The Canadian Journal of Statistics
Vol. 31, No. 2, 2003, Pages 115–128
La revue canadienne de statistique

The historical functional linear model

Nicole MALFAIT and James O. RAMSAY

Key words and phrases: Finite element method; functional data analysis; functional linear model.

MSC 2000: Primary 62J05; secondary 62G08.

Abstract: The authors develop a functional linear model in which the values at time t of a sample of curves $y_i(t)$ are explained in a feed-forward sense by the values of covariate curves $x_i(s)$ observed at times $s \leq t$. They give special attention to the case $s \in [t-\delta, t]$, where the lag parameter δ is estimated from the data. They use the finite element method to estimate the bivariate parameter regression function $\beta(s,t)$, which is defined on the triangular domain $s \leq t$. They apply their model to the problem of predicting the acceleration of the lower lip during speech on the basis of electromyographical recordings from a muscle depressing the lip. They also provide simulation results to guide the calibration of the fitting process.

Le modèle linéaire fonctionnel historique

Résumé : Les auteurs décrivent un modèle linéaire fonctionnel dans lequel les valeurs au temps t d'un échantillon de courbes $y_i(t)$ sont expliquées par les valeurs observées aux temps $s \leq t$ de courbes covariables $x_i(s)$. Ils accordent une attention particulière au cas où $s \in [t-\delta, t]$, δ représentant un paramètre de délai estimé à partir des données. Ils emploient la méthode des éléments finis pour estimer la fonction paramètre $\beta(s,t)$ bivariée définie sur le domaine triangulaire $s \leq t$. Ils appliquent leur modèle à la prévision de courbes d'accélération de la lèvre inférieure d'un locuteur à partir d'enregistrements électromyographiques d'un muscle abaissant celle-ci. Ils présentent aussi des résultats de simulation pouvant guider le processus de calibration intervenant dans l'ajustement du modèle.

1. INTRODUCTION

This paper considers a functional regression problem in which a function $x_i(s)$, $s \in [0, S]$, is used as the independent variable to explain the variation in a response curve $y_i(t)$, $t \in [0, T]$, through a linear model, where $i = 1, \ldots, N$. Besse & Cardot (1996), Bosq (2000), Cardot, Ferraty & Sarda (1999), James (2002) and Ramsay & Silverman (1997) considered the case

$$y_i(t) = \alpha(t) + \int_0^S x_i(s)\beta(s,t)\, ds + \varepsilon_i(t), \tag{1}$$

where $\varepsilon_i(t)$ is a residual function and $\beta(s,t)$ is the bivariate regression coefficient function that must be estimated from the data. The limits of integration in this model are fixed, and $y_i(t)$ is potentially influenced by $x_i(s)$ at any value of s. For example, if s and t are both times over the same interval ($S = T$), then $y_i(t)$ can be affected by $x_i(s)$ for a future time $s > t$.

Model (1) would be reasonable if the processes were periodic, but otherwise it may be illogical. We look at the situation in which the influence of the carrier function $x(s)$ on the outcome $y(t)$ is of a feed-forward nature, so that the behaviour of y at time t depends only on the behaviour of x at times $s \leq t$. We call this model the *historical linear model.* As an illustration, let us consider that an indicator of a patient's recovery $y(t)$ may depend linearly on the time course of a treatment variable $x(s)$ and that this relation logically only involves times $s \leq t$. Moreover, there may be some reason to suppose that only treatments at times $s \in [t-\delta, t]$ for some lag $\delta \geq 0$ are likely to have an impact on $y(t)$.

2. PREDICTING LIP ACCELERATION FROM EMG RECORDINGS

In this paper, we apply this model to data from a speech production experiment. While an overview of this application appears in Ramsay & Silverman (2002), it will be discussed in detail here.

Speaking is what our species does best. In conversation, an English speaker easily pronounces 14 phonemes per second. This rate appears to be limited by the cognitive aspects of language rather than by the physical ability to perform the articulatory movements. Taking into account the muscles of the thoracic and abdominal walls, the neck and face, the larynx and pharynx, as well as the oral cavity, there are over 100 muscles that must be centrally controlled for speech production (Lenneberg 1984).

The timing of the activation of different groups of muscles is a central issue for the anatomy and physiology of speech. A noninvasive method for collecting data on muscle activation comes from electromyography (EMG), which exploits the fact that muscle contractions are accompanied by electrochemical changes. A resting muscle is isopotential and generates no current, but when it is stimulated, the resulting activation travels along the muscle as waves of action potentials; and if electrodes are attached to the skin over the muscle, they can pick up these potentials. The time taken for a neural signal to be transduced into muscle contraction can vary considerably, but for the facial muscles, a representative value is about 50 msec.

Figure 1 displays the data that are analyzed in this paper. A subject was required to say the syllable "bob" $N = 32$ times. The duration of the syllable varied, but was time-normalized to 700 msec. The movement of the center of the lower lip is shown in the upper panel. The middle panel shows the accelerations of the center of the lower lip, denoted as $y_i(t)$, $i = 1, \ldots, N$, and accelerations, by Newton's second law, reflect the force applied to tissue by muscle contraction. The lower panel displays the so-called linear envelopes of EMG activities, denoted by $x_i(s)$, recorded in the primary muscle depressing the lower lip, the depressor labii inferior (DLI).

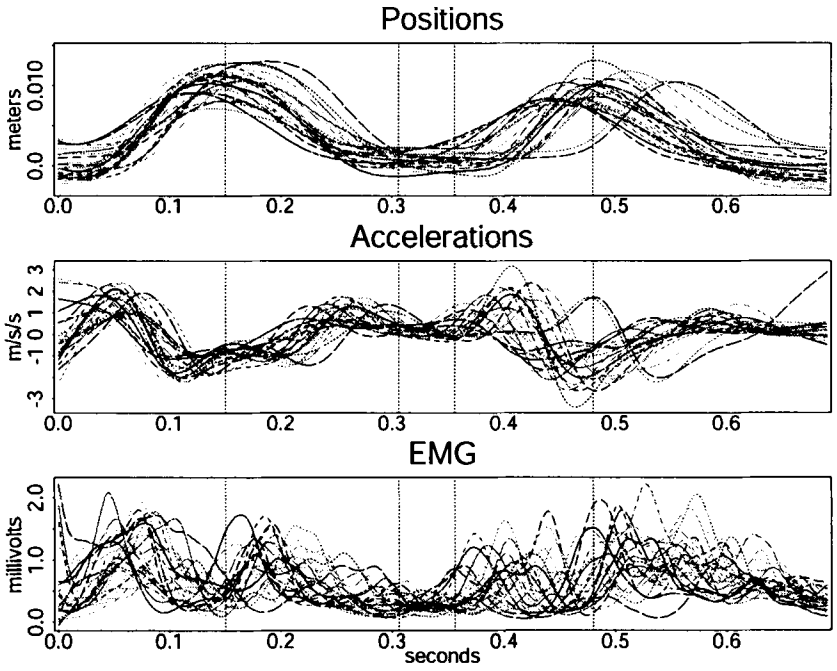

FIGURE 1: The upper two panels show the position and acceleration of the center of the lower lip, respectively, during the articulation of the syllable "bob." The lower panel displays the linear envelopes of the EMG activity associated with the depressor labii inferior (DLI) muscle. The vertical dotted lines separate distinct segments in the motion of the lower lip.

The lower lip trajectory can be segmented into roughly five epochs, separated by dotted lines in Figure 1. The central interval corresponds to the /o/, during which the lip is stationary. To produce each /b/, the lip moves up and down. The DLI muscle plays two roles: antagonist, when it brakes the movement during the ascending phases, and agonist when it accelerates the lip dur-

ing the descending phases. Antagonist episodes are reflected by EMG bursts as the acceleration crosses zero while moving from a positive to a negative phase, and agonist activity, by bursts at the start of the lip's descents.

3. THE HISTORICAL LINEAR MODEL

Let 0 and T indicate initial and final times for a set of records $y_i(t)$, and let δ indicate a time lag, beyond which we conjecture there is no feed-forward type influence of $x(s)$ on $y(t)$. That is, $y(t)$ is influenced by $x(s)$ for $s_0(t) \leq s \leq t$, with $s_0(t) = \max(0, t - \delta)$.

We will assume that $x(s)$ influences $y(t)$ linearly according to the following model that integrates this influence from $s_0(t)$ to t:

$$y_i(t) = \alpha(t) + \int_{s_0(t)}^{t} x_i(s)\beta(s,t)\,ds + \varepsilon_i(t), \quad t \in [0,T]. \tag{2}$$

Here the function $\alpha(t)$ is a fixed intercept function that allows for the relationship between the mean of the $y_i(t)$ and the mean of the $x_i(s)$, but it cannot model their covariation. The residual function $\varepsilon_i(t)$ reflects the inability of the linear model to fit the data completely, and we assume that $\mathrm{E}\{\varepsilon_i(t)\} = 0$ with $\mathrm{cov}\{\varepsilon_i(t), \varepsilon_j(t)\} = 0, i \neq j$.

Hastie & Tibshirani (1993), Staniswalis & Lee (1998), and others have considered the *pointwise model*, also called the *varying coefficient model* or perhaps the *contemporary model*, which is a special case of the historical linear model in which $\delta \to 0$,

$$y_i(t) = \alpha(t) + x_i(t)\beta(t) + \varepsilon_i(t), \tag{3}$$

in which the regression function $\beta(t)$ depends only on t. More generally, (1), (2) and (3) are all special cases of

$$y_i(t) = \alpha(t) + \int_{\Omega_t} x_i(s,t)\beta(s,t)\,ds + \varepsilon_i(t),$$

in which the domain of integration Ω_t is an arbitrary function of t.

A central question in this research is then whether β depends on both s and t, as opposed to only t in the pointwise model, and how far back, δ, this dependency goes. The regression coefficient function $\beta(s,t)$ is defined on the subset of a triangle where $s \leq t$, as illustrated in Figure 2.

In order to simplify our discussion of the estimation of model (2), let us first drop the intercept function $\alpha(t)$. This can be done without loss of generality since from the normal equations,

$$\alpha(t) = \bar{y}(t) - \int_{s_0(t)}^{t} \bar{x}(s)\beta(s,t)\,ds,$$

and consequently, we obtain by substitution

$$y_i^*(t) \;=\; \int_{s_0(t)}^{t} x_i^*(s)\beta(s,t)\,ds + \varepsilon_i(t), \tag{4}$$

where $x_i^*(s) = x_i(s) - \bar{x}(s)$. In order to further simplify the expressions, we drop the asterisk in what follows.

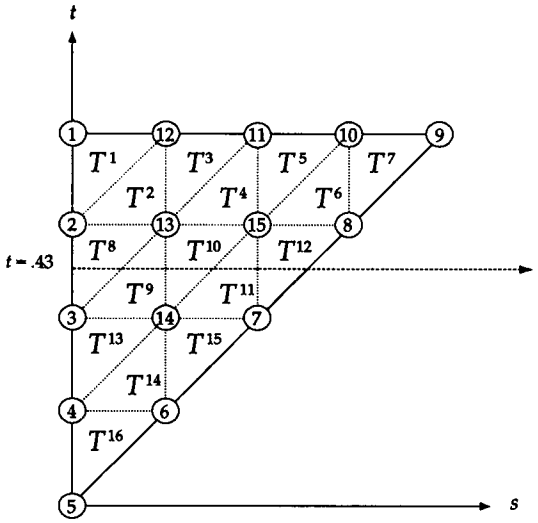

FIGURE 2: The large triangle is the domain for the regression function $\beta(s,t)$ in the historical linear model. The horizontal axis s is time for the independent variable $x(s)$, and the vertical axis t is time for the dependent variable $y(t)$. The smaller triangles within the larger triangle, labelled $T^1,\ldots,T^{16}$, are finite elements used to approximate $\beta(s,t)$. The fifteen vertices of these triangles, called *nodes*, correspond to basis functions $\phi_k(s,t)$. The horizontal dashed line shows which elements play a role in determining the behaviour of $y(t)$ at $t = 0.43$.

3.1. *A finite-dimensional version.*

The historical linear model, and in fact any linear model involving integration over functional covariates, is identifiable but essentially inestimable since we have a model with effectively an infinite number of covariates, each corresponding to a value of s, available to fit a finite amount of data. Ramsay & Silverman (1997) discuss this issue and point out the necessity of some regularization or roughness penalty on $\beta(s,t)$ in the estimation process. A simple way of achieving this is to replace $\beta(s,t)$ by an expansion in terms of a limited number of basis functions, so that we can assure stable estimation by controlling the number of basis functions used.

Consequently, we let the regression function $\beta(s,t)$ be approximated by an expansion $\hat{\beta}(s,t)$ in terms of K known basis functions $\phi_k(s,t)$, namely

$$\hat{\beta}(s,t) = \sum_{k=1}^{K} b_k \phi_k(s,t).$$

Defining

$$\psi_{ik}(t) = \int_{s_0(t)}^{t} x_i(s)\phi_k(s,t)\,ds,$$

we have the alternative formulation

$$y_i(t) = \sum_{k=1}^{K} b_k \int_{s_0(t)}^{t} x_i(s)\phi_k(s,t)\,ds + \int_{s_0(t)}^{t} x_i(s)\varepsilon_a(s,t)\,ds + \varepsilon_i(t) = \sum_{k=1}^{K} b_k \psi_{ik}(t) + \varepsilon_i'(t), \quad (5)$$

where $\varepsilon_a(s,t) = \beta(s,t) - \hat{\beta}(s,t)$ is the approximation error, and $\varepsilon_i'(t)$ combines the random and the approximation errors.

Let $\boldsymbol{y}(t)$ and $\boldsymbol{e}(t)$ be the vectors of length N containing the values $y_i(t)$ and $\varepsilon_i'(t)$, respectively. Let also $\boldsymbol{\Psi}(t)$ be the $N \times K$ matrix containing values of $\psi_{ik}(t)$ and let the coefficient vector $\boldsymbol{b}$ be $(b_1,\ldots,b_K)'$. Then we have the matrix expression of (4):

$$\boldsymbol{y}(t) = \boldsymbol{\Psi}(t)\boldsymbol{b} + \boldsymbol{e}(t).$$

We wish to minimize the L_2 criterion

$$\text{SSE} = \int_0^T \sum_{i=1}^N \{\varepsilon_i'(t)\}^2 \, dt, \tag{6}$$

and this is equivalent to solving the normal equations

$$\left\{\int_0^T \boldsymbol{\Psi}'(t)\boldsymbol{\Psi}(t)\, dt\right\}\boldsymbol{b} = \int_0^T \boldsymbol{\Psi}'(t)\boldsymbol{y}(t)\, dt. \tag{7}$$

Moreover, if the vector $\boldsymbol{x}(s)$ of covariate function values $x_i(s)$ can be expressed as the basis function expansion $\boldsymbol{C}\boldsymbol{\eta}(s)$, where $\boldsymbol{\eta}(s)$ is a vector of length M of basis function values $\eta_m(s)$, $\boldsymbol{C}$ is the $N \times M$ matrix of coefficients for the expansion, and the $M \times K$ matrix $\boldsymbol{\Gamma}(t)$ contains the values

$$\gamma_{mk}(t) = \int_{s_0(t)}^t \eta_m(s)\phi_k(s,t)\, ds,$$

then we have

$$\left\{\int_0^T \boldsymbol{\Gamma}'(t)\boldsymbol{C}'\boldsymbol{C}\boldsymbol{\Gamma}(t)\, dt\right\}\boldsymbol{b} = \int_0^T \boldsymbol{\Gamma}'(t)\boldsymbol{C}'\boldsymbol{y}(t)\, dt. \tag{8}$$

From the structure of these equations, we see that the coefficient matrix can be expected to be nonsingular if K is not too large, even if the number M of basis functions used in the expansion of the covariates is less than K.

3.2. A multivariate linear model approximation.

The normal equations (7) and (8) may imply an unacceptable level of computation because of the need to approximate the integrals involved. However, we can approximate the model in the form (5) by a multivariate linear model based on an approximation of the integrals by proceeding as follows. Evaluating $y_i(t)$, $i = 1, \ldots, N$ at a finite set of time points t_q, $q = 0, \ldots, Q$, for each record i leads to

$$\mathrm{E}(\boldsymbol{y}_i) = \boldsymbol{\Psi}_i\boldsymbol{b}, \tag{9}$$

where

$$\boldsymbol{y}_i = \begin{bmatrix} y_i(t_0) \\ \vdots \\ y_i(t_Q) \end{bmatrix}, \quad \boldsymbol{\Psi}_i = \begin{bmatrix} \psi_{i1}(t_0) & \cdots & \psi_{ik}(t_0) & \cdots & \psi_{iK}(t_0) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \psi_{i1}(t_Q) & \cdots & \psi_{ik}(t_Q) & \cdots & \psi_{iK}(t_Q) \end{bmatrix},$$

and $\boldsymbol{b} = (b_1, \ldots, b_K)'$.

By stacking these matrices $\boldsymbol{y}_i$ and $\boldsymbol{\Psi}_i$ on top of each other to obtain $N(Q+1) \times 1$ and $N(Q+1) \times K$ matrices $\boldsymbol{y}$ and $\boldsymbol{\Psi}$, respectively, we obtain the least squares estimator of $\boldsymbol{b}$ defined by the normal equations

$$\boldsymbol{\Psi}'\boldsymbol{\Psi}\boldsymbol{b} = \boldsymbol{\Psi}'\boldsymbol{y}. \tag{10}$$

The summation implied by $\boldsymbol{\Psi}'\boldsymbol{\Psi}$ effectively approximates both the integration and the summation involved in the stationary equations (7).

We found that equally-spaced time points $Q = 4n$ gave a satisfactory level of precision with acceptable computational overhead, where n is the number of intervals in $[0, T]$ used for the triangulation shown in Figure 2 and described in detail below. Note that in (13) and (14), the equally-spaced t_i can be replaced by suitable quadrature points designed to approximate a modification of (11) where dt is replaced by $f(t)\, dt$, $f(t)$ being the design density of the t_i.

3.3. Using finite elements to approximate $\beta(\mathbf{s}, \mathbf{t})$.

To approximate the bivariate regression function $\beta(s,t)$ we use the finite element method. This method, which is widely used in the numerical solution of partial differential equation boundary-value problems, can also be viewed as a method of interpolation over two-dimensional regions.

First, we partition each axis into n equal intervals, each of length λ, and then split each square grid into two triangles by the diagonal parallel to the line $s = t$. This divides our domain into n^2 congruent triangular elements T^e, as illustrated in Figure 2 where $n = 4$ and 16 triangles are constructed. We then discretize our possible choices of lag δ to $m\lambda$, $0 \leq m \leq n$, and discard all triangles containing points more than δ units from the diagonal in the horizontal direction. This leaves $m(2n - m)$ triangular elements covering the domain of $\beta(s,t)$ and $K = (m+1)(n+1-m/2)$ nodes.

Each basis function ϕ_k is a tent shaped function, as illustrated in Figure 3, which corresponds to one of the remaining nodes. Over its support, which consists in the six elements having node k as a vertex, ϕ_k is piecewise linear and continuous. It equals 1 at node k, and vanishes at the boundary of its support. If a node k is on the boundary of the domain of $\beta(s,t)$, then the corresponding ϕ_k can be extended beyond the boundary in an arbitrary way since its value in this region is never used. The ϕ_k's are generally not orthogonal to each other, but we will see that the linear equation systems defined by (7) are sparse and therefore are solvable rapidly. We now consider in more detail how to work with this basis system.

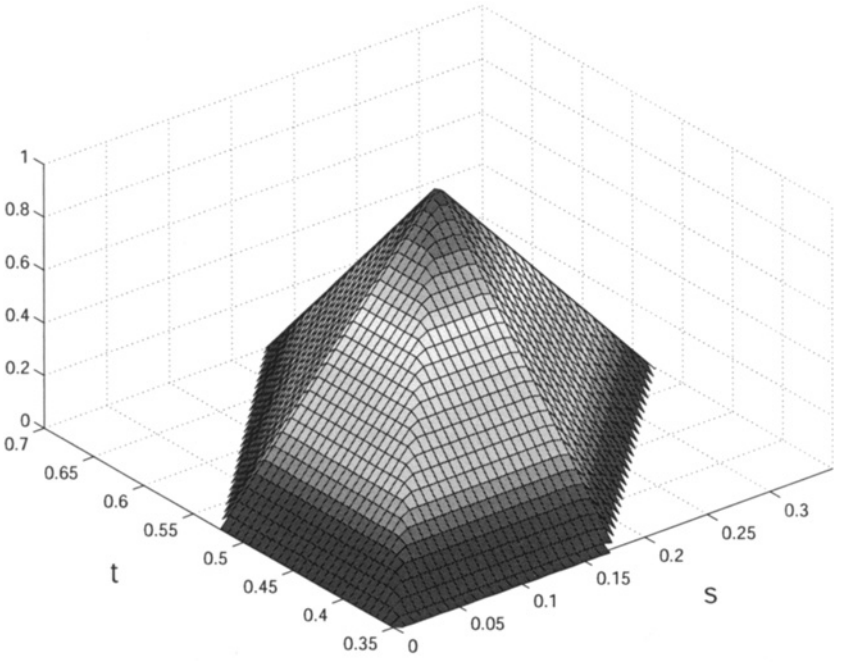

FIGURE 3: The basis function $\phi_{13}(s,t)$ corresponding to the 13th node in the triangulation that is shown in Figure 2.

We approximate β over each element T^e by a linear function $\hat{\beta}$. Any linear function on T^e can be written as a sum of three particular linear functions, ϕ^e_ν, $\nu = 1,2,3$, each associated with one of the three vertices of T^e. These functions are defined by the property that ϕ_ν equals 1 at the νth vertex of T^e and 0 at the other two vertices. Then, in this local notational system,

$$\hat{\beta}(s,t) = \sum_{\nu=1}^{3} b^e_\nu \phi^e_\nu(s,t), \quad (s,t) \in T^e.$$

The basis functions ϕ^e_ν associated with the vertices of triangle T^e can be defined by the area coordinates for that triangle. For an arbitrary point $(s,t) \in T^e$ shown in Figure 4, the area coordinates L^e_ν, $\nu = 1,2,3$, for its three vertices are defined as the ratio of the triangular subarea

A^e_ν to the whole area A^e.

$$L^e_1(s,t) = \frac{A^e_1(s,t)}{A^e}, \quad L^e_2(s,t) = \frac{A^e_2(s,t)}{A^e}, \quad L^e_3(s,t) = \frac{A^e_3(s,t)}{A^e},$$

where $A^e = \sum_{\nu=1}^3 A^e_\nu(s,t)$. Each $L^e_\nu(s,t) = 1$ at vertex ν but is equal to zero at each other vertex, and for all $(s,t) \in T^e$, $\sum_{\nu=1}^3 L^e_\nu(s,t) = 1$.

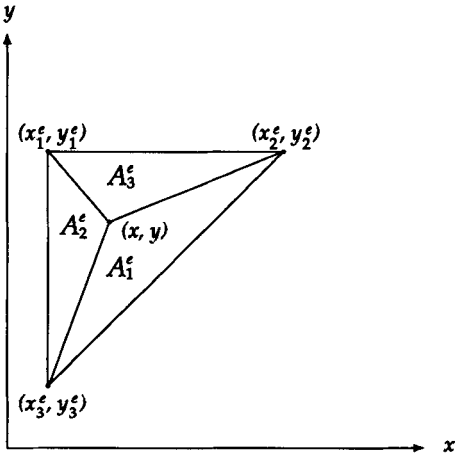

FIGURE 4: The area coordinates associated with point (s,t) in a triangular element are the ratios of each of the areas A^e_ν shown to the total area A^e of the triangle.

To express the local basis functions ϕ^e_ν in terms of the global Cartesian coordinates (s,t), we use the fact that

$$L^e_1(s,t) = \frac{A^e_1(s,t)}{A^e} = \det \begin{bmatrix} 1 & s^e_1 & t^e_1 \\ 1 & s^e_2 & t^e_2 \\ 1 & s & t \end{bmatrix} \Big/ \det \begin{bmatrix} 1 & s^e_1 & t^e_1 \\ 1 & s^e_2 & t^e_2 \\ 1 & s^e_3 & t^e_3 \end{bmatrix},$$

where (s^e_1, t^e_1), (s^e_2, t^e_2), (s^e_3, t^e_3) are the vertices of T^e. This can also be expressed as

$$\begin{bmatrix} L^e_1 \\ L^e_2 \\ L^e_3 \end{bmatrix} = \frac{1}{2A^e} \begin{bmatrix} s^e_2 t^e_3 - s^e_3 t^e_2 & t^e_2 - t^e_3 & s^e_3 - s^e_2 \\ s^e_3 t^e_1 - s^e_1 t^e_3 & t^e_3 - t^e_1 & s^e_1 - s^e_3 \\ s^e_1 t^e_2 - s^e_2 t^e_1 & t^e_1 - t^e_2 & s^e_2 - s^e_1 \end{bmatrix} \begin{bmatrix} 1 \\ s \\ t \end{bmatrix}$$

which gives the local basis functions $\phi^e_\nu = L^e_\nu$ on T^e.

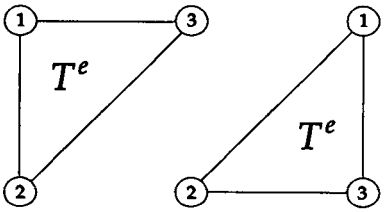

FIGURE 5: The nodes defining the 15 finite element basis functions shown in Figure 2 are numbered as shown in the left triangle for odd-numbered triangular elements $e = 1, 3, \ldots, 15$, and as shown in the right panel for even-numbered elements $e = 2, 4, \ldots, 16$.

We then obtain our global interpolant $\hat{\beta}(s,t) = \sum_{k=1}^{K} b_k \phi_k(s,t)$ by fitting together the local approximations in the following way. Our nodal points being the vertices of the triangular elements, we have in our illustration in Figure 2 a total of 15 nodes for the whole domain, each of which can be designated in two ways: globally and locally. If we designate them globally as shown in Figure 2, and locally as shown in Figure 5, we obtain the correspondence of numbering listed in Table 1.

So, for example, the linear global basis function ϕ_{13} is constructed by "patching" together the local basis functions ϕ_3^2, ϕ_2^3, ϕ_2^4, ϕ_3^8 ϕ_1^9 and ϕ_1^{10}, as shown in Figure 3. Because a linear function in one dimension is uniquely determined by specifying its values at two points, the local basis functions that are patched together coincide everywhere on the boundaries common to two adjacent elements. Thus, ϕ_{13} is continuous over its support, consisting of the elements T^2, T^3, T^4, T^8, T^9 and T^{10}. Using the correspondence listed in Table 1, we can express the local interpolant on element T^1, for example, as

$$\begin{aligned} \hat{\beta}(s,t) &= b_1^1\phi_1^1(s,t) + b_2^1\phi_2^1(s,t) + b_3^1\phi_3^1(s,t) \\ &= b_1\phi_1(s,t) + b_2\phi_2(s,t) + b_{12}\phi_{12}(s,t), \quad (s,t) \in T^1. \end{aligned}$$

When $\delta = m = 0$, $\hat{\beta}(s,t)$ reduces to a polygonal spline representation of the variation on the diagonal $\beta(t,t)$ with $n+1$ basis functions.

TABLE 1: Correspondence between the local and the global numbering of nodes. A number in the table indicates the local number of a node as shown in Figure 5.

	Global node numbers														
Elements	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
T^1	1	2										3			
T^2		2										1	3		
T^3											3	1	2		
T^4											1		2	3	
T^5										3	1				2
T^6								3		1					2
T^7								2	3	1					
T^8		1	2										3		
T^9			2										1	3	
T^{10}													1	2	3
T^{11}							3							2	1
T^{12}							2	3							1
T^{13}			1	2										3	
T^{14}				2		3								1	
T^{15}						2	3							1	
T^{16}				1	2	3									

Finally, we can now calculate the relations between the local and global basis functions. Let S_{kt} be the set of triangular elements intersected by the integration line defined by fixed t and belonging to the support of ϕ_k, that is, having node k as a vertex. Let $s_{ae}(t)$ and $s_{be}(t)$ be the extreme left and the extreme right points, respectively, belonging to both a given element T^e in

S_{kt} and to the integration line. Then

$$\psi_{ik}(t) = \int_{s_0(t)}^{t} x_i(s)\phi_k(s,t)\,ds = \sum_{T^e \in S_{kt}} \int_{s_{ae}(t)}^{s_{be}(t)} x_i(s)\phi_k(s,t)\,ds.$$

Thus, for instance, to calculate $\psi_{i13}(0.43)$, since ϕ_{13} is defined on T^2, T^3, T^4, T^8, T^9 and T^{10}, we carry out the integration piecewise over the elements T^8, T^9 and T^{10} along the horizontal line of height $t = 0.43$.

This finite element basis system has a number of advantages over the common practice of representing bivariate functions using tensor product basis functions, such as tensor product B-splines. B-spline basis functions are defined over a rectangular grid, equivalent to using rectangular elements, and do not conform naturally to the triangular domain needed here. Moreover, the design matrix in (9) defining estimated values for the b_k is sparse and becomes increasingly so as the triangulation becomes finer. This is because the number of basis functions having positive values for a given t is determined by counting the number of nodes next to the corresponding line of integration. For example, with the triangulation shown in Figure 2, this number is seven for $t = 0.43$. Thus, as the number of triangular elements increases, the proportion of basis functions having positive values for a given t decreases. As a consequence, sparse matrix computational methods can greatly speed up the solution of this equation. Finally, the evaluation of any basis function over an element involves the evaluation of a function linear in s and t and is therefore very fast.

4. ASSESSING FIT

The assessment of fit by comparing two historical linear models, one embedded within the other, raises a number of issues, and our proposals in this section are only intended to be preliminary to further work. Note also that we describe the approximation to the historical linear model, not the model itself.

The historical linear model maps a parameter space of dimension K_1, denoted by G_1, into a response space H by means of a linear operator L. The image of this mapping is denoted by $\widehat{Y}_1$. The space G_1 here consists of regression functions $\beta(s,t)$ representable by the finite element basis, the dimensionality K_1 of which equals the number of nodes determined by the mesh density and lag $\delta = m\lambda$. A model assessment involves comparing a fit based on all of G_1 with that provided by a subspace G_0 of dimensionality K_0, with an image $\widehat{Y}_0$.

The dependent variable space H consists of N independently sampled functions y_i, each function being representable within a function space of dimension, say, $\dim(Y)$. If we use centered functions to fit model (4), the dimensionality of H is $\dim(H) = (N-1)\dim(Y)$; otherwise it is $N\dim(Y)$. Dimensionality $\dim(Y)$ in turn is the minimum of the number of basis functions used to represent each y_i and the number of sampling points at which each function is observed in the raw data. However, if a roughness penalty or regularization procedure is employed, assessing $\dim(Y)$ is more complex. Discussions of this issue can be found in Hastie & Tibshirani (1990) and Ramsay & Silverman (1997).

If the linear operator L is of full rank, then the dimension of the subspace of H containing the image $\widehat{Y}_1 = L\beta_1$ for $\beta_1 \in G_1$ is K_1; and for the reduced model $\beta_0 \in G_0$, the dimensionality of the subspace containing $\widehat{Y}_0$ is K_0. It can be shown in this general context from the nature of the inner product defined on Hilbert space H that if the full and reduced models are fitted by minimizing $\|Y - \widehat{Y}\|^2$, then H is partitioned into the direct sum of three orthogonal spaces containing $Y - \widehat{Y}_1$, $\widehat{Y}_1 - \widehat{Y}_0$ and $\widehat{Y}_0$, respectively and the dimensions of these subspaces are $\dim(H) - K_1$, $K_1 - K_0$ and K_0, respectively.

The procedure described above for estimating β will in general be of full rank, although in any actual application the design matrix Ψ should certainly be checked for nonsingularity. Although our fitting method does not precisely minimize least squares criterion (6), it seems

appropriate to assume the orthogonal decompositions described above and their associated degrees of freedom are approximately correct provided that the number of sampling points Q is sufficiently large.

Consequently we may be guided in model selection by

$$R^2 = 1 - \mathrm{SSE}_1/\mathrm{SSE}_0, \tag{11}$$

and, moreover,

$$F = \frac{(\mathrm{SSE}_0 - \mathrm{SSE}_1)/(K_1 - K_0)}{\mathrm{SSE}_1/(\dim(H) - K_1)}$$

can be inspected as a means of assessing the importance of the global improvement in fit in going from the simpler model to the more powerful alternative. This use of the F-ratio as a model selection tool is based on the orthogonality of the subspaces of H containing $Y - \widehat{Y}_1$ and $\widehat{Y}_1 - \widehat{Y}_0$, and the assumption that when the reduced model holds, the lack of fit per degree of freedom in the two associated subspaces of H should be approximately equal. One should probably stop short, however, of taking tabled values of the F-distribution as descriptions of its reduced model behaviour since the numerator and denominator may not be assumed to have normalized chi-square distributions.

However, this type of global analysis may be of only limited interest if it seems clear that the improvement in fit depends strongly on t. Consequently, in addition to these two scalar measures of fit, we will want to display measures of fit as functions of t, for instance,

$$\mathrm{RMSE}(t) = \sqrt{\mathrm{SSE}(t)/N}\,, \tag{12}$$

where $\mathrm{SSE}(t) = \sum_{i=1}^{N}\{y_i(t) - \hat{y}_i(t)\}^2$, which allows for closer time related analyses.

5. RESULTS FOR THE LIP/EMG DATA

A first issue is how fine to make the triangulation, which controls the smoothness and the amount of detail in the estimated $\beta(s,t)$. We used a preliminary calibration step in which we tested the method on a data set created from our actual EMG functions $x_i(s)$ using these intercept and regression functions:

$$\alpha(t) = 2\sin(40t) + 2\cos(20t), \quad \beta(s,t) = 0.5\sin(30s) + 0.5\cos(30t). \tag{13}$$

For the model involving integration on the whole triangular domain ($\delta = m\lambda = T$), we computed the true dependent variable functions $y_i(t)$ with $\varepsilon_i(t) = 0$. We used triangulations into n^2 elements, $n = 4, 6, \ldots, 26$, where n is the number of intervals into which the horizontal and vertical boundaries of the triangular domain are divided. This corresponds to from 15 to 378 nodes, and from 16 to 676 triangular elements.

Using squared correlations R^2 defined in (11), we found, as expected, that the finer the mesh the better the fit. Figure 6 shows $\hat{\beta}(s,t)$ for 190 nodes and 324 elements, for which $R^2 = 0.997$. These results reassured us that the variation among the 32 EMG records was, in principle, sufficient to support estimation of a regression function with considerable local detail. For the analysis of our actual data, we chose a triangulation of the whole domain into 196 elements or 120 nodes, and this gave $R^2 = 0.991$ for this calibration problem. This triangulation corresponds to dividing $[0,T]$ into $n = 14$ intervals with lengths $\lambda = 50$ msec, roughly equal to the delay mentioned above necessary for a neural signal to be transduced into muscle contraction.

The second issue is the width of the domain of integration δ, which in this case is set at the discrete values $m\lambda$. From Figure 7, we see that the fit steadily improves as we enlarge the domain of integration from $\delta = 0$ ($R^2 = 0.17$) up to $\delta = 6\lambda$ ($R^2 = 0.46$), corresponding to 300 msec, but does not increase substantially with larger values.

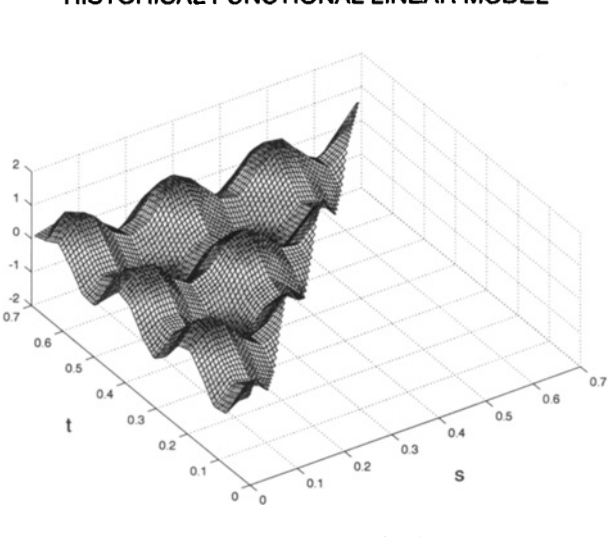

FIGURE 6: The estimate of the regression function $\beta(s,t)$ for the calibration problem (13), corresponding to $R^2 = 0.997$.

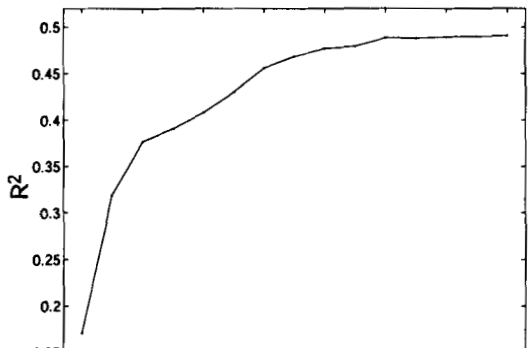

FIGURE 7: The squared correlation measure (11) as a function of lag $\delta = m\lambda$.

Figure 8 shows the estimate of the bivariate regression function $\beta(s,t)$, corresponding to $\delta = 6\lambda$. The ratio of the largest singular value of design matrix Ψ to its smallest singular value was 16.1, indicating that this estimate is well-determined by the data. We estimated the standard error of $\beta(s,t)$ by drawing 1000 bootstrap samples from among the 32 acceleration functions. Figure 9 shows estimated β values, with 95% pointwise confidence bands, for $\beta(s,t)$ along the diagonal $s = t$.

The shape of the estimate of $\beta(s,t)$ indicates, as expected, that the muscle activation is the most influential around the lip closure times, when the signal-to-noise ratio is the largest. Also, there is a ridge of influence along the diagonal, corresponding to the transduction delay of about 50 msec. But peaks deeper in the triangular domain reveal covariation of EMG events and lip displacements separated by longer delays. For instance, the peak at about $s = 250$ msec and $t = 500$ msec cannot be interpreted in terms of immediate causal relationship between muscle activation and lip acceleration. Rather it can be explained as follows. The agonist depressing activity of the DLI ends at around 250 msec, and this determines the position reached by the lower lip for the production of the /o/. The amplitude of the upward movement that follows depends on this position and consequently, so does the deceleration during the second part of this movement that is necessary to allow for an adequate closing of the lips for the second /b/. This illustrates how the pronunciation of a phoneme is determined by the chain of phonemes within which it is embedded.

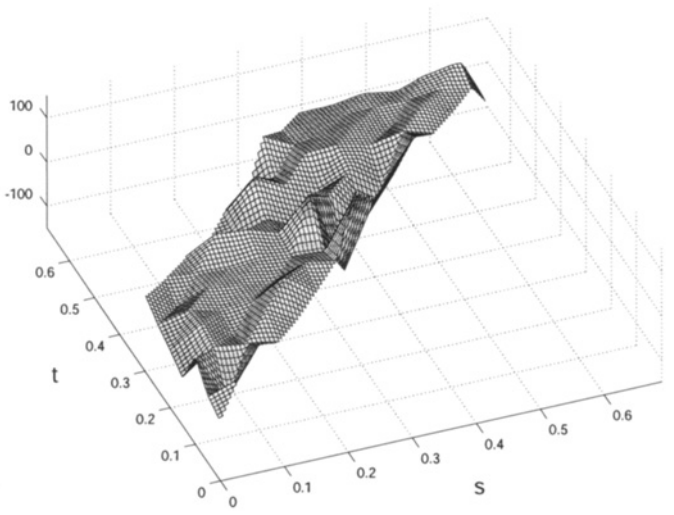

FIGURE 8: The estimated regression function $\beta(s,t)$ for lag $\delta = 6\lambda$ for the actual data.

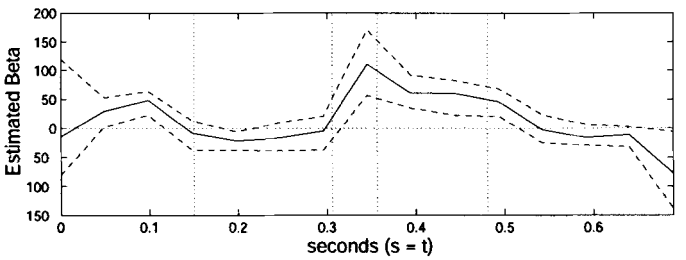

FIGURE 9: The solid line is the value of the estimated regression function along the diagonal, $\beta(t,t)$. The dashed lines indicate plus and minus two standard errors estimated by bootstrapping.

Next, we specifically examined how well this version of the model performs in comparison to the one using lag $\delta = 0$, that is, reduced to the pointwise model (3). Because there is inevitably less fitting power for values of $t < \delta$ than for t where the full history of EMG is available, we chose to consider only values of t going from 6λ to T.

R^2 values computed over this range of t are 0.57 and 0.22 for lags $\delta = 6\lambda$ and $\delta = 0$, respectively. Plotting the corresponding error functions RMSE (t), defined in (12), in the top panel of Figure 10 shows that the model using the wider lag is superior specifically around the second closure time, making clear that the explanatory value of the past behaviour of the EMG curves varies with the events to be explained. The small failure of the RMSE (t) for $\delta = 6\lambda$ to be greater than that for $\delta = 0$ at $t = 0.36$ is due to the use of discrete values of t to estimate the models.

The bottom panel of Figure 10 shows F-ratio values integrated over each λ-wide intervals. Here, 100 basis functions were used to represent each of the $N = 32$ acceleration functions from data sampled at 501 points without regularization, so we have $\dim(Y) = 100$ and $\dim(H) = 3100$. With $m = 14$, we have $K_1 = 84$ and $K_0 = 15$, so the degrees of freedom are 69 and 3016 for the numerator and the denominator respectively, and the critical value is $F_{.05} = 1.27$.

The data for this application and annotated Matlab code for the analyses may be obtained from web site www.springer-ny.com through the entry for Ramsay & Silverman (2002).

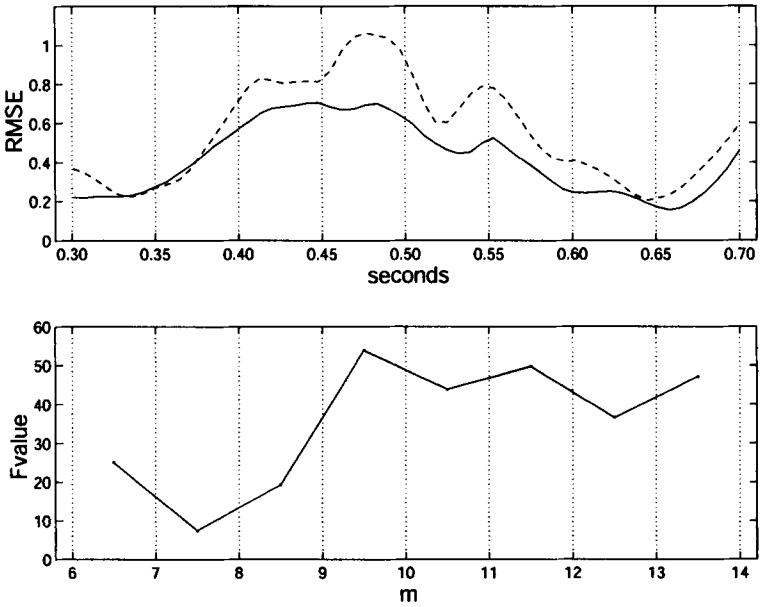

FIGURE 10: The top panel shows the error function RMSE (t) for the models with lag $\delta = 6\lambda$ (solid line) and $\delta = 0$ (dashed line). The bottom panel shows F-ratio values integrated over each λ-wide intervals. In both panels, the vertical dotted lines indicate boundaries between elements.

6. DISCUSSION AND CONCLUSIONS

Our application of the historical linear model (2) described the contribution of a facial muscle to the articulation of a phoneme. We could see that, in addition to the straightforward causal effects involving conversion of electrochemical into mechanical energy, more complex mediate influences take place over delays as long as 300 msec.

At least two groups of muscles applying force in different directions are generally necessary to control the motion of a body part, and three groups of muscles contribute to the motion of the lower lip. Our model, by using the activation pattern of only one single group of muscles, could explain only a limited part of the variation observed in the lip accelerations, and this contributed to the modest R^2 values that we obtained. Including the influences of the other two groups of muscles would have introduced an additional integral for each set of explanatory functions. However, we judged that the number of replications available could only support a single functional independent variable.

As with any linear model, there should be some careful consideration of whether the variation in the covariate $x(s)$ is sufficient to support a reasonably detailed fit to the observed dependent variable $y(t)$. In the multivariate context, this would correspond to checking the design matrix for near singularities. We felt that a calibration experiment such as the one that we used, based on the actual covariates, offers a useful indication of whether a known regression function $\beta(s,t)$ can be recovered, and we found our results to be encouraging.

It is possible to envisage a number of approaches to the representation of the regression function $\beta(s,t)$ and to the estimation of the parameters or coefficients defining this representation. We chose the finite element approach because it conforms naturally to the triangular domain for $\beta(s,t)$, is easy to evaluate, permits an arbitrary amount of detail in the estimate, and leads to a sparse matrix linear equation defining the estimate. These are important assets when the estimate needs to have complex features. The finite element method has already proven itself in other

areas of applied mathematics, and there are well-developed software tools available in languages like Matlab for tasks such as the triangulation of the domain (MathWorks 1995).

A next step is to add the possibility of regularizing or smoothing the estimate of $\beta(s,t)$ or the prediction of $y_i(t)$ through the use of roughness penalties. These roughness penalties will require the use of basis functions $\phi_k(s,t)$ of higher order, but again these have already been well worked out in the finite element analysis literature.

ACKNOWLEDGEMENTS

We wish to express our deep appreciation to Dr Sorana Froda, Université du Québec à Montréal, who co-supervised the Master's thesis by Malfait (1999) on which this paper is based, and who read several drafts. We also thank Professor V. Gracco, McGill University, for making the data available. The research was supported by a grant to the second author and to Dr Froda from the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

P. Besse & H. Cardot (1996). Approximation spline de la prévision d'un processus fonctionnel autorégressif d'ordre 1. *The Canadian Journal of Statistics*, 24, 467–487.

D. Bosq (2000). *Linear Processes in Functions Spaces: Theory and Applications.* Lecture Notes in Statistics 149, Springer, New York.

H. Cardot, F. Ferraty & P. Sarda (1999). Functional linear model. *Statistics & Probability Letters*, 45, 11–22.

T. Hastie & R. J. Tibshirani (1993). Varying-coefficient models. *Journal of the Royal Statistical Society Series B*, 55, 757–796.

G. M. James (2002). Generalized linear models with functional predictors. *Journal of the Royal Statistical Society Series B*, 64, 411–432.

E. H. Lenneberg (1984). *Biological Foundations of Language.* Wiley, New York.

N. Malfait (1999). *Estimation par la méthode des éléments finis dans un modèle linéaire fonctionnel.* Unpublished Master's thesis, Université du Québec à Montréal, Montréal (Québec), Canada.

Mathworks, Inc. (1995). *Partial Differential Equation Toolbox User's Guide.* The MathWorks, Inc., Natick, MA.

J. O. Ramsay & B. W. Silverman (1997). *Functional Data Analysis.* Springer, New York.

J. O. Ramsay & B. W. Silverman (2002). *Applied Functional Data Analysis.* Springer, New York.

J. G. Staniswalis & J. J. Lee (1998). Nonparametric regression analysis of longitudinal data. *Journal of the American Statistical Association*, 93, 1403–1418.

Received 3 May 2002
Accepted 8 April 2003

Nicole MALFAIT: malfait@motion.psych.mcgill.ca
Department of Psychology, McGill University
Montréal (Québec), Canada H3A 1B1

James O. RAMSAY: ramsay@psych.mcgill.ca
Department of Psychology, McGill University
Montréal (Québec), Canada H3A 1B1