stage units are the repeated measurements, and the specific nonlinear curve of each individual is described. The specified curve include eventual effect of covariates. In the second stage, an overall model is estimated where the units are all the individuals. To characterize inter-individual differences in learning processes and growth curves, our second objective, we partition individuals into homogeneous groups. Previous studies have addressed the issue of individual differences by focusing on individual data (Cohen et al. 2008), clustering to identify groups of subjects (Lee and Webb 2005), using hierarchical random-effect models to allow for continuous variation between subjects (Rouder and Lu 2005; Shiffrin et al. 2008). Different methods for clustering longitudinal data have been proposed, including variants of k-means (Tarpey and Kinateder 2003; Genolini et al. 2015) and various model-based classification methods relying on mixture models (Muthén and Shedden 1999; James and Sugar 2003; Paddock and Savitsky 2013), or Bayesian Gaussian mixture model (Vimal et al. 2020). We propose to partition individuals according to the posterior distributions of the parameters describing the curve, rather than according only to the temporal pattern (Genolini et al. 2015). The method then captures both the entire behaviour through time and the uncertainties about the parametric time curve. The main innovation of our method is the computation of the overlap between posterior distributions as the measure of similarities used to partition individuals, which provides a novel approach for clustering longitudinal data.

The paper is organized as follows. In Sect. 2 we introduce a hierarchical Bayesian framework for individual longitudinal data, in the particular case of growth and learning curves. A Bayesian approach allows individual data to be modeled jointly, with an aggregate model at the second stage. On the other hand, the flexibility of a Bayesian approach can be exploited easily to include linear and nonlinear component of the independent variables, error distribution for the response variables other than Gaussian and effects of covariates. Bayesian approaches also explicitly model the variation in parameters across individuals, making them superior to approaches that model each individual separately. The natural correlation among the measurements within an individual is thus taken into account. In Sect. 3 we study the posterior distributions of the individual parameters. The logical foundation is that the task of comparing units becomes the task of comparing the distribution of parameters of individual trajectories. Following a Bayesian approach, individuals are compared according to their individual posterior distributions. Thus, individual differences or similarities are investigated through the distance among posterior distributions. Section 3.1 is dedicated to the

description of two methods to measure the overlap of multivariate distributions. Once the overlap of the posterior distributions is defined as a measure of distance (or similarities) between individuals, they can be partitioned through a hierarchical agglomerative clustering approach. Sections 4 and 5 focus on two example applications of the method to growth curve modeling in agriculture and motor learning. In Sect. 6 a sensitivity analysis is performed, model evaluation is presented and results on simulated data are shown. In Sect. 7 finite mixture modeling is compared with the proposed method. Finally, in Sect. 8 we discuss the significance of the methods introduced.

2 Bayesian hierarchical model for longitudinal data

We propose a two stage Bayesian hierarchical model for longitudinal data. According to our model, at the first stage the individual response variable Y is assumed distributed according to a distribution belonging to the exponential family, in particular Y_{ij} is the response variable of individual i at trial (or time) j.

The expected value, or a function g of the expected value (g is any of the usual link function) is a function of j, either time o trial, and, eventually, of P-dimensional vector of covariates X_{ij}:

$$g\big(E(Y_{ij})\big) = f(j, X_{ij}, \alpha_i, \beta) \tag{1}$$

Function f can be any linear or non linear function, α is the matrix of the individual parameters (α_i), each row describes the longitudinal curve of the i-subject, while β is the vector of parameters linked to the covariates effect, β^p is the parameter describing the effect of the p-th covariate, that we assume the same between subjects but it can easily be generalized to be subject specific. In the review paper of Craig et al. (2002), a list of different learning curves as a function of repetition, as trial number or time, is provided. This list includes curves with a variety of shapes: linear, quadratic, cubic, power law, exponential, double exponential, logarithmic, Gompertz, Weibul, log-linear, log-log linear and logistic. The advantage of the Bayesian approach is its flexibility, as it allows to fit all the listed curves, as function $f()$ in (1), and to naturally take into account random effects.

The examples illustrated in Sects. 4 and 5 focus on two particular cases of longitudinal data: growth curve and learning curve. In the illustrated examples, the response variable is distributed, respectively as a Gaussian or a Binomial. The non-linear function is a logistic function, such that the vector parameters (α_i) is generally linked with the asymptote, the value at the inflection point of the curve and a scale parameter, in the first case, and the starting performance level, the rate of learning and final performance level, in the second case.

Prior distributions on the individual parameters and the parameters linked with covariates effects need to be specified. We present two possible sets of prior distributions. Defining prior distributions corresponds to specifying the second stage of the hierarchical model, which relates individual performance to an average performance across all the subjects, described by vector of parameters (a).

The following prior distributions allow the component of the parameter (α_i) to be correlated.

$$\alpha_i \sim MN(a, \Sigma_\alpha) \tag{2}$$

$$\Sigma_\alpha \sim Wishart(\Omega, \tau) \tag{3}$$

The third level is defined as:

$$a \sim MN(\underline{0}, \Psi) \tag{4}$$

Moreover for each component p of the vector β we assume the following prior distributions:

$$\beta^p \sim N(0, \sigma_p^2) \tag{5}$$

$$\begin{aligned} p &= 1, \ldots, P \\ \sigma_\beta^2 &\sim InvGamma(v_{\sigma_\beta}, \lambda_{\sigma_\beta}) \end{aligned} \tag{6}$$

Component of the parameter (a) are the overall population parameters, describing the overall learning or growth curve model. The posterior distribution of the matrix Σ will indicate the strength of the correlation.

The following set of prior distributions replaces (2) and (3), does not consider the possible correlation between the parameters and simplifies the computations:

$$\alpha_i^h \sim N(a^h, \sigma_{\alpha^h}^2) \tag{7}$$

$$\begin{aligned} &for\ h = 1, 2, 3 \\ &\sigma_{\alpha^h}^2 \sim InvGamma(v_{\sigma_{\alpha^h}}, \lambda_{\sigma_{\alpha^h}}) \end{aligned} \tag{8}$$

Then, the third level is now defined as:

$$a^h \sim N(0, \tau^2) \tag{9}$$

When the distribution of Y_{ij} has a scale parameter, σ_i^2, the following distribution is considered:

$$\sigma_i^2 \sim InvGamma(v_\sigma, \lambda_\sigma) \tag{10}$$

The distribution in (10) allows modelling intra-individual variability, σ_i^2 is defined to be subject-specific. Intra-individual variability represents individuals' variability around their own mean, which is necessary to model to characterize individual behavior (Wiley et al. 2014). While $(\sigma_{\alpha^h}^2)$ in (8) or possibly Σ_α in (2) represent the between-subjects variance.

3 Posterior distributions and cluster analysis

In the Bayesian framework, prior distributions need to be updated to posterior distributions via appropriate likelihood functions. However, closed-form exact expressions for most of the relevant joint and marginal posterior distributions are not available. Instead, we rely here on sampling-based approximations to the distributions of interest via Markov chain Monte Carlo (MCMC) methods: we use a Gibbs sampler or a Metropolis-within-Gibbs algorithm to explore the posterior.

The algorithm is implemented using Just another Gibbs sampler (JAGS), using an interface of the CRAN R-packages. JAGS is a program for Bayesian graphical modelling developed by Plummer (2003). It is written in C++ and it allows for an object-oriented programming style that is extremely useful in this context. JAGS has the advantages of running on all platforms and of interfacing with R. The algorithm is based on adaptive rejection Metropolis sampling and it generalizes adaptive rejection sampling including a Hastings Metropolis algorithm step to address non log-concave full conditional distributions. After a sufficient number of burn-in iterations, the remaining samples from the MCMC simulations are used to obtain the distribution of any function of the parameters of interest. To assess the stability of the final estimates, multiple MCMC simulations are run with different initial values and starting points. The convergence of the MCMC samples of the parameters after excluding the initial burn-in samples are monitored using the R package CODA.

These models allow to asses both the within-subject effects (e.g., assessing change over time for one subject) and between-subjects effects. The next two subsections illustrates two methods to measure the distance between posterior distributions, that will be interpreted as distance between subjects. The similarities and differences between the individual parameters posterior distributions will be used to assess the existence of partitions.

3.1 Overlap multivariate distribution as a distance between subjects

According to model described in previous section each subject is characterized by a vector of parameters: (α_i). The closer are the values (α_i) to $(\alpha_{i'})$, the more similar are subject i and subject i'. Instead of comparing only the posterior point estimates of the individual temporal behaviour, through the distance between the vectors (α_i) and $(\alpha_{i'})$, we compare the whole distribution of the parameters, taking into account the relative uncertainties. Within each individual fit, the posterior distributions of each component of the parameter (α_i) might be strongly correlated, thus a tool to compare multivariate distributions is necessary.

To quantify distances between the multivariate distributions of the subject i and the subject $i\prime$, we define two measures of overlap among multivariate distributions. They are both defined in terms of the pairwise distances between all points in the two distributions.

1. The first one has been introduced by Gutierrez-Pena and Walker (2019). Let us consider two multivariate distributions f^1 and f^2, and let us assume that we

observe $\mathbf{x^1}$ and $\mathbf{x^2}$, respectively from the two distributions. We indicate with $\mathbf{x^j}$ a matrix of s_j observations and p variables. In our context, s_j is the number of samples simulated from the posterior distribution while K is the number of individual parameters describing the time curve. We consider one matrix $\mathbf{Z} = [\mathbf{x^1}; \mathbf{x^2}]$. For each row i of the matrix, for $i = 1, \ldots, s_1 + s_2$, we look for the *closest* vector among the others $s_1 + s_2 - 1$ vectors, where closeness is defined according to the Euclidean distance. Function v associates to each index i an index $i\prime$ such that z_i and $z_{i'}$ are the closest vectors among all pairs (i, k) with $k = 1, \ldots, s_1 + s_2$, i.e. $v(i) = i'$ if the vector $z_{i'}$ is the closest to z_i between all the vectors z_k. The function δ is a binary function that associates to each row the value 1 if the closest vector is a point from the same distribution, that is:

$$\delta_i = \begin{cases} 1 & v(i) \leq s_1 \quad \text{and} \quad i \leq s_1 \\ 1 & v(i) > s_1 \quad \text{and} \quad i > s_1 \\ 0 & v(i) > s_1 \quad \text{and} \quad i \leq s_1 \\ 0 & v(i) \leq s_1 \quad \text{and} \quad i > s_1 \end{cases}$$

The following indexes are defined:

$$m_1 = \sum_{i=1}^{s_1} \delta(i) \quad \text{and} \quad m_2 = \sum_{i=1+s_1}^{s_1+s_2} \delta(i).$$

However, if f^1 and f^2 overlap completely we would expect roughly $s_1/(s_1 + s_2)$ of the observations from f^1 to be matched with observations from the same distribution f^2. In this case, m_1 would be approximately $s_1^2/(s_1 + s_2)$. A similar argument can be made for the observations from f^2. On the other hand, if f^1 and f^2 do not have any point of intersection, m_1 would be equal to s_1.

As most approaches in the literature, we would like our measure of overlap to take the value 0 for no overlap and the value 1 for a complete overlap. We define index DO_1, the degree of overlap of f^1 and f^2 as:

$$DO_1 = min\left\{1, \frac{s(s_1 - m_1)}{s_1 s_2}\right\} \qquad (11)$$

where $s = s_1 + s_2$

2. As a second measure, we propose the following empirical comparison between multivariate distributions. If we compute all the pairwise distance of the observation in Z, the distance matrix D can be interpreted as a block matrix

$$D = \begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix}$$

where D_{11} is a $s_1 \times s_1$ matrix containing the pairwise distances between the observations from f^1, D_{22} is a $s_2 \times s_2$ matrix containing the pairwise distances between the observations from f^2, D_{12} is a $s_1 \times s_2$ matrix containing the cross-distances between observations from f^1 and f^2. If f^1 and f^2 do not overlap at all, we expect the distances in the matrices D_{11} and D_{22} to be smaller than those in

the matrix D_{12} (or equivalently D_{21}). Conversely, if f^1 and f^2 overlap completely, the distances in D_{11} and D_{22} would be of the same order as those in D_{12}. The degree of overlap is then defined as the empirical overlap of the two univariate distributions, following Inman and Bradley (1989): distribution g_A is the distribution of all the elements in D_{11} and D_{22}, while g_B is the distribution of the elements in D_{12}.

$$DO_2 = \int_R min(g_A(x), g_B(x))dx \qquad (12)$$

3.2 Partitioning individuals

Individuals are clustered into homogeneous groups by means of an agglomerative hierarchical approach. As a measure of similarities and dissimilarities between individuals, one of the two measures described above is used. Classical algorithms such as agglomerative hierarchical clustering or the k-means algorithm are popular but they only explore a nested subset of partitions or require specifying the number of clusters apriori (Hartigan and Wong 1979).

There is a vast literature devoted to the issue of choosing the number of clusters, as illustrated in McLachlan and Peel (2000), but there is no consensus on a reliable method to determine the *true* number of clusters in a dataset (Everitt et al. 2001). Although various indices have been suggested, none is completely satisfactory and this issue remains practically unsolved. The main objective of the search for clusters is, on one hand, homogeneity (elements in the same cluster should all look like each other) and, on the other hand, separation (elements in different clusters should look different from each other). Generally, model parsimony is preferred, using as few clusters as possible. Bayesian model selection or different information criteria can be used for this purpose, but the selection of the number of clusters is beyond the aim of the current paper. In Vimal et al. (2020) difference between clusters is evaluated through appropriate measures of distance.

Once defined a partition of the subjects, it is possible to define a hierarchical Bayesian modeling under a formulation that permits the borrowing of strength across clusters, by shrinking parameters at the second stage towards a common parameters. There has been many studies on how to define model to allow information borrowing within and across clusters, (Xu et al. 2020; Quintana and Iglesias 2003; Leon-Novelo et al. 2012). The second stage of the hierarchical model, specified in (2)-(3) or, alternatively in (7)-(8), can be redefined taken into account the partition, such as:

$$\begin{aligned} \alpha_i &\sim MN(a_m, \Sigma_\alpha) \\ if \ \ i &\in m - th \, cluster \end{aligned} \qquad (13)$$

or, equivalently

$$\alpha_i^h \sim N(a_m^h, \sigma_{\alpha^h}^2)$$
$$for \ h = 1, 2, 3 \tag{14}$$
$$if \ i \in m - th \, cluster$$

4 Application to growth of soybean plants

The example discussed in this Section illustrates an important area of application of non linear mixed effect models: growth curve data. We considered the soybean data described in Davidian and Giltinan (1995) and in Pinheiro and Bates (2000). The response variable is the average leaf weight of 6 plants of two different genotypes chosen at random from each of 8 experimental plots and measured at approximately weekly intervals, between two and eleven weeks after planting. The two genotypes of soybeans considered are: Plant Introduction ♯416937 (P), an experimental strain, and Forrest (F), a commercial variety. Moreover, the experiment was carried out over three different planting years: 1988, 1989 and 1990.
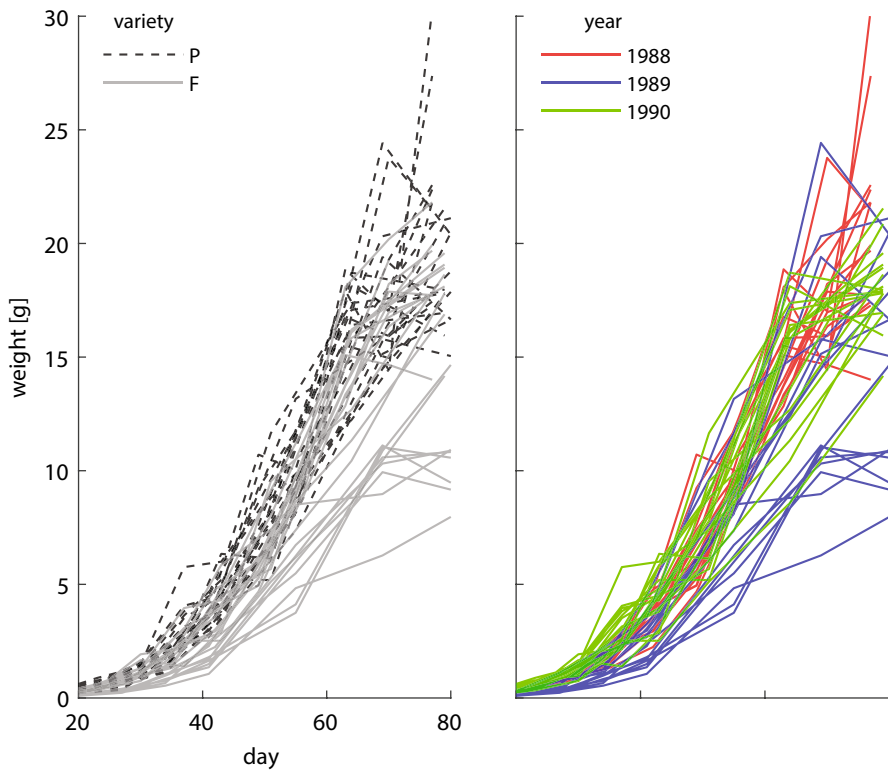
Fig. 1 Curves for soybean growth. Average leaf weight per plant of two soybean genotypes (*left*) and for three different planting years (*right*) as a function of time since planting. Within each year data were obtained on eight plots of each variety of soybean

The growth pattern of leaf weight (in grams) for two genotypes of soybeans is illustrated in Fig. 1 as the temporal trajectories of the average leaf weight for individual plots. It is apparent that the experimental strain yielded heavier leaves that the commercial variety, at least on average. There also seem to be differences between planting years, with a wider spread of the curves in 1989. Previous analyses of these data have focused on the average growth curves (Davidian and Giltinan 1995; Pinheiro and Bates 2000), we propose an individual model for each subject. We assume a Gaussian distribution for the weight of each plot and the same logistic model as that in Pinheiro and Bates (2000) to explain the relation between the expected value and the time (measured in days).

The following model is considered, with i indicating the plot and j the time:

$$Y_{ij} \sim N(\mu_{ij}, \sigma_i^2) \tag{15}$$

$$\mu_{ij} = \frac{\alpha_i^1}{1 + exp((\alpha_i^2 - j)/\alpha_i^3)} \tag{16}$$

In the following equation, possible effect of covariates is added:

$$Y_{ij} \sim N(\mu_{ij}, \sigma_i^2)$$
$$\mu_{ij} = \frac{\alpha_i^1 + \sum_p \beta_1^p X_{ij}^p}{1 + exp\left(\frac{\alpha_i^2 + \sum_p \beta_2^p X_{ij}^p - j}{\alpha_i^3 + \sum_p \beta_3^p X_{ij}^p}\right)} \tag{17}$$

The covariates used are *Variety* and *Year* that does not depend on time, so we remove the index j in X_{ij}^h .

X_i, as a 5 dimensional variable is defined as dummy variables from variable Z_i defines as:

$$Z_i = \begin{cases} 1 & if \quad Variety = P \quad and \quad Year = 1988 \\ 2 & if \quad Variety = P \quad and \quad Year = 1989 \\ 3 & if \quad Variety = P \quad and \quad Year = 1990 \\ 4 & if \quad Variety = F \quad and \quad Year = 1988 \\ 5 & if \quad Variety = F \quad and \quad Year = 1989 \\ 6 & if \quad Variety = F \quad and \quad Year = 1990 \end{cases}$$

We explore two different models: expected value illustrated in (16) and expected value illustrated in (17) and different set of prior distributions.

When a model without covariates is fitted, Figure 2 shows a comparison of the fit of the individual model assuming prior distributions as in (7)–(10) and the fit of an average model, where the hypothesis is that all the plants have the same growth curve. In the Figure it is also shown the fit of the overall model from the second stage of the hierarchical model (green line).

The posterior distribution of α_i^h are compared between subjects through the percentage of overlap. Parameters α_i^1 show the lowest percentage of overlap
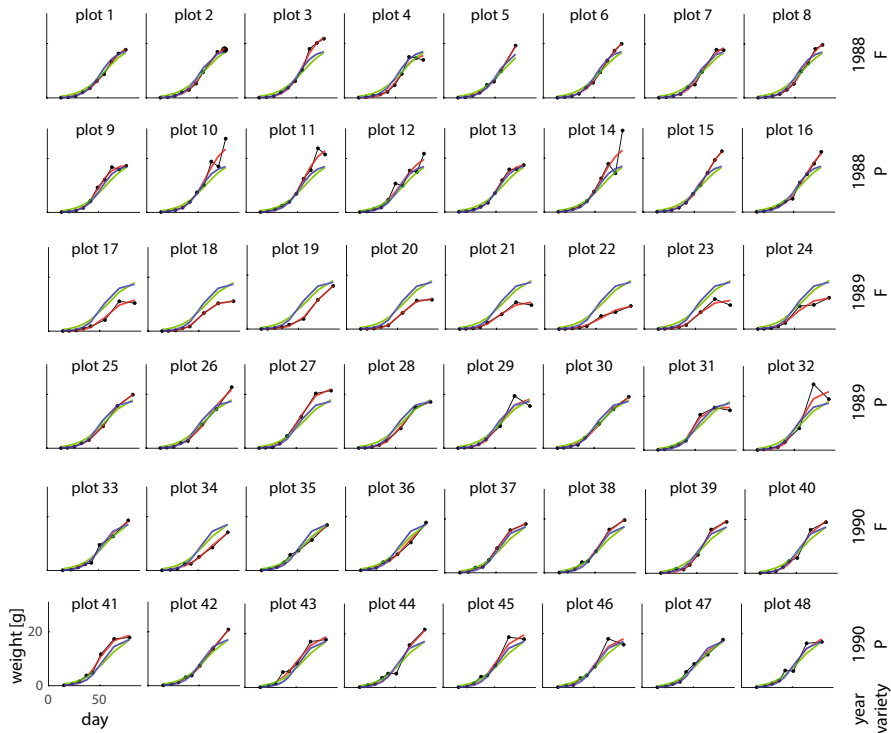
Fig. 2 Soybean growth fitted curves. Observed soybean growth data (*black*) and curves fitted with the individual model (*red*), the average model (*blue*) and the average model (*green*) from the second stage of the hierarchical model

between individual specific distribution, (416 comparisons lower than 0.05, while 301 comparisons lower than 0.01), while parameters α_i^3 show the highest percentage of overlap between individual specific distribution (3 comparisons lower than 0.05, while 750 comparisons greater than than 0.5). Implement the cluster analysis of Sect. 3.2, using distance DO_1 defined in (11), we obtain the results shown in Fig. 3. The first four columns represent the partitioning according to 3 up to 6 groups. The last two columns represent year and genotype. In each column, the color identifies the cluster, year, or genotype. Each row represents a plant. The clusters do not match specific years or genotypes, but it is possible to observe similarities. For instance, plants from 16 to 23, corresponding to plants of genotype *F* planted in 1989, belong to same cluster.

When model in (17) is considered and 5-dummy variables are added, only β_1^4, corresponding to $Z_i = 5$, has a credible interval that does not contain 0, confirming partially the results obtained with partitioning subjects. We apply the procedure shown in Sect. 3, and the partition found does not provide further information.

Once subjects are partitioned according to partition in second column in Fig. 3, a hierarchical bayesian model defined as in (15) and (16) with second stage as in (14) is considered. In Fig. 4 the four fitted curves from the second stage of the
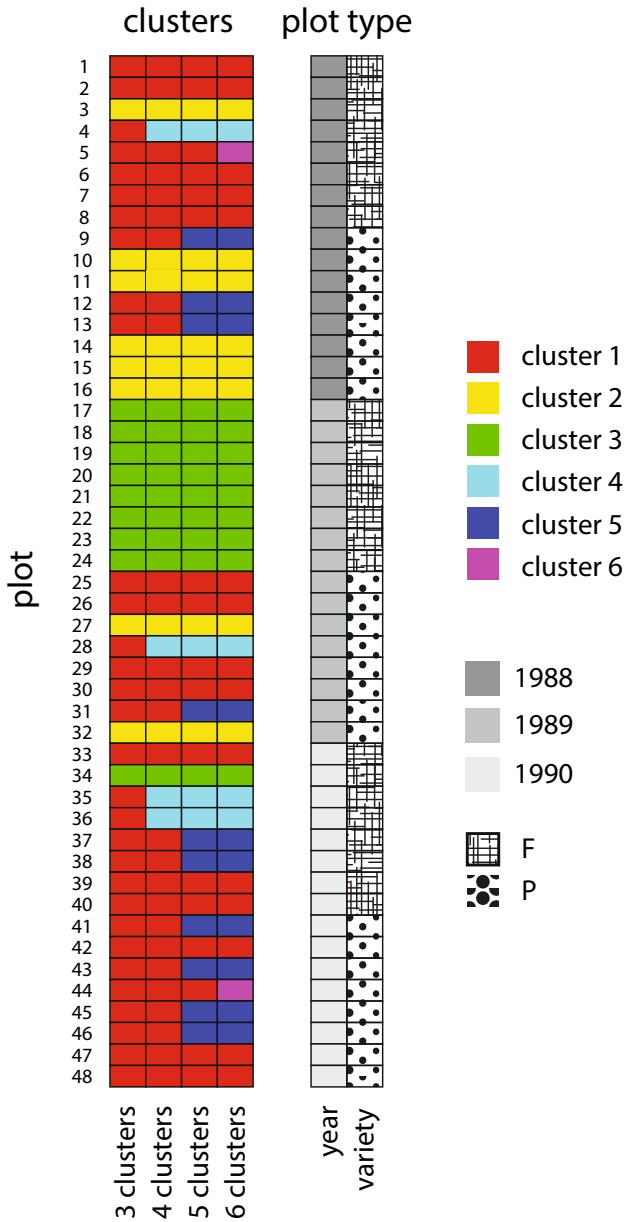
Fig. 3 Clustering of soybean data. The first four columns show results of cluster analysis when searching from 3 to 6 clusters. Column 5 indicates planting year and column 6 indicates genotype. Different categories are represented by different colors
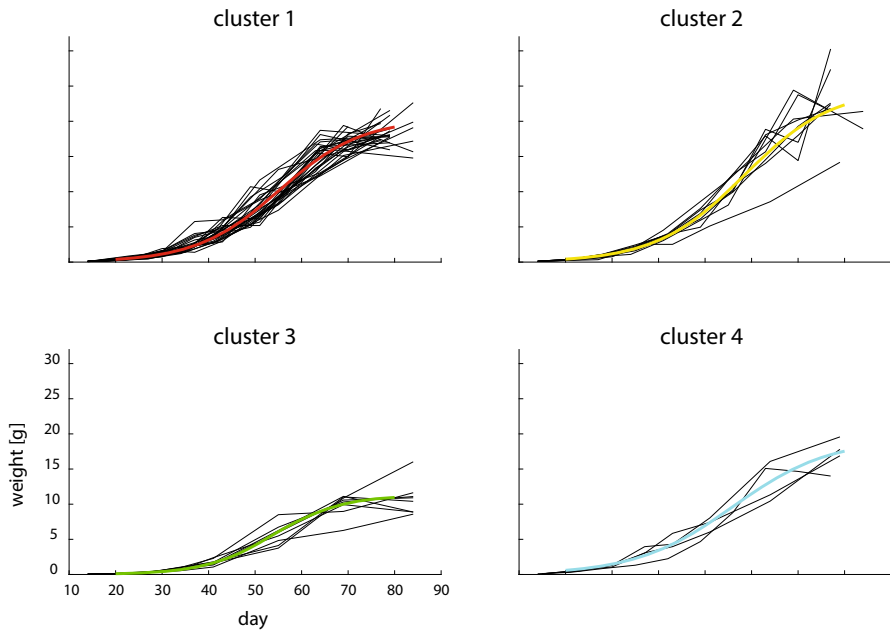
Fig. 4 Soybean growth fitted curves. Average model from the second stage of the hierarchical model: subjects partitioned in 4 clusters

hierarchical model are illustrated and it shows how the clustered structure will help 'borrow strength'

5 Application to motor learning after extended practice with virtual surgeries

The proposed algorithm is applied to experimental data describing the performance of subjects learning a novel motor task, i.e., generating novel muscle activation patterns to displace a cursor in a virtual environment to compensate for a perturbation in the forces generated at the hand by arm muscles simulated using myoelectric control ('virtual surgery', Berger et al. (2013)). Before the perturbation the observed muscle activation patterns (i.e. time-series of vectors representing the level of activation of a set of muscles) required to move the cursor in different directions could be described as linear combinations of specific sets of time-invariant muscle activation vectors, or spatial 'muscle synergies'. The aim of the experiment was to compare how participants adapted to two types of perturbations across multiple sessions. Seventeen right-handed subjects participated in the experiments after giving written informed consent. All procedures were conducted in conformance with the Declaration of Helsinki.

The task consisted in quickly and precisely moving a cursor over a virtual desktop to reach one of eight targets equally spaced on a circle in the horizontal plane. The cursor displacement from a central rest position was proportional to the

sum of the muscle pulling forces estimated, in real-time, from the activity of 15 shoulder and elbow muscles of the right arm, collected with bipolar wireless surface electromyographic sensors. The distance of the targets from the rest position corresponded to a force magnitude equal to 20% of the maximum voluntary force. Participants performed 3 sessions in different days, each composed of three phases: baseline, perturbation, and washout. During the perturbation phase the muscle pulling could be altered according to one of two types of virtual surgeries. One type was incompatible with the muscle synergies identified in each subjects in the baseline phase and required learning of new synergies to perform the task. The second type was compatible with the muscle synergies and only required new synergy combinations. In each session, during the perturbation phase each participant performed 288 reaching trials, subdivided in 36 sets of trials cycling through the 8 targets, in random order. The seventeen participants were randomly assigned to one of two groups who were exposed to either an incompatible virtual surgery (participants 1–9) or a compatible virtual surgery (participants 10–17).

To assess and compare the ability of the participants to overcome compatible and incompatible surgeries and, thus, to improve their performance, the fraction of trials in which the cursor reached the target during each cycle in the perturbation phase was considered. Let Y_{ijkh} be the binary response variable of subject i at cycle (time) j, h is the target index for each cycle, and k the experimental session in one of three different days ($k = 1, 2, 3$). We assume that the relation between success/failure probability and the time is non linear, as the following:

$$Y_{ijkh} \sim Bernoulli(p_{ijk})$$
$$logit(p_{ijk}) = \alpha_{ik}^1 + \alpha_{ik}^2 \frac{1}{1 + exp(-\alpha_{ik}^3(j - t_k^0))} \tag{18}$$

Equation (18) is an extension of Equation (1) since it depends on an extra index, k, representing the day (experimental session). Parameters α_{ik}^3 are constrained to be positive. According to this model each subject is characterized by nine parameters, as $(\alpha_{ik}^1, \alpha_{ik}^2, \alpha_{ik}^3)$ depend also on k, the day of the experimental session, with $k = 1, 2, 3$. The area of the overlapping region between posterior distributions of each pair of subjects is empirically computed. The parameters (α_{ik}^3) show high degree of overlap between individual distributions. In contrast, parameters (α_{i1}^2), linked to the final performance at the end of each session, show the highest inter-individual differences.

We run a second model in which t_k^0, that represents the inflection point of the learning curve at day k, is also a random variable, with the following prior distributions:

$$t_{i1}^0 \sim Uniform(1, 36)$$
$$t_{i2}^0 \sim Uniform(37, 72)$$
$$t_{i3}^0 \sim Uniform(73, 108)$$

Assuming that the inflection point may vary across subjects, i.e. t_{ik}^0, Equation (18) can be rewritten as:

$$logit(p_{ijk}) = \alpha_{ik}^1 + \alpha_{ik}^2 \frac{1}{1 + exp(-\alpha_{ik}^3(j - t_{ik}^0))} \tag{19}$$

Moreover, parameters (α_{ik}^3) encounter problems in reaching convergence, we run the following model, letting α_k^3 be constant for each subject:

$$logit(p_{ijk}) = \alpha_{ik}^1 + \alpha_{ik}^2 \frac{1}{1 + exp(-\alpha_k^3(j - t_{ik}^0))} \tag{20}$$

Again, as in the previous example, we include eventual effect of covariates, assuming trial and day independent, and modify the equation in (19) as follows:

$$logit(p_{ijk}) = (\alpha_{ik}^1 + \sum_p \beta_{1k}^p X_i^p) + \frac{(\alpha_{ik}^2 + \sum_p \beta_{2k}^p X_i^p)}{1 + exp(-(\alpha_{ik}^3 + \sum_p \beta_{3k}^p X_i^p)(j - t_{ik}^0))} \tag{21}$$

We introduce one covariate, X, binary variable indicating compatible and incompatible, model in (21) is rewritten as:

$$logit(p_{ijk}) = (\alpha_{ik}^1 + \beta_{1k} X_i) + \frac{(\alpha_{ik}^2 + \beta_{2k} X_i)}{1 + exp(-(\alpha_{ik}^3 + \beta_{3k} X_i)(j - t_{ik}^0))} \tag{22}$$

Two different sets of prior distributions can be assumed, as shown in Sect. 2:

$$\alpha_{ik}^h \sim N(a_k^h, \sigma_{\alpha^h}^2) \tag{23}$$

$$\sigma_{\alpha^h}^2 \sim InvGamma(v_{\sigma_{\alpha^h}}, \lambda_{\sigma_{\alpha^h}})$$
$$for\ i = 1, \ldots, n \qquad h = 1, 2, 3 \tag{24}$$

$$a_k^h \sim N(0, \tau^2) \tag{25}$$

$$\beta_{hk} \sim N(0, \sigma_{\beta_{hk}}^2) \tag{26}$$

$$\sigma_{\beta_{hk}}^2 \sim InvGamma(v_{\sigma_\beta}, \lambda_{\sigma_\beta})$$
$$for\ h = 1, 2, 3 \qquad k = 1, 2, 3 \tag{27}$$

In alternative, the following prior can be assumed:

$$(\alpha_{ik}) \sim MN(a_k, \Sigma) \tag{28}$$

$$\Sigma \sim Wishart(\Omega, \tau) \tag{29}$$

$$a_k \sim MN(\underline{0}, \Psi) \tag{30}$$

As an alternative, we used a spline approach to explain the motor learning and the performance at each cycle. We follow a Bayesian implementation as presented in Stenglein and Van Deelen (2016) and Crainiceanu et al. (2005), where it is provided background, implementation, and R (Yan et al. 2000) and WinBUGS (Spiegelhalter et al. 2003) code for Bayesian analysis of penalized spline regression. We modified their code for our model, and refer to Ruppert et al. (2003) and Crainiceanu et al. (2005) for additional details. The general methodology of semiparametric modeling using the equivalence between penalized splines and mixed models is presented in Ruppert et al. (2003), we focus on low-rank thin-plate splines which tend to have very good numerical properties. The model is as follows:

$$Y_{ijkh} \sim Bernoulli(p_{ijk})$$
$$logit(p_{ijk}) = \beta_{ik}^0 + \sum_{h=1}^{H} \beta_{ik}^h \mid j - \kappa_h \mid^3 \tag{31}$$

where (β_{ik}^h) is the vector of regression coefficients, and $\kappa_1 < \kappa_2 < \ldots < \kappa_H$ are fixed knots. We consider $H = 10$, and κ_h is the sample quantile of x's corresponding to probability $h/(H + 1)$.

In Fig. 5 the observed success probability values (black lines and markers) of each subject is compared with the individual logistic non linear model (red lines), with prior distributions in (23)–(27), the the overall logistic non linear model (blue lines) and the fitted curve according to splines interpolation as in (31).

Subjects are then partitioned according to the method illustrated in Sect. 3.1. Partitions are obtained using either distance DO_1, in (11) or DO_2, in (12). The partitions are slightly dependent on the distance or the prior distributions assumed.
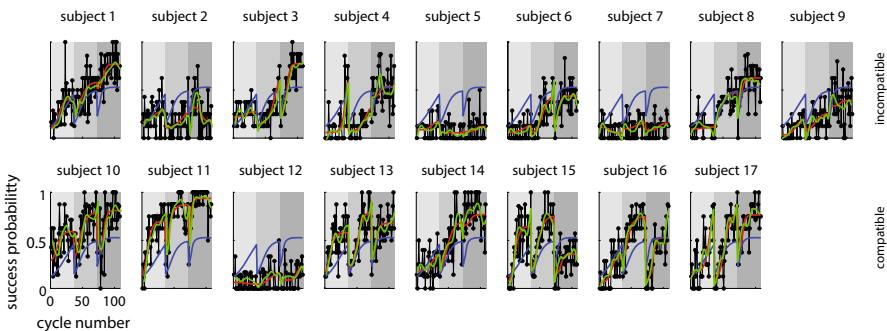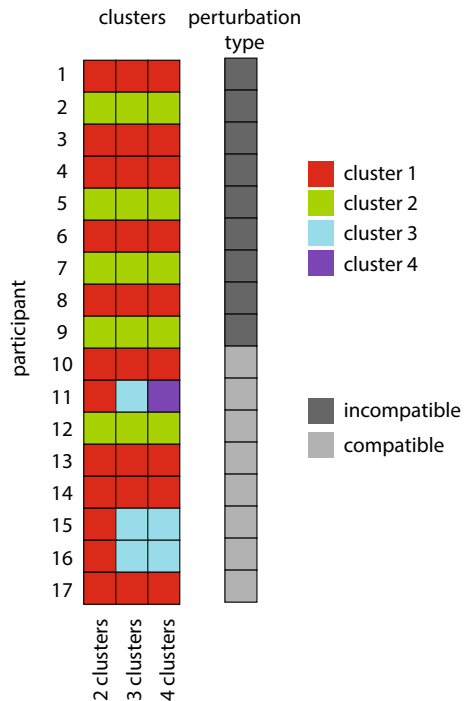

Fig. 5 Curves for learning a novel motor skill. Observed probability of success over time (cycles, *black*), individual logistic fitted model (*red*), individual splines fitted model (*green*) and overall logistic model (*blue*). Each plot presents a different subject. Success probability is computed over cycles of 8 trials performed in 3 sessions (cycles $1 - 36$, $37 - 72$, $73 - 108$, different gray background) practicing a reaching task after a virtual surgery

Using the distance defined in (11) we obtain a partition of the participants into two groups. Subjects (2, 5, 7, 9, 12) belong to one group, and they are characterized by a low final performance (mean performance in the last three cycles of the last session less than 0.33). Subject 15 also has a low final performance, but it is assigned to the second group. However, as our method takes into account the whole learning curve and the uncertainty of the estimated parameters, subject 15 has other distinguishing features. For example, in Fig. 5 it can be appreciated that subject 15, differently for the subjects in the first group, has a higher and more variable performance on the first two days.

Increasing the number of clusters, features characterizing further partitions of subjects into groups can be highlighted, as shown in Fig. 6. Subject 15 is matched with subject 16, who has a better final performance. The two subjects are characterized by the highest initial slope on the second day. With three partitions, the third cluster also contains subject 11, who has unique learning curve parameters, the best overall performance and, again, good initial performance the second day. Indeed with four partition, the addition cluster contains only the subject 11. In Fig. 7 the fitted curves from the second stage of the hierarchical model after partitioning the subjects are shown. The model is defined in (20), and the prior distributions defined as in (14). Moreover, a partition of subjects is searched from distance of posterior distributions of parameters of the the splines fit. The concordance between the partition obtained from distribution of parameters β describing the splines and the partition presented in the paper is above 73%.

Fig. 6 Clustering of learning motor skill. The first three columns show results of cluster analysis when searching from 2 to 4 clusters. Column 4 indicates compatible and incompatible subjects. Different categories are represented by different colors
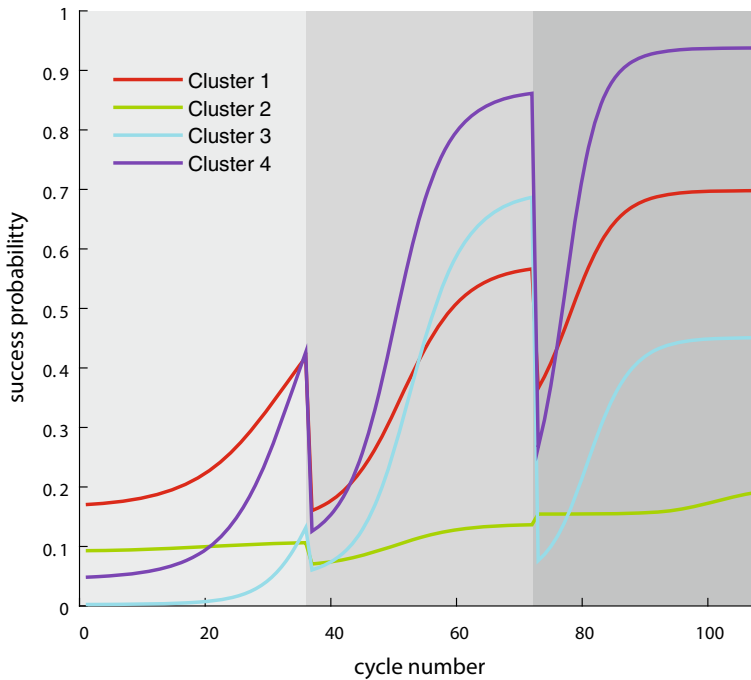
Fig. 7 Learning a novel motor skill fitted curves. Average model from the second stage of the hierarchical model: subjects partitioned in 4 clusters

6 Sensitivity analysis

Updating the prior distribution in (10) to the posterior distribution, through the likelihood as in the model presented in (15) and (16), allows the posterior estimation of σ_i^2. Posterior distributions of σ_i^2 provide hints on differences in individual variability. Different prior distributions for the variance can be defined as in Demirhan and Kalaylioglu (2015). However, alternative prior distributions, such as a joint prior distribution for the variance or a multivariate normal distribution on the logarithm of the variance, do not improve the model.

In the second example (Sect. 5), we encounter some problems in the convergence of the posterior estimates when lower hyperparameters for the prior distribution of the precision parameters in (7)–(8) are assumed. We then choose hyperparameters $(v_\star, \lambda_\star)$ equal to (10, 0.1) and τ equal to 0.01. On the other hand, in the first example (Sect. 4), where fewer parameters are estimated and sample size is larger, posterior estimates result less sensitive to the choice of $(v_\star, \lambda_\star)$ and we let them to be equal to (0.01, 0.01). In the second example, furthermore, parameters α_{ik}^3 are very sensitive to the hyperparameters chosen, and reducing to just three parameters α_k^3 improves either the fit and the diagnostic on the chains.

Bayesian models can be evaluated and compared in several ways (Gelfand and Dey 1994; Wasserman 2000; Gelman et al. 2014). Tables 1 and 2 show

Table 1 Soybean Growth: Model Comparison

Model	Log Likel	# of par	$\sum_i \epsilon^2$	WAIC	LOOIC
Individual Logistic priors as in (7)–(8)	− 488.72	$3 \times (n+1)$	387.8	36650.4 (204.4)	31481 (66.6)
Individual Logistic priors as in (2)–(3)	− 492.28	$3 \times (n+1)$	393.5	37135.9 (272.1)	31584.3 (56.5)
Individual Logistic Partitioned in 4 clusters	− 488.02	$3 \times (n+4)$	390.1	36846.8 (215.6)	31371.8 (51.6)
Overall Logistic	− 805.39	3	1618.2	49096.8 (26.4)	49118.2 (33.6)
Individual Logistic with covariates	− 517.14	$3 \times (n+1) + 10$	450.1	41610.0 (465.8)	33392.4 (91.2)

Table 2 Motor Learning Example: Model Comparison

Model	Log Likel	# of par	$\sum_i \epsilon^2$	WAIC	LOOIC
Logistic Overall	− 8534.596	12	122.93	512321.4 (26.6)	512356 (24))
Prior as in (23)–(27) $\tau = 0.001$	− 7032.077	$12 \times n$	31.68	426594.4 (79.2)	423563.8 (30.6)
Prior as in (23)–(27) $\alpha_{ik}^3 = \alpha_k^3$ and $\tau = 0.01$	− 7032.728	$9 \times n + 3$	33.46	425327.1 (82)	423374.4 (45.2)
Priors as in (28)–(30)	− 7034.666	$12 \times n$	33.07	425710.4 (165.8)	423521.4 (75.6)
Logistic with covatiates as in (22)	− 6797.361	$12 \times n + 9$	31.78	413905.4 (158.6)	409756.2 (67.2)
Splines (10 knots)	− 7019.182	$30 \times n$	28.75	429697.8 (139.8)	423426.2 (38.2)

comparisons between different models. The second column of the Tables reports the log likelihood. Although it plays an important role in statistical model comparison, it also has some drawbacks, among them its dependence on the number of parameters and on the sample size. To overcome the limitations of the log likelihood, a reasonable alternative is to evaluate a model through the log predictive density and its accuracy. Log pointwise predictive density (*lppd*) for a single value y_i is defined as (Vehtari et al. 2017);

$$lppd = \sum_{i=1}^{n} log p(y_i \mid y) = \sum_{i=1}^{n} log \int p(y_i \mid \theta) p(\theta \mid y) d\theta$$

To compute *lppd* in practice, we can evaluate the expectation using draws from the posterior simulation.

Watanabe-Akaike Information Criteria (WAIC) (Watanabe and Opper 2010), is shown in fifth column of the Tables. WAIC, as already introduced with deviance

information criterion (DIC) (Burnham 1998), estimates the effective number of parameters to adjust for overfitting. Compared to Akaike Informarion Criteria (AIC) (Burnham 1998) and DIC, WAIC has the desirable property of averaging over the posterior distribution rather than conditioning on a point estimate and is defined as

$$WAIC = lppd - p_{WAIC}$$

and

$$p_{WAIC} = \sum_{i=1}^{n} Var_{\theta|y}(log(y_i \mid \theta))$$

As with AIC and DIC, we define WAIC as -2 times the previous expression. WAIC is particularly helpful for models with hierarchical and mixture structures in which the number of parameters increases with sample size.

In Bayesian cross-validation, the data are repeatedly partitioned into a training set y_{train} and a holdout set $y_{holdout}$, and then the model is fit to y_{train} with the fit evaluated using an estimate of the log predictive density of the holdout data. The Bayesian leave-one-out cross-validation information criterion (LOOIC) is defined as:

$$LOOIC = -2 \times \sum_{i=1}^{n} log(p(y_i \mid y_{-i})) = -2 \times \sum_{i=1}^{n} log \int p(y_i \mid \theta)\pi(\theta \mid y_i)d\theta$$

where y_{-i} represents the data without the ith data point. Approximate LOOIC can be computed easily using importance sampling as described in Vehtari et al. (2017). Each prediction is conditioned on $n-1$ data points, which causes under-estimation of the predictive fit. For large n the difference is negligible. In Column 6 it is illustrated the point estimates and standard errors of *LOOIC*. Fourth column of the Tables shows sum of squared errors for each fit.

Table 1 shows the comparisons between the models presented in Sect. 4, in particular the hierarchical bayesian individual logistic model defined as in (15) and (16) with second stage as in (14), overall model and individual logistic model with covariates as in (17). Models are all very similar in terms of model evaluation, a part for the overall model. Prior distribution as in (7)–(8) perform slightly better than prior distribution as in (2)–(3). The hierarchical bayesian model with subjects partitioned in four groups results a good model to explain the data. The fit from the second stage model as shown in Fig. 4 facilitates a parsimonious modeling.

Table 2 shows the comparisons between the models presented in Sect. 5, in particular, the overall logistic model, the individual logistic model with different priors, the individual logistic model with covariate effect included as in (22), and the splines model with 10 knots. The models show similar results in terms of measures of evaluation. Increasing the number of parameters, makes it harder to reach convergence in the simulation chains. The second line presents a good performance in terms of model evaluation but the chains show troubles in reaching convergence. The third line in Table 2 shows the individual logistic model assuming prior as in (23)–(27) and $\alpha_{ik}^3 = \alpha_k^3$ and $\tau = 0.01$. Individual curves from

this model are shown in Fig. 5, since the curves are considered the best one to explain the data.

To test the performance of the proposed partition procedure, we simulated 300 data sets. The Bayesian hierarchical model is run and the resulting clustering performance is assessed. Each dataset consists of 20 subjects belonging to two different groups. For each time t_j data are sampled from two Gaussian distribution, with a four parameter logistic function as the expected value. The response variable Y_{ijk}^g, observed at repetition k and time t_j, for subject i belonging to group g is distributed according the following parametric curve:

$$Y_{ijk}^g \sim N(\mu_{ij}^g, \sigma_i^2)$$
$$logit\left(\mu_{ij}^g\right) = \alpha_i^g + \beta_i^g \frac{1}{1 + exp(-\gamma_i^g(t_j - t_i^{0g}))} \tag{32}$$

the parameters are sampled from the following distributions:

$$\alpha_i^g \sim N(-2.1972, 0.05) \quad g = 1, 2$$
$$\beta_i^1 \sim N(3.044, 0.05)$$
$$\beta_i^2 \sim N(4.3944, 0.05)$$
$$\gamma_i^1 \sim N(\delta_1, 0.01)$$
$$\gamma_i^2 \sim N(\delta_2, 0.01)$$
$$\delta_g \sim N(0.25, 0.2)$$
$$t_i^{01} \sim N(24, 5)$$
$$t_i^{02} \sim N(30, 5)$$

The parameters are chosen such that the starting point of both group is on average 0.1, while the final point for the first group is on average 0.7 and for the second group is on average 0.9. For each simulated dataset, we assessed how well our procedure can recover the groups from the posterior estimates. The concordance between the generating partition and the partition obtained with our procedure is satisfactory and, as expected, concordance decreases as the within subject variance increases. By assuming $\sigma_i \sim U(0.05, 2)$ average concordance is 0.87 and it decreases to 0.71 when $\sigma_i \sim U(2, 5)$.

7 Comparison with finite mixture model

Finite mixture models represent an adaptive class of statistical models that gained strong interest in recent years. Finite mixture models and latent class analysis have been extensively used for providing a convenient framework for clustering and classification (McLachlan and Peel 2000; Melnykov and Maitra 2010).

We implement the estimation of finite mixture model for the dataset illustrated in Sect. 4, and compare the classification obtained with the classification shown in Sect. 4.

The idea is to assume that K latent groups exist, The distribution of each Y is given by;

$$h(y \mid t, \alpha, \sigma^2) = \sum_{k=1}^{K} \pi_k f_k(y \mid t, \alpha_k, \sigma_k^2)$$

$$\pi_k \geq 0 \quad \sum_{k=1}^{K} \pi_k = 1$$

Each subject is a member of one of the K classes with the following posterior probability:

$$P(k \mid y) = \frac{\pi_k f_k(y \mid t, \alpha_k, \sigma_k^2)}{\sum_{k=1}^{K} \pi_k f_k(y \mid t, \alpha_k, \sigma_k^2)} \tag{33}$$

We implement the estimation using the R package flexmixNL (Omerovic 2019). flexmixNL was developed as an extension of package flexmix (Leisch 2004). Mixtures of generalized nonlinear models are fitted for a given number of components and component- specific starting values for the regression coefficients. The component-specific parameters are assumed to vary over all components. The EM algorithm might either fail to converge or only converge to a local optimum depending on the initial values selected. Thus the fitting procedure is applied several times with different starting values and the solution with the best log-likelihood value is retained.

Due to the possible multimodality of the log-likelihood function, the convergence of the EM algorithm is sensitive to well chosen initial values. Furthermore the fitting requires a properly chosen number of components K. Prior knowledge on possible groups within the population may indicate a suitable number of components. If the number of latent groups present is unknown, the final choice of K can be derived in a data-driven way by means of model selection criteria based on the penalisation of the log-likelihood function (such as the AIC or the corrected AIC for small data sets, the BIC and the WAIC).

The posterior estimates of the probability in (33) has a satisfactory concordance with the clusters shown in Sect. 4, indicating that mixture model and our procedure provide similar clustering information. Three groups classification provide a 96% of concordance, four groups classification provide a 75% concordance and, finally, five groups classification has 83% concordance.

Alternatively the Bayesian hierarchical semiparametric proposal for longitudinal data in Paddock and Savitsky (2013) can be implemented. A Dirichlet prior distributions can be assumed.

8 Conclusion

Studies involving multiple individuals are usually performed with the implicit assumption that individual data are generated by the same unique model, and that discrepancies among individuals are due to irrelevant individual characteristics.

Therefore, these studies attempt to reduce the variability by pooling together or averaging data collected from different individuals. However, this approach is not always optimal, because variability may be a consequence of major differences among individuals, which would be hidden by pooling or averaging data from different individuals.

To characterize inter-individual differences in learning and growth curves, specific cases of longitudinal data, where intra-individual correlations may be present and may be related to individual-specific features, we used a two stage Bayesian nonlinear model. Mixed-effects modeling has a long tradition in statistical applications (Pinheiro and Bates 1995, 2000), and nonlinear mixed-effects models provide a powerful and useful tool for analyzing repeated-measures data in different fields of research. Bayesian nonlinear mixed-effects models have also been successfully presented in literature (Lachos et al. 2013). While inter-individual variability is commonly investigated as a continuous function of some parameters, other applications may require a discrete subdivision of individuals into separate clusters. In this study we exploit the two stage Bayesian nonlinear model to introduce a novel approach for clustering different individuals, taking into account a large number of parameters that influence the output. The separation of individuals in groups allows the investigation of the differences among clusters or the specific characteristics of a single cluster. Moreover, partitioning individuals may open novel research approaches related to the investigation of the differences across individuals of different clusters.

Clustering of longitudinal data have been presented (Genolini et al. 2015) and well motivated (de Cassia et al. 2018). The innovation here is the introduction of a new metric, such that all the uncertainties from the fitted nonlinear mixed model can be taken into account. The posterior distributions of the parameters defining the temporal pattern are in fact used as a tool to partition individuals. One advantage of the proposed method is that response variables do not have to be measured at the same time points, nor at the same number of time points, as already present in Paddock and Savitsky (2013). This can be an advantage compared with most of the existing multivariate growth curve models that assume identically time-structured or balanced response variables (Hwang and Takane 2004).

We presented application to two real data set, one related to learning curves for motor skills and one related to growth curve modeling in agriculture. However, our model may be generalized to many different applications and different types of nonlinear parametric curves, with relatively limited effort from a computational point of view, since the code is written in JAGS. Being able to identify, classify, and predict individual differences, through a parametric curve, has recognised advantages (Vimal et al. 2020). Moreover, being able to partition different individuals according to their time-dependent behavior may be useful in many fields. For example, in the biomedical field, assigning a patient to a specific cluster might lead to the identification, and the consequent administration, of the most efficient personalized treatment.

References

Becher H, Kauermann G, Khomski P et al (2009) Using penalized splines to model age-and season-of-birth-dependent effects of childhood mortality risk factors in rural burkina faso. Biometrical J 51(1):110–122

Berger DJ, Gentner R, Edmunds T, et al (2013) Differences in adaptation rates after virtual surgeries provide direct evidence for modularity. Journal of Neuroscience 33(30):12,384–12,394. https://doi.org/10.1523/JNEUROSCI.0122-13.2013, https://www.jneurosci.org/content/33/30/12384, https://arxiv.org/abs/https://www.jneurosci.org/content/33/30/12384.full.pdf

Burnham KP (1998) Model selection and multimodel inference. A practical information-theoretic approach

Cafarelli B, Calculli C, Cocchi D (2019) Bayesian hierarchical nonlinear models for estimating coral growth parameters. Environmetrics 30(5):e2559. https://doi.org/10.1002/env.2559, https://onlinelibrary.wiley.com/doi/abs/10.1002/env.2559, e2559 env.2559, https://arxiv.org/abs/onlinelibrary.wiley.com/doi/pdf/10.1002/env.2559

Chapman CJ (1919) The learning curve in type writing. J Appl Psychol 3(3):252–268

Cohen AL, Sanborn AN, Shiffrin RM (2008) Model evaluation using grouped or individual data. Psychon Bull Rev 15(4):692–712. https://doi.org/10.3758/PBR.15.4.692

Craig RR, Wallace S, Garthwaite PH et al (2002) Assessing the learning curve effect in health technologies: lessons from the non-clinical literature. Int J Technol Assess Health Care 18(1):1–10

Crainiceanu CM, Ruppert D, Wand MP (2005) Bayesian analysis for penalized spline regression using winbugs. J Stat Softw 14(14):1–24. https://doi.org/10.18637/jss.v014.i14, https://www.jstatsoft.org/index.php/jss/article/view/v014i14

Davidian M, Giltinan D (1995) Nonlinear Models for Repeated Measurement Data. Chapman and Hall

Davidian M, Giltinan DM (2003) Nonlinear models for repeated measurement data: An overview and update. J Agric Biol Environ Stat 8(4):387–419. https://doi.org/10.1198/1085711032697

de Cassia Oliveira Barboza R, de Lima Silva F, Hongyu K (2018) Cluster analysis of the estimates from growth curves. Biodiversidade 17:39–47

Demirhan H, Kalaylioglu Z (2015) Joint prior distributions for variance parameters in Bayesian analysis of normal hierarchical models. J Multivar Anal 135:163–174. https://doi.org/10.1016/j.jmva.2014.12.013, http://www.sciencedirect.com/science/article/pii/S0047259X15000020

Duncan TE, Duncan SC, Strycker LA (2006) An Introduction to Latent Variable Growth Curve Modeling: Concepts, Issues, and Applications, 2nd edn. Lawrence Erlbaum Associates, Mahwah, NJ

Estes W (1956) The problem of inference from curves based on group data. Psychol Bull 53(2):134–140. https://doi.org/10.1037/h0045156

Everitt B, Landau S, Leese M (2001) Cluster Analysis, 4th edn. Arnold, London

Fong Y, Rue H, Wakefield J (2010) Bayesian inference for generalized linear mixed models. Biostatistics 11(3):397–412. https://doi.org/10.1093/biostatistics/kxp053, https://arxiv.org/abs/academic.oup.com/biostatistics/article-pdf/11/3/397/18604192/kxp053.pdf

Gallistel C, Fairhurst S, Balsam P (2004) The learning curve: Implications of a quantitative analysis. Proceedings of the National Academy of Sciences 101(36):13,124–13,131. https://doi.org/10.1073/pnas.0404965101, https://www.pnas.org/content/101/36/13124, https://arxiv.org/abs/https://www.pnas.org/content/101/36/13124.full.pdf

Gelfand AE, Dey DK (1994) Bayesian model choice: asymptotics and exact calculations. J R Stat Soc 56(3):501–514

Gelman A, Hwang J, Vehtari A (2014) Understanding predictive information criteria for bayesian models. Stat Comput 24(6):997–1016

Genolini C, Alacoque X, Sentenac M et al (2015) kml and kml3d: R packages to cluster longitudinal data. J Stat Softw 65:1–34

Ghosh M, Kim D, Maiti T (1997) Hierarchical Bayesian analysis of longitudinal data. Sankhya 59(3):326–334

Green PJ, Silverman BW (2019) Nonparametric regression and generalized linear models: a roughness penalty approach. Chapman and Hall/CRC

Gutierrez-Pena E, Walker S (2019) An efficient method to determine the degree of overlap of two multivariate distribution. In: Antoniano-Villalobos I, Mena R, Mendoza M et al (eds) Selected Contributions on Statistics and Data Science in Latin America, Proceedings in Mathematics and Statistics, vol 301. Springer. https://doi.org/10.1007/978-3-030-31551-1_5

Hartigan JA, Wong MA (1979) Algorithm as 136: A k-means clustering algorithm. J R Stat Soc: Series C (Applied Statistics) 28(1):100–108. http://www.jstor.org/stable/2346830

Hwang H, Takane Y (2004) A multivariate reduced-rank growth curve model with unbalanced data. Psychometrika 69:65–79

Inman HF, Bradley ELJ (1989) The overlapping coefficient as a measure of agreement between probability distributions and point estimation of the overlap of two normal densities. Commun Stat Theory Methods 18:3851–3874. https://doi.org/10.1016/j.tics.2010.05.001

James G, Sugar CA (2003) Clustering for sparsely sampled functional data. J Am Stat Assoc 98(462):397–408

Lachos VH, Castro LM, Dey DK (2013) Bayesian inference in nonlinear mixed-effects models using normal independent distributions. Comput Stat Data Anal 64:237–252. https://doi.org/10.1016/j.csda.2013.02.011, https://www.sciencedirect.com/science/article/pii/S0167947313000558

Lee MD, Webb MR (2005) Modeling individual differences in cognition. Psychon Bull Rev 12(4):605–621. https://doi.org/10.3758/BF03196751

Leisch F (2004) Flexmix: A general framework for finite mixture models and latent class regression in r. J Stat Softw 11(8):1–18. https://doi.org/10.18637/jss.v011.i08, https://www.jstatsoft.org/v011/i08

Leon-Novelo L, Bekele BN, Müller P et al (2012) Borrowing strength with nonexchangeable priors over subpopulations. Biometrics 68(2):550–558

Lestari B, Budiantara I, Sunaryo S et al (2012) Spline smoothing for multi-response nonparametric regression model in case of heteroscedasticity of variance. J Math Stat 8(3):377–384

McLachlan G, Peel D (2000) Finite Mixture Models. John Wiley and Sons, New York

Melnykov V, Maitra R (2010) Finite mixture models and model-based clustering. Statistics Surveys 4:80–116. https://doi.org/10.1214/09-SS053

Muthén B, Shedden K (1999) Finite mixture modeling with mixture outcomes using the EM algorithm. Biometrics 55(2):463–469. https://doi.org/10.1111/j.0006-341X.1999.00463.x

Newell A, Rosenbloom PS (1993) Mechanisms of Skill Acquisition and the Law of Practice. MIT Press, Cambridge, MA, USA, pp 81–135

Omerovic S (2019) flexmixNL: Finite Mixture Modeling of Generalized Nonlinear Models. R package version 0.0.1

Oravecz Z, Muth C (2018) Fitting growth curve models in the Bayesian framework. Psychon Bull Rev 25(1):235–255. https://doi.org/10.3758/s13423-017-1281-0

Paddock SM, Savitsky TD (2013) Bayesian hierarchical semiparametric modelling of longitudinal post-treatment outcomes from open enrolment therapy groups. J R Stat Soc 176(3):795–808

Pinheiro JC, Bates DM (1995) Approximations to the log-likelihood function in the nonlinear mixed-effects model. J Comput Graph Stat 4(1):12–35. http://www.jstor.org/stable/1390625

Pinheiro JC, Bates DM (2000) Mixed-Effects Models in S and S-PLUS. Springer, New York

Plummer M (2003) Jags: A program for analysis of Bayesian graphical models using Gibbs sampling. In: Hornik K, Leisch F, Zeileis A (eds), Proceedings of the 3rd International Workshop on Distributed Statistical Computing. p 1–10

Quintana FA, Iglesias PL (2003) Bayesian clustering and product partition models. J R Stat Soc 65(2):557–574

Raudenbush S (2001) Comparing personal trajectories and drawing causal inferences from longitudinal data. Annu Rev Psychol 52:501–525

Rice N, Leyland A (1996) Multilevel models: applications to health data. J Health Serv Res Policy 1:154–164

Rouder JN, Lu J (2005) An introduction to Bayesian hierarchical models with an application in the theory of signal detection. Psychon Bull Rev 12(4):573–604. https://doi.org/10.3758/BF03196750

Ruppert D, Wand MP, Carroll RJ (2003) Semiparametric Regression. Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, https://doi.org/10.1017/CBO9780511755453

Samuels R (2012) Science and human nature. Royal Institute of Philosophy Supplement 70:1–28

Shiffrin RM, Lee MD, Wagenmakers EJ et al (2008) A survey of model evaluation approaches with a focus on hierarchical Bayesian methods. Cognitive Sci 32:1248–1284

Spiegelhalter DJ, Thomas A, Best N, et al (2003) Winbugs version 1.4 user manual. MRC Biostatistics Unit, Cambridge https://www.mrc-bsu.cam.ac.uk/bugs

Stenglein JL, Van Deelen TR (2016) Demographic and component allee effects in southern lake superior gray wolves. PLoS One 11(3):e0150,535

Tarpey T, Kinateder KKJ (2003) Clustering functional data. J Classification 20:93–114

Vehtari A, Gelman A, Gabry J (2017) Practical bayesian model evaluation using leave-one-out cross-validation and waic. Stat Comput 27(5):1413–1432

Villarroel L, Marshall G, Baron AE (2009) Cluster analysis using multivariate mixed effects models. Statistics Med 28:2552–2565

Vimal V, Zheng H, Hong P et al (2020) Characterizing individual differences in a dynamic stabilization task using machine learning. Aerosp Med Hum Perform 91(6):479–488. https://doi.org/10.3357/AMHP.5552.2020

Wasserman L (2000) Bayesian model selection and model averaging. J Math Psychol 44(1):92–107. https://doi.org/10.1006/jmps.1999.1278, https://www.sciencedirect.com/science/article/pii/S0022249699912786

Watanabe S, Opper M (2010) Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. J Mach Learn Res 11(12)

Welham S (2009) Smoothing spline models for longitudinal data. Longitudinal Data Analysis pp 253–289

Wiley JF, Bei B, Trinder J, et al (2014) Variability as a predictor: A Bayesian variability model for small samples and few repeated measures. arXiv preprint arXiv:1411.2961

Xu G, Zhu H, Lee JJ (2020) Borrowing strength and borrowing index for bayesian hierarchical models. Comput Stat Data Anal 144(106):901

Yan W, Hunt L, Sheng Q, et al (2000) R: Development core team (2005): R: a language and environment interaction for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, www.R-project.org

Zhang Z, Hamagami F, Wang L et al (2007) Bayesian analysis of longitudinal data using growth curve models. Int J Behav Dev 31(4):374–383

Authors and Affiliations

Maura Mezzetti[1] (ID) **· Daniele Borzelli[2] · Andrea d'Avella[2,3]**

✉ Maura Mezzetti
   maura.mezzetti@uniroma2.it

   Daniele Borzelli
   daniele.borzelli@unime.it

   Andrea d'Avella
   andrea.davella@unime.it

[1] Department of Economics and Finance, Università "Tor Vergata", Via Columbia 2,
   Rome 00133, Italy

[2] Department of Biomedical and Dental Sciences and Morphofunctional Imaging, University of
   Messina, Messina, Italy

[3] IRCCS Fondazione Santa Lucia, Rome, Italy