Semi-parametric Nonlinear Mixed Effects Models and Their Applications

Chunlei Ke and Yuedong Wang *

April 3, 2000

Abstract

NonLinear Mixed effects Models (NLMM) and SElf-MOdeling nonlinear Regression models (SEMOR) are often used to fit repeated measures data. They use a common function shared by all subjects to model variation within each subject and some fixed and/or random parameters to model variation between subjects. The parametric NLMM may be too restrictive and the semi-parametric SEMOR model ignores correlations within each subject. In this paper we propose a class of Semi-parametric Nonlinear Mixed effects Models (SNMM) which extend NLMM, SEMOR and many other existing models in a natural way. A SNMM assumes that the mean function depends on some parameters and nonparametric functions. The parameters provide interpretable data summary. The non-parametric functions provide flexibility to allow data to decide some unknown/uncertain components, such as the shape of the mean response over time. A second stage model with fixed and random effects is used to model the parameters. Smoothing splines are used to

*Chunlei Ke (email: kechlei@pstat.ucsb.edu) is a graduate student, and Yuedong Wang (email: yuedong@pstat.ucsb.edu) is an Associate Professor in the Department of Statistics and Applied Probability, University of California, Santa Barbara, California 93106. This research was supported by NIH Grants R01 GM58533. Address for correspondence: Yuedong Wang, Department of Statistics and Applied Probability, University of California, Santa Barbara, California 93106.

model the nonparametric functions. Covariate effects on parameters can be built into the second stage model and covariate effects on nonparametric functions can be constructed using smoothing spline ANOVA decompositions. We extend the two-step procedure in NLMMs to estimate parameters and nonparametric functions using penalized likelihood. As a by-product, we propose a new spline smoothing algorithm for data as a noisy map of nonlinear functionals which extends the Gauss-Newton method to infinite dimensional spaces. A Bayesian formulation of SNMMs and Bayesian confidence intervals for the nonparametric function f are also discussed. We examine the proposed estimation and inference procedures through a large-scale simulation study. Applications of SNMMs are illustrated with analyses of two datasets.

KEY WORDS: L-spline; Nonlinear mixed effects models; Penalized likelihood; Repeated measures; Self-modeling nonlinear regression; Smoothing spline; Smoothing spline ANOVA.

1 Introduction

Repeated measures data arise in many areas of investigation such as agriculture, pharmacokinetics, epidemiology, medicine and social science. They are generated by observing each of a number of subjects repeatedly under varying conditions where the subjects are assumed to constitute a random sample from a population of interest. Common types of repeated measures data are longitudinal data with observations made at different times and spatial data where observations are made at different spatial locations.

We begin by reviewing NonLinear Mixed effects Models (NLMM) and SElf-MOdeling nonlinear Regression (SEMOR) models, which have proved invaluable in analyzing longitudinal data. The model proposed in this paper is a generalization of both of these.

Lindstorm and Bates (1990) proposed the following NLMM:

$$\begin{aligned} y_{ij} &= \eta(\boldsymbol{\phi}_i; t_{ij}) + e_{ij}, \quad i = 1, \ldots, m, \quad j = 1, \cdots, n_i, \\ \boldsymbol{\phi}_i &= \boldsymbol{A}_i \boldsymbol{\beta} + \boldsymbol{B}_i \boldsymbol{b}_i, \quad \boldsymbol{b}_i \overset{iid}{\sim} N(\mathbf{0}, \sigma^2 \mathbf{D}), \end{aligned} \tag{1}$$

where m is the number of subjects; n_i is the number of observations from subject i; y_{ij} is the jth observation on subject i; η is a known function of a covariate t_{ij}; $\boldsymbol{\phi}_i$ is a vector of parameters which may act nonlinearly on η; the e_{ij}'s are random errors, $\boldsymbol{e}_i = (e_{i1}, \cdots, e_{in_i})^T \sim N(\mathbf{0}, \sigma^2 \boldsymbol{\Lambda}_i)$ and $\boldsymbol{e}_i$ are mutually independent; $\mathbf{A}_i$ and $\mathbf{B}_i$ are design matrices; $\boldsymbol{\beta}$ is a vector of population parameters (fixed effects); $\boldsymbol{b}_i$ are random effects for subject i, and $\boldsymbol{b}_i$ and $\boldsymbol{e}_i$ are independent. At the first stage in model (1), the NLMM assumes a *common* parametric function η that relates the conditional mean of the response variable to the covariate t and parameters $\boldsymbol{\phi}_i$. At the second stage, effects of other covariates on $\boldsymbol{\phi}_i$ are modeled with fixed and random effects. NLMMs provide flexible tools to model both the mean and variance-covariance structures. They can easily incorporate covariate effects into the second stage model. Parameters $\boldsymbol{\phi}_i$ usually have meaningful interpretations. Available software to fit NLMMs includes the Splus function `nlme` by Pinheiro and Bates (1999) and the SAS procedure `PROC NLMIXED`.

For most NLMMs, we can write the first stage model in (1) as:

$$y_{ij} = \eta(\boldsymbol{\phi}_i, f; t_{ij}) + e_{ij}, \quad i = 1, \ldots, m, \quad j = 1, \cdots, n_i, \tag{2}$$

where f is a known function. For example, taking $f = 1 - e^{-t}$, we can write the model $\eta(\boldsymbol{\phi}_i, t) = \phi_{1i}(1 - e^{-\phi_{2i}(t - \phi_{3i})})$ in Section 6.2 as $\eta(\boldsymbol{\phi}_i, f; t) = \phi_{1i} f(\phi_{2i}(t - \phi_{3i}))$ with $\boldsymbol{\phi}_i = (\phi_{1i}, \phi_{2i}, \phi_{3i})^T$. The form (2) of a NLMM allows us to compare a NLMM with a SEMOR model and a Semi-parametric Nonlinear Mixed effects Model (SNMM) introduced in this paper. It also allows us to use a SNMM to test if a NLMM is appropriate by testing a particular functional form of f. See remarks in Section 2.1 and applications in Section 6 for more details.

The functions η and f in a NLMM are assumed to be known. Usually they are derived from an approximation to a mechanistic model such as a compartmental model under some simplifying assumptions. The resulting models may or may not be appropriate for a dataset because the assumptions may be too restrictive and/or the approximations may be too crude. As the popularity of nonlinear mixed effects models increases, there is an increasing need for more sophisticated model building and diagnostic techniques.

An alternative approach to model longitudinal data is SEMOR (Lawton, Sylvestre, and Maggio, 1972). A SEMOR model assumes that a *common* curve f exists for all subjects and that a particular subject's response curve is some parametric transformation of the common curve. Specifically, it assumes that

$$y_{ij} = \eta(\boldsymbol{\phi}_i, f; t_{ij}) + e_{ij}, \quad i = 1, \cdots, m, \quad j = 1, \cdots, n_i, \tag{3}$$

where f is the common curve; η is a known mean function; $\boldsymbol{\phi}_i$ are *deterministic* parameters which define the transformation for subject i. Notice that we present the SEMOR model in a more general form (3) than those in literature. It allows η to be nonlinear in f. The SEMOR model (3) is in the same form as the first stage model of a NLMM as restated in (2). However, while f is known in (2), the common curve f in (3) is unknown and usually estimated nonparametrically. A SEMOR model combines advantages of a parametric model, where parameters provide efficient and interpretable data summaries, and of a nonparametric model where the functional form is flexible and decided by data. SEMOR methods have been applied successfully to a number of areas, such as human growth (Stutzle et al, 1980), spirometry, spectrophometric and sensitometric analyses (Lawton, Sylvestre and Maggio, 1972), human circadian rhythms (Wang and Brown, 1996), bioassay (Guardabasso, Rodbard and Munson, 1987) and curve registration (Ramsay and Li, 1998). Also see Altman (1996). Nevertheless, SEMOR models have the following drawbacks due to the deterministic nature of $\boldsymbol{\phi}_i$: (1) they ignore correlations among observations; (2) they have the number of parameters proportional to the number of subjects; and (3) it is difficult to investigate covariate effects on parameters and/or the common curve. Ad hoc second stage analyses are often used to investigate covariate effects.

Usually the observed subjects represent a random sample from a population and inferences are about the population. Therefore, as in a NLMM, it is more natural to treat the parameters $\boldsymbol{\phi}_i$ in (3) as random effects. The SNMM model we propose in this paper is a natural combination and extension of models (1) and (3). We include mixed effects in a SEMOR to model variations within and between subjects, covariate effects and variance-covariance structure simultaneously. When certain components such as the common function are unknown or difficult to specify, we will leave them unspecified and estimate them nonparametrically using smoothing splines. Lindstorm (1995) included random shift and scale parameters into a shape invariant model, a special case of SEMOR models where η is linear in f. She modeled the common curve using a free-knot spline. Since the number of knots is fixed and common for all subjects, her model reduces to a NLMM and existing programs for NLMM can be used directly. We will consider the general SEMOR models where η is a known function of f, but where f may act nonlinearly on η, and use general smoothing splines to model the unknown function(s). Thus, besides longitudinal data, we can deal with many forms of data in a unified fashion. For example, thin-plate splines can be used for spatial data and Smoothing Spline ANOVA (SS ANOVA) models can be used for spatio-temporal data. We can conduct hypothesis tests using L-splines and SS ANOVA decompositions as illustrated by examples in Section 6. The complexity of the function f will be decided by data-based criteria. Our spline estimate of f depends on random effects, and thus can pool information across subjects more efficiently. SNMM does not reduce to any parametric model, which imposes challenges to estimation and inference (see Section 3).

Wang (1998) proposed Nonparametric Mixed effects Models (NMM) which model the fixed and random effects nonparametrically using SS ANOVA decompositions. The SNMMs are extensions of NMMs by allowing the mean response η to depend on both the parameters $\boldsymbol{\phi}_i$ and the nonparametric function f nonlinearly.

In the next section, we introduce and discuss SNMMs. In Section 3 we propose a two-step estimation procedure based on penalized likelihood. Implementation of this procedure using Splus function `nlme` and Fortran routine `RKPACK` is described in Section 4. Section 5 discusses inference procedures for f. We apply SNMMs to Canadian temperature data and CO2 uptake data in Section 6. Simulation results are reported in Section 7. In Section 8 the paper ends with discussion including possible further development in this area.

2 Semi-parametric Nonlinear Mixed Effects Models

2.1 The Model

We define a class of SNMMs as

$$
\begin{aligned}
&y_{ij} = \eta(\boldsymbol{\phi}_i, f; t_{ij}) + e_{ij}, \quad j = 1, \cdots, n_i, \quad i = 1, \cdots, m, \\
&\boldsymbol{\phi}_i = \mathbf{A}_i \boldsymbol{\beta} + \mathbf{B}_i \boldsymbol{b}_i, \quad \boldsymbol{b}_i \overset{iid}{\sim} N(\mathbf{0}, \sigma^2 \mathbf{D}),
\end{aligned}
\tag{4}
$$

where y_{ij} is the jth response on subject i; t_{ij} is a covariate which could be a scalar or a vector; η is a known function of t_{ij} which depends on a parameter vector $\boldsymbol{\phi}_i$ of length r and an unknown function f; $\boldsymbol{\beta}$ is a p-vector of fixed effects; $\boldsymbol{b}_i$ is q-vector of random effects associated with subject i; $\mathbf{A}_i$ and $\mathbf{B}_i$ are design matrices of sizes $r \times p$ and $r \times q$ for the fixed and random effects respectively; the random errors $\boldsymbol{e}_i = (e_{i1}, \cdots, e_{in_i})^T$ are mutually independent, distributed as $N(\mathbf{0}, \sigma^2 \mathbf{\Lambda_i})$ and independent of the random effects.

Let $N = \sum_{i=1}^m n_i$, $\boldsymbol{y}_i = (y_{i1}, \cdots, y_{in_i})^T$, $\boldsymbol{y} = (\boldsymbol{y}_1^T, \cdots, \boldsymbol{y}_m^T)^T$, $\boldsymbol{\phi} = (\boldsymbol{\phi}_1^T, \cdots, \boldsymbol{\phi}_m^T)^T$, $\boldsymbol{\eta}_i(\boldsymbol{\phi}_i, f) = (\eta(\boldsymbol{\phi}_i, f; t_{i1}), \cdots, \eta(\boldsymbol{\phi}_i, f; t_{in_i}))^T$, $\boldsymbol{\eta}(\boldsymbol{\phi}, f) = (\boldsymbol{\eta}_1(\boldsymbol{\phi}_1, f)^T, \cdots, \boldsymbol{\eta}_m(\boldsymbol{\phi}_m, f)^T)^T$ and $\boldsymbol{b} = (\boldsymbol{b}_1^T, \cdots, \boldsymbol{b}_m^T)^T$. Then the overall model can be written in the following matrix form

$$
\begin{aligned}
&\boldsymbol{y} | \boldsymbol{b} \sim N(\boldsymbol{\eta}(\boldsymbol{\phi}, f), \sigma^2 \mathbf{\Lambda}), \\
&\boldsymbol{\phi} = \boldsymbol{A}\boldsymbol{\beta} + \boldsymbol{B}\boldsymbol{b}, \;\; \boldsymbol{b} \sim N(\mathbf{0}, \sigma^2 \tilde{\boldsymbol{D}}),
\end{aligned}
\tag{5}
$$

where $\mathbf{\Lambda} = diag(\mathbf{\Lambda}_1, \cdots, \mathbf{\Lambda}_m)$, $\mathbf{A} = diag(\boldsymbol{A}_1, \cdots, \boldsymbol{A}_m)$, $\mathbf{B} = diag(\boldsymbol{B}_1, \cdots, \boldsymbol{B}_m)$ and $\tilde{\mathbf{D}} = diag(\mathbf{D}, \cdots, \mathbf{D})$. In the following we assume that $\mathbf{\Lambda}$ and $\mathbf{D}$ depend on an unknown parameter vector $\boldsymbol{\theta}$.

Remarks:

1. The name SNMM comes from the facts that the model contains both parameters and a nonparametric function and that the mean function η may be nonlinear in both parameters $\boldsymbol{\phi}_i$ and f. As in NLMM and SEMOR models, the parameters provide efficient data summary and the nonparametric function provides a tool to allow some unknown/uncertain components f flexible and to be decided by data.

2. Clearly, caution must be exercised to ensure the identifiablity of a SNMM (Kneip and Gasser, 1988). In this paper we assume this model is identifiable. Identifiability conditions depend on specific applications. We will show ways to make a model identifiable in the examples in Section 6.

3. We have adopted assumptions and notations as in Lindstorm and Bates (1990). It is trivial to extend our models to more general forms as is true for NLMMs. For example, $\boldsymbol{\phi}_i$ can be nonlinear in $\boldsymbol{\beta}$ and $\boldsymbol{b}_i$ as in Davidian and Giltinan (1995).

4. A NLMM is a special case of a SNMM when f is known. Therefore, SNMMs can be used to test if a specific nonlinear function f in (2) is appropriate (see Section 6). Using a nonparametric function may simplify the form of a model since some nuisance parameters can be absorbed into the nonparametric functions.

5. It is obvious that SEMOR models are special cases of the SNMM. Therefore, the computer software we develop can also be used to fit SEMOR models.

6. For simplicity of notation, we assume that η depends on one unknown function f. In general η may depend on multiple unknown functions. Domains for these functions may be the same or different. See Case 2 in Section 2.3 for an interesting example.

7. As in nonlinear mixed effects models, covariance matrices $\boldsymbol{\Lambda}$ and $\tilde{\boldsymbol{D}}$ can be modeled with many kinds of structures. See Davidian and Giltinan (1995) and Pinheiro and Bates (1999).

2.2 Functional Space for f

In a SNMM (4), f is thought of as a parameter just like $\boldsymbol{\phi}_i$. The difference is that $\boldsymbol{\phi}_i$ belongs to a finite dimensional space (usually $\mathcal{R}^r$) while f, as a function, belongs to an infinite dimensional space. Which space to use for f depends on factors such as (1) domain of the function, (2) prior knowledge, (3) constraints for identifiability, and (4) purpose of the analysis.

Usually f depends on the covariate t only with all possible values of t as the domain of f. For example, if the domain is a continuous interval $[a,\ b]$ and we assume f is smooth, we may use the Sobolev space

$$W_m([a,\ b]) = \{f : f^{(i)} \text{ absolute continuous}, \quad i = 1, \cdots, m-1, \quad \int_a^b (f^{(m)})^2 < \infty\} \tag{6}$$

as the functional space for f. If we know a priori that f is a periodic smooth function of a continuous variable t, then the domain is $\mathcal{C}$, the unit circle, and the following periodic Sobolev space

$$\begin{aligned} & W_m(per) \\ = \; & \{f : f^{(i)} \text{ absolute continuous}, \quad f^{(i)}(0) = f^{(i)}(1), \quad i = 1, \cdots, m-1, \quad \textstyle\int_0^1 (f^{(m)})^2 < \infty\} \end{aligned} \tag{7}$$

may be used as the model space for f.

Sometimes we only know some characteristics rather than a specific form of f. For example, we may know that f is monotone, or close to a function that satisfies certain differential equations. We can model these characteristics conveniently by using appropriate functional spaces, transformations (which are easy to deal with in our setting because we allow the mean to be a nonlinear functional of f), constraints and L-splines (Wahba, 1990; Ansley, Kohn and Wong, 1993; Heckman and Ramsay, 2000). In addition, putting the favored functions into the null space can lead to a spline estimate with smaller bias (Heckman and Ramsay, 2000).

Constraints to make a SNMM identifiable usually may be written as conditions on f. We often begin with an appropriate model space such as the Sobolev space and apply these conditions to reduce the model space to a proper subspace. For example, in order to make f and the constant μ identifiable in Section 6.1, we reduce the model space from $W_2(per)$ to $W_2(per) \ominus span\{1\}$, where $span\{1\}$ denotes all the constant functions.

Commonly used splines such as cubic splines and periodic splines may be employed to model f for the purpose of estimation. However, if our aim is hypothesis tests, the null hypothesis might determine the type of splines. For example, partial splines or L-splines may be used.

The nonparametric function f may also depend on other covariates such as treatment group. In this case the domain is the tensor product of the treatment group and the domain of t. We can then build SS ANOVA models to test treatment effects on f. Applications that use SS ANOVA decompositions to model covariate effects on f will be presented elsewhere.

We will show how to choose model spaces in Section 6 using examples. For generality, we denote the domain of the function f as $\mathcal{T}$, an arbitrary index set. We will assume f belongs to some Reproducing Kernel Hilbert Space (RKHS) $\mathcal{H}$ (Aronszajn, 1950; Wahba, 1990). Since the domain is arbitrary, we can deal with many situations in a unified fashion. Interesting examples of $\mathcal{T}$ are $\mathcal{T} = [a, b]$ for longitudinal data, $\mathcal{T} = \mathcal{C}$ for circular data, $\mathcal{T} = \mathcal{R}^2$ for spatial data, and $\mathcal{T} = \{1, \cdots, k\} \otimes [a, b]$ for longitudinal data with different common curves for each of k treatment groups.

2.3 Several Special Cases

In the model (4), η may be nonlinear in both $\boldsymbol{\phi}_i$ and f. In the following, we discuss several special cases of the SNMM.

Case 1: η is linear in both $\boldsymbol{\phi}_i$ and f, then

$$\eta(\boldsymbol{\phi}_i, f; t) = \alpha(t) + \boldsymbol{\delta}_i^T(t)\boldsymbol{\phi}_i + \gamma(t)f(t),$$

where $\alpha(t)$, $\boldsymbol{\delta}_i(t)$ and $\gamma(t)$ are known functions. In the special case where $\alpha(t) = 0$, $\gamma(t) = 1$ and $\boldsymbol{\phi}_i = \boldsymbol{B}_i\boldsymbol{b}_i$, the model becomes

$$\boldsymbol{\eta}(\boldsymbol{\phi}, f) = \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{f},$$

where $\boldsymbol{Z}_i = (\boldsymbol{\delta}_i(t_{i1}), \cdots, \boldsymbol{\delta}_i(t_{in_i}))^T\boldsymbol{B}_i$, $\boldsymbol{Z} = diag(\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_m)$ and $\boldsymbol{f} = (f(t_{11}), \cdots, f(t_{mn_m}))^T$. This is a special case of the NMM proposed in Wang (1998). In this case, SNMM could be thought of as an extension of the NMMs to the nonlinear case.

Case 2: If η is linear in f conditional on $\boldsymbol{\phi}_i$, then

$$\eta(\boldsymbol{\phi}_i, f; t) = \alpha(\boldsymbol{\phi}_i; t) + \delta(\boldsymbol{\phi}_i; t)f(\gamma(\boldsymbol{\phi}_i; t)),$$

where α, δ and γ are known parametric functions. In the special case where $\alpha(\boldsymbol{\phi}_i; t) = \mu + a_i$, $\delta(\boldsymbol{\phi}_i; t) = b_i$ and $\gamma(\boldsymbol{\phi}_i; t) = (t - c_i)/d_i$ with $\boldsymbol{\phi}_i = (\mu, a_i, b_i, c_i, d_i)^T$, the model becomes

$$\eta(\boldsymbol{\phi}_i, f; t) = \mu + a_i + b_i f((t - c_i)/d_i),$$

the simplest and most popular SEMOR model: the shape invariant model. In this case, the parameters $\boldsymbol{\phi}_i$ can be thought of as location and scale parameters and f can be thought as a shape parameter. Another special case is to take $\delta(\boldsymbol{\phi}_i; t) = 1$ and $\gamma(\boldsymbol{\phi}_i; t) = t$. The resulting model extends the partial spline model (Wahba, 1990) to allow random effects in the parametric part.

Next assume η depends on multiple functions $f_1, \cdots, f_l$. If η is linear in $f_1, \cdots, f_l$ conditional on $\boldsymbol{\phi}_i$, then

$$\eta(\boldsymbol{\phi}_i, f_1, \cdots, f_l; \boldsymbol{t}) = \alpha(\boldsymbol{\phi}_i; \boldsymbol{t}) + \sum_{k=1}^{l} \delta_{ik}(\boldsymbol{\phi}_i; \boldsymbol{t}) f_k(\gamma_k(\boldsymbol{\phi}_i; \boldsymbol{t})),$$

where α, δ_{ik} and γ_k are known parametric functions. $\boldsymbol{t} = (t^1, \cdots, t^l)^T$ is a vector, whose elements may represent the same or different variables. In the special case where $\alpha = \alpha_0$ and $\delta_{ik}(\boldsymbol{\phi}_i; \boldsymbol{t}) = x_{ik}$ for a known predictor x_{ik} which could depend on $\boldsymbol{t}$, we have

$$\eta(\boldsymbol{\phi}_i, f; \boldsymbol{t}) = \alpha_0 + \sum_{k=1}^{l} f_k(t^k) x_{ik},$$

which is the varying coefficient model in Hastie and Tibshirani (1993). This is also a functional linear model stated in Ramsay and Silverman (1997).

3 Estimation

We need to estimate $\boldsymbol{\beta}$, f, $\boldsymbol{\theta}$, σ^2, and $\boldsymbol{b}$. In a SNMM, both $\boldsymbol{\phi}$ and f may enter the mean function η nonlinearly. Furthermore, the design points (knot locations) for f may depend on random effects. Therefore the spline estimate of f depends on estimates of the random effects which in turn depend on f in a complicated way. As a consequence, SNMM can not be reduced to any parametric model and care has to be taken for estimation.

For NLMMs, Lindstorm and Bates (1990) proposed a two-step procedure, including a Pseudo Data (**PD**) step and a Linear Mixed Effects (**LME**) step with conditional linearization, to estimate parameters and random effects. In the following, we extend this two-step procedure to fit SNMMs.

We still use **PD** and **LME** to name the two steps in our procedure even though they do not bear the original meanings completely. We propose the following two-step procedure:

PD step. With $\boldsymbol{\theta}$ fixed, f, $\boldsymbol{\beta}$ and $\boldsymbol{b}$ are estimated as minimizers of Penalized log-Henderson Likelihood

$$PHL(\boldsymbol{\beta}, \boldsymbol{b}, f|\boldsymbol{\theta}) = (\boldsymbol{y} - \eta(\boldsymbol{\phi}, f))^T \boldsymbol{\Lambda}^{-1}(\boldsymbol{y} - \eta(\boldsymbol{\phi}, f)) + \boldsymbol{b}^T \tilde{\boldsymbol{D}}\boldsymbol{b} + N\lambda J(f), \tag{8}$$

where the first two terms are negative log Henderson likelihood (Robinson, 1990); the third term, $J(f)$, is a penalty for the function f. The form of $J(f)$ depends on specific applications and prior knowledge. For example, $J(f) = \int (f''(t))^2 dt$ for fitting a cubic spline and $J(f) = \int (Lf)^2 dt$ for fitting a L-spline with a linear differential operator L. See Section 6 for more details. λ is a smoothing parameter which controls the balance between the goodness-of-fit and the penalty $J(f)$. We choose λ using a data-adaptive criterion such as generalized cross validation (GCV), unbiased risk (UBR) or Generalized Maximal Likelihood (GML) (Wahba, 1990) at each iteration in our algorithms. Denote the estimates by $\hat{\boldsymbol{b}}$, $\hat{\boldsymbol{\beta}}$ and $\hat{f}$.

LME step. Let $l(\boldsymbol{y}; \boldsymbol{\beta}, f, \boldsymbol{\theta}) = log \int p(\boldsymbol{y}|\boldsymbol{b})p(\boldsymbol{b})d\boldsymbol{b}$, the log-marginal likelihood, where $p(\boldsymbol{y}|\boldsymbol{b})$ and $p(\boldsymbol{b})$ are density functions of $\boldsymbol{y}$ conditional on $\boldsymbol{b}$ and of $\boldsymbol{b}$ respectively. f, $\boldsymbol{\beta}$, $\boldsymbol{\theta}$ and σ^2 are estimated by minimizing the Penalized log-marginal Likelihood

$$PL(\boldsymbol{\beta}, f, \boldsymbol{\theta}, \sigma^2) = -l(\boldsymbol{y}; \boldsymbol{\beta}, f, \boldsymbol{\theta}) + \frac{N\lambda}{2} J(f). \tag{9}$$

Since the marginal likelihood does not have a closed form, we use the conditional first-order approximation. Expand η with respect to $\boldsymbol{b}$ at $\hat{\boldsymbol{b}}$

$$\boldsymbol{y} - \eta(\boldsymbol{\phi}, f) = \boldsymbol{y} - \eta(\boldsymbol{A\beta} + \boldsymbol{Bb}, f) \approx \boldsymbol{y} - (\eta(\boldsymbol{A\beta} + \boldsymbol{B}\hat{\boldsymbol{b}}, f) + \hat{\boldsymbol{Z}}(\boldsymbol{b} - \hat{\boldsymbol{b}})),$$

where $\hat{\boldsymbol{Z}}_i = \partial\eta/\partial\boldsymbol{b}_i^T|_{\hat{\boldsymbol{b}},\hat{\boldsymbol{\beta}},\hat{f}}$ and $\hat{\boldsymbol{Z}} = diag(\hat{\boldsymbol{Z}}_1, \cdots, \hat{\boldsymbol{Z}}_m)$. Then approximately

$$\boldsymbol{y}|\boldsymbol{b} \stackrel{.}{\sim} N(\eta(\boldsymbol{A\beta} + \boldsymbol{B}\hat{\boldsymbol{b}}, f) - \hat{\boldsymbol{Z}}\hat{\boldsymbol{b}} + \hat{\boldsymbol{Z}}\boldsymbol{b}, \sigma^2\boldsymbol{\Lambda}),$$

and

$$\boldsymbol{y} \stackrel{.}{\sim} N(\eta(\boldsymbol{A\beta} + \boldsymbol{B}\hat{\boldsymbol{b}}, f) - \hat{\boldsymbol{Z}}\hat{\boldsymbol{b}}, \sigma^2\hat{\boldsymbol{V}}), \tag{10}$$

where $\hat{\boldsymbol{V}} = \boldsymbol{\Lambda} + \hat{\boldsymbol{Z}}\tilde{\boldsymbol{D}}\hat{\boldsymbol{Z}}^T$. We will estimate $\boldsymbol{\theta}$, f, $\boldsymbol{\beta}$ and σ^2 by minimizing the Penalized Approximate log-marginal Likelihood

$$\begin{aligned} & PAL(\boldsymbol{\beta}, f, \boldsymbol{\theta}, \sigma^2) \\ = \ & log|\sigma^2\hat{\boldsymbol{V}}| + \sigma^{-2}(\boldsymbol{y} - \eta(\boldsymbol{A\beta} + \boldsymbol{B}\hat{\boldsymbol{b}}, f) + \hat{\boldsymbol{Z}}\hat{\boldsymbol{b}})^T\hat{\boldsymbol{V}}^{-1}(\boldsymbol{y} - \eta(\boldsymbol{A\beta} + \boldsymbol{B}\hat{\boldsymbol{b}}, f) + \hat{\boldsymbol{Z}}\hat{\boldsymbol{b}}) + N\tilde{\lambda}J(f). \end{aligned} \tag{11}$$

We can also define the penalized approximate restricted likelihood similar to Lindstorm and Bates (1990).

As in the case of NLMMs (Lindstorm and Bates, 1990), it is straightforward to show (proof omitted here) that for $\tilde{\lambda} = \lambda/(\sigma^{ML})^2$ the two steps (8) and (11) give the same estimates of $\boldsymbol{\beta}$ and f. That is $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}^{ML}) = \boldsymbol{\beta}^{ML}$, and $\hat{f}(\boldsymbol{\theta}^{ML}) = f^{ML}$. Notice that it is not difficult to implement other approximations such as Laplace approximations for the marginal likelihood.

The approximate MLE and REstricted Maximum Likelihood (REML) estimate of σ^2 underestimate due to the loss of degree of freedoms for estimating f. One simple correction for the REML estimate is

$$\hat{\sigma}^2_{CREML} = \hat{\sigma}^2_{REML}\frac{N-p}{N-p-df(f)},$$

where p is the dimension of $\boldsymbol{\beta}$ and $df(f)$ is a properly defined degree of freedoms for f. We choose $df(f)$ as the trace of the hat matrix evaluated at convergence.

4 Implementation

In this section we implement our two-step procedure using the Splus program `nlme` (Pinherio and Bates, 1999) and Fortran program `RKPACK` (Gu, 1989). `nlme` fits NLMMs using the conditional linearization method. `RKPACK` fits the general smoothing spline model described in Wahba

(1990). Since η may depend on f nonlinearly, we first introduce a new iterative algorithm for smoothing spline estimates with data as a noisy map of nonlinear functionals, and then present two similar algorithms for SNMMs.

4.1 Spline Smoothing with Nonlinear Functionals

Consider the model

$$y_i = \mathcal{N}_i g + \epsilon_i, \quad i = 1, \cdots, n, \tag{12}$$

where $\mathcal{N}_i$ is a known nonlinear functional of g, g is an unknown function and belongs to a RKHS $\mathcal{H}$, and $\epsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$.

Nonlinear functionals often appear in physics and biology, for example in remote sensing experiments where functions of interest are observed indirectly (O'Sullivan and Wahba, 1985). Model (12) can also be used to fit spline models with constraints by transforming these models into unconstrained nonlinear models. For example, if the unknown function f is monotone, we can represent it as $f(t) = c_1 + c_2 \int_0^t exp(\int_o^s g(u)du)ds$ (Ramsay, 1998), which is a nonlinear functional of an unconstrained function g.

We estimate g by solving

$$\min_{g \in \mathcal{H}} \left\{ \sum_{i=1}^n (y_i - \mathcal{N}_i g)^2 + n\lambda J(g) \right\}.$$

Assume that the Fréchet differential of $\mathcal{N}_i$ with respect to g exists and is bounded at the current estimate $\hat{g}$. Define $\mathcal{L}_i = \partial \mathcal{N}_i / \partial g|_{g=\hat{g}}$. We approximate $\mathcal{N}_i g$ by its first order Taylor expansion at the current estimate $\hat{g}$ (Lusternik and Sobolev, 1974): $\mathcal{N}_i g \approx \mathcal{N}_i \hat{g} + \mathcal{L}_i (g - \hat{g})$. Then (12) can be approximated by

$$\tilde{y}_i = \mathcal{L}_i g + \epsilon_i, \ i = 1, \cdots, n, \tag{13}$$

where $\tilde{y}_i = y_i - \mathcal{N}_i \hat{g} + \mathcal{L}_i \hat{g}$. Since $\mathcal{L}_i$ is a linear and bounded functional (Lusternik and Sobolev, 1974), the spline estimate of the model (13) has an explicit solution and can be calculated using existing software such as `RKPACK`. This suggests the following algorithm to estimate g:

1. Choose an initial estimate g_0;
2. Denote the current estimate by $\hat{g}$. Calculate $\tilde{y}_i = y_i - \mathcal{N}_i\hat{g} + \mathcal{L}_i\hat{g}$, $i = 1, \cdots, n$;
3. Fit model (13) using a program such as `RKPACK` to derive a new estimate $\hat{g}$;
4. Repeat 2 and 3 until convergence.

This algorithm could be seen as an extension of the Gauss-Newton method to infinite dimensional spaces. Simulations indicate that the algorithm works well and converges quickly for commonly used nonlinear functionals. O'Sullivan and Wahba (1985) proposed an algorithm by first approximating g using a finite collection of basis functions and then estimating coefficients by minimizing a parametric penalized least square criterion. It is unclear what bases should be chosen and how many terms should be included, which could be difficult for multivariate functions. Our algorithm avoids this problem and is easy to implement with the existing softwares for smoothing splines.

4.2 The Algorithms for SNMMs

We now describe our algorithms for fitting SNMMs. Since we can estimate f at either the **PD** step or the **LME** step, we have the following two similar algorithms.

Algorithm 1: estimate f at **LME** step:

1. choose initial values $\boldsymbol{b}^0$, $\boldsymbol{\beta}^0$, $\boldsymbol{\theta}^0$ and f^0;
2. `nlme` step: given current estimates f^k, $\boldsymbol{b}^k$, $\boldsymbol{\beta}^k$, $\boldsymbol{\theta}^k$ and fixing f as f^k, we have a NLMM. We can estimate $\boldsymbol{\beta}$, $\boldsymbol{b}$, $\boldsymbol{\theta}$, σ^2, $\hat{\boldsymbol{Z}}$ and $\hat{\boldsymbol{V}}$ by calling `nlme` with $\boldsymbol{b}^k$, $\boldsymbol{\beta}^k$ and $\boldsymbol{\theta}^k$ as initial values. Denote the new estimates as $\boldsymbol{\beta}^{k+1}$, $\boldsymbol{b}^{k+1}$, $\boldsymbol{\theta}^{k+1}$, $(\sigma^2)^{k+1}$, $\hat{\boldsymbol{Z}}^{k+1}$ and $\hat{\boldsymbol{V}}^{k+1}$. Note that $\hat{\boldsymbol{Z}}^{k+1}$ and $\hat{\boldsymbol{V}}^{k+1}$ are evaluated at f^k, $\boldsymbol{b}^{k+1}$, $\boldsymbol{\beta}^{k+1}$ and $\boldsymbol{\theta}^{k+1}$. This step corresponds to the **PD** step and part of **LME** step;
3. `RKPACK` step: given $\boldsymbol{b}^{k+1}$, $\boldsymbol{\beta}^{k+1}$, $\boldsymbol{\theta}^{k+1}$, $\hat{\boldsymbol{Z}}^{k+1}$ and $\hat{\boldsymbol{V}}^{k+1}$,

(a) if η is linear in f: $\eta(\boldsymbol{\phi}_i, f; t) = \alpha(\boldsymbol{\phi}_i; t) + \delta(\boldsymbol{\phi}_i; t) f(\gamma(\boldsymbol{\phi}_i; t))$, then we have:

$$\tilde{\boldsymbol{y}}_i = \tilde{\boldsymbol{\Delta}}_i f(\boldsymbol{\gamma}_i^{k+1}) + \tilde{\boldsymbol{e}}_i, \tag{14}$$

where $\boldsymbol{\phi}_i^{k+1} = \boldsymbol{A}_i \boldsymbol{\beta}^{k+1} + \boldsymbol{B}_i \boldsymbol{b}_i^{k+1}$, $\boldsymbol{\alpha}_i^{k+1} = (\alpha(\boldsymbol{\phi}_i^{k+1}; t_{i1}), \cdots, \alpha(\boldsymbol{\phi}_i^{k+1}; t_{in_i}))^T$, $\tilde{\boldsymbol{y}}_i = (\hat{\boldsymbol{V}}_i^{k+1})^{-1/2}(\boldsymbol{y}_i + \hat{\boldsymbol{Z}}_i^{k+1} \boldsymbol{b}_i^{k+1} - \boldsymbol{\alpha}_i^{k+1})$, $\boldsymbol{\gamma}_i^{k+1} = (\gamma(\boldsymbol{\phi}_i^{k+1}; t_{i1}), \cdots, \gamma(\boldsymbol{\phi}_i^{k+1}; t_{in_i}))^T$, $f(\boldsymbol{\gamma}_i^{k+1}) = (f(\gamma(\boldsymbol{\phi}_i^{k+1}; t_{i1})), \cdots, f(\gamma(\boldsymbol{\phi}_i^{k+1}; t_{in_i})))^T$, $\tilde{\boldsymbol{\Delta}}_i = (\hat{\boldsymbol{V}}_i^{k+1})^{-1/2} diag(\delta(\boldsymbol{\phi}_i^{k+1}; t_{i1}), \cdots, \delta(\boldsymbol{\phi}_i^{k+1}; t_{in_i}))$, and $\tilde{e}_{ij} \overset{iid}{\sim} N(0, \sigma^2)$. Denote $\mathcal{H}$ as the model space for f and suppose that $\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2$, where $\mathcal{H}_1 = span\{\phi_1(t), \cdots, \phi_M(t)\}$ and $\mathcal{H}_2$ is a RKHS with reproducing kernel $R_2(s,t)$. We estimate f in (14) by minimizing

$$\sum_{i=1}^{m} (\tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{\Delta}}_i f(\boldsymbol{\gamma}_i^{k+1}))^T (\tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{\Delta}}_i f(\boldsymbol{\gamma}_i^{k+1})) + N\lambda J(f). \tag{15}$$

The solution can be represented by (Wahba, 1990)

$$f(t) = \sum_{i=1}^{M} d_i \phi_i(t) + \sum_{i=1}^{N} c_i R_2(\tilde{t}_i, t),$$

where $\tilde{\boldsymbol{t}} \triangleq (\tilde{t}_1, \cdots, \tilde{t}_N) = (\boldsymbol{\gamma}_1^T, \cdots, \boldsymbol{\gamma}_m^T)^T$. We need to solve $\boldsymbol{c} = (c_1, \cdots, c_N)^T$ and $\boldsymbol{d} = (d_1, \cdots, d_M)^T$. Let $\tilde{\boldsymbol{\Delta}} = diag(\tilde{\boldsymbol{\Delta}}_1, \cdots, \tilde{\boldsymbol{\Delta}}_m)$, $\boldsymbol{T} = \{\phi_v(\tilde{t}_i)\}_{i=1}^{N}{}_{v=1}^{M}$, $\boldsymbol{\Sigma} = \{R_2(\tilde{t}_i, \tilde{t}_j)\}_{i=1}^{N}{}_{j=1}^{N}$, $\tilde{\boldsymbol{T}} = \tilde{\boldsymbol{\Delta}} \boldsymbol{T}$ and $\tilde{\boldsymbol{\Sigma}} = \tilde{\boldsymbol{\Delta}} \boldsymbol{\Sigma} \tilde{\boldsymbol{\Delta}}^T$. Then (15) becomes

$$(\tilde{\boldsymbol{y}} - (\tilde{\boldsymbol{T}} \boldsymbol{d} + \tilde{\boldsymbol{\Sigma}} \tilde{\boldsymbol{c}}))^T (\tilde{\boldsymbol{y}} - (\tilde{\boldsymbol{T}} \boldsymbol{d} + \tilde{\boldsymbol{\Sigma}} \tilde{\boldsymbol{c}})) + N\lambda \tilde{\boldsymbol{c}} \tilde{\boldsymbol{\Sigma}} \tilde{\boldsymbol{c}}, \tag{16}$$

where $\tilde{\boldsymbol{c}} = \tilde{\boldsymbol{\Delta}}^{-T} \boldsymbol{c}$. We can solve $\boldsymbol{d}$ and $\tilde{\boldsymbol{c}}$ from (16) by calling `RKPACK`, and then let $\boldsymbol{c} = \tilde{\boldsymbol{\Delta}}^T \tilde{\boldsymbol{c}}$ to form f^{k+1}.

(b) if η is nonlinear in f, after approximating the marginal likelihood by (10), we further expand η with respect to f at f^k

$$\boldsymbol{y}_i = \boldsymbol{\eta}_i(\boldsymbol{A}_i \boldsymbol{\beta}^{k+1} + \boldsymbol{B}_i \boldsymbol{b}_i^{k+1}, f^k) + \mathcal{L}_i(f - f^k) - \hat{\boldsymbol{Z}}_i^{k+1} \boldsymbol{b}_i^{k+1} + \boldsymbol{e}_i^*,$$

where $\mathcal{L}_i = \partial \boldsymbol{\eta}_i / \partial f|_{f^k, \boldsymbol{b}_i^{k+1}, \boldsymbol{\beta}^{k+1}}$ and $Cov(\boldsymbol{e}_i^*) = \sigma^2 \hat{\boldsymbol{V}}_i^{k+1}$. Then we transform the data and call `RKPACK` as in (a). After several iterations or at convergence, we get f^{k+1}.

This step corresponds to part of the **LME** step;

4. repeat 2 and 3 until convergence.

Algorithm 2: estimate f at **PD** step:

1. choose initial values $\boldsymbol{b}^0$, $\boldsymbol{\beta}^0$, $\boldsymbol{\theta}^0$ and f^0;

2. `nlme` step: given current estimates f^k, $\boldsymbol{b}^k$, $\boldsymbol{\beta}^k$ and $\boldsymbol{\theta}^k$, call `nlme` with f fixed as f^k and $\boldsymbol{b}^k$, $\boldsymbol{\beta}^k$ and $\boldsymbol{\theta}^k$ as initial values. Denote the new estimates as $\boldsymbol{b}^{k+1}$, $\boldsymbol{\beta}^{k+1}$, $\boldsymbol{\theta}^{k+1}$ and $(\sigma^2)^{k+1}$. This step corresponds to **LME** step and part of the **PD** step;

3. `RKPACK` step: given $\boldsymbol{b}^{k+1}$, $\boldsymbol{\beta}^{k+1}$ and $\boldsymbol{\theta}^{k+1}$,

   (a) if η is linear in f: $\eta(\boldsymbol{\phi}_i, f; t) = \alpha(\boldsymbol{\phi}_i; t) + \delta(\boldsymbol{\phi}_i; t) f(\gamma(\boldsymbol{\phi}_i; t))$, then call `RKPACK` to fit the transformed data: $\tilde{\boldsymbol{y}}_i = \tilde{\boldsymbol{\Delta}}_i f(\boldsymbol{\gamma}_i^{k+1}) + \tilde{\boldsymbol{e}}_i$ as in Step 3(a) of Algorithm 1, where $\tilde{\boldsymbol{y}}_i = (\hat{\boldsymbol{\Lambda}}_i^{k+1})^{-1/2}(\boldsymbol{y}_i - \boldsymbol{\alpha}_i^{k+1})$, $\tilde{\boldsymbol{\Delta}}_i = (\hat{\boldsymbol{\Lambda}}_i^{k+1})^{-1/2} diag(\delta(\boldsymbol{\phi}_i^{k+1}; t_{i1}), \cdots, \delta(\boldsymbol{\phi}_i^{k+1}; t_{in_i}))$, $\boldsymbol{\alpha}_i^{k+1}$, $\boldsymbol{\phi}_i^{k+1}$, $\boldsymbol{\gamma}_i^{k+1}$, $f(\boldsymbol{\gamma}_i^{k+1})$ and $\tilde{e}_{ij}$ are the same as in 3(a) of Algorithm 1;

   (b) if η is nonlinear in f, we expand η at f^k

$$\boldsymbol{y}_i = \boldsymbol{\eta}_i(\boldsymbol{A}_i \boldsymbol{\beta}^{k+1} + \boldsymbol{B}_i \boldsymbol{b}_i^{k+1}, f^k) + \mathcal{L}_i(f - f^k) + \boldsymbol{e}_i^*,$$

   where $\mathcal{L}_i = \partial \boldsymbol{\eta}_i / \partial f|_{f^k, \boldsymbol{b}_i^{k+1}, \boldsymbol{\beta}^{k+1}}$ and $Cov(\boldsymbol{e}_i^*) = \sigma^2 \hat{\boldsymbol{\Lambda}}_i^{k+1}$. Then we transform the data and call `RKPACK` as in (a). After several iterations or at convergence, we get f^{k+1}.

   This step corresponds to part of the **PD** step;

4. repeat 2 and 3 until convergence.

In Algorithm 1, f is to be estimated by maximizing the approximate penalized marginal likelihood, i.e. to minimize (11) jointly with respect to $\boldsymbol{\beta}$, $\boldsymbol{\theta}$ and f. So at the `RKPACK` step, we need an inner iteration to estimate f and $\boldsymbol{\beta}$, $\boldsymbol{\theta}$ alternately for several iterations or until convergence. Our simulations indicate that iterating once, i.e. no inner iterations, is usually enough. Similarly in Algorithm 2, f is estimated from minimizing (8) jointly with $\boldsymbol{b}$ and $\boldsymbol{\beta}$, and the inner iterations are not necessary. At each iteration in both algorithms, the `nlme` step should be run until convergence. However, our simulations suggest that several iterations, say 5, are enough. Both algorithms perform very well and give similar estimate at convergence in our intensive simulations. In general, Algorithm 1 is faster and Algorithm 2 is more stable.

As in any iterative procedure, good initial values are very important. For SNMMs, the parameter space may have multiple local minima. Therefore reasonable initial values are essential to both algorithms. Estimates of a SEMOR model often provide good initial values. If a nonlinear function form approximates f well, we can fit a NLMM first and use its estimates as initial values. Multiple fits with different initial values are recommended. Notice that it is trivial to implement our estimation procedure using other related softwares. More details on our algorithms and their performance will be reported elsewhere. We have developed Splus programs for the above algorithms through the interface of Splus with Fortran and C. These programs are available upon request and will be posted to public sites.

5 Inferences

Inferences about parameters such as $\boldsymbol{\beta}$, $\boldsymbol{\theta}$ and $\boldsymbol{b}$ are similar to those in NLMMs. We will focus on inference for the nonparametric function f. For smoothing spline models, Bayesian confidence intervals are often used, and are shown to possess good frequentist properties (Wahba, 1990; Nychka, 1988; Gu, 1992; Gu and Wahba, 1994; Nychka, 1988; Wang and Wahba, 1995). Wang and Wahba (1995) compared the performance of Bayesian confidence interval and bootstrap confidence interval, and concluded that both have similar interpretation as the Average Coverage Property (ACP). In this section we first introduce a Bayesian formulation for a SNMM and then discuss how to construct the Bayesian confidence interval for f.

For a SNMM (4), suppose $f \in \mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2$, where $\mathcal{H}_1$ is spanned by $\phi_1(t), \cdots, \phi_M(t)$, and $\mathcal{H}_2$ is a RKHS with reproducing kernel $R_2(s,t)$. As in the general smoothing spline models

(Wahba, 1990), we specify a prior on f as

$$F(t) = \sum_{i=1}^{M} d_i \phi_i(t) + \tau^{1/2} Z(t), \tag{17}$$

where $d_i \overset{iid}{\sim} N(0, a)$ and $Z(t)$ is a mean zero Gaussian process with $\mathrm{Cov}(Z(t), Z(s)) = R_2(s,t)$. Consider the following Bayesian model with the prior specified in (17)

$$\begin{aligned} &\boldsymbol{y} = \boldsymbol{\eta}(\boldsymbol{\phi}, F) + \boldsymbol{e}, \\ &\boldsymbol{\phi} = \boldsymbol{A\beta} + \boldsymbol{Bb},\ \boldsymbol{b} \sim N(\boldsymbol{0}, \sigma^2 \tilde{\boldsymbol{D}}),\ \boldsymbol{e} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{\Lambda}). \end{aligned}$$

Since there is no closed form for the posterior distribution of F, we expand η with respect to $\boldsymbol{b}$ at the final estimate $\hat{\boldsymbol{b}}$

$$\boldsymbol{y} \approx \boldsymbol{\eta}(\boldsymbol{A\beta} + \boldsymbol{B}\hat{\boldsymbol{b}}, F) + \boldsymbol{Z}(\boldsymbol{b} - \hat{\boldsymbol{b}}) + \boldsymbol{e}, \tag{18}$$

where $\boldsymbol{Z}$ is the first derivative of $\boldsymbol{\eta}$ with respect to $\boldsymbol{b}$ evaluated at $\hat{\boldsymbol{b}}$ and $\hat{F}$. Then we have

$$\boldsymbol{y}|F \mathrel{\dot\sim} N(\boldsymbol{\eta}(\boldsymbol{A\beta} + \boldsymbol{B}\hat{\boldsymbol{b}}, F) - \boldsymbol{Z}\hat{\boldsymbol{b}}, \sigma^2(\boldsymbol{Z}\tilde{\boldsymbol{D}}\boldsymbol{Z}^T + \boldsymbol{\Lambda})). \tag{19}$$

We assume that $F(t)$ has differentiable path almost surely if the derivative of η involves the derivative of F. For cubic splines, this condition is satisfied because $D^2 F = dW$, where $W(\cdot)$ is the Wiener process (Wahba, 1990).

Firstly we assume η is linear in F: $\eta(\boldsymbol{\phi}_i, F; t) = \alpha(\boldsymbol{\phi}_i; t) + \delta(\boldsymbol{\phi}_i; t) F(\gamma(\boldsymbol{\phi}_i; t))$. Let $\boldsymbol{\alpha} = (\alpha(\boldsymbol{A}_1\boldsymbol{\beta} + \boldsymbol{B}_1\hat{\boldsymbol{b}}_1, t_{11}), \cdots, \alpha(\boldsymbol{A}_m\boldsymbol{\beta} + \boldsymbol{B}_m\hat{\boldsymbol{b}}_m, t_{mn_m}))^T$, $\boldsymbol{\Delta} = diag(\delta(\boldsymbol{A}_1\boldsymbol{\beta} + \boldsymbol{B}_1\hat{\boldsymbol{b}}_1, t_{11}), \cdots, \delta(\boldsymbol{A}_m\boldsymbol{\beta} + \boldsymbol{B}_m\hat{\boldsymbol{b}}_m, t_{mn_m}))$, $\tilde{\boldsymbol{t}} = (\gamma(\boldsymbol{A}_1\boldsymbol{\beta} + \boldsymbol{B}_1\hat{\boldsymbol{b}}_1, t_{11}), \cdots, \gamma(\boldsymbol{A}_m\boldsymbol{\beta} + \boldsymbol{B}_m\hat{\boldsymbol{b}}_m, t_{mn_m}))^T$ and $\boldsymbol{F} = (F(\tilde{t}_1), \cdots, F(\tilde{t}_N))^T$. Then $\boldsymbol{\eta}(\boldsymbol{A\beta} + \boldsymbol{B}\hat{\boldsymbol{b}}, F) = \boldsymbol{\alpha} + \boldsymbol{\Delta F}$. For fixed $\boldsymbol{\beta}$, $\boldsymbol{\theta}$ and $\lambda = \sigma^2/(N\tau)$, as in Wahba (1990), we can show that for any t,

$$\begin{aligned} &\lim\nolimits_{a\to\infty}\ \mathrm{E}(F(t)|\boldsymbol{y}) = \hat{f}(t), \\ &\lim\nolimits_{a\to\infty}\ \mathrm{Var}(\boldsymbol{F}|\boldsymbol{y}) = \sigma^2 \boldsymbol{H}, \end{aligned}$$

where $\hat{f}$ minimizes

$$(\boldsymbol{y} - \eta(\boldsymbol{A\beta} + \boldsymbol{B}\hat{\boldsymbol{b}}, f) + Z\hat{\boldsymbol{b}})^T (\boldsymbol{Z}\tilde{\boldsymbol{D}}\boldsymbol{Z}^T + \boldsymbol{\Lambda})^{-1} (\boldsymbol{y} - \eta(\boldsymbol{A\beta} + \boldsymbol{B}\hat{\boldsymbol{b}}, f) + Z\hat{\boldsymbol{b}}) + N\lambda J(f),$$

and $\boldsymbol{H}$ is the hat matrix. Notice that both $\hat{f}$ and $\boldsymbol{H}$ depend on $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$.

For the Algorithm 1, the Bayesian confidence interval can be calculated at convergence by replacing $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ with their estimates. We need to take into account the loss of degrees of freedom incurred for estimating $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$. One simple way is to use the dimensions of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ as the loss of degrees of freedom, and then count them when calculating confidence interval for f. Another way is to use the generalized degrees of freedom in Ye (1998). From our small scale of simulations, the two approaches give similar results. For the Algorithm 2, the Bayesian confidence interval at convergence is too liberal if variances of $\boldsymbol{b}$ are large since the first component of the variance in (19) is excluded. One solution is to fit `RKPACK` again to maximize the marginal likelihood after convergence as the `RKPACK` step in Algorithm 1.

If η is nonlinear in F, we can first expand η with respect to F to the first order as in Section 4.1 and then apply the above method.

Due to the complexity of the model and algorithm, simple correction sometimes does not work very well. Our intensive simulations (see Section 7) indicate the (corrected) Bayesian confidence interval can provide a quick rough check. For precise confidence intervals, a bootstrap method is recommended. We will explain the bootstrap confidence intervals in Section 6.2.

6 Applications

6.1 Canadian Temperature Data

The dataset, downloaded from the website http://www.psych.mcgill.ca/faculty/ramsay/fda.html, includes mean monthly temperaures at 35 Canadian weather stations. The left panel in Figure 1 shows the plots of mean temperature from 4 selected weather stations.

Ramsay and Silverman (1997) and Ramsay and Li (1998) used this dataset to illustrate usefulness of Functional Data Analysis (FDA). Among others, they addressed the following two questions: how to measure the departure from a sinusoidal model and how to register the curves

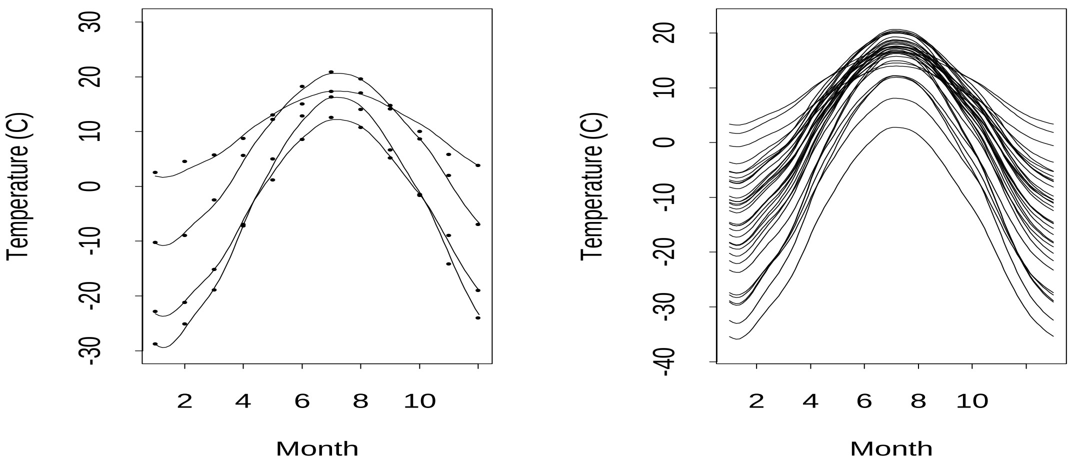

Figure 1: Left: plot of temperatures for 4 selected stations. Dots are observations. Lines are fits by SNMM. Right: plot of fitted curves after alignment for all 35 stations.

for further analyses. Registration of curves is also called dynamic time warping in engineering, which is often a pre-processing step in FDA (Ramsay and Li, (1998)). In the following, we will approach these questions with methods proposed in this paper. We assume that mean temperatures over time at all weather stations share a common shape function. Variation between stations are modeled by vertical shift and scale and horizontal shift parameters. That is, we assume

$$y_{ij} = \mu + b_{1i} + e^{b_{2i}} f(t_{ij} - \frac{e^{b_{3i}}}{1 + e^{b_{3i}}}) + e_{ij}, \quad i = 1, \cdots, 35, \quad j = 1, \cdots, 12, \tag{20}$$

where y_{ij} is the temperature of the jth month at station i; t_{ij} is the time point scaled to $[0, 1]$; $\boldsymbol{b}_i = (b_{1i}, b_{2i}, b_{3i})^T \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{D})$, where $\boldsymbol{D}$ is an unstructured covariance matrix; e_{ij}'s are error terms modeled by an AR(1) within-subject correlation structure with lag 1 autocorrelation coefficient ρ and variance σ^2; μ, fixed, is the annual mean temperature for all stations; b_{1i} is the vertical shift; $e^{b_{2i}}$ is the amplitude; and $e^{b_{3i}}/(1 + e^{b_{3i}})$ is the horizontal shift of station i. Since the mean temperature is a periodic function with period equal to one year, f is periodic with period equal to 1. Thus we assume $f \in W_2(per)$. To make the constant μ and f identifiable, we need the side condition $\int_0^1 f = 0$. This is equivalent to assuming that $f \in W_2^0(per) = W_2(per) \ominus span\{1\}$, where $span\{1\}$ represents all constant functions.

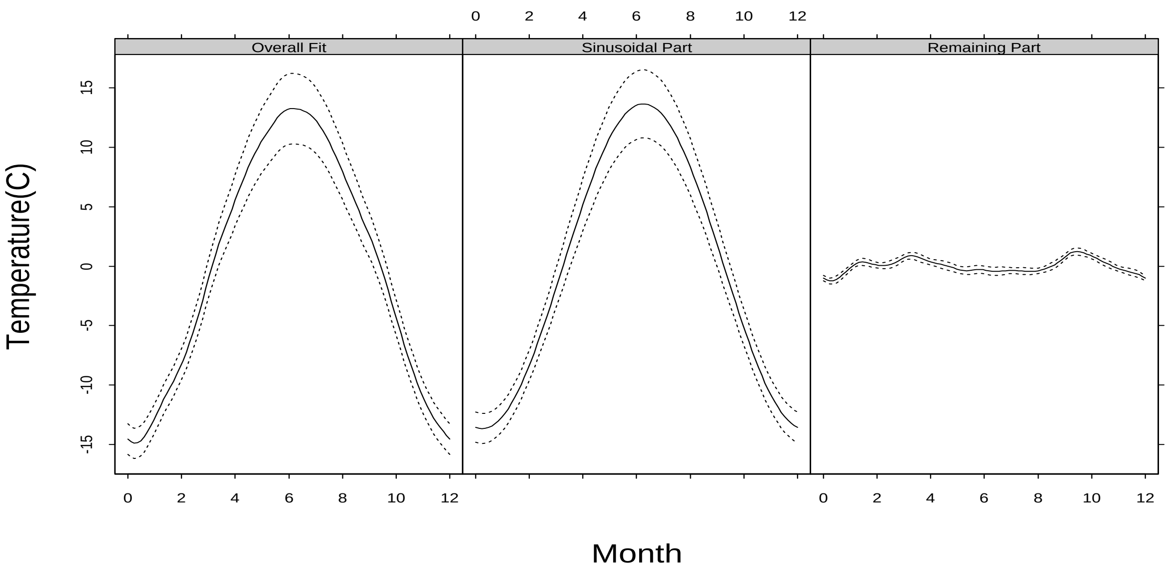

Figure 2: The overall fit of the common shape function (left), and its projection onto $\mathcal{H}_1$ (center) and $\mathcal{H}_2$ (right). Solid lines are fitted values. Dotted lines are 95% Bayesian confidence limits.

Since the sinusoidal form provides rough approximation to the temperature profile, we use the following L-spline, which puts sine and cosine functions into the null space. Let L be the linear differential operator $L = D^2 + (2\pi)^2$, where $D^2 f$ is the second derivative of f. Let $\mathcal{H}_1 = kerL = span\{sin2\pi t, cos2\pi t\}$. Define a linear operator $\mathcal{B} : W_2^0(per) \rightarrow \mathcal{R}^2$ such that $\mathcal{B}f = (\int_0^1 f(t)sin2\pi tdt, \int_0^1 f(t)cos2\pi tdt)^T$. Let $\mathcal{H}_2 = ker\mathcal{B}$. Define inner products on $\mathcal{H}_1, \mathcal{H}_2$, and $W_2^0(per)$ as $< f, g >_1 = (\mathcal{B}f)^T(\mathcal{B}g)$, $< f, g >_2 = \int_0^1 (Lf)(Lg)dt$, and $< f, g > = < f, g >_1 + < f, g >_2$. Then it is easy to check that $W_2^0(per) = \mathcal{H}_1 \oplus \mathcal{H}_2$. The reproducing kernel of $\mathcal{H}_2$

is

$$R_2(s,t) = \sum_{v=2}^{\infty} \frac{2}{(2\pi)^4(1-v^2)^2} cos(2\pi v(s-t)).$$

Let P_i be the projection operator from $W_2^0(per)$ onto $\mathcal{H}_i$, $i = 1, 2$. The hypothesis that f is a sinusoidal function is equivalent to H_0: $P_2 f = 0$. To test H_0, we fit model (20) using (8) with the penalty $J(f) = \int_0^1 (Lf)^2 dt = ||P_2 f||_2^2$, which measures the *deviation* of f from the sinusoidal function. See Wang and Brown (1996) for more details on this decomposition.

The resulting fits are plotted in the left panel of Figure 1 for the four selected stations. Fits for other stations are similar. The estimated correlation coefficient is $\hat{\rho} = 0.698$ with approximate 95% confidence interval $(0.03, 1)$. Thus there is evidence that serial correlation does exist in the data. Figure 2 shows the overall fit of the common curve and its projections on subspaces $\mathcal{H}_1$ (sinusoidal part) and $\mathcal{H}_2$ (remaining part) together with their 95% Bayesian confidence intervals. The small departure from sinusoidal form is statistically significant, especially in Spring and Fall. A formal test can be done by using a bootstrap method (Wang and Brown, 1996; also Section 6.2). Our conclusion here is comparable to Ramsay and Silverman (1997), but their approach is on individual stations and does not provide a formal way to test. By removing the horizontal shift in each curve, we can easily align all curves. Therefore our method can be used for curve registration. The key differences between our approach and that in Ramsay and Silverman (1997) are that we assume a common function for all stations and use all observations to estimate this nonparametric function. Therefore, our approach is appropriate for situations where the number of observations for each individual is small and where it is reasonable to assume a common shape for all individuals. The right panel in Figure 1 displays the fitted curves for all 35 stations after aligning them together by removing shifts. These aligned data can be used for further analysis.

6.2 CO_2 Uptake Data

This dataset comes from a study of cold tolerance of a C_4 grass species, *Echinochloa crus-galli*, described in Potvin, Lechowicz and Tardif (1990) and Pinherio and Bates (1999). A total of twelve four-week-old plants, six from Quebec and six from Mississippi, were divided into two groups: control plants that were kept at 26^{o}C and chilled plants that were subject to 14 hours of chilling at 7^{o}C. After 10 hours of recovery at 20^{o}C, CO_2 uptake rates (in $\mu mol/m^2 s$) were measured for each plant at seven concentrations of ambient CO_2 in increasing, consecutive order. Plots of observations are shown in Figure 3 as circles. The objective of the experiment was to evaluate the effect of plant type and chilling on the CO_2 uptake.

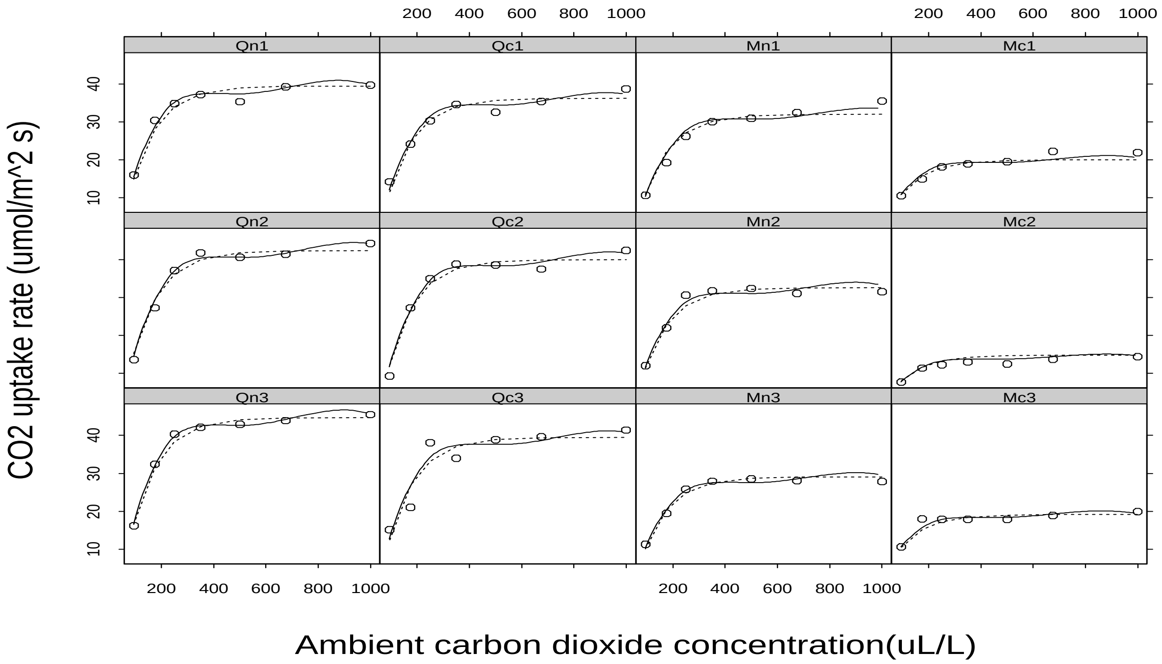

Figure 3: Plots of the data and fitted curves. Circles are observations. Solid lines represent SNMM fits. Dotted lines represent NLMM fits. In the strip name for each plot, “Q” indicates Quebec, “M” Mississippi, “c” chilling, “n” control, and “1”, “2”, “3” the replicate number.

Pinherio and Bates (1999) gave detailed analyses of this dataset based on NLMMs using their software `nlme`. They reached the following model:

$$
\begin{aligned}
&U_{ij} = \psi_{1i}\{1 - e^{-e^{\phi_{2i}(C_j - \phi_{3i})}}\} + \epsilon_{ij}, \quad i = 1, \cdots, 12, \quad j = 1, \cdots, 7,\\
&\psi_{1i} = \beta_{11} + \beta_{12} Type + \beta_{13} Treatment + \beta_{14} Treatment : Type + b_{1i},\\
&\phi_{2i} = \beta_{21} + b_{2i},\\
&\phi_{3i} = \beta_{31} + \beta_{32} Type + \beta_{33} Treatment + \beta_{34} Treatment : Type + b_{3i},
\end{aligned}
\tag{21}
$$

where U_{ij} denotes the CO_2 uptake rate of plant i at the jth CO_2 ambient concentration; *Type* equals 0 for plants from Quebec and 1 for plants from Mississippi, *Treatment* equals 0 for chilled plants and 1 for control plants; ψ_{1i}, $e^{\phi_{2i}}$ and ϕ_{3i} denote respectively the asymptotic uptake rate, the uptake growth rate, and the maximum ambient CO_2 concentration at which no uptake is verified for plant i; C_j denotes the jth ambient CO_2 level; $\boldsymbol{b}_i = (b_{1i},\, b_{2i},\, b_{3i}) \overset{iid}{\sim} N(\boldsymbol{0}, \sigma^2 \boldsymbol{D})$ and $\boldsymbol{D}$ is diagonal; $\epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$. Fits of model (21) are plotted in Figure 3 as dotted lines. Based on Model (21), one may concluded that CO2 uptake is higher for plants from Quebec and that chilling, in general, results in lower uptake, and its effect on Mississippi plants is much bigger than that on Quebec plants. These conclusions are comparable to the results in Potvin et al (1990).

We aim to use this dataset to demonstrate how to build SNMMs with covariates, and how to test if a NLMM is appropriate. Since $\psi_{1i} > 0$, we reparemeterize (21) by $\psi_{1i} = e^{\phi_{1i}}$. As an extension of (21), we fit the following SNMM

$$
\begin{aligned}
&U_{ij} = e^{\phi_{1i}} f(e^{\phi_{2i}}(C_j - \phi_{3i})) + \epsilon_{ij},\\
&\phi_{1i} = \beta_{12} Type + \beta_{13} Treatment + \beta_{14} Treatment : Type + b_{1i},\\
&\phi_{2i} = \beta_{21} + b_{2i},\\
&\phi_{3i} = \beta_{31} + \beta_{32} Type + \beta_{33} Treatment + \beta_{34} Treatment : Type + b_{3i},
\end{aligned}
\tag{22}
$$

where $f \in W_2([0, T])$ for some fixed $T > 0$ and the second stage model is paralleled to (21). In order to test the hypothesis that the parametric model (21) is appropriate, we construct the following L-spline to model f. The hypothesis is H_0: $f \in span\{1 - e^{-t}\}$. Let $L = D + D^2$, where

D^j denotes the jth derivative operator. The kernel space of L is $\mathcal{H}_1 = span\{1, e^{-t}\}$. Define a linear operator $\mathcal{B} : W_2[0,T] \to \mathcal{R}^2$ such that $\mathcal{B}f = (f(0), f'(0))$. Let $\mathcal{H}_2 = ker\mathcal{B}$. Define inner products on $\mathcal{H}_1, \mathcal{H}_2$, and $W_2[0,T]$ as $< f, g >_1 = (\mathcal{B}f)^T(\mathcal{B}g)$, $< f, g >_2 = \int_0^T (Lf)(Lg)dt$ and $< f, g > = < f, g >_1 + < f, g >_2$. Then it is easy to check that $W_2[0,T] = \mathcal{H}_1 \oplus \mathcal{H}_2$. The reproducing kernel of $\mathcal{H}_2$ is

$$R(s,t) = \min(s,\ t) + e^{-t} + e^{-s} - e^{\min(s,\ t)-s} - e^{\min(s,\ t)-t} - \frac{e^{-s-t}}{2} + \frac{e^{2\min(s,\ t)-s-t}}{2}.$$

Note that β_{11} in (21) is excluded from (22) to make f free on the vertical scale. We need the side conditions that $f(0) = 0$ and $f(t) \neq 0$ for $t \neq 0$ to separate β_{31} from f. The first condition reduces $\mathcal{H}_1$ to $\mathcal{H}_1 = span\{1 - e^{-t}\}$ and is satisfied by all functions in $\mathcal{H}_2$. We do not enforce the second condition because it is satisfied by all reasonable estimates. Thus the null space for f becomes $\mathcal{H}_1^0 = span\{1 - e^{-t}\}$ and the model space is $\mathcal{H}_1^0 \oplus \mathcal{H}_2$. The penalty is chosen as $J(f) = \int (Lf)^2$.

With the initial values chosen from the NLMM fit, our program converged after 8 iterations. We obtained similar estimates with other initial values. We also fitted models with the unstructured covariance matrix for random effects, AR(1) within-subject correlations and covariate effects on ϕ_2. None of these models improve fits significantly. The estimates are comparable to the nonlinear fits and conclusion is similar to that based on (21). Figure 3 contains the fitted values from model (22) as solid lines. It is clear that SNMM and NLMM fits are close to each other. To test if the parametric NLMM (21) is appropriate, we look at the projection of f onto $\mathcal{H}_1^0$ (the parametric part) and $\mathcal{H}_2$ (the remaining part). The 95% Bayesian confidence intervals for the common curve f, the parametric part and the remaining part are shown in Figure 4. We also used a parametric bootstrap procedure (Laird and Louis, 1987) to find the confidence interval for f and its projections. $\boldsymbol{b}_i^*$ and ϵ_{ij}^* were generated from the estimated distributions $N(0, \hat{\sigma}^2\hat{\boldsymbol{D}})$ and $N(0, \hat{\sigma}^2)$ respectively. Then $\boldsymbol{y}^*$ were generated from (22). The above SNMM model was fitted to the simulated data. 100 replicates were taken. The normal interval method in Wang and Wahba (1995) is used to construct confidence intervals. The bootstrapping confidence intervals are also shown in Figure 4. The sample size is small in this case. Therefore Bayesian confidence intervals may not be reliable (see Section 7). In spite of some difference between the two sets of confidence intervals, the conclusion remains the same: the remaining part is not significantly different from 0. Therefore, model (21) is appropriate for this dataset.

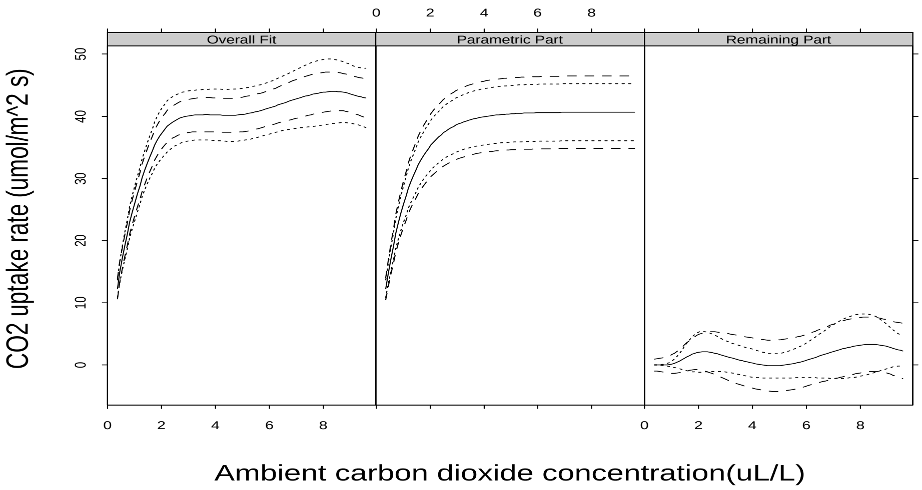

Figure 4: The overall fit of the common curve (left), its projection onto $\mathcal{H}_1^0$ (center) and $\mathcal{H}_2$ (right). Solid lines are fitted values. Dash lines are 95% Bayesian confidence limits. Dotted lines are 95% bootstrap confidence limits.

Heckman and Ramsay (2000) discussed methods to estimate coefficients in the differential operator L for L-spline smoothing. From this example, we can see that SNMM can serve as an alternative to their method. In spirit the SNMM approach is the same as the PL (penalized likelihood) approach in Heckman and Ramsay. Our method estimates the coefficients in L and f iteratively. In addition, we allow these coefficients to depend on covariates.

7 Simulations

In this section a group of simulations is conducted to assess the performance of our estimation procedure and Bayesian confidence intervals. Data are generated from the following model

$$y_{ij} = 1 + b_{1i} + e^{b_{2i}} 2\sin(2\pi(\ j/n - \frac{e^{b_{3i}}}{1+e^{b_{3i}}} \ )) + e_{ij}, \quad i = 1, \cdots, n, \quad j = 1, \cdots, m,$$

where $e_{ij} \overset{iid}{\sim} N(0, \sigma^2)$ and $\boldsymbol{b}_i = (b_{i1}, b_{i2}, b_{i3})^T \overset{iid}{\sim} N(\mathbf{0}, \sigma^2 \boldsymbol{D})$. We use a factorial design with the following options for each parameter: (1) $\sigma = 1, 0.4$; (2) $m = 10$, 20; (3) $n = 10$, 20; (4) $\boldsymbol{D}$ is diagonal with $diag(\boldsymbol{D}) = (0.25, 0.16, 0.04)$ (low) or $(1, 0.25, 0.16)$ (high). We use $W_2^0(per)$ as the model space for f. Other simulations including non-periodic function f and unstructured $\boldsymbol{D}$ were also conducted and conclusions are similar. Algorithm 2 is used in all simulations. Results based on Algorithm 1 are similar. We use GML to select $\hat{\lambda}$ in `RKPACK` step and REML method in `nlme` step in all simulations. 100 replicates are taken for each combination. In all of our simulations, 3 replicates selected very small smoothing parameters (interpolation) and they are excluded from the following presentation. SEMOR fits are used as initial values.

Tables 1 and 2 summarize the performances of estimates of parameters μ, σ and variance components (MSE denotes the mean square error of estimators). The first column of both tables denotes the simulation setting. REML estimation of σ and its correction in Section 3 are both included in Table 1. Clearly the REML estimate underestimates and the correction provides some improvement. The MSE of $\hat{\mu}$ increases with $\boldsymbol{D}$ and σ and decreases with m and n. The MSE of $\hat{\sigma}$ depends more on the total number of observations, while for the estimates of variance components it depends more on the number of subjects.

Figure 5 includes a summary of $\hat{f}$ for a selection of six simulation settings. For each setting we calculate the integrated mean square errors for each $\hat{f}$, and plot $\hat{f}$'s corresponding to the minimum, 1/4 quantile, median, 3/4 quantile and maximum integrated mean square errors. The estimate of f is affected by the total number of observations, $\boldsymbol{D}$ and σ.

Figure 6 shows the boxplot of ACP coverage of Bayesian confidence intervals. Similar to

Table 1: Mean, Bias and MSE for Estimates of μ and σ. $\hat{\sigma}_{REML}$ is the REML estimate. $\hat{\sigma}_{CREML}$ is the corrected estimate proposed in Section 3.

				$\hat{\mu}$			$\hat{\sigma}_{REML}$			$\hat{\sigma}_{CREML}$		
m	n	σ	$\boldsymbol{D}$	Mean	Bias	MSE	Mean	Bias	MSE	Mean	Bias	MSE
10	10	1	high	0.983	-0.017	0.110	0.937	-0.063	0.012	0.975	-0.025	0.009
			low	0.986	-0.014	0.032	0.947	-0.053	0.012	0.982	-0.018	0.009
		0.4	high	0.993	-0.007	0.016	0.370	-0.030	0.002	0.389	-0.011	0.002
			low	0.993	-0.007	0.005	0.371	-0.029	0.002	0.389	-0.011	0.001
	20	1	high	0.994	-0.006	0.110	0.968	-0.032	0.004	0.989	-0.011	0.003
			low	0.987	-0.013	0.032	0.969	-0.031	0.004	0.989	-0.011	0.003
		0.4	high	0.993	-0.007	0.016	0.382	-0.018	0.001	0.392	-0.008	0.001
			low	0.996	-0.004	0.005	0.382	-0.018	0.001	0.393	-0.007	0.001
20	10	1	high	0.987	-0.013	0.069	0.956	-0.044	0.006	0.978	-0.022	0.004
			low	0.988	-0.012	0.022	0.963	-0.037	0.005	0.984	-0.016	0.004
		0.4	high	0.993	-0.007	0.011	0.383	-0.017	0.001	0.393	-0.007	0.001
			low	0.996	-0.004	0.003	0.384	-0.016	0.001	0.394	-0.006	0.001
	20	1	high	0.984	-0.016	0.051	0.989	-0.011	0.002	1.000	0.000	0.002
			low	0.997	-0.003	0.015	0.989	-0.011	0.002	1.001	0.001	0.002
		0.4	high	0.994	-0.006	0.010	0.390	-0.010	0.000	0.396	-0.004	0.000
			low	0.995	-0.005	0.003	0.390	-0.010	0.000	0.396	-0.004	0.000

Table 2: Mean, Bias and MSE for Estimates of Variance Components. D_{11}, D_{22} and D_{33} are diagonal elements of $\boldsymbol{D}$.

				$\hat{D}_{11}$			$\hat{D}_{22}$			$\hat{D}_{33}$		
m	n	σ	$\boldsymbol{D}$	Mean	Bias	MSE	Mean	Bias	MSE	Mean	Bias	MSE
10	10	1	high	1.163	0.163	0.444	0.244	-0.006	0.024	0.163	0.003	0.009
			low	0.276	0.026	0.035	0.159	-0.001	0.008	0.039	-0.001	0.001
		0.4	high	1.177	0.177	0.352	0.292	0.042	0.036	0.164	0.004	0.009
			low	0.297	0.047	0.031	0.188	0.028	0.017	0.041	0.001	0.001
	20	1	high	1.053	0.053	0.244	0.233	-0.017	0.018	0.147	-0.013	0.007
			low	0.259	0.009	0.018	0.133	-0.027	0.005	0.039	-0.001	0.001
		0.4	high	1.099	0.099	0.235	0.257	0.007	0.021	0.154	-0.006	0.008
			low	0.280	0.030	0.019	0.161	0.001	0.009	0.038	-0.002	0.001
20	10	1	high	1.107	0.107	0.175	0.240	-0.010	0.011	0.145	-0.015	0.004
			low	0.281	0.031	0.016	0.152	-0.008	0.005	0.035	-0.005	0.001
		0.4	high	1.111	0.111	0.191	0.256	0.006	0.013	0.153	-0.007	0.004
			low	0.283	0.033	0.017	0.163	0.003	0.006	0.036	-0.004	0.000
	20	1	high	1.045	0.045	0.189	0.230	-0.020	0.006	0.145	-0.015	0.003
			low	0.269	0.019	0.017	0.144	-0.016	0.003	0.036	-0.004	0.000
		0.4	high	1.040	0.040	0.146	0.249	-0.001	0.007	0.155	-0.005	0.004
			low	0.258	0.008	0.012	0.159	-0.001	0.003	0.040	0.000	0.000

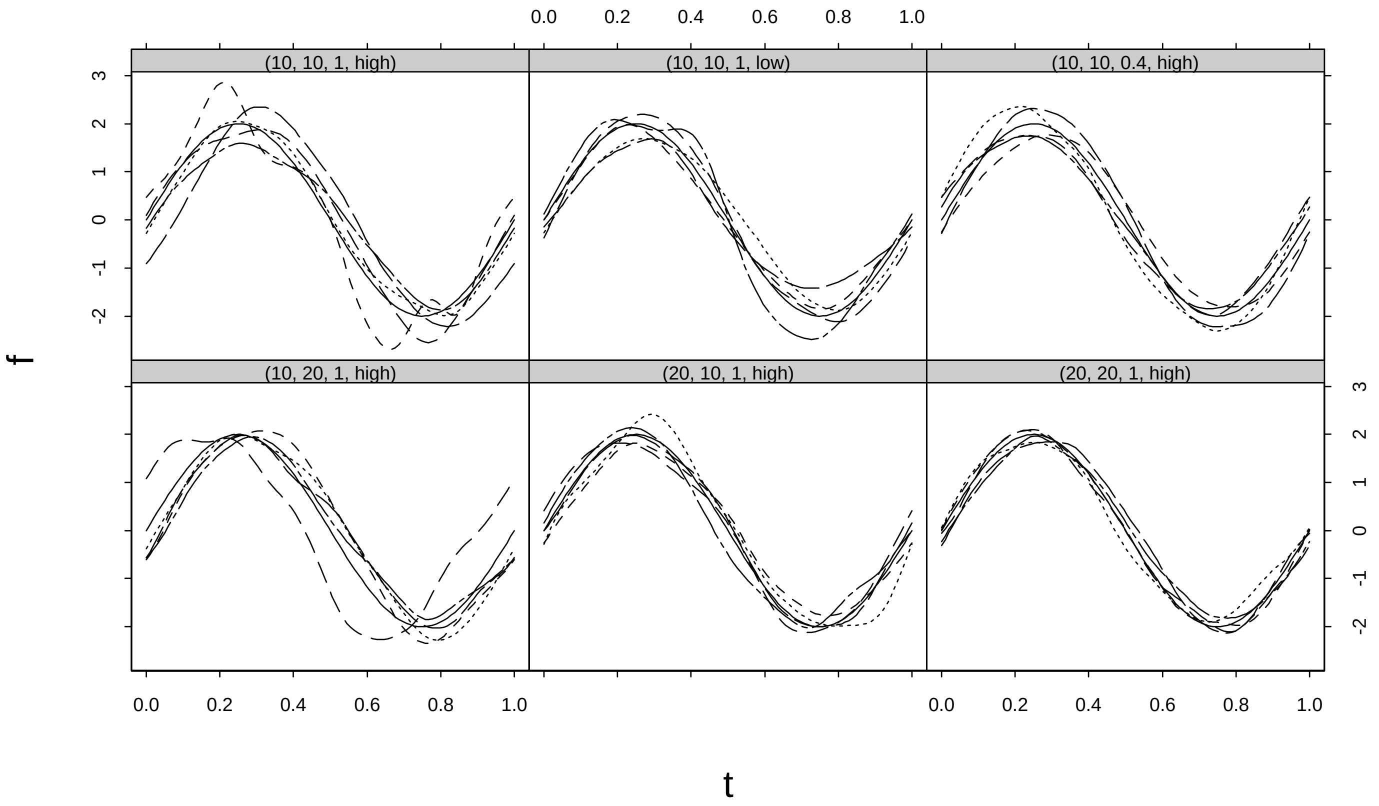

Figure 5: Estimates of f. Solid lines are true functions. Dotted and dashed lines are estimates corresponding to the five quantiles of the integrated MSE.

the point estimate of f, the covariance matrix $\boldsymbol{D}$, σ and the total number of observations all have big effects on the coverage. The ACP decreases with the decrease of sample size, the increase of variance components and the increase of σ^2. The Bayesian confidence interval from simulations is not completely satisfactory, and can only be used as a quick informal check. More research is necessary to improve the performance of Bayesian confidence intervals. For more precise statements, a bootstrap approach is recommended. A nonparametric bootstrap (Laird and Louis, 1987) can be used when the number of subjects is not small. Otherwise, a parametric bootstrap may be prefered. See Laird and Louis (1987) for discussion of bootstrap methods for mixed effects models. We conducted small scale simulations for parametric bootstrap procedures. These simulations indicate that they provide correct coverage.

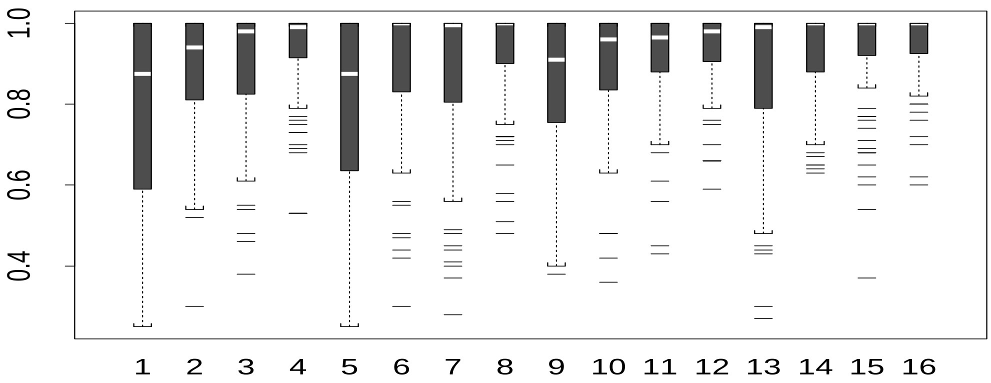

Figure 6: Boxplot of the ACP of the Bayesian confidence interval. The first 4 plots correspond to settings $(10, 10, 1, high)$, $(10, 10, 1, low)$, $(10, 10, 0.4, high)$ and $(10, 10, 0.4, high)$ respectively. The remaining plots are in the same order but with (n, m) replaced by $(10, 20)$, $(20, 10)$ and $(20, 20)$ in every four boxplots.

8 Conclusions

In this paper we have proposed a class of SNMMs for repeated measures data. They are very general and include NLMMs, SEMORs and many existing models as special cases. While SNMMs inherit the advantages of NLMMs and SEMOR models with interpretable parameters and flexible nonparametric functions, they overcome the drawbacks in these models. As the general linear models do for classical regression, ANOVA and ANCOVA models, the general form of SNMMs provides a unified framework for estimation, inference and software implementation. We have presented an estimation procedure based on penalized likelihood and first-order approximation. Data-adaptive methods such as GCV, GML and UBR are used to choose smoothing parameters. This procedure is easy to implement with the existing software, such as `nlme` and `RKPACK`. Approximate Bayesian confidence intervals provide a quick check even though it is not completely satisfactory. Further research on inference and model building techniques for SNMMs are in progress. Bayesian approaches with Markov chain Monte Carlo can be employed to fit SNMMs and construct confidence intervals.

We have shown how to fit a SEMOR with random effects, how to test if a NLMM is appropriate and how to register curves. Other potential applications include correlation of cores in geography (Clark and Thompson, 1979), bioassay in medicine (Guardabasso, Rodbard and Munson, 1987), accelerated degradation test in engineering (Meeker, Escobar and Lu, 1998) and modeling human growth (Stutzle et al, 1980).

References

[1] Altman, N. S. (1996). Semi-parametric methods for longitudinal data analysis. *Proceedings of the Biometrics Section of the American Statistical Association.*

[2] Ansley, C. F., Kohn, R., and Wong, C. (1993). Nonparametric spline regression with prior information. *Biometrika* **80**, 75-88.

[3] Aronszajn, N. (1950). The theory of reproducing kernel. *Transactions of the American Mathematical Society* **68**, 337-404.

[4] Clark, R. M. and Thompson, R. (1979). A new statistical approach to the alignment of time series. *Geophysical Journal of the Royal Astronomical Society* **58**, 593-607.

[5] Davidian, M.and Giltinan, D. (1995). *Nonlinear Models for Repeated Measurement Data.* Chapman & Hall, London.

[6] Gu, C. (1989). RKPACK and its application: Fitting smoothing spline models. *Proceedings of the Statistical Computing Section, American Statistical Association*, 42-51.

[7] Gu, C. (1992). Penalized likelihood regression: a Bayesian analysis. *Statistica Sinica* **2**, 255-264.

[8] Gu, C. and Wahba, G. (1993). Smoothing spline ANOVA With component-wise Bayesian confidence intervals. *Journal of Computational and Graphical Statistics* **55**, 353-368.

[9] Guardabasso, V., Rodbard, D. and Munson, P. J. (1987). A model-free approach to estimation of relative potency in dose-response curve analysis. *American Journal of Physiology* **252**, 357-364.

[10] Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society, Series B* **55**, 757-796.

[11] Heckman, N. and Ramsay, J. O. (2000). Penalized regression with model-based penalties. To appear in *Canadian Journal of Statistics.*

[12] Kneip, A. and Gasser, T. (1988). Convergence and consistency results for self-modeling nonlinear regression. *The Annals of Statistics* **16**, 82-112.

[13] Laird, N. M. and Louis, T. A. (1987). Empirical Bayes confidence intervals Based on Bootstrap Samples. *Journal of the American Statistical Association* **82**, 739-750.

[14] Lawton, W. H., Sylvestre, E.A. and Maggio, M.S. (1972). Self modeling nonlinear regression. *Technometrics* **14**, 513-532.

[15] Lindstorm, M. J. (1995). Self modeling with random shift and scale parameters and a free-knot spline shape function. *Statistics in Medicine* **14**, 2009-2021.

[16] Lindstorm, M. J. and Bates, D. M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics* **46**, 673-687.

[17] Lusternik, L.A. and Sobolev, V. J. (1974). *Elements of Functional Analysis*. Hindustan Publishing Co., New Delhi.

[18] Meeker, W. Q., Escobar, L. A. and Lu, C. J. (1998). Accelerated regradation tests: modeling and analysis. *Technometrics* **40**, 89-99.

[19] Nychka, D. (1988). Bayesian confidence intervals for smoothing splines. *Journal of the American Statistical Association* **83**, 1134-1143.

[20] Pinheiro, J. and Bates, D. (1995). Approximations to the log-likelihood function in the Nonlinear Mixed-effects Model. *Journal of Computational and Graphical Statistics* **4**, 12-35.

[21] Pinheiro, J. and Bates, D. (1999). User's Guide for `nlme`. Available at: http://www.nlme.stat.wisc.edu.

[22] Potvin, C., Lechowicz, M. J. and Tardif, S. (1990). The statistical analysis of ecophysiological response curves obtained from experiments involving repeated measures. *Ecology* **71**, 1389-1400.

[23] Ramsay, J. O. (1998). Estimating smooth monotone function. *Journal of the Royal Statistical Society, Series B* **60**, 365-375.

[24] Ramsay, J. O. and Li, X. (1998). Curve registration. *Journal of the Royal Statistical Society, Series B* **60**, 351-363.

[25] Ramsay, J. O. and Silverman, B. W. (1997). *Functional Data Analysis*. New York:Springer.

[26] Stutzle, W., Gasser, T., Molinari, L., Largo, R. H., Prader, A. and Huber, P.J. (1980). Shape-invariant modeling of human growth. *The Annals of Human Biology* **7**, 507-528.

[27] Tsaneva, N., Topalova, M. and Beraha, R. (1990). Relation of MESOR to circadian amplitude as a criterion of mental load. *Chronobiology: its role in clinical medicine, general biology, and agriculture*, D. K. Hayes, J. E. Pauly and R. J. Reiter Eds, 99-103.

[28] Wahba, G. (1990). *Spline Models for Observational Data*. SIAM, CBMS-NSF Regional Conference Series in Applied Mathematics, V59, Philadelphia.

[29] Wang, Y. (1998). Mixed-effects smoothing spline ANOVA. *Journal of the Royal Statistical Society, Series B* **60**, 159-174.

[30] Wang, Y. and Brown, M. M. (1996). A flexible model for human circadian rhythms. *Biometrics* **53**, 588-696.

[31] Wang, Y., Guo, W. and Brown, M. M. Spline smoothing for bivariate data with applications to association between hormones. To appear in *Statistica Sinica*.

[32] Wang, Y. and Wahba, G. (1995). Bootstrap confidence intervals for smoothing spline estimates and their comparison to Bayesian confidence intervals. *Journal of Statistical Computation and Simulation* **51**, 263-279.

[33] Ye, J. (1998). On measuring and correcting the effects of data mining and model selection. *Journal of the American Statistical Association* **93**, 120-131.