Fast Bayesian Functional Data Analysis: Application to basal body temperature data

James M. Ciera[1], **Bruno Scarpa**[1] **and David B. Dunson**[2]
[1]Department of Statistical Science, University of Padua,
Via Cesare Battisti 241, 35121 Padova, Italy.

[2]Department of Statistical Science, Duke University,
Box 90251, Durham, NC 27708-0251 USA.

Abstract: In many clinical settings, it is of interest to monitor a bio-marker over time for a patient in order to estimate that patient's trajectory and to ident

ify or predict clinically important features. For example, these features may correspond to a low or high point in the trajectory or to a sudden change. There is a need for fast algorithms for estimating functional trajectories while borrowing information from other patients about the shape and location of features in the function. Borrowing of information is crucial when observations are sparse and the interest is in prediction. In this paper, we presents an application of a fast approximate Bayes functional data analysis relying on spareness-favoring hierarchical priors for P-spline basis coefficients. The proposed method is used to rapidly estimate individual-specific functions. We present an application to basal body temperature (bbt) data.

Keywords: Bio-marker, MAP estimation, Ovulation, Relevance vector machine, Sparsity; Splines.

Department of Statistical Sciences
University of Padua
Italy

UNIVERSITA
DEGLI STUDI
DI PADOVA
DIPARTIMENTO
DI SCIENZE
STATISTICHE

Contents

Department of Statistical Sciences
Via Cesare Battisti, 241
35121 Padova
Italy

tel: +39 049 8274168

fax: +39 049 8274170

http://www.stat.unipd.it

Corresponding author:
James Ciera
tel: +39 049 827 4168
ciera@stat.unipd.it
http://www.stat.unipd.it/ciera

Fast Bayesian Functional Data Analysis: Application to basal body temperature data

James M. Ciera[1]**, Bruno Scarpa**[1] **and David B. Dunson**[2]
[1]Department of Statistical Science, University of Padua,
Via Cesare Battisti 241, 35121 Padova, Italy.

[2]Department of Statistical Science, Duke University,
Box 90251, Durham, NC 27708-0251 USA.

Abstract: In many clinical settings, it is of interest to monitor a bio-marker over time for a patient in order to estimate that patient's trajectory and to ident

ify or predict clinically important features. For example, these features may correspond to a low or high point in the trajectory or to a sudden change. There is a need for fast algorithms for estimating functional trajectories while borrowing information from other patients about the shape and location of features in the function. Borrowing of information is crucial when observations are sparse and the interest is in prediction. In this paper, we presents an application of a fast approximate Bayes functional data analysis relying on spareness-favoring hierarchical priors for P-spline basis coefficients. The proposed method is used to rapidly estimate individual-specific functions. We present an application to basal body temperature (bbt) data.

1   Introduction

In the recent years, there has been an increased interest on researches involving data that is characterized with high dimension. This results from experimental and observational studies where data are collected repeatedly from many subjects over a period of time. In such studies, data for different subjects are commonly characterized with similar patterns in trajectories. This characteristic can be used to study clinically important features. By observing patterns for measurements from a particular bio-maker over time, an investigator can easily identify or predict the occurrence of an event or identify unusual patterns. For example in reproductive studies, tracking measurement patterns of hormonal level or daily basal body temperature among women can help to identify or predict early pregnancy loss and occurrence of the ovulation day (Collins, 1996; Dunson, *et al*, 1999; Bigelow and Dunson, 2008).

In many studies it might be convenient to rapidly predict features of interest based on data extracted from large database that involve many subjects. Unfortunately, due to heterogeneity among subjects, data collection problems, data entry and storage errors, it is common to have subjects that have sparse and unequally spaced measurements. Therefore, raising a motivation to adopt a fast method that can accommodate those data problems while estimating or predicting curve features. We target to use a fast Bayesian methodology based on functional data analysis (FDA) that is commonly used to analyze longitudinal and highly dimensional data (Ramsay and Silverman, 2005).

Our research is motivated by the basal body temperature (bbt) data from the European fecundability study (Colombo and Masarotto, 2000). The study consists of daily bbt measurements from 880 women that contributed temperature measurements from at least one menstrual cycle. The data are characterized by missing temperature measurements in some days and there is variability in cycle lengths among women such that, majority have cycles that ranges between 20 to 40 days. Besides measurement errors which is a common problem with many longitudinal data, unequal cycle lengths and data sparsity pose a great hindrance to analyze the data using most available standard software.

A standard bbt curve from a healthy ovulating female has a biphasic pattern. This is characterized by a low plateau during the follicular phase, a temperature dip that occur prior to ovulation, and a sharp rise immediately after ovulation which is subsequently followed by a luteal phase plateau (Scarpa and Dunson, 2006). Several studies have suggested that ovulation day corresponds to the low point prior to the rise in basal body temperature (Marshall, 1979; Colombo and Masarotto, 2000). It is important to identify the ovulation day, since there is less probability of conception if an intercourse occurs outside of the six-day fertile interval ending on the ovulation day (Dunson et al., 1999). Therefore, to identify these features it is of interest to estimate a smooth trajectory in bbt over the menstrual cycle based on the available cycles data, while borrowing information flexibly across cycles in the data base.

Functional data analysis (FDA) is an ideal tool to use in estimation of smooth trajectories resulting from the bbt data characterized with sparseness and unequal cycle lengths (Ramsay and Silverman, 2005). The main aim of FDA is to explore and highlight important features of a curve. A trajectory may consist of one or several segments weighted using functional coefficients (Ramsay and Silverman, 2005). Unfortunately, FDA relies on relatively large number of basis functions and estimation of functional coefficients becomes time consuming activity using standard software. A common approximation procedure in FDA is to consider only a subset of carefully chosen basis functions that can be used for approximation purpose. However, it can be difficult to choose the basis functions in advance, motivating the use of adaptive methods that allow uncertainty in basis function selection (Bigelow and Dunson, 2007; Johnson and Rosen, 2008).

A lots of literature exist in selection of basis functions, for example the use of Multivariate adaptive regression spline (MARS) proposed by Friedman (1991). Bayesian methods can accommodate uncertainty in selection of basis functions. For example, recent approaches in Bayesian that incorporated MARS have been proposed by Denison *et al*, (2002). The method has good prediction performance but the poste-

rior sampling is based on reversible jump Markov Chain Monte Carlo (RJMCMC). The use of RJMCMC involves MCMC which is an extremely computer intensive method and involves hours to implement in data sets involving thousands of subjects. Hence, raising a practical motivation for fast approximate Bayes approaches that bypass MCMC while maintaining some of the benefits of a Bayesian analysis. Recent MARS extensions that bypasses the MCMC algorithms involves the use of an empirical Bayes approach for selecting basis functions and knots and is based on frequentist model selection methods and Bayesian approach Sakamoto (2007).

The article seeks to approximate the bbt projectiles using Multi-Task Relevant Vector machine (MT-RVM) method an extension of Relevant Vector machine (RVM) method (Tipping, 2001; Ji, *et al*, 2009). RVM is a fast Bayesian method based on Empirical Bayes methodology and has featured mostly in Machine Learning especially in signal reconstruction and compressive sensing. Relevant Vector machine is one among several methods that promote sparseness in estimation of functional coefficients. Other similar methods include; Least Absolute Shrinkage and Selection Operator (LASSO), Support Vector Machine (SVM) (Tibshirani, 1996; Burges, 1998). Sparseness is a property where the fitted model retains the least number of basis functions (by having non-zero coefficients), while all the other basis functions are pruned by setting their corresponding coefficients to zero (Tzikas, *et al*, 2005). This property provides a natural mechanism in variable selection leading to a sparse model that is fast to compute.

The subsequent sections are as follows; Section 2 introduces a simple functional model. Section 3 provides an implementation of the MCMC and MT-RVM estimation procedures and presents results based on comparison between the two methods using simulated and the bbt data from European fecundability study. Section 4 contains a discussion.

2   Methods

2.1   A functional data analysis model

We consider observations from the ith woman with response vector $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{iT_i})'$ consisting of bbt measurements and covariate vector $\boldsymbol{z}_i = (z_{i1}, \ldots, z_{iT_i})'$ representing observation day. A functional model is represented as,

$$y_{it} = f_i(z_{it}) + \epsilon_{it}, \quad \epsilon_{it} \sim N(0, \sigma_\epsilon^2), \quad t = 1, \ldots, T_i, \quad i = 1, \ldots, N. \qquad (1)$$

where $f_i(.)$ is a smooth function for subject i, ϵ_{it} is a measurement error and N is the number of women. The smoothing function can be described as a linear combination of M basis functions $f_i(z_{it}) = \sum_{j=1}^{M} \beta_{ij} \varphi_j(z_{it}) = \mathbf{x}_{it}' \boldsymbol{\beta}_i$ where $\boldsymbol{x}_{it} = (x_{it1}, \cdots, x_{itM})'$ are the values of the basis functions at z_{it}, parameter β_{ij} is the coefficient for the jth basis function $\varphi_j(.)$ and $\boldsymbol{\beta}_i = (\beta_{i1}, \cdots, \beta_{iM})'$. The basis functions $\boldsymbol{\varphi} = \{\varphi_j\}_{j=1}^{M}$ can be generated using numerous methods that have been discussed in the literature (e.g. Hastie, et al. 2001; Ruppert, et al. 2003). Conditionally on the basis $\boldsymbol{\varphi}$, expression (1) can be expressed in the form of linear random effects model.

The priors are $\beta_{ij} \sim N(0, \alpha_j^{-1})$, $\sigma_\epsilon^{-2} \sim Gamma(a, b)$ and $\alpha_j \sim Gamma(c, d)$ where $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_M)'$ and σ_ϵ^{-2} are shared among the subjects. The hyper-

parameters $\boldsymbol{\alpha}$ and σ_ϵ^{-2} are computed from the data as *maximum a posteriori* (MAP) estimates. Since these parameters are shared among the subjects, all the subjects contribute to estimation of $\boldsymbol{\alpha}$ and σ_ϵ^{-2} leading to borrowing of strength across subjects in estimating subject-specific functions. To promote sparseness over the model coefficients $\boldsymbol{\beta}_i$, the hyper-parameters c and d are set close to zero leading to a distribution with a large spike concentrated at zero and a heavy right tail. The basis functions for which α_j is in the right tail have coefficients that are strongly shrunk toward zero resulting in sparseness.

2.2 Posterior Estimates

The commonly used approach to compute the joint posterior density $p(\boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma_\epsilon^{-2}|\boldsymbol{Y})$ cannot be implemented since the computation of the posterior density require a normalization that cannot be expressed analytically. An alternative approach is to compute the posterior density based on the conditional distribution

$$p(\boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma_\epsilon^{-2}|\boldsymbol{Y}) = p(\boldsymbol{\alpha}, \sigma_\epsilon^{-2}|\boldsymbol{Y}) \prod_{i=1}^N p(\boldsymbol{\beta}_i|\boldsymbol{y}_i, \boldsymbol{\alpha}, \sigma_\epsilon^{-2}), \tag{2}$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \cdots \boldsymbol{\beta}_N)$. The density function $p(\boldsymbol{\beta}_i|\boldsymbol{y}_i, \boldsymbol{\alpha}, \sigma_\epsilon^{-2})$ is the posterior distribution for the random coefficients $\boldsymbol{\beta}_i$, while $p(\boldsymbol{\alpha}, \sigma_\epsilon^{-2}|\boldsymbol{Y})$ is the posterior density for the variance components $\boldsymbol{\alpha}$ and σ_ϵ^{-2}.

The posterior density for the random coefficients $\boldsymbol{\beta}_i$ is a multivariate normal distribution

$$p(\boldsymbol{\beta}_i|\boldsymbol{Y}, \boldsymbol{\alpha}, \sigma_\epsilon^{-2}) = \boldsymbol{N}(\boldsymbol{\beta}_i; \hat{\boldsymbol{\mu}}_i, \hat{\boldsymbol{\Sigma}}_i), \tag{3}$$

where $\hat{\boldsymbol{\mu}}_i = \sigma_\epsilon^{-2} \hat{\boldsymbol{\Sigma}}_i \boldsymbol{X}_i' \boldsymbol{y}_i$ is the mean vector and $\hat{\boldsymbol{\Sigma}}_i = (\boldsymbol{A} + \sigma_\epsilon^{-2} \boldsymbol{X}_i' \boldsymbol{X}_i)^{-1}$ is the covariance matrix such that $\boldsymbol{A} = diag\{\alpha_1, \cdots, \alpha_M\}$ and $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{iM})'$.

Since it is impossible to express the posterior density for the variance components $p(\boldsymbol{\alpha}, \sigma_\epsilon^{-2}|\boldsymbol{Y})$ analytically, we use an Empirical Bayes approach to compute the posterior estimates for $\boldsymbol{\alpha}$ and σ_ϵ^{-2}. These estimates are computed as MAP estimates as will be discussed in the next section. The density function $p(\boldsymbol{\alpha}, \sigma_\epsilon^{-2}|\boldsymbol{Y}) \propto p(\boldsymbol{Y}|\boldsymbol{\alpha}, \sigma_\epsilon^{-2})p(\boldsymbol{\alpha})p(\sigma_\epsilon^{-2})$ and both $p(\boldsymbol{\alpha})$ and $p(\sigma_\epsilon^{-2})$ are assumed to be Gamma density. To compute the estimates for $\boldsymbol{\alpha}$ and σ_ϵ^{-2}, we assume that the modes for $p(\boldsymbol{\alpha}, \sigma_\epsilon^{-2}|\boldsymbol{Y})$ and $p(\boldsymbol{Y}|\boldsymbol{\alpha}, \sigma_\epsilon^{-2})p(\boldsymbol{\alpha})$ are equivalent and hence the MAP estimates for $p(\boldsymbol{\alpha}, \sigma_\epsilon^{-2}|\boldsymbol{Y})$ are equivalent to the MLE estimates from $p(\boldsymbol{Y}|\boldsymbol{\alpha}, \sigma_\epsilon^{-2})$ (Ji et al, 2009).

2.3 Empirical Bayes method

Expressing the posterior density for the variance components $p(\boldsymbol{\alpha}, \sigma_\epsilon^{-2}|\boldsymbol{Y})$ is difficult analytically and the MAP estimates for $\boldsymbol{\alpha}$ and σ_ϵ^{-2} are computed from the marginal likelihood $p(\boldsymbol{Y}|\boldsymbol{\alpha}, \sigma_\epsilon^{-2})$, obtained after integrating out $\boldsymbol{\beta}_i$ from $p(\boldsymbol{Y}|\boldsymbol{\beta}_i, \sigma_\epsilon^{-2})$ such that

$$p(\boldsymbol{Y}|\boldsymbol{\alpha}, \sigma_\epsilon^{-2}) = \int \prod_{i=1}^N p(\boldsymbol{Y}|\boldsymbol{\beta}_i, \sigma_\epsilon^{-2})p(\boldsymbol{\beta}_i|\boldsymbol{\alpha})p(\boldsymbol{\alpha})d\boldsymbol{\beta}_i$$

This results to a normal density function $p(\boldsymbol{Y}|\boldsymbol{\alpha}, \sigma_\epsilon^{-2}) = N(\boldsymbol{y}_i; \boldsymbol{0}, \boldsymbol{C}_i)$ where the covariance matrix $\boldsymbol{C}_i = \sigma_\epsilon^2 \boldsymbol{I}_{T_i} + \sum_{j=1}^M \alpha_j^{-1} \boldsymbol{x}_{ij} \boldsymbol{x}_{ij}'$. The expressions for the estimates of $\boldsymbol{\alpha}$ and σ_ϵ^{-2} are obtained from the log-likelihood function $\ell(\boldsymbol{\alpha}, \sigma_\epsilon^{-2}) = \sum_{i=1}^N \log N(\boldsymbol{y}_i; \boldsymbol{0}, \boldsymbol{C}_i)$. We differentiate the log-likelihood $\ell(\boldsymbol{\alpha}, \sigma_\epsilon^{-2})$ with respect to parameters $\boldsymbol{\alpha}$ and σ_ϵ^{-2} respectively and equating the resulting score equations to zero. This results to

$$\hat{\alpha}_j = \frac{N}{\sum_{i=1}^N \mu_{ij}^2 + \boldsymbol{\Sigma}_{i,jj}}, \qquad j = 1, \cdots, M \tag{4}$$

$$\hat{\sigma}_\epsilon^{-2} = \frac{\sum_{i=1}^N ||\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\mu}_i||^2}{\sum_{i=1}^N (T_i - M - \boldsymbol{\alpha}^{-1} \boldsymbol{\Sigma}_{i,jj})}. \tag{5}$$

The estimates for $\boldsymbol{\alpha}$ and σ_ϵ^{-2} are inserted into $\hat{\boldsymbol{\mu}}_i$ and $\hat{\boldsymbol{\Sigma}}_i$ in equation (3) leading to an interactive procedure alternating between estimation of the parameters in equation (3) and (4-5) respectively.

However, two related problems arise while implementing the above empirical Bayes approach when the number basis functions is large. These problems are: estimability problems leading to lack of convergence and the computation process require large amount of time especially when dealing with large data sets. Such computation difficulties are commonly encountered when the dimension of the basis functions is large relative to the sample size, which is a common practice in many functional analysis work. For example, when M is large (e.g. $M > 10$), the inversion of $M \times M$ covariance matrix $\boldsymbol{\Sigma}_i$ becomes impossible leading to estimability problems resulting to lack of convergence of the procedure. Moreover, when the dataset consists of thousands of subjects, the computation process may take days.

Potentially these problems can be solved by using a MAP estimation approach that includes a proper prior to induce a penalty in the procedure that leads to shrinkage towards the prior and regularization. However, such an approach will be sensitive to hyper-parameter choice. An alternative approach is to adapt a fast algorithm that leads to a reduced model with dimension $m \times m$ for $\hat{\boldsymbol{\Sigma}}_i$ where $m <<$ M. The RVM iterative algorithm can generate such a sparse model and will be discussed in the next section.

2.4   A Fast Empirical Bayes method

Conditioning on the MLE estimates for σ_ϵ^{-2}, a fast approach to compute the elements of $\boldsymbol{\alpha}$ can be done sequentially. The algorithm is based on the dependence of the kth component of $\boldsymbol{\alpha}$ upon the log likelihood function

$$\ell(\boldsymbol{\alpha}, \sigma_\epsilon^{-2}) = -1/2 \sum_{i=1}^N \{T_i \log(2\pi) + \log |\boldsymbol{C}_i| + \boldsymbol{y}_i \boldsymbol{C}_i^{-1} \boldsymbol{y}_i\}. \tag{6}$$

However, the presence of matrix $\boldsymbol{C}_i$ in the log likelihood function $\ell(\boldsymbol{\alpha}, \sigma_\epsilon^{-2})$ makes it impossible to express the log-likelihood function into two parts -one containing the kth component and the other one that does not. Hence, to allow such decomposition, we first decompose the variance matrix $\boldsymbol{C}_i$ into two part -with and without the

contribution of the kth basis function. This leads to $\boldsymbol{C}_i = \boldsymbol{C}_{i,-k} + \alpha_k^{-1}\boldsymbol{x}_{ik}\boldsymbol{x}_{ik}'$ where $\boldsymbol{C}_{i,-k}$ is the part that does not have the contribution of the kth basis function. The resulting decomposed log-likelihood function is

$$\ell(\boldsymbol{\alpha}, \sigma_\epsilon^{-2}) = \ell(\boldsymbol{\alpha}_{-k}, \sigma_\epsilon^{-2}) + \frac{1}{2}\sum_{i=1}^{N}\left(\log\alpha_k - \log|\alpha_k + s_{ik}| + \frac{q_{ik}^2}{\alpha_k + s_{ik}}\right)$$

where $\ell(\boldsymbol{\alpha}_{-k}, \sigma_\epsilon^{-2})$ is the part without the contribution of the kth basis function, $s_{ik} = \boldsymbol{x}_{ik}'\boldsymbol{C}_{i,-k}^{-1}\boldsymbol{x}_{ik}$ and $q_{ik} = \boldsymbol{x}_{ik}'\boldsymbol{C}_{i,-k}^{-1}\boldsymbol{y}_i$. Differentiating $\ell(\boldsymbol{\alpha}, \sigma_\epsilon^2)$ with respect to α_k and setting the result to zero yield the score equations,

$$\frac{\ell(\boldsymbol{\alpha}, \sigma_\epsilon^{-2})}{\partial\alpha_k} = \sum_{i=1}^{N}\frac{s_{ik}^2/\alpha_k + s_{ik} - q_{ik}^2}{2(\alpha_k + s_{ik})^2} = 0.$$

The solutions for the score equations are infeasible to express analytically except for a trivial case when $\alpha_k = \infty$. The exact solutions require finding the zeros of a polynomial of degree $2N - 1$ which is computationally expensive. An alternative method to avoid such computation complexities is to assume that $\alpha_k << s_{ik}$, leading to the approximate estimate

$$\hat{\alpha}_k \cong \begin{cases} \frac{N}{\sum_{i=1}^{N}(q_{ik}^2 - s_{ik})/s_{ik}^2} & if \ \sum_{i=1}^{N}\frac{(q_{ik}^2 - s_{ik})}{s_{ik}^2} > 0, \\ \infty & otherwise. \end{cases} \qquad (7)$$

where $s_{ik} = \boldsymbol{x}_{ik}'\boldsymbol{C}_{i,-k}^{-1}\boldsymbol{x}_{ik}$, $q_{ik} = \boldsymbol{x}_{ik}'\boldsymbol{C}_{i,-k}^{-1}\boldsymbol{y}_i$ and $\boldsymbol{C}_{i,-k}$ is the component of $\boldsymbol{C}_i$ without the contribution of the kth basis function. The estimate for $\hat{\sigma}_\epsilon^{-2}$ is as expressed in equation (3). For a justification of this type of approximation, refer to Ji, et al. (2009).

We first start with an empty model and select the basis function that has the largest impact on the log-likelihood $\ell(\boldsymbol{\alpha}, \sigma_\epsilon^{-2})$. The subsequent steps on selection of the remaining basis functions involves three operations on $\boldsymbol{X}_{ik}$. These are: addition, deletion or updating $\hat{\mu}_{ik}$ operations. Addition occurs when $\sum_{i=1}^{N}\frac{(q_{ik}^2 - s_{ik})}{s_{ik}^2} > 0$ and $\boldsymbol{X}_{ik}$ is not in the model, while an update occurs when $\boldsymbol{X}_{ik}$ is already in the model and $\sum_{i=1}^{N}\frac{(q_{ik}^2 - s_{ik})}{s_{ik}^2} > 0$. We delete the basis function $\boldsymbol{X}_{ik}$ from the model when $\sum_{i=1}^{N}\frac{(q_{ik}^2 - s_{ik})}{s_{ik}^2} < 0$. The values from the previous iteration are used to update vectors $\boldsymbol{s}$ and $\boldsymbol{q}$. The final sparse linear model tends to have most of the $\alpha_j = \infty$, which corresponds to $\beta_{ij} = 0$ for all i and $j = 1, \cdots, M$.

3   Results

We present results from the bbt data and also evaluate the performance of the MT-RVM method relative to a classical MCMC based method as the number of observations increases. In both applications, the O'Sullivan-type penalized splines (Wand and Ormerod, 2008) are used to generate the basis functions $\boldsymbol{\varphi}$ based on the standardized values of time covariate ($\boldsymbol{z}_i$). In total we generated 29 columns for matrices $\boldsymbol{X}_i$, where the first and the second columns contained values of 1's and $\boldsymbol{z}_i$, respectively.

3.1   Application to the bbt data

We considered data from 500 women and select cycles that had an identified ovulation day. The study enrolled women aged between 18 and 40 years, were not taking hormonal medications or drugs affecting fertility, and had no known impairment of fecundity. The participants kept daily records of basal body temperature, and recorded the days during which intercourse and menstrual bleeding occurred. For more details about the study protocol, refer to Colombo and Masarotto (2000).

A standard bbt curve has biphasic shape and is characterized with three phases representing the pre-ovulation, ovulation and post-ovulation periods. The ovulation day is commonly identified using the three over six rule (Colombo and Masarotto, 2000) or by identifying the day that correspond to a dip that is followed by a sharp rise in the bbt curve. In reality, the classic bbt pattern is difficult to replicate and wide fluctuations in bbt, with many false nadirs and peaks, are commonly observed. Fluctuations result from a host of factors, other than hormonal fluctuations, that affect a woman's bbt: amount of sleep, sleep disturbances, ambient bedroom temperature, convection currents, food ingestion and emotional state (Colombo and Masarotto, 2000).

Many methods have been proposed to estimate the shape of the bbt curves. For example, Scarpa and Dunson (2008) proposed a Bayesian semi-parametric model based on nonparametric contamination of a linear mixed effects model. Implementation of their approach relies on a highly computationally -intensive MCMC algorithm, while our current goal is instead to obtain a fast approximate Bayes approach that can be implemented much more rapidly, while obtaining smooth estimates of the bbt trajectories.

We generated the basis function and estimated the curves using the MT-RVM and the classical Bayesian method. Figure 1 presents estimated curves for one cycle using the two methods where the continuous and dotted curves represent the curves generated by the MT-RVM and MCMC based methods respectively. The grey region is the credible band for the classical Bayesian method.

We note that the MCMC based curve is generated by estimating 29 non-zero basis coefficients while the approximation by the MT-RVM method is based on estimates for 4 non-zero basis coefficients. On time factor, the MCMC based method takes an average of 19.35 seconds to generate a curve while the MT-RVM method takes an average of 0.11 seconds to select the relevant basis coefficients and generate a curve.

3.2   Effects of adding more observations

It might be of interest to evaluate the performance of the RVM method with the increase of the number of observations. Potentially, we expect that the gap between the curves from the two methods to narrow down as the number of observations increases. Similarly, the estimated non-zero random coefficients from the two methods are also expected to be identical at after a certain number of observations. However, we do not know the threshold number of observations when the two curves or non-zero parameter estimates look identical.

To assess the performance of the MT-RVM with the increase in the number of observations per cycle, we generate 30 biphasic curves that mimic the shape of the
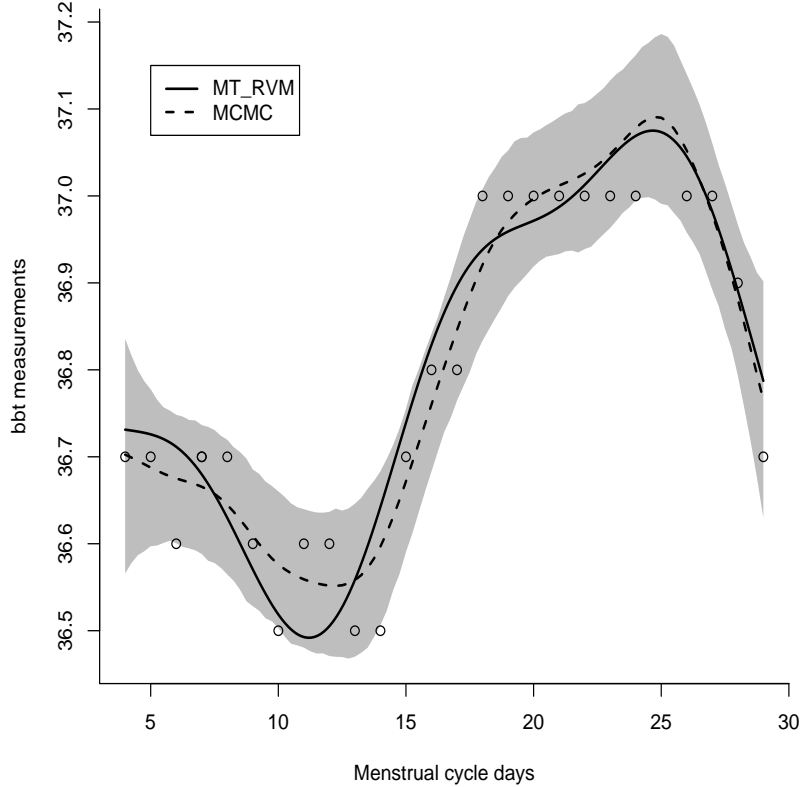
Figure 1: Trajectories estimated using the MT-RVM and MCMC methods.

bbt curve. The curves are generated using a sine function,

$$y_{it} = v_i + \rho_i z_{it} \sin(10 z_{it} - r_i) + \epsilon_{it}, \quad t = 1, 2, \ldots 27, \quad i = 1, 2, \ldots 30,$$

where covariate $z_{it} \sim unif(0, 1)$, while $v_i \sim unif(-1, 1)$ and $r_i \sim unif(-1, 1)$ are the vertical and horizontal shift parameters and $\rho_i \sim unif(0.5, 1.5)$ controls the amplitude of the curves. Each curve had 27 observations and we generated the basis functions using the method used in the previous section.

The computation of the basis coefficients is based on the two methods -an MCMC based and the MT-RVM methods. To compare the estimation of the basis coefficients from the two methods as the number of observations increases, we computed Reconstructive Error defined as

$$RE_l = \frac{1}{N} \sum_{i=1}^{N} \frac{||\boldsymbol{\beta}_i^{RVM} - \boldsymbol{\beta}_i^{MCMC}||}{||\boldsymbol{\beta}_i^{RVM}||}, \qquad l = 27, \ldots 227.$$

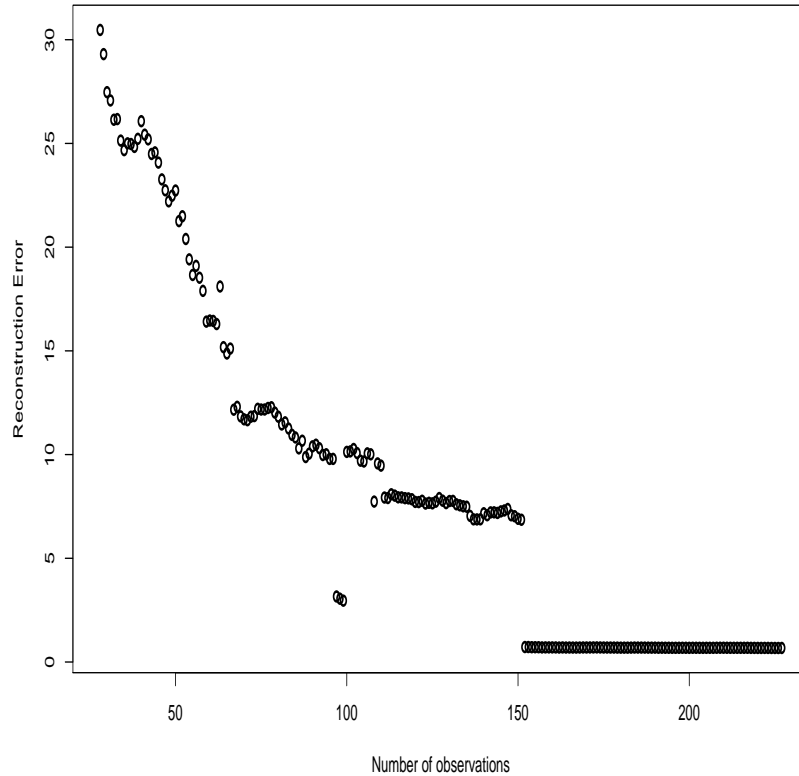Reconstructive error is a measure that captures the differences between the two sets

Figure 2: A plot for RE against the number of observations.

of parameters estimates from the two methods. After the computation of the coefficient estimates and the reconstructive error for the initial model, we subsequently simulated additional 200 observations for each cycle. After each increment, we re-computed the basis functions $\boldsymbol{X}_i$ for the new data, computed the basis coefficients using the two methods and then re-computed the reconstructive error value. We plot the RE_l against the number of observations (l) as shown in figure 2.

From the plots, it is evident that an increase in the number of observations leads to a gradual decrease in the reconstructive error but the decreasing trend reaches to a constant value after about 150 observations. However, the reconstructive error curve remains constant at a non-zero RE value since most of the non-relevant basis coefficients from the MCMC based method are non-zero while their corresponding basis coefficients from the MT-RVM method are zero.

4   Concluding note

In this paper we have demonstrated the use of Multi-task Relevant Vector Machine (MT-RVM) -a fast Empirical Bayes method- as an alternative to computer intensive methods that rely on MCMC. The method is fast and can be used to rapidly approximate non-linear curves for massive datasets. The discussed Multi-task Relevant Vector Machine approach has been used in machine learning especially in signal reconstruction and compressive sensing applications Ji et al, (2009). However, the application of this approach has not featured in applications that involve smoothing non-linear curves using penalized spline basis functions. In particular, the use of RVM procedure to smoothen non-linear bbt and hormonal data curves that feature commonly in many reproductive studies.

Multi-task Relevant Vector Machine method allows shrinking the subject's random coefficients toward zero, yielding a sparse linear model where majority of the random coefficients are zero and a few are non-zero. The method allows automatic selection and estimation of the relevant (non-zero) random coefficients to model functional data. The selection procedure is achieved by specifying independent priors for the random coefficients which facilitates the implementation of the MT-RVM algorithm that aids in the selection of the relevant basis functions (Ji et al, 2009).

Similar data smoothing approaches have been used in other contexts. For example, Brumback and Rice, (1998), used penalized smoothing spline mixed-model to generate smooth curves for multi-level progesterone data. Ramsay and Silverman, (2005) used Functional Principal Component Analysis (FPCA) on linear mixed models to smoothen non-linear data. However, these approaches do not allow fast computation since the computation relies on slow and computational intensive MCMC algorithms. The use of the MT-RVM method helps to bypass the use of MCMC and effectively unites the problem of variable (basis function) selection and coefficient estimation which leads to a fast model building process that simultaneously addresses the two problems within a single step at a less computational cost. The approach can easily be implemented into more complex high dimensional and hierarchical functional data to yield a sparse model that can be used to rapidly estimate functional trajectories.

The advantage of our approach is not only on computational speed but it also allows for better generalization performance which leads to sparse generalized linear mixed models. This aspect can provide inference for a wide variety of models at a moderate computational cost. For example, this approach can easily be extended to accommodate multiple predictors, linear mixed effect models and probit models where multiple binary categorical outcomes can be handled using data augmentation (Albert and Chib, 1993). Currently, we are generalizing the implementation of the MT-RVM method into linear mixed model to generate sparse LME models. Such generalization can easily be extended into LME models that can handle hierarchical data where a woman can have data from multiple cycles.

References

Albert, J., and Chib, S. (1993). Bayesian Analysis of Binary and Polychotomous Response Data. *Journal of the American Statistical Association*, **88**, 669–679.

Bigelow, J.L., and Dunson, D.B. (2008). Bayesian adaptive regression splines for hierarchical data. *Biometrics*, **63**,724-732.

Brumback B.A. and Rice J.A., (1998). Smoothing spline models for the analysis of nested and crossed samples of curves (with discussion), *Journal of the American Statistical Association*, **93**,961-94.

Burges, C.J.C. (1998). *A Tutorial on Support Vector Machines for Pattern Recognition.* Kluwer Academic Publishers, Boston.

Colombo, B. and Masarotto, G. (2000). Daily fecundability: First results from a new data base. *Demographic Research*, **3**:5.

Crainiceanu, C.M, Ruppert, D., (2004). Likelihood ratio tests in linear mixed models with one variance component. *Journal of the Royal Statistical Society B* **66** (1), 165-185.

Daniels, M.J. and Kass, R.E. (1999). Nonconjugate Bayesian Estimation of Covariance matrices and its use in hierarchical models. *Journal of the American Statistical Association*, **94**, 1254-1263.

Dunson, D. B., Weinberg, C. R., Perreault, S. D., and Chapin, R.E. (1999). Summarizing the motion of self-propelled cells: applications to sperm motility. *Biometrics*, **55**, 537-543.

George, E. I. and McCulloch, R. E. (1993), Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, **88**, 881-889.

Hastie, T., Tibshirani, R. and Friedman, J. (2001), *The Elements of Statistical Learning; Data mining, Inference and Prediction*, Springer Verlag, New York.

Green, P.J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82**, 711-732.

Guo, W.S. (2004). Functional data analysis in longitudinal settings using smoothing splines. *Statistical Methods in Medical Research*, **13**, 49-62.

Ji, S., Dunson, D.B. and Carin, L. (2009). Multi-task compressive sensing. *IEEE Transactions on Signal Processing*, **57**(1), 92-106.

Kinney, S. and Dunson, D.B. (2007). Fixed and random effects selection in linear and logistic models. *Biometrics*, **63**, 690-698.

Krivobokova, T., Crainiceanu, C.M. and Kauermann, G. (2008). Fast adaptive penalized splines. *Journal of Computational and Graphical Statistics*, **17**, 1-20.

Lin, X.H. and Zhang, D.W. (1999). Inference in generalized additive mixed models by using smoothing splines. *Journal of the Royal Statistical Society B*, **61**, 381-400.

Morris, J.S. and Carroll, R.J. (2006). Wavelet-based functional mixed models. *Journal of the Royal Statistical Society, Series B*, **68**(2) 179-199.

Pauler, D.K., Wakefield, J.C. and Kass, R.E. (1999). Bayes factors and approximations for variance component models, *Journal of the American Statistical Association*, **94**, 1242-1253.

Pourahmadi, M. (2007). Cholesky decompositions and estimation of a covariance matrix: orthogonality of variance-correlation parameters. *Biometrika*, **94**(4), 1006-101.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional data analysis.* Springer Verlag, New York.

Ruppert, D. and Carroll, R. J. (2000). Spatially adaptive penalties for spline fitting. *Australia and New Zealand Journal of Statistics*, **42**, 205-223.

Ruppert, D., Wand, M. P. and Carroll, R. J. (2003). *Semi-parametric Regression.* Cambridge University Press.

Sakamoto, W. (2007). Selecting basis functions and knots with an empirical Bayes method. *Computational Statistics*, **22**(1), 583-597.

Scarpa, B., and Dunson, D.B. (2008). Bayesian hierarchical functional data analysis via contaminated informative priors. *Biometrics*, to appear.

Smith, M. and Kohn, R. (1996). Nonparametric regression via Bayesian variable selection. *Journal of Econometrics*, **75**, 317-344.

Thompson, W.K. and Rosen, O. (2008). A Bayesian model for sparse functional data. *Biometrics*, **64**, 54-63.

Tibshirani, R. J. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society B*, **58**, 267-288.

Tipping, M. E. (2001). Sparse Bayesian learning and the relevance vector machine. *Journal of Machine Learning Research*, **1**, 211-244.

Tzikas, D.G., Wei, L., Likas, A., Yang, Y., and Galatsanos, N.P. (2006), A tutorial on relevance vector machines for regression and classification with applications. *EURASIP News Letter*, **17**.

Wand, M.P. and Ormerod, J.T. (2008), On O'Sullivan penalised splines and semi-parametric regression. *Australian and New Zealand Journal of Statistics*, **50**, 179-198.

Working Paper Series
Department of Statistical Sciences, University of Padua

You may order paper copies of the working papers by emailing wp@stat.unipd.it

Most of the working papers can also be found at the following url: http://wp.stat.unipd.it

Department of Statistical Sciences
University of Padua
Italy