Functional Components of Variation in Handwriting

J. O. RAMSAY

Functional data analysis techniques are used to analyze a sample of handwriting in Chinese. The goals are (a) to identify a differential equation that satisfactorily models the data's dynamics, and (b) to use the model to classify handwriting samples taken from differential individuals. After preliminary smoothing and registration steps, a second-order linear differential equation, for which the forcing function is small, is found to provide a good reconstruction of the original script records. The equation is also able to capture a substantial amount of the variation in the scripts across replication. The cross-validated classification process is 100% effective for the samples analyzed.

KEY WORDS: Classification; Differential equations; Dynamic model; Functional data analysis; Longitudinal data; Penalized nonparametric regression; Principal differential analysis; Registration; Smoothing spline; Time warping.

1. INTRODUCTION

Handwriting, such as that of "statistical science" in Chinese in Figure 1, displays complex variation both within and across individuals. At a casual level of observation, it is not hard to see why. The hand is a complex biomechanical system, and the movements made by the fingers, wrist, and forearm are produced by the contraction of dozens of muscle groups. The system exhibits various kinds of harmonic or oscillatory behavior, is subject to forces of elastic and viscous components of muscle, and must work in the context of gravitational and inertial forces. These mechanical components vary in many ways from one individual to another, as well as within an individual.

Moreover, handwriting is the consequence of a complex control process that activates muscle contraction. Trains of spike potentials cascade down a network of nerve fibers and arrive at motoneuron junctions to trigger or release contraction of muscle fibers. These neural events vary in timing and amplitude, due in part to the variation in the activity at the cortical level and in part to variation in transmission properties along the neural pathways.

Models for the handwriting process have been researched extensively (see, e.g., Faure, Keuss, Lorette, and Vinter 1994). A recent approach of particular interest to statisticians is that of Plamondon and Guerfali (1998).

1.1 Why a Linear Differential Equation?

This article models this complex variation by a linear differential equation estimated from replicated handwriting samples from a single individual. A number of steps are required as preliminaries, and these are typical of a functional data analysis, as described by Ramsay and Silverman (1997).

Exploratory techniques such as principal components analysis will certainly offer some insight into data such as these. But because handwriting is the result of a mechanical process, it necessarily has derivatives, and in particular we can expect from basic physics that acceleration will provide a way to see the forces involved. A differential equation describes processes by finding relationships among derivatives. For example, the position $x(t)$ of an undamped spring is determined by the second-order linear homogeneous differential equation $D^2x(t) = -\omega^2 x(t)$, with $\omega/(2\pi)$ the frequency of oscillation. More complex mechanical systems subject to frictional and other forces can be described in terms of second-order linear equations of a more general nature.

The handwriting system is subject to external control through forces applied by muscle contractions, and it seems reasonable to suppose that we can more clearly see this control system at work through aspects of a differential equation model than through models taking only position into account, just as the spring equation is a useful starting point for studying springs subject to external loading forces f, and described by the nonhomogeneous equation $\omega^2 x(t) + D^2x(t) = f(t)$.

A differential equation also exploits the intrinsic smoothness of the process by taking explicit account of derivative behavior. This is not to say, of course, that we do not need to allow for observational noise that may perturb the measured pen position in a nonsmooth manner.

Finally, I indicate in Section 6 that a differential equation model permits a rather richer error or stochastic structure to account for observed variation in handwriting samples than is available through traditional smooth-signal-plus-error models. That is, there may be stochastic aspects of derivative behavior in addition to stochastic exogenous influences.

1.2 Outline of the Steps Involved

After a preliminary inspection of the data, the first task is to estimate derivatives. Of the various methods available, I used a spline smoothing module that permits smooth derivative estimates for an arbitrary derivative $D^m x$, and also allows additional smoothing in the neighborhood of the end points to stabilize derivative estimates.

Jim Ramsay is Professor, Department of Psychology (E-mail: *ramsay@psych.mcgill.ca*), McGill University, Montreal, Quebec, Canada H3A 1B1. The author thanks Xiaochun Li and Xiaohui Wang, who acted as subjects in the experiments, and David Ostry, in whose laboratory the experiments were conducted. The research was supported by grant APA320 from the Natural Science and Engineering Research Council of Canada. The S-PLUS and Matlab functions used in these analyses are available free of charge by ftp from site *ego.psych.mcgill.ca/pub/ramsay/FDAfuns*.

Journal of the American Statistical Association
March 2000, Vol. 95, No. 449, Applications and Case Studies

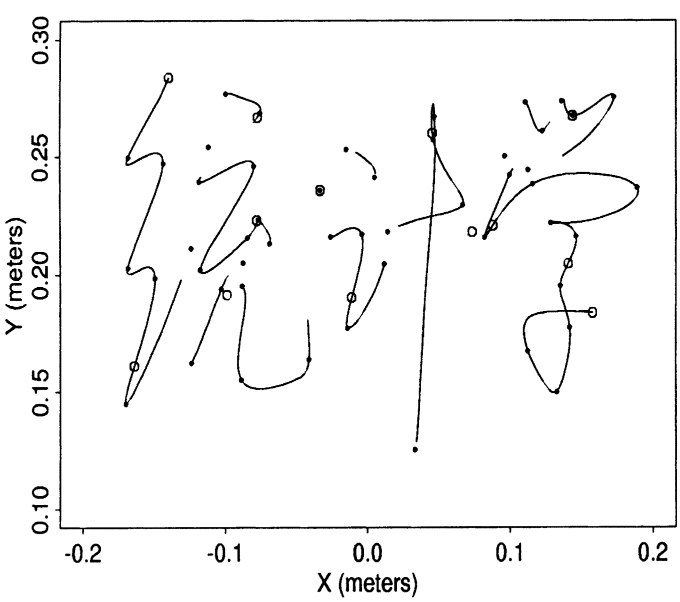

Figure 1. The Average, After Registration, of 50 Writings of "Statistical Science" in Chinese. The dots indicate the positions of peaks in the magnitude of the average acceleration vector, and circles indicate the times 0 (.5) 6.0 in seconds.

An essential next step is registration of the data. This involves aligning coordinate functions by allowing for smooth monotone but nonlinear transformations of clock times so as to allow for temporal variation across handwriting samples. A registration technique of Ramsay and Li (1998) based on a smooth monotone function family considered by Ramsay (1998) is used here.

Once the curves are registered, so that temporal variation is largely removed, the next step is to model amplitude variation from one sample to another by a linear differential equation. For these data, a homogeneous equation of the second order in velocity was estimated, using a technique called principal differential analysis (PDA) described by Ramsay (1996).

The final stage is the interpretation of the equation and evaluation of how well solutions of the equation can recover the original individual handwriting samples. The equation is then applied to the task of discriminating between handwriting samples produced by this subject and those of a second person.

2. THE DATA AND PRELIMINARY ANALYSES

2.1 The Experiment

The handwriting sample consists of 50 replications by one person writing the Chinese characters for "statistical science," displayed in Figure 1. Each replicate consists of three coordinate functions, corresponding to lateral and vertical movement in the writing surface and lifting of the pen off the surface. The notation $x_i(t)$ is used in the following to denote any one of the coordinate functions for record i. When a specific coordinate is specified, $x_{ik}(t)$ is used to indicate the kth coordinate function, $k = 1$ for $X, k = 2$ for Y, and $k = 3$ for Z.

The script samples were recorded by an OPTOTRACK tracking system that recorded the position of a small infrared-emitting diode fastened near the writing tip of the pen 400 times per second. Samples varied in duration, and a preliminary normalization and linear interpolation were performed so that each record comprised 601 values of the three coordinate functions corresponding to 6 seconds, roughly the average duration of the samples. Small rescalings and rotations were also applied to each record in the X-Y plane to produce maximum congruence with the mean record. Techniques for computing these were given by Ramsay, Styan, and ten Berge (1984) and Ramsay (1990). Figure 2 shows the three coordinate functions after these preliminary analyses.

2.2 Smoothing and Derivative Estimation

The OPTOTRACK system is rather accurate, so that the signal-to-noise ratio was high. The standard error of pen position was estimated at about .4 mm, which is about 10^{-3} times the variation of the script in the X-direction. Nevertheless, the signal is very complex, and estimates of derivatives up to order three were required. Consequently, the sampling rate of 400 Hz was essential, and slight smoothing of the raw data was also required.

Because derivatives up to order three were needed, a method that would ensure a reasonably smooth third derivative was essential. Smoothing splines, using the P-spline module available by anonymous ftp from statlib.stat.cmu.ca, and penalized regression splines as described by Ramsay and Silverman (1997) were candidate techniques, and the latter approach was used, partly because the same basis expansion was used later to estimate the linear differential equation. A A package of functions for functional data analysis, called FDA funs and available in both S-PLUS and MATLAB versions by anonymous ftp from ego.psych.mcgill.ca/pub/ramsay/FDAfuns, made these analyses especially convenient. But smoothing splines and local polynomial estimates of derivatives were computed as well, and results were negligibly different from what is reported here using penalized regression splines.

After some experimentation, it was felt that the detail in the records and their derivatives up to order three were adequately represented by approximating each set of observed coordinate values in terms of a set of 105 order-six B-spline functions defined by 101 equally spaced knots, $0, .06, \ldots, 5.94, 6.00$. Because an order-six spline is piecewise of degree five, the third derivative is reasonably

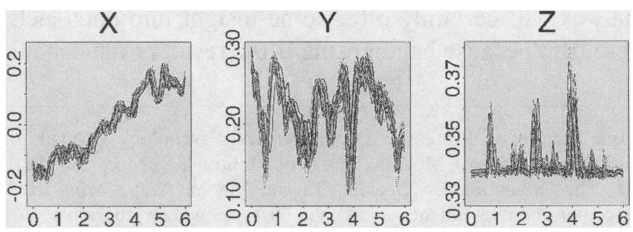

Figure 2. The 50 Coordinate Functions After Initial Normalization and Interpolation to 601 Equally Spaced Values. The time scale is in seconds, and the position scale is in meters.

smooth. Knot refinement methods might have been useful, but I did not explore this option. This expansion did not need a penalty on the roughness of D^5 integrated across [0, 6], but the higher derivatives were unstable for the two or three data points at each end of the record, due largely to additional variability contributed by the pen striking the writing surface at the beginning and end of each script. The roughness penalty

$$P(x) = \int_0^6 (e^{\alpha t} + e^{\alpha(6-t)})[D^4 x(t)]^2 \, dt, \tag{1}$$

where $\alpha = \ln(.1)/6$, smoothed the fourth derivative toward 0, and hence acceleration toward linearity and the weight functions chosen to decay to .1 beyond the extreme 10% parts of the interval caused the penalty to be applied locally. This process was easy to implement within the FDA function package FDAfuns. The quality of the approximation of the data for the first record is displayed in Figure 3.

2.3 Looking at Acceleration

Figure 4 displays the magnitude of the acceleration vector, $(D^2X + D^2Y + D^2Z)^{1/2}$, for the first record. This plot, typical of all records, indicates some rather remarkable characteristics of handwriting. About 50 acceleration pulses are clearly visible, corresponding to about 8 pulses per second. These are separated by near-zero acceleration events, indicating that the many muscles involved are acting in a highly synchronized fashion. In comparison, an expert typist can produce about 500 keystrokes per minute, or around 8 per second, each involving an upstroke and a downstroke. This rate is also comparable to the number of phonemes per second produced in normal speech. Thus this rate of event production ranks with the best that we can achieve. The magnitude of the acceleration pulses is also worth noting; if sustained, an acceleration of 20 m/s^2 would put a satellite in orbit in about 7 minutes, or accelerate a car from 0 to 60 miles per hour in about 1 second.

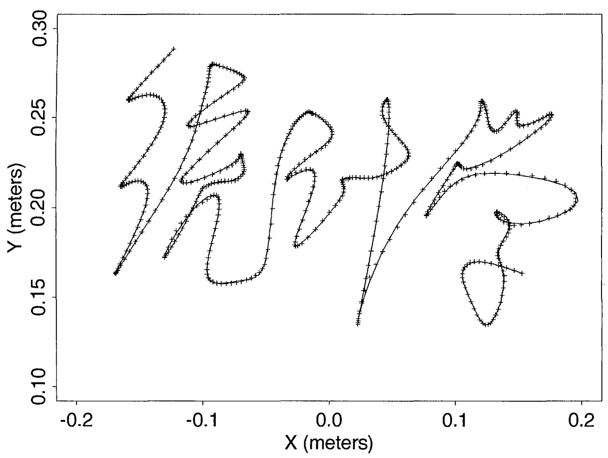

Figure 3. The Points Indicate the Observed Values for the First Record, and the Line Shows the Estimate of the Smooth Curve Achieved by Using Least Squares Approximation by Order Nine B-spline Basis Functions Defined by 101 Equally Spaced Knots.

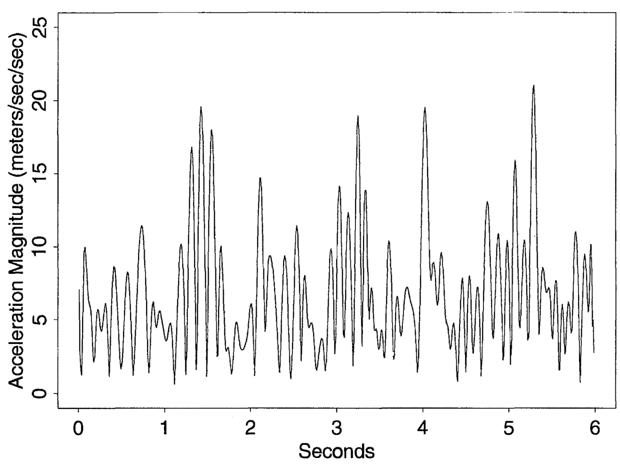

Figure 4. The Magnitude of the Acceleration Vector for the First Record.

3. REGISTRATION OF THE SAMPLES

The data show that although individual records displayed the level of detail in Figure 4, the timing of the events involved varied substantially from record to record. If, for example, one attempted to estimate the average acceleration function by simple cross-sectional averaging, then the time variation would cause the sharp acceleration peaks and troughs to smear and thus be unrepresentative of individual curves. This temporal variation is to be expected; higher velocities are exhibited for some characters and records than for others, due perhaps to factors such as speed, arousal, arm position, and motivation level.

Consequently, the registration process seeks for record i a smooth strictly monotone function h_i common to all three coordinates for record i, such that, for example, x_1 at time $h_1(t)$ is exhibiting about the same behavior as x_2 at time $h_2(t)$, so that the values $x_i[h_i(t)]$ can be viewed as comparable. Function h_i is called a warping function in the engineering literature (Sankoff and Kruskal 1983), and can be viewed as a transformation from clock time to system time.

The warping functions used in this analysis were developed by Ramsay (1998) and have the form

$$h_i(t) = C_i \int_0^t \exp\left[\int_0^u w_i(v)dv\right] du. \tag{2}$$

The function w_i is completely unconstrained, except for being integrable, and thus easier to estimate than h_i itself. The constant C_i is is defined by the constraint $h_i(6) = 6$. The identity transformation corresponds to $w_i = 0$, and a constant value for w_i implies an exponential transformation. In fact, Ramsay (1998) showed that all twice-differentiable strictly monotone functions can be put in this form.

The registration algorithm used involves the following features:

1. Each w_i is expressed as a piecewise constant function defined by a set of 20 equally spaced knots.
2. Each sample is registered toward a single target function x_0 by minimizing the smallest eigenvalue of the cross-

product matrix

$$\sum_{k=1}^{3}\begin{bmatrix}\int x_{0k}^2(t)\,dt & \int x_{0k}(t)x_{ik}(t)\,dt\\ \int x_{0k}(t)x_{ik}(t)\,dt & \int x_{ik}^2(t)\,dt\end{bmatrix}\tag{3}$$

with respect to the values c_{ikl} specifying the height of w_{ik} over interval l. If the smallest eigenvalue is near 0, then the warped coordinate and the target have a nearly linear relationship. A penalty term was also appended to this criterion to control the size of function w_i, thus ensuring that the monotone functions were kept smooth.

3. Initially, the simple cross-sectional mean is used for x_0. But after a first registration cycle, this is replaced by the cross-sectional mean of the warped functions, and a second registration cycle is performed.

This process, called Procrustes fitting, can be repeated until convergence in the cross-sectional mean is achieved. This usually requires only two cycles in practice. Further technical details were provided by Ramsay and Li (1998). Figure 5 displays the warping deformations $h_i(t) - t$.

Figure 1 was constructed from the cross-sectional mean resulting from two iterations of the Procrustes process, applied to the velocity curves. Figure 6 displays all 50 acceleration magnitude functions before and after registration. Registration reveals a remarkable stability in both timing and amplitude for the first character, but rather less stability for the second and third characters. This reflects real variability in how these were formed, rather than any failure of the registration process.

4. THE DIFFERENTIAL EQUATION MODEL

4.1 The Model

The differential equation model proposed for the data is as follows. Let L be the linear differential operator defined by

$$Lx(t) = w_1(t)Dx(t) + w_2(t)D^2x(t) + \cdots + D^m x(t). \tag{4}$$

The model is

$$Lx_i = f_i. \tag{5}$$

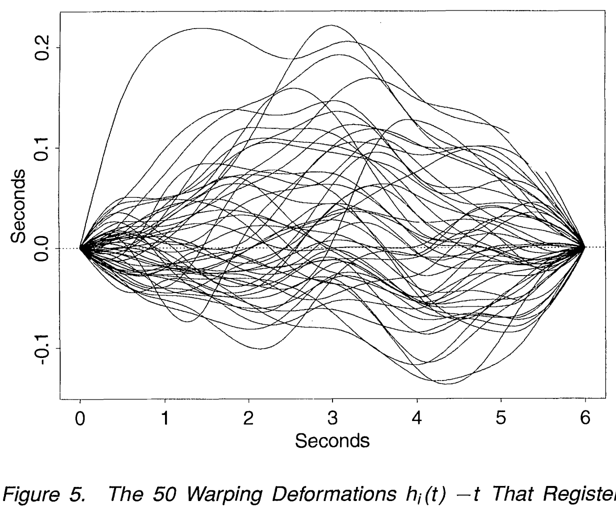

Figure 5. The 50 Warping Deformations $h_i(t) - t$ That Registered the Velocity Functions.

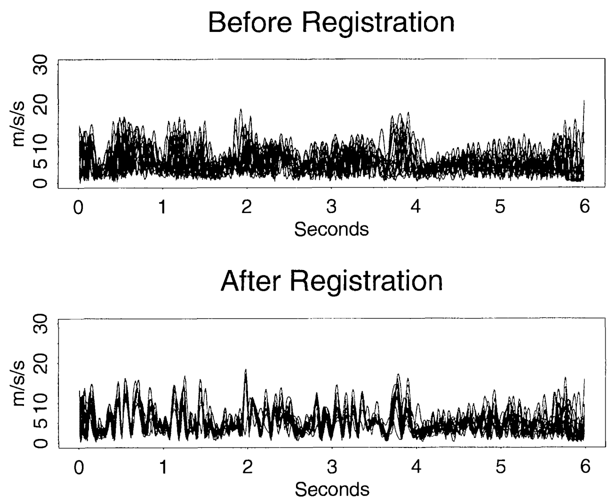

Figure 6. The 50 Acceleration Magnitude Functions Before and After Registration.

The operator L is defined by the $m-1$ coefficient functions w_j, which weight the first two derivatives in a manner that varies over time t. Note that if we regard t as fixed, then the model is a standard regression model, with dependent variable $D^m x_i(t)$ and regression coefficients $-w_j(t)$. The function f, termed the forcing function, corresponds to the residual term in regression analysis. It reflects variation that cannot be explained by the homogeneous differential equation $Lx_i = 0$, and thus its size should be assessed in relation to that of $D^m x$. A separate operator $L_k, k = 1, \ldots, 3$ is required for each coordinate, but in the following discussion subscript k is omitted.

Equation (5) can be viewed as a linear differential equation of order $m-1$ in velocity Dx, and the use of velocity as the function to be modeled here ensures that the results will be translation invariant. Moreover, to further simplify the analysis, the linear trend in the X-coordinate was first removed by regressing these functions on time and replacing them by their residuals. In effect, this places the origin of the system within the hand–arm structure rather than in a fixed point, and this could be viewed as a reasonable change of coordinates. This transformation also centers the velocity functions on 0.

The order m of the operator (4) can be chosen in part by comparing values of a fit measure such as (7) for various values. Moreover, for a fixed m, there is also the possibility of omitting certain coefficients functions w_j. However, the theory of linear differential equations tells us that certain choices will be illogical for these data; $m = 2$ implies strictly positive velocity, and $m = 3$ combined with $w_1 = 0$ implies strictly monotone velocity. The simplest model that has a chance of succeeding for this problem is $m = 3$ and $w_2 = 0$.

4.2 Estimating the Equation

The PDA approach developed by Ramsay (1996) proceeds by expanding each coefficient function in terms of a fixed number of basis functions. Let $\phi_q, q = 1, \ldots, Q$ be a set of Q such basis functions, and let ϕ denote the Q-dimensional vector function $(\phi_1, \ldots, \phi_Q)^t$. Then it is assumed that

$$w_j \approx \sum_q \mathbf{c}_{j_q} \phi_q, \qquad j = 1, \ldots, m-1, \tag{6}$$

where $(m-1)Q$ coefficients $\mathbf{c}_{j_q}$ define the approximations and require estimation from the data. Let vector $\mathbf{c}$ contain these coefficients, where index q varies inside index j. For these data, where a high level of detail is required in each w_j, the basis functions ϕ_q were the same 105 order-six B-spline functions used to approximate the raw observations.

The fitting criterion to be minimized for each coordinate is

$$\hat{F}(\mathbf{c}|x) = \sum_{i=1}^{50} \int_0^6 f_i^2(t)\, dt. \tag{7}$$

That is, I aim to minimize the L^2 norm of the residual functions f_i for the homogeneous model $Lx_i = 0$. This criterion is quadratic in the coefficients, and expansion of (7) yields t,

$$\hat{F}(\mathbf{c}|x) = C + \mathbf{c}^t \mathbf{R} \mathbf{c} + 2\mathbf{c}^t \mathbf{s}, \tag{8}$$

where constant C does not depend on $\mathbf{c}$. Matrix $\mathbf{R}$ is of order $(m-1)Q$ and symmetric and contains order Q submatrices of the form

$$\mathbf{R}_{j_1, j_2} = \int \phi(t)\phi^t(t) E[D^{j_1} x D^{j_2} x](t)\, dt, \tag{9}$$

where

$$E[D^{j_1} x D^{j_2} x](t) = 50^{-1} \sum_i^{50} D^{j_1} x_i D^{j_2} x_i(t). \tag{10}$$

Similarly, $(m-1)Q$-dimensional vector $\mathbf{s}$ contains subvectors $\mathbf{s}_j = \int \phi(t) E[D^j x D^m x](t)$. These integrations are carried out in practice by numerical methods. Finally, $\mathbf{c}$ is the solution of the equation $\mathbf{s} = -\mathbf{Rc}$.

4.3 The Results

If the homogeneous equation Lx_i is successful, then the size of the f_i's should be small relative to that of $D^m x$ and should exhibit white noise characteristics. A squared multiple correlation measure of fit is defined by

$$R^2 = \left[\sum_{k=1}^{3} \sum_{i=1}^{50} \int_0^6 (D^m x_{ik})^2(t) - f_{ik}^2(t)\, dt \right] \div \sum_{k=1}^{3} \sum_{i=1}^{50} \int_0^6 (D^m x_{ik})^2(t)\, dt. \tag{11}$$

For $m = 3$ and 4 and full complements of weight functions, R^2 was .952 and .994. For the simpler model $m = 3, w_2 = 0, R^2 = .239$. I opted for the second-order ($m = 3$) model because the fit was good, a second-order equation is usual for closed mechanical systems, and because leaving some signal in the forcing functions might highlight interesting effects not represented by the homogeneous equation.

Figure 7 displays the mean forcing function for the X-coordinate, along with its pointwise 95% confidence limits.

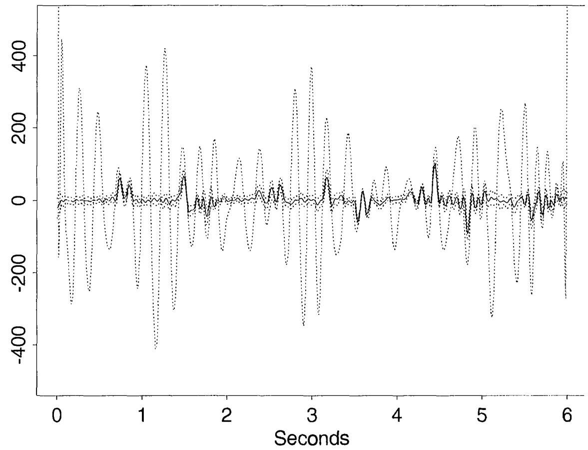

Figure 7. The Average Forcing Function f for the X-Coordinate (Solid Line) and Pointwise 95% Confidence Limits (Dotted Lines). The most widely varying dotted line indicates the average third derivative for comparison purposes.

Also displayed for comparison purposes is the average of $D^3 x$. On average, the forcing function is small relative to the size of the third derivative over most of the time interval. However, there are several sharply localized events where the average forcing function deviates strongly from 0. These points correspond to especially sharp changes in direction that the homogeneous second-order equation cannot accommodate.

Figure 8 shows the coefficient functions for the X-coordinate. A comparison can be made with what would been seen were the system exhibiting purely harmonic motion and thus behaving like a spring or pendulum. In that case, $w_2 = 0$, and w_1 takes on a positive constant value. In fact, w_2, although obviously variable in important ways, is not greatly different from 0, and w_1 has values varying around about 700. This corresponds to a frequency of $\sqrt{700/(2\pi)} \approx 10$ cps, or around 20 strokes per second. In fact, this is about what is shown in the script. Again, results for the other two coordinates are rather similar.

The solution space to the nonhomogeneous equation $Lx = f_i$ is spanned by the three linearly independent functions u_j satisfying the homogeneous equation $Lx = 0$,

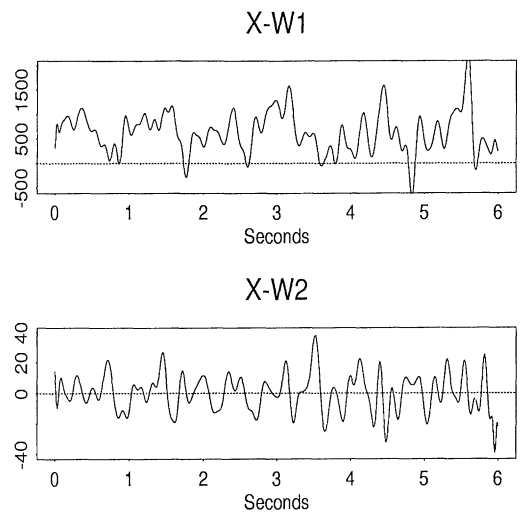

Figure 8. The Estimated Weight Coefficient Functions w_1 and w_2 Defining Linear Differential Equations (4) for the X-Coordinate.

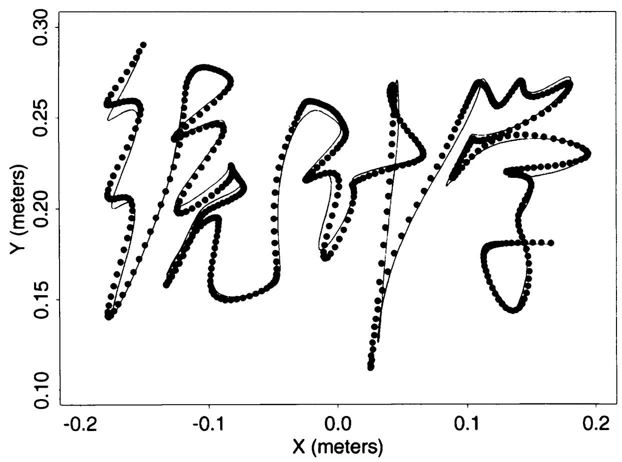

Figure 9. The Observed Script for the Last Record (X-Y Plot Only) is Given by the Points, and the Approximation to This Script Based on Solving the Homogeneous Differential Equation is the Solid Line.

among which is $u_1 = 1$, and by the function $\int G(t, w)\bar{f}(w)\, dw$, where in this case $\bar{f}$ is the mean forcing function averaged across all 50 records and shown in Figure 7, and G is the Green function associated with L. Standard differential equation-solving techniques, such as the methods of Picard or Runge–Kutta, can be used to estimate these solution functions u_j. One can then use this, along with the mean forcing function, $\bar{f}$, as a basis for reconstituting the original functions. Because this reconstruction is based on three functions estimated from the data—namely w_1, w_2, and $\bar{f}$—it can be said that the reconstruction has three functional degrees of freedom. Figure 9 displays both the original X-Y coordinates and the reconstituted approximation for the last record. The figure shows that the approximation is fairly satisfactory.

The equation must not only capture the features of a typical record, but also should model the record-to-record variation. To see this, I computed the function

$$R^2(t) = \frac{\mathrm{var}[\hat{X}(t)] + \mathrm{var}[\hat{Y}(t)]}{\mathrm{var}[X(t)] + \mathrm{var}[Y(t)]}, \tag{12}$$

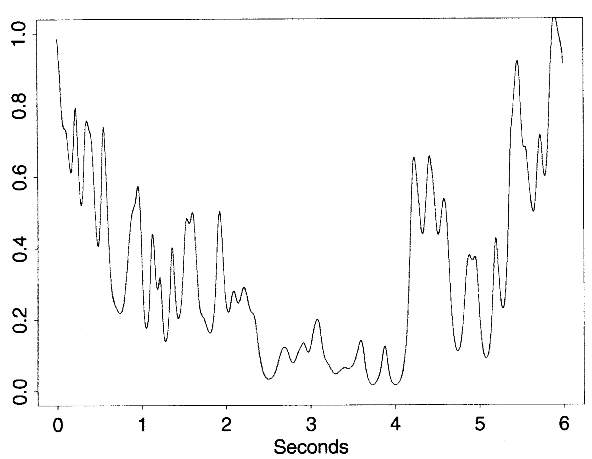

Figure 10. The Ratio of the Total Cross-Record Variation in the Reconstructed X-Y Coordinates to That in the Observed Coordinates.

with the variances taken over the 50 records and the approximated coordinate functions $\hat{X}$ and $\hat{Y}$ the reconstructions of the coordinates using the differential equation. Figure 10 indicates that the success of the model varied considerably over the three characters. It was able to reproduce around 50% of the variation in the first and third characters, but was too rigid for the central character.

Because it is the task of the estimated differential operator to minimize the size of the forcing function, it is primarily these f_i's, and thus the results in Figure 7, that should be taken as evidence whether or not the model has worked. That the recovered position functions look good is encouraging, but if this were the main aim, one would expect a functional version of principal components analysis, which uses only position information, to do a better job. That is, the differential equation model is a dynamic model linking together the variation in time of four orders of derivatives, rather than a static model designed merely to capture the shape characteristics of the position curves.

5. CLASSIFICATION BY THE DIFFERENTIAL OPERATOR

A likely application of the estimated differential operators is to the authentication of handwriting by using the dynamic characteristics of a sample. Although forgery of a static signature is common, it seems hard to imagine how a train of features unfolding at around 10 per second could be reproduced consistently by anyone but the original author, whose neural and motor control systems have been tuned to the task over thousands of trials and years of practice.

A sample of 18 records for the same Chinese characters were collected from a second subject, and the same smoothing and registration procedures were applied. The differential operators L for the X- and Y-coordinates were estimated from a random sample of 32 records for the original subject, and these were then applied to that subject's remaining 18 records, as well as to the 18 records

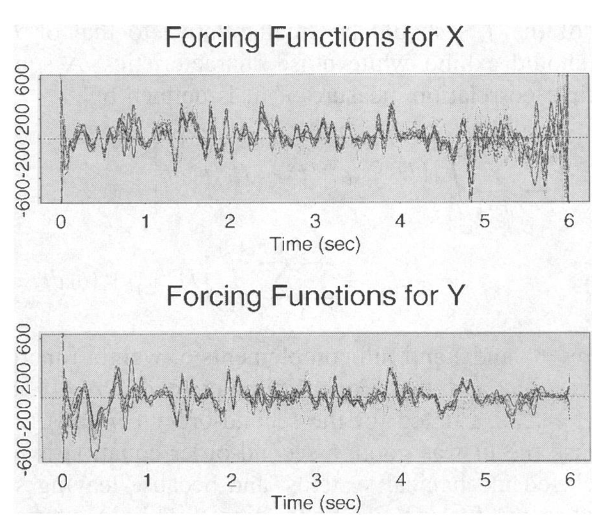

Figure 11. Forcing Functions L_1x_{i1} and L_2y_{i2} Computed From a Sample of 18 Records From a Second Subject, With the Differential Operators Estimated From the First Subject's Data.

from the second subject. Figure 11 shows the forcing functions Lx for the second subject. Note that these forcing functions do not display the white-noise character shown in Figure 7. They are also substantially larger in size; the measures

$$\mathrm{RMS}_i = \sqrt{\mathrm{median}([L_1 x_{i1}(\mathbf{t})]^2 + [L_2 y_{i2}(\mathbf{t})]^2)},$$

where $\mathbf{t}$ is the vector 601 equally spaced sampling points, vary from 31 to 46 $\mathrm{m/s}^3$ for the first subject and from 88 to 119 $\mathrm{m/s}^3$ for the second subject. Thus any threshold positioned between 46 and 88 would have classified the records perfectly.

6. DISCUSSION AND CONCLUSIONS

The second-order differential equation model has three functional degrees of freedom for each coordinate, corresponding to the two weight functions and the forcing function. These seem to do a reasonable job of accounting for these 50 actual curves, so that a fair amount of data compression has been achieved.

A number of modifications of this model seem plausible, but perhaps beyond the capacity of this small sample to support. The model used here has treated each coordinate separately, but it seems possible that one could aim at coupled systems of equations, so that variation in X can be modeled as connected with variation in Y, for example. This is possible to do in principle with PDA, but the existing model seems to be about as complex as required for such a modest sample.

One may also ask whether the correct coordinate system was used. Is the Euclidean system the right one, or would some aspects of a polar coordinate system show more simplicity? However, it can at least be said that the order of the operator L annihilating the curves has been identified, and this would remain invariant with respect to any one-to-one change of coordinates.

It was indicated earlier that a differential equation offers a rich structure for modeling stochastic variation. I have considered implicitly or explicitly (a) the usual observational error superimposed on a smooth function, (b) a stochastic forcing function that varies from record to record, (c) the possibility that the weight functions w_j are stochastic, and (d) a stochastic aspect to time itself, as indicated in the deformation functions in Figure 5. Moreover, I considered variation from one writer to another. I would welcome the development of sharper methods for estimating these components of variation in these and other types of functional data.

Finally, this study has also shown the large amount of useful information available in derivatives. A linear differential equation system is only one possible approach to exploiting this information, but what seems to be best illustrated by these data can be seen directly in Figure 4 without any recourse to a model. The action in functional data may be more obvious at the level of derivatives than at the level of the curves themselves.

[Received July 1998. Revised October 1999.]

REFERENCES

Faure, C., Keuss, P., Lorette, G., and Vinter, A. (Eds.) (1994), *Advances in Handwriting and Drawing: A Multidisciplinary Approach*, Paris: Europia.

Plamondon, R., and Guerfali, W. (1998), "The Generation of Handwriting With Delta-Lognormal Synergies," *Biological Cybernetics*, 78, 119–132.

Ramsay, J. O. (1990), "MATFIT: A FORTRAN Subroutine for Comparing Two Matrices in a Subspace," *Psychometrika*, 55, 551–553.

——— (1996), "Principal Differential Analysis: Data Reduction by Differential Operators," *Journal of the Royal Statistical Society*, Ser. B, 58, 495–508.

——— (1998), "Estimating Smooth Monotone Functions," *Journal of the Royal Statistical Society*, Ser. B, 60, 365–375.

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

Ramsay, J. O., Styan, G. P. H., and ten Berge, J. (1984), "Matrix Correlation," *Psychometrika*, 49, 403–424.

Ramsay, J. O., and Li, X. (1998), "Curve Registration," *Journal of the Royal Statistical Society*, Ser. B, 60, 351–363.

Sankoff, D., and Kruskal, J. B. (Eds.) (1983), *Time Warps, String Editors, and Macromolecules: The Theory and Practice of Sequence Comparison.* Reading, MA: Addison-Wesley, pp. 125–160.