

# Functional models of growth for landmark data \*

John T. Kent, Kanti V. Mardia, Richard J. Morris and Robert G. Aykroyd  
Department of Statistics, University of Leeds, Leeds LS2 9JT, UK

## Abstract

Growth models for shape are investigated for landmark data. First the data are given a Euclidean representation using Procrustes tangent coordinates. Roughness penalties are defined for directions of growth in space and for rates of growth in time. These penalties are then combined to give a joint space-time roughness penalty. Transforming the data to bases defined by principal warps in space and time facilitates model specification and fitting, both for parametric and nonparametric models. A generalized cross validation criterion can be used to choose a smoothing parameter for a nonparametric smoothing spline-type model. Fitted models can be interpreted either just in terms of the finite set of landmarks at the finite set of data times, or in terms of a deformation of space which varies continuously through time. The methods are illustrated on a set of rat data.

## 1 Introduction

The purpose of this paper is to construct growth models for the shapes of biological objects. For simplicity we work with objects in  $M = 2$  dimensions, but the methodology extends to higher dimensions. The basic approach is to construct and fit a time-varying deformation of  $\mathbb{R}^M$  which in particular deforms the object of interest. There are two main ingredients in this approach:

- (a) a roughness penalty on functions in space, specifying the smoothness of possible directions of growth;
- (b) a roughness penalty on functions in time, specifying the smoothness of possible rates of growth.

These penalties, together with fixed reference sets of landmarks and times, respectively, determine finite dimensional spaces of functions of space and time. The direct product of these spaces will form the framework for our modelling approach in space-time.

Since growth is usually dominated by increasing size, it is helpful to look at changes in size and shape separately. For the purposes of this paper we ignore changes in size and limit attention to changes in the shape of the object. Recall that the “shape” of an object comprises all the geometric information about the object except for location, rotation and size (e.g. Dryden and Mardia, 1998).

This paper extends earlier work by Morris et al. (1999a,b, 2000) and Kent et al. (2000). The outline of the paper is as follows. In Section 2, we describe how to represent the data in a form suitable for fitting growth models. In Section 3 roughness penalties on space and time are

---

\*This paper is dedicated to F. James Rohlf on the occasion of his 65th birthday.

constructed and combined to give a joint space-time penalty. Also, a transformation of the data is given which facilitates the fitting of these smoothing models. Section 4 discusses various sorts of parametric and nonparametric models and illustrates their application on a set of growth data for rat skulls.

## 2 Representing the data

Suppose landmark data are available on different individuals at a common set of ages, taking the form of a 4-way array  $\{x_{nkmh}\}$  where

$n = 1, \dots, N$  labels different individuals,  $k = 1, \dots, K$  labels different landmarks,  
 $m = 1, \dots, M$  labels different coordinates,  $h = 1, \dots, H$  labels different times  $t_1, \dots, t_H$ .

It is convenient to represent these data as a collection  $\{x_{nh}\}$  of  $K \times M$  matrices. In general bold-face will be reserved for  $K \times M$  matrices.

Since the shape of an object determines its coordinates only up to a similarity transformation, it is necessary to reduce the data to just the shape information. We do this using Procrustes tangent coordinates about a centered and scaled “mean” configuration  $\mu$ , say. A convenient choice for  $\mu$  is the generalized Procrustes estimate based on all  $NH$  configurations, but the exact choice does not matter. Let  $v_{nh}(K \times M)$  denote the (centered rather than Helmertized) Procrustes tangent coordinates of the data  $x_{nh}$ .

Next, assuming all  $N$  individuals are i.i.d., we take a sample average of the Procrustes coordinates to get averaged data  $\bar{v}_h$ . It is this form of the data to which we wish to fit a growth model. For this purpose it is convenient to rewrite the data as a  $KM \times H$  matrix  $W$ , say, with  $h$ th column of  $W$  defined by stacking the  $M = 2$  columns of  $\bar{v}_h$  on top of one another.

## 3 Details of roughness penalties

For simplicity of presentation, we focus on two specific examples of roughness penalties, though other choices are also possible.

For a real-valued function of time,  $\psi(t)$ ,  $t \in \mathbb{R}^1$ , we shall use the cubic spline penalty

$$Q_1(\psi) = \int \left( \frac{d^2\psi}{dt^2} \right)^2 dt \quad (3.1)$$

where the integral is over the whole line. Note that the nullspace of this penalty is 2-dimensional, spanned by the functions  $1, t$ . Let  $t_1, \dots, t_H$  denote the common set of times at which the data are observed. Associated with these times is an  $H \times H$  symmetric positive semidefinite “bending energy” matrix  $B$ , say, of rank  $H - 2$ . The eigenvectors and eigenvalues of  $B$  play an important role. Associated with each standardized eigenvector  $\gamma$ , say, of  $B$  is a function  $b(t)$ , say, such that

1.  $b(t_h) = \gamma_h$ , so that  $b$  interpolates the values in the eigenvector, and
2.  $b$  minimizes the penalty  $Q_1(b)$  over all such interpolating functions, and  $Q_1(b) = \beta$ , say, the corresponding eigenvalue.

The eigenvectors corresponding to the nonzero eigenvalues are called “principal warp vectors” and the corresponding functions are called “principal warp functions”.

For our modelling purposes it is useful to combine all the eigenvectors (excluding the constant vector) into an  $H \times (H - 1)$  column orthonormal matrix  $G$ , say. Denote the corresponding eigenvalues of  $B$  by  $\beta_h$ ,  $h = 1, \dots, H - 1$ , arranged in nondecreasing order. Note  $\beta_1 = 0$ . The columns of  $G$  roughly correspond to the effects of orthogonal polynomials at the data times.

A similar construction can be carried out in space. For a real-valued function in  $M = 2$  spatial dimensions,  $\phi(s)$ ,  $s \in \mathbb{R}^2$ , we shall use the thin plate spline penalty,

$$Q_2(\phi) = \int \left( \left( \frac{\partial^2 \phi}{\partial s_1^2} \right)^2 + 2 \left( \frac{\partial^2 \phi}{\partial s_1 \partial s_2} \right)^2 + \left( \frac{\partial^2 \phi}{\partial s_2^2} \right)^2 \right) ds \quad (3.2)$$

where the integral is over all of  $\mathbb{R}^2$ . Note that the nullspace of this penalty is 3-dimensional, spanned by the functions  $1, s[1], s[2]$ , where square brackets denote the coordinates of  $s$ . Let  $\{s_1, \dots, s_K\}$  denote a fixed collection of sites in  $\mathbb{R}^2$ . (In our applications we shall use the  $K$  rows of  $\mu$ .) Then a  $K \times K$  bending energy matrix can be constructed with analogous properties to the above paragraph, though this time of rank  $K - 3$ .

Since a deformation can be constructed from a pair of functions  $\mathbb{R}^2 \rightarrow \mathbb{R}^1$ , we combine together two copies of the eigenvectors. After removing 4 degrees of freedom for the constraints in Procrustes tangent space, we are left with a  $2K \times (2K - 4)$  column orthonormal matrix  $F$ , say. The first two columns of  $F$  represent linear functions orthogonal to the similarity transformations (that is, Bookstein's "uniform" component; see Bookstein (1995) and Mardia (1995, Section 7). The remaining columns of  $F$  come in pairs:  $\begin{bmatrix} \gamma & 0 \\ 0 & \gamma \end{bmatrix}$ , where  $\gamma$  is a  $K \times 1$  eigenvector of the bending energy matrix for the thin-plate spline. Also, let  $\alpha_{2j-1} = \alpha_{2j}$ ,  $j = 1, \dots, K - 2$  denote the corresponding eigenvalues in nondecreasing order, each listed twice. Note  $\alpha_1 = \alpha_2 = 0$ .

Then  $W$  can be written in the form

$$W = \nu 1_H^T + F A G^T, \quad \text{that is,} \quad F^T W G = A = (a_{jh}). \quad (3.3)$$

The intercept  $\nu$  is of no interest and depends on the choice of  $\mu$ . The matrix of coefficients  $A((2K - 4) \times (H - 1))$  is of key importance for specifying different models. If the eigenvectors in  $F$  and  $G$  are interpolated to yield a pair of space-time functions  $(\Phi_1(s, t), \Phi_2(s, t))$  then the penalty function

$$Q_3(\Phi_1, \Phi_2) = \sum_{i=1}^2 \int \left( \left( \frac{\partial^4 \Phi_i}{\partial s_1^2 \partial t^2} \right)^2 + 2 \left( \frac{\partial^4 \Phi_i}{\partial s_1 \partial s_2 \partial t^2} \right)^2 + \left( \frac{\partial^4 \Phi_i}{\partial s_2^2 \partial t^2} \right)^2 \right) ds dt, \quad (3.4)$$

where the integral is over  $s \in \mathbb{R}^2$ ,  $t \in \mathbb{R}$ , reduces to

$$\sum_{j=1}^{2K-4} \sum_{h=1}^{H-1} a_{jh}^2 \alpha_j \beta_h,$$

which depends just on the coefficients and the eigenvalues.

## 4 Various models for the rat data

We consider a set of rat growth data described and analyzed in Bookstein (1991). The data are obtained from a two-dimensional midsagittal section of the calvarium (the skull without the lower jaw). There is complete information on  $N = 18$  rats at  $H = 8$  times (or ages) on  $K = 8$

landmarks. To facilitate the model fitting, we replace the actual age by a “pseudo-age” given by the average centroid size at each age. For the purposes of this paper we shall ignore any differences between the individual rats.

For this rat dataset, let  $A = F^T W G$  denote the  $12 \times 7$  doubly rotated version of the Procrustes tangent matrix for the mean data. The value of  $A$  is given as follows, where the rows label the eigenvectors in space and the columns label the eigenvectors in time.

The matrix  $A$

-0.128	0.005	-0.003	0.010	0.004	-0.003	0.003
0.012	-0.049	0.003	0.007	0.004	-0.001	-0.001
0.079	-0.008	0.013	0.003	-0.002	-0.003	0.002
0.064	-0.007	0.011	-0.001	-0.001	-0.001	0.000
-0.057	0.013	-0.005	-0.004	0.001	0.002	0.001
-0.004	0.002	0.002	-0.003	-0.002	0.000	-0.001
0.016	0.003	-0.002	0.000	0.001	-0.001	-0.003
0.021	-0.006	0.004	-0.003	-0.001	0.000	-0.001
0.006	-0.008	-0.009	-0.004	-0.001	0.003	0.002
-0.013	-0.001	-0.001	0.001	-0.001	0.000	0.000
0.018	-0.003	-0.002	-0.001	0.001	0.000	-0.002
-0.037	0.003	-0.002	0.000	0.000	0.000	-0.001

Some important models and their application to the rat data are described below, with the fitted parameter matrix denoted  $\hat{A}$ . For each fitted model, the residual sum of squares (RSS) comparing a given model to the full model is quoted. A selection of fitted models is plotted in Figure 1.

1. *Parametric models, full rank.* These models can be specified in terms of the nonzero entries in  $\hat{A}$ . For example, the notation [1:6,1:2] will mean that the block of entries specified by the first 6 rows and first two columns will be allowed to be nonzero.
  - (a) full = [1:12, 1:7]. (RSS= 0) In this case there is no data reduction (other than the averaging over individuals in the first place). See Figure 1(a).
  - (b) linlin = [1:2,1]. (RSS = 0.01986) This is the simplest possible model, linear in space (a two-dimensional subspace) and linear in time (one-dimensional). Thus this model consists of a constant growth rate in a single direction arising from a linear transformation in two dimensions. This model captures some of the main features of the data, but fails to capture the curvature of the paths, and does not fit well at several landmarks.
  - (c) fulllin = [1:12, 1]. (RSS = 0.00363) In this case there is full flexibility in growth direction, but the growth trajectory is still linear in this single direction. In Figure 1(b) we see that the main pattern of growth is captured, but the fit is not very good at the left-hand landmarks and the curvature of the paths is not captured at all. Overall we see that under this model the top landmarks move downwards and slightly inwards; the bottom landmarks move up and outwards.
  - (d) 2quad = [1:6, 1:2]. (RSS = 0.00350) This model is specified by a pair of growth directions (lying in the space spanned by the linear terms plus the first two principal warps in space), one direction progressing at a linear rate and the second according

to the first principal warp in time (which is roughly like a quadratic). This model was suggested in Kent et al. (2000), but the following model now seems preferable.

- (e)  $sp = [1:12, 1] + [1:2, 2]$ . (RSS = 0.00120) This model is called “special” because the nonzero parameter values do not form a rectangular block. It captures the curvature in the paths through a term which is second order in time and linear in space. Figure 1(c) shows that this model yields a good fit to the data.
- (f)  $linfull = [1:2, 1:7]$ . (RSS = 0.01723) This model allows growth in the two linear directions in space, but at arbitrary rates in time. The growth paths are curved but do

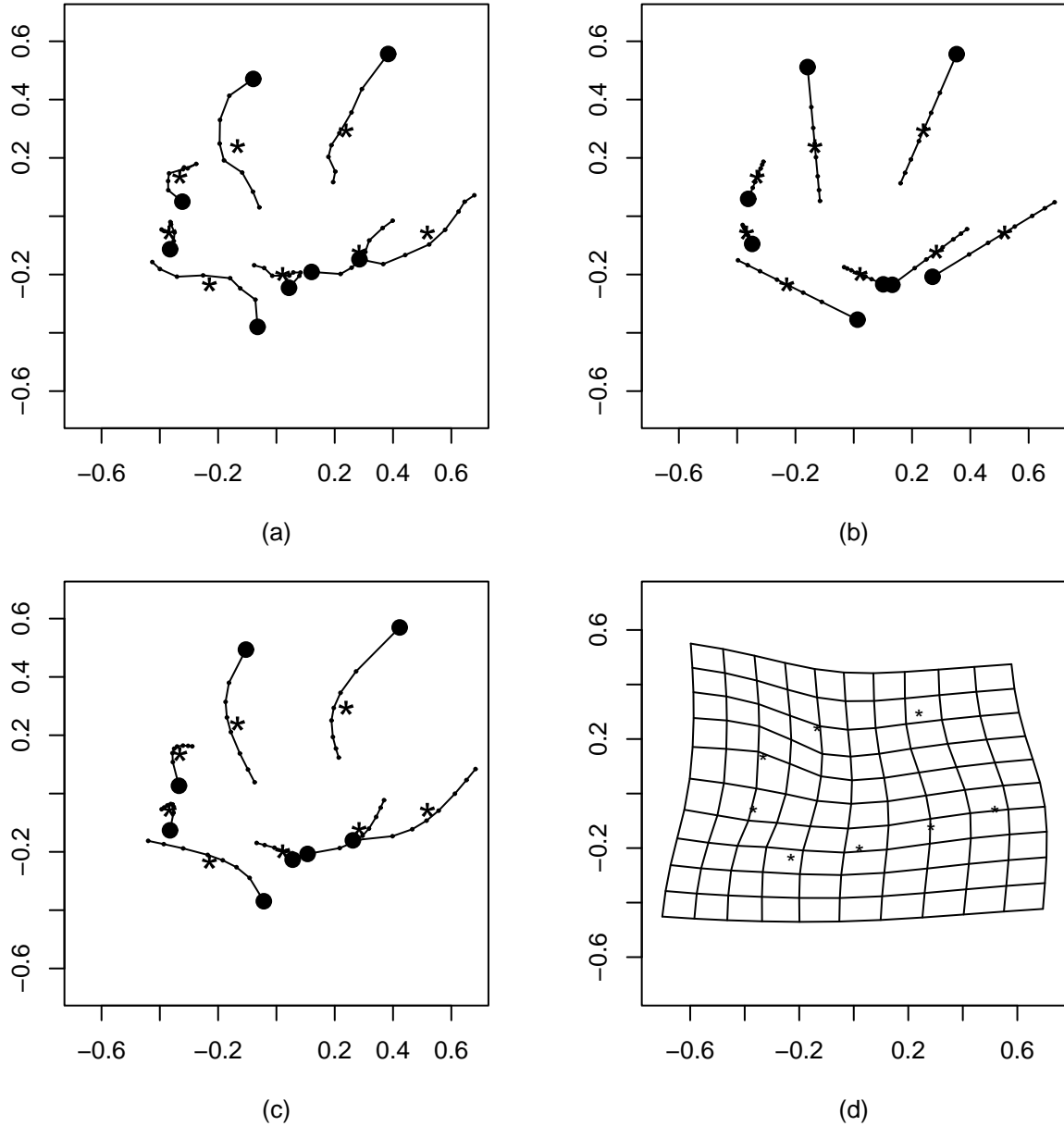


Figure 1: Fitted growth models for the rat data: (a) full, (b) fulllin, (c) sp, with the growth patterns blown up by a factor of 5 for clarity. Each “\*” represents a landmark of the the Procrustes mean shape  $\mu$ , and each closed circle represents the position of a landmark at the initial time. Part (d) shows the grid deformation, without an expansion factor, between the initial and final times for the sp model.

not match the data very well.

(g)  $\text{null} = [1:12, 1] + [1:2, 2:7] = \text{fulllin} + \text{linfull}$ . (RSS = 0.00100) This is the most general model that has 0 roughness penalty and fits the data surprisingly well. Visually the fit is similar to the special model.

2. *Parametric models, reduced rank*. Start with a rectangular block of coefficients in  $A$ , take a singular value decomposition, and retain the dominant components. The simplest such model which fits the data well can be written as  $\text{full.2} = [1:12, 1:7, 2]$ ; i.e. take the dominant 2 components of the singular value decomposition of the whole matrix  $A$ , with RSS = 0.00045. This fitted model is very closely related to the exploratory analysis based on two principal components in Le and Kume (2000).
3. *Nonparametric*. A nonparametric spline-like model can be defined by minimizing a combination of a goodness of fit term plus  $\lambda$  times the penalty term, where the minimization is over the finite dimensional vector space generated by products of the principal warp functions (including linear functions) for space and time, respectively. The terms of the fitted model are given by the estimates  $\hat{a}_{jh} = a_{jh} / (1 + \lambda \alpha_j \beta_h)$ . The null model is a special case as  $\lambda \rightarrow \infty$ . A generalized cross-validation criterion can be defined and leads to the optimal smoothing parameter  $\lambda = .00071$  for which RSS = 0.00004. This value is very small suggesting that little smoothing has taken place, so that the nonparametric fit is very close to the full model

## Discussion

A glance at the matrix  $A$  shows that the largest values fall in the first column and the first two entries in the second column. This observation suggests that the special parametric model will provide a good-fitting parsimonious model, as verified by the RSS value. The null model includes the special model and yields a similar fit. Further, the reduced rank model  $\text{full.2}$  also captures these effects and yields a similar fit. Since the nonparametric model includes the null model, it also gives a good fit. However, the lack of strong structure in the  $[3:12, 2:7]$  block of data seems to lead to a small smoothness parameter  $\lambda$  and to over-fitting of the data.

Overall the most useful fitted model seems to be the special parametric model. The growth can be decomposed into two components: a linear (in time) growth in a general spatial direction, together with a “quadratic” component (in time) which is restricted to a linear transformation in space. An illustration of the deformation involved between the starting and final times under this model is plotted (with no expansion factor) in Figure 1(d).

## Acknowledgement

This work was supported by a research grant from the Engineering and Physical Sciences Research Council.

## References

- [1] Bookstein, F.L. (1991) *Morphometric Tools for Landmark Data*. Cambridge University Press.

- [2] Bookstein, F.L. (1995) A standard formula for the uniform shape component in landmark data. *NATO Proceedings*.
- [3] Dryden, I.L. and Mardia, K.V. (1998) *Statistical Shape Analysis*. Wiley, Chichester.
- [4] Kent, J.T., Mardia, K.V., Morris, R.J. and Aykroyd, R.G. (2000) Procrustes growth models for shape. In *Proceedings of the First Joint Statistical Meeting New Delhi, India*.
- [5] Le, H. and Kume, A. (2000) Detection of shape changes in biological features. *J. Microscopy* 200, 140–147.
- [6] Mardia, K.V. Shape advances and future perspectives. In *Proceedings in Current Issues in Statistical Shape Analysis*, 57-75. Edited by K. V. Mardia and C. A. Gill. Leeds University Press.
- [7] Morris, R.J., Kent, J.T., Mardia K.V., Fidrich, M., Aykroyd, R.G. and Linney, A. (1999a) Analysing growth in faces. In *Proc. Conf. Imaging Science, Systems and Technology 1999, Las Vegas*, pp. 404–410.
- [8] Morris, R.J., Kent, J.T., Mardia, K.V., Fidrich, M., Aykroyd, R.G. and Linney, A. (1999b) Exploratory analysis of facial growth. In *Proceedings in Spatial Temporal Modelling and its Applications*. Leeds University Press, pp 39–42.
- [9] Morris, R.J., Kent, J.T., Mardia, K.V. and Aykroyd, R.G. (2000) A parallel growth model for shape. In *Proceedings Medical Imaging Understanding and Analysis 2000*, pp. 171–174.