This is a repository copy of *A Statistical Analysis of the Cardioid Radial Growth Model*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/185518/

Version: Accepted Version

Book Section:

A statistical analysis of the cardioid radial growth model

John T. Kent[1], Kanti V. Mardia[1,2], Luigi Ippoliti[3], and Pasquale Valentini[3]

`j.t.kent@leeds.ac.uk, k.v.mardia@leeds.ac.uk,`
`ippoliti@unich.it, pvalent@unich.it`

[1]Department of Statistics, University of Leeds LS2 9JT, United Kingdom
[2]Department of Statistics, University of Oxford, Oxford OX1 3LB, United Kingdom
[3]Department of Economics, University G. d'Annunzio, Chieti-Pescara, Italy

Abstract

A new two-parameter "full exponential cardioid" radial growth model for two-dimensional geometric objects is proposed and analyzed. The model depends additionally on two rotation parameters and on two seeds about which the growth is centered, plus a choice of three possible assumptions about statistical errors. If the seeds are assumed known, the remaining parameters can be estimated in closed form. Comparisons are given to earlier approaches. Two examples are given, one for a set of simulated data and one for a set of rat calvarial data.

1 Introduction

This paper revisits the Todd and Mark (1981a) "revised cardioid strain (RCS) radial growth model", a simple mechanistic model for craniofacial growth. Ramanathan et al. (2009) give a recent summary of its history and applications. A key property of the model is that the growth rate is greater near the bottom of the head than at the top so that it captures some of the key characteristics of real growth for humans and other animals. Further, the model gives a very simple description of growth, with only one growth parameter (plus 6 registration parameters, often treated as known) to fit.

The two key assumptions of the model are that (a) there is a "seed" inside the skull about which growth is centered, and (b) the rate of growth at a particular point on the head depends only on the angle from vertical. Various attempts have been made to justify

the model in terms of physical principles. For example, Todd and Mark (1981a) motivate the model using hydrostatics and gravity.

The cardioid growth model can be considered as a very simplified version of the pattern theoretic growth model of Grenander et al. (2007) and Portman (2009, p. 19), which is based on infinitesimal growth patterns about a seed. However, the pattern theoretic growth model is much richer, albeit more complicated, because cumulative growth involves a series of iterated diffeomorphisms and many different seeds.

The use of the cardioid model as a "real" description of biological growth has been hugely controversial. See especially the Letter to the Editor by Bookstein (1981) in the same volume as Todd and Mark (1981a); there is also a rejoinder by the authors (Todd and Mark, 1981b). One of the main objections is the existence of a constant seed. Another is the simplistic assumption that growth depends on a single parameter. The full exponential cardioid (FEC) radial growth model developed below in (4) includes two parameters for growth.

The RCS model was firstly used to characterize craniofacial growth. However, this model has been also found useful to effectively approximate ageing on frontal photographs of faces. See, for example, Miyoshi and Hyodo (2006); Ramanathan and Chellappa (2006); Ramanathan et al. (2009) and Yamaguchi and Oda (1999), especially for female faces and faces that appear childlike.

One of the main successful uses of the cardioid model has been in psychological experiments, where experimenters artificially age outlines or images of human heads using this model, with the aim of getting subjects to visually react to the perceived age; see, for example, Yamaguchi and Oda (1999). For this purpose it is not necessary for the model to be fully accurate biologically. It is only necessary that the subjects perceive appropriate differences in age as the image is altered.

The purpose of this paper is to develop statistical shape methodology to assess the strengths and weaknesses of the cardioid growth model. To facilitate the statistical analysis, we emphasize a modified version of the growth model, which differs from the original RCS model in three ways.

(a) Growth is modelled on a log scale rather than a linear scale.

(b) There are two parameters to model growth (essentially an intercept and slope parameter) instead of a single slope parameter in the original RCS model.

(c) Explicit assumptions are introduced to model the statistical error. Three possibilities are described.

It is a pleasure to include this paper in a volume dedicated to C R Rao's 100th birthday as it relates to two research areas where he has made substantial contributions. He did pioneering work in growth starting from Rao (1958) and subsequently he worked on shape analysis based on landmark data starting from Rao and Suryawanshi (1996).

2 Radial growth models for two-dimensional objects

Let X and Y denote two geometric objects in the plane where every point $x'_j \in X$ has a unique counterpart $y'_j \in Y$, $j = 1, \ldots, J$. A radial growth model states that after appropriate centering and rotation, the X and Y configurations are related by a simple parametric transformation in polar coordinates.

It is convenient to represent points in the plane by complex numbers. Given "seeds" μ, $\nu \in \mathbb{C}$ define "centered" points by

$$x_j = x'_j - \mu, \quad y_j = y'_j - \nu$$

with polar coordinates

$$x_j = x'_j - \mu = r_j \exp(i\theta_j), \quad y_j = y'_j - \nu = s_j \exp(i\phi_j). \tag{1}$$

Thus r_j, s_j are the radial components and θ_j, ϕ_j are the angular components of the centered data.

For graphical purposes, let an angle θ be measured counterclockwise from vertical. Thus $\theta = 0$ points upwards and $\theta = \pi/2$ points to the left. Thus a complex number with positive real part lies in the upper half-plane and a complex number with positive imaginary part lies in the left half-plane.

Let a function $M(\theta) = \exp\{L(\theta)\}$, taking an angle to a positive number, be called a *radial deformation function*, following Grenander et al. (2007). The specific choice

$$M(\theta; a_0, b) = \exp\{L(\theta)\} = \exp(a_0 - b\cos\theta) \tag{2}$$

is called the *full exponential cardioid (FEC)* radial deformation function. The minus sign is chosen so that if b is positive, then $M(\theta; a_0, b) > M(0; a_0, b)$ for $\theta \neq 0$; in particular, the growth rate is smallest in the upwards vertical direction and largest in the downwards vertical direction.

Consider a two-dimensional side or sagittal view of a human head, with a seed inside the head but near the top of the skull. Suppose the head has been rotated so that the direction from the seed to the top of the head points upwards (the *preferred orientation*), with angle $\theta = 0$. Then the FEC radial deformation function can capture the property that for babies and children, the growth rate is greater near the bottom of the head than at the top.

In general, the configurations X and Y may need to be rotated to their preferred orientations by angles α and β, say, before the deformation function can be applied. Given a radial deformation function, a general *radial growth model* from X to Y is defined by

$$e^{-i\beta}y_j = M(\theta_j - \alpha)e^{-i\alpha}x_j = M^*(\theta_j)e^{-i\alpha}x_j \tag{3}$$

where the angles α and β are nuisance orientation parameters and $M^*(\theta) = M(\theta - \alpha)$ is the *adapted* version of the M function.

For the FEC radial growth model, the adapted log radial deformation function can be written

$$L^*(\theta) = L(\theta - \alpha) = a_0 - b\cos(\theta - \alpha) = a_0 - a_1 \cos\theta - a_2 \sin\theta, \text{ say.} \tag{4}$$

Note that

$$b = (a_1^2 + a_2^2)^{1/2} \text{ and } \alpha = \text{atan2}(a_2, a_1) \tag{5}$$

can be recovered from a_1 and a_2. Here atan2, a function found in many computing languages, is a version of the atan function, modified to ensure that the result is in the correct quadrant, so that $(a_1, a_2) = b(\cos\alpha, \sin\alpha)$.

For any radial growth model there are 6 *registration parameters* (two complex seeds and two orientation parameters). For the FEC radial growth model, there are additionally two *growth parameters*, a_0 and b. Under any radial growth model, the angular part does not change, $\phi_j - \beta = \theta_j - \alpha$; it is only the radial part that changes. That is, the angular part of $e^{-i\beta}y_j$ is the same as that of $e^{-i\alpha}x_j$; the radial part s_j of y_j depends on the radial part r_j of x_j through the radial deformation function. Although the phrase "growth" model is used for simplicity, shrinkage can occur when $M(\theta) < 1$.

If b is near 0, then the FEC radial growth model can be approximated by the *full linear cardioid (FLC)* radial growth model with deformation function

$$M_{\text{FLC}}(\theta; k_1, k_2) = k_1 - k_2 \cos\theta, \tag{6}$$

with $k_1 = \exp(a_0)$, $k_2 = b\exp(a_0)$. If θ is allowed to range around the whole circle then the constraints $k_1 > 0$ and $k_1 + |k_2| > 0$ are needed to ensure that $M_{\text{FLC}}(\theta) = M_{\text{FLC}}(\theta; k_1, k_2)$ is always positive. However, in many applications the possible values of θ_j for points in X lie in a smaller arc $\theta^{(0)} < \theta < \theta^{(1)}$, say, and in such a situation the constraints on k_1 and k_2 can be relaxed somewhat.

The full linear cardioid radial growth model (6) includes two one-parameter special cases.

(a) The case $k_1 = 1$ and $k_2 = k_{\text{CS}}$ in (6), with parameter k_{CS}, is known as the *cardioid strain (CS)* model (Shaw et al., 1974). The analogous restriction for the FEC model (2) is $a_0 = 0$.

(b) The case $k_1 = 1 + k_{\text{RCS}}$ and $k_2 = k_{\text{RCS}}$, with parameter k_{RCS}, is known as the *revised cardioid strain (RCS)* model (Todd and Mark, 1981a). The analogous restriction for the FEC model (2) is $a_0 = b$.

These models are called "strain" models because a physical justification can be attempted in terms of mechanical strain. For the other models there is no claim of any physical motivation.

If growth is viewed as a continuous activity, then the growth model of equation (6) requires the choice of a unit time interval. After n time units with constant registration parameters, the model has overall radial deformation function $M_{\text{FLC}}(\theta; k_1, k_2)^n$ and it is

4

not defined for non-integer values of time. On the other hand, a continuous version of (2) can be defined for all values of time $t \geq 0$ through a time-dependent radial deformation function $M_t(\theta) = \exp\{tL(\theta)\}$. Hence the second model feels more natural when the choice of a unit time interval is arbitrary.

A comparison between radial deformation functions for the RCS and the FEC growth models is given in Figure 1 for different choices of k_{RCS}, with $a_0 = b = k_{\mathrm{RCS}}$. The differences are small for small k_{RCS} and increase for larger k_{RCS}.
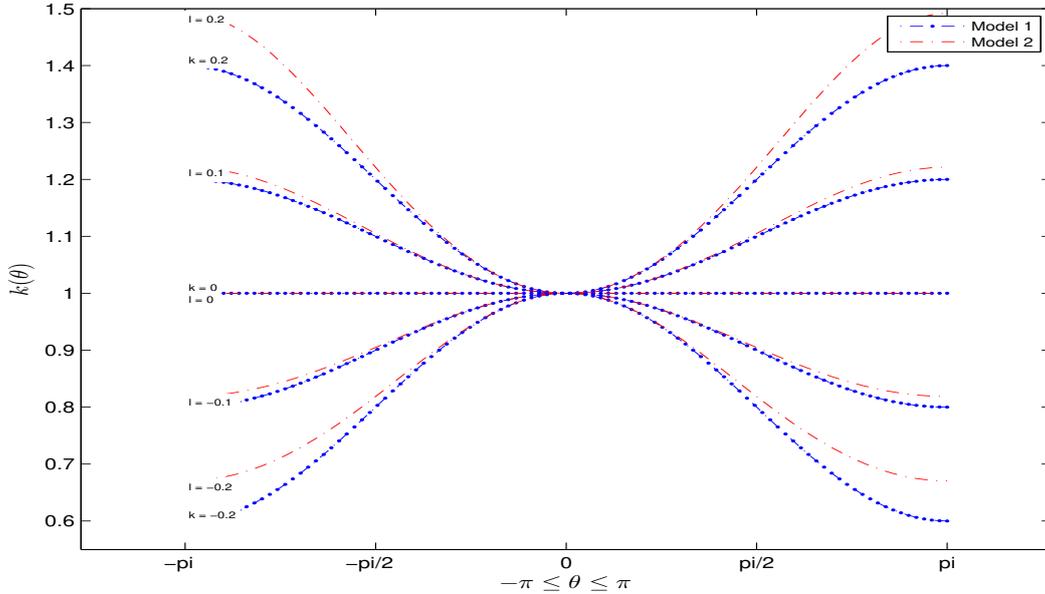

Figure 1: Radial deformation functions with $k_{\mathrm{RCS}} = -0.2, -0.1, 0.0, 0.1, 0.2$. Model 1 = RCS model (blue); Model 2 = analogous FEC model (red).

The choice of the model specification may be based on biological/scientific arguments. For example, the CS model shows no growth in the direction $\theta = 90°$ while the RCS shows no growth in the direction $\theta = 0°$. When there are no scientific reasons to prefer one version over the other, it can be more useful to work with the full model, either in continuous (2) or discrete (6) form. This paper emphasizes the continuous models.

The study of growth models can viewed as part of the subject of statistical shape analysis. Technically, the *shape* of an object consists of the information that remains after location, rotation and size effects have been removed. However, in growth models it is important to retain information abut the size of an object. Hence growth models can be more accurately described as examples of *size-and-shape* analysis (e.g. Dryden and Mardia, 2016, p. 66).

For convenience here is a reminder of the key abbreviations in the paper for various radial growth models:

CS : cardioid strain
RCS: revised cardioid strain
FLC: full linear cardioid, equation (6)

FEC: full exponential cardioid, equation (2).

3 Fitting the FEC growth model for landmark data

3.1 Landmark data

In general, an object in the plane can be represented either by a finite set of landmarks or by a continuous outline. In this section we focus on the landmark case. The outline case is covered below.

Thus the two objects are described in terms of J landmarks, $X = \{x'_1, \ldots, x'_J\}$ and $Y = \{y'_1, \ldots, y'_J\}$ with x'_j known to correspond to y'_j, $j = 1, \ldots, J$.

3.2 Statistical models

There are several ways to introduce statistical errors into the FEC radial growth model. In each case the ϵ_j are assumed to follow independent complex normal distributions $CN(0, \sigma^2)$, so the real and imaginary parts are independent $N(0, \sigma^2)$. We assume σ^2 is "small" in each case. Recall the polar coordinates θ_j, ϕ_j for the centered landmarks are given in (1). In addition the notation $L_j^* = L^*(\theta_j) = L(\theta_j - \alpha)$ in (4) is used for the adapted log radial FEC value at the data points to simplify the formulas.

- (multiplicative errors)

$$e^{-i\beta} y_j = e^{-i\alpha} e^{L_j^*} x_j \left(1 + \epsilon_j\right). \tag{7}$$

- (additive-in-x errors)

$$e^{-i\beta} y_j = e^{-i\alpha} e^{L_j^*} \left(x_j + \epsilon_j\right). \tag{8}$$

- (additive-in-y errors)

$$e^{-i\beta} y_j = e^{-i\alpha} \left(e^{L_j^*} x_j + \epsilon_j\right). \tag{9}$$

  The model with additive-in-y errors is closest in character to a standard regression model.

To fit these models it is convenient to write complex numbers in polar coordinates and to look at the resulting models for the log radial and angular components. In particular, write $\epsilon_j = \epsilon_{j1} + i\epsilon_{j2}$ in terms of its real and imaginary components and note that the radial and angular components satisfy

$$|1 + \epsilon_j| = \left\{(1 + \epsilon_{j1})^2 + \epsilon_{j2}^2\right\}^{1/2} \approx 1 + \epsilon_{j1},$$

so that $\log |1 + \epsilon_j| \approx \epsilon_{j1}$, and

$$\arg(1 + \epsilon_j) = \operatorname{atan2}(\epsilon_{j2}, 1 + \epsilon_{j1}) \approx \epsilon_{j2},$$

6

where the error in the approximations is $O(|\epsilon_j|^2)$. Hence the multiplicative error model can be written approximately as

$$\log(s_j/r_j) = L_j^* + \epsilon_{j1} \tag{10}$$
$$\phi_j - \theta_j = \psi + \epsilon_{j2} \tag{11}$$

where $\psi = \beta - \alpha$. In (10) the left-hand side of the equation can be viewed as the response variable in a linear regression and the right-hand side contains the regressor variables. The second equation (11) involves angles with small random errors about a common mean. It is convenient to approximate the the normal distribution for ϵ_{j2} by a von Mises distribution.

The von Mises distribution $VM(\psi, \kappa)$ with mean direction ψ and concentration parameter κ has density

$$f(\eta) = \frac{1}{2\pi I_0(\kappa)} \exp\{\kappa \cos(\eta - \psi)\}, \quad 0 \le \eta < 2\pi$$

(e.g. Mardia and Jupp, 2000). The normalizing constant $I_0(\kappa)$ is a modified Bessel function. For large concentration parameter κ,

$$I_0(\kappa) \approx e^\kappa/(2\pi\kappa)^{1/2}, \tag{12}$$

and the von Mises distribution is approximately the same as a normal distribution with variance $\sigma^2 = 1/\kappa$. Hence the angles $\phi_j - \theta_j$ are approximately i.i.d. $VM(\psi, \kappa)$, a von Mises distribution with mean direction ψ and concentration parameter $\kappa = 1/\sigma^2$.

The calculations for the additive models are similar but a bit more involved. Start with the additive-in-x error model. Since the complex normal distribution is invariant under rotations of the complex plane about the origin, $\epsilon_j' = \exp(-i\theta_j)\epsilon_j$ is also $CN(0, \sigma^2)$. Then

$$|x_j + \epsilon_j| = |r_j + \epsilon_j'| = \left\{\left(r_j + \epsilon_{j1}'\right)^2 + \epsilon_{j2}'^2\right\}^{1/2} \approx r_j + \epsilon_{j1}',$$

so that $\log|r_j + \epsilon_j'| \approx \log r_j + \epsilon_{j1}'/r_j$, and

$$\arg(x_j + \epsilon_j) = \theta_j + \arg(r_j + \epsilon_j') = \theta_j + \operatorname{atan2}(\epsilon_{j2}', r_j + \epsilon_{j1}') \approx \theta_j + \epsilon_{j2}'/r_j,$$

Hence the additive-in-x error model can be written approximately as

$$\log(s_j/r_j) = L_j + \epsilon_{j1}'/r_j \tag{13}$$
$$\phi_j - \theta_j = \psi + \epsilon_{j2}'/r_j. \tag{14}$$

Thus (13) represents a weighted regression model, where the jth term has variance σ^2/w_j in terms of the weights

$$w_j = r_j^2 = |x_j|^2. \tag{15}$$

Similarly, in (14), the angles $\phi_j - \theta_j$ are independently distributed from a von Mises distribution with a common mean direction ψ and with jth concentration parameter $\kappa_j = w_j / \sigma^2$.

The expansion for the additive-in-y error model is similar, except the weights are now given by $\left(e^{L_j^*} |x_j|\right)^2$. However, $e^{L_j^*} |x_j|$ is not observed, so it is approximated by $|y_j|$ to give the weights

$$w_j = s_j^2 = |y_j|^2 \tag{16}$$

which are used below for estimation.

Assuming for the moment that σ^2 is known, the parameters of both the radial and the angular models can be estimated by maximum likelihood. Details are given in the next section.

4 Estimation

In this section details are given for estimating the parameters of the FEC model, either in its multiplicative or additive form. The estimation procedure takes the same form in all three cases, but with different "weight terms". Define

$$w_j^{(1)} = 1, \quad w_j^{(x)} = |x_j|^2 = r_j^2, \quad w_j^{(y)} = |y_j|^2 = s_j^2. \tag{17}$$

For the multiplicative model, the weight term is $w_j = w_j^{(1)}$. For the additive models in x and y, the weight terms are $w_j = w_j^{(x)}$ and $w_j = w_j^{(y)}$, respectively.

If the seeds μ and ν are known, the estimation can be carried out in closed form. The details for the radial and angular parts are given in the next two subsections, and combined in the following subsection. Finally the estimation of the seeds is discussed.

4.1 Estimation for the radial model

The model for the log radial component is a linear regression model where the response variable $v_j = \log(s_j/r_j)$ has mean $L_j = a_0 + a_1 \cos\theta_j + a_2 \sin\theta_j$ and normally distributed $N(0, w_j^{-1}\sigma^2)$ error, $j = 1, \ldots, J$. Hence the parameters a_0, a_1, a_2 can be estimated by minimizing the weighted sum of squares

$$\sum w_j (v_j - L_j)^2.$$

The minimum sum of squares RSS_1, say, is given by

$$\text{RSS}_1 = v^T (W - H) v.$$

Here $H = WX(X^T W X)^{-1} X^T W$ is the weighted "hat" matrix based on the $J \times 3$ design matrix

$$X = \begin{bmatrix} \mathbf{1} & \mathbf{c} & \mathbf{s} \end{bmatrix}$$

where $\mathbf{1}$ is a vector of ones, and $\mathbf{c}$ and $\mathbf{s}$ are vectors with entries $-\cos\theta_j$, $-\sin\theta_j$. The matrix $W = \mathrm{diag}(w_j)$ is a diagonal matrix containing the weights. The parameter estimates are given by

$$\begin{bmatrix} \hat{a}_0 \\ \hat{a}_1 \\ \hat{a}_2 \end{bmatrix} = (X^T W X)^{-1} X^T W \mathbf{v},$$

where $\mathbf{v}$ is a vector containing the v_j.

4.2   Estimation for the angular model

The angular differences $\phi_j - \theta_j = \eta_j$, say, can be modelled using a normal distribution

$$\eta_j = \phi_j - \theta_j \sim N(\psi, \sigma_j^2/w_j) \mod 2\pi.$$

However, since angles are only defined up to a multiple of 2π, it is more convenient, and nearly equivalent, to express the model in terms of the von Mises distribution

$$\eta_j \sim VM(\psi, \kappa_j),$$

with mean direction ψ and concentration parameter $\kappa_j = w_j\kappa$, where $\kappa = 1/\sigma^2$.

If the weights κ_j are treated as known, the log likelihood for ψ becomes

$$\sum \kappa_j \cos(\eta_j - \psi) - K = \sum \kappa_j \{\cos\eta_j \cos\psi + \sin\eta_j \sin\psi\} - K$$
$$= (\sum \kappa_j)\{\overline{C}\cos\psi + \overline{S}\sin\psi\} - K,$$

where $K = \sum \log\{2\pi I_0(\kappa_j)\}$, and

$$\overline{C} = \sum(w_j \cos\phi_j)/\sum w_j, \quad \overline{S} = \sum(w_j \sin\phi_j)/\sum w_j.$$

The maximizing value of ψ is $\mathrm{atan2}(\overline{S}, \overline{C})$. If we define

$$\mathrm{RSS}_2 = 2(\sum w_j)(1 - \overline{R}) = 2\sigma^2(\sum \kappa_j)(1 - \overline{R})$$

where $\overline{R} = \{\overline{C}^2 + \overline{S}^2\}^{1/2}$ is the weighted resultant length, then the maximized log-likelihood becomes

$$-\frac{1}{2\sigma^2}\mathrm{RSS}_2 - K + \sum \kappa_j.$$

Once ψ and α have been estimated, then β can be estimated using the identity $\psi = \beta - \alpha$.

4.3   Overall estimation

Maximum likelihood estimation for the overall model requires several additional considerations.

(a) (Known seeds) Assume the seeds μ and ν are known. Define an overall "residual sum of squares"

$$\mathrm{RSS} = \mathrm{RSS}_1 + \mathrm{RSS}_2. \tag{18}$$

Then using the approximation (12), the log-likelihood maximized over the parameters a_0, a_1, a_2, ψ (or equivalently, over a_0, b, α, β) is given up to a constant term by

$$l = -\frac{1}{2}\{\mathrm{RSS}/\sigma^2 + 2J\log\sigma^2 - 2\sum\log w_j\}. \tag{19}$$

Further, maximizing (19) over σ^2 yields a profile log-likelihood

$$l = l(\mu, \nu) = -\frac{1}{2}\{2J + 2J\log(\mathrm{RSS}/(2J)) - 2\sum\log w_j\}, \tag{20}$$

depending just on the seeds μ and ν.

(b) (Standardization) Unfortunately, the log-likelihood in (20) is not suitable for comparing different seeds. The reason is that the response variables $v_j = \log(s_j/r_j)$ in (13) and $\eta_j = \phi_j - \theta_j$ in (14) measure changes in *relative* positions of the final landmark y_j with respect to the seed ν. To appreciate the problem for fixed data values x'_j, y'_j, $j = 1, \ldots, J$, let the seeds take the form $\mu = \nu = c > 0$, where for simplicity attention is restricted to the case c real. Then for large c, $v_j \approx (|y'_j| - |x'_j|)/c$ and $\eta_j \approx \mathrm{Im}(x'_j - y'_j)/c$. For the multiplicative model, $w_j = 1$ and so $\mathrm{RSS} = O(1/c)$ for large c. For both of the additive models, $w_j \approx c^2$ and $\mathrm{RSS} = O(1)$. In all three cases, (20) is approximately $-\frac{1}{2}\{-2J\log(c^2)\} = J\log(c^2) \to \infty$ as $c \to \infty$. That is, the likelihood is maximized at the singular solution $c = \infty$.

The solution is to scale the response variables v_j and η_j to $|y_j|v_j$ and $|y_j|\eta_j$ so that they measure changes in *absolute* positions instead of relative positions. The effect on the log-likelihood is to include an extra term. The profile log-likelihood for the scaled responses becomes

$$l^{\mathrm{scaled}} = l^{\mathrm{scaled}}(\mu, \nu) = -\frac{1}{2}\{2J + 2J\log(\mathrm{RSS}/(2J)) - 2\sum\log w_j + 2\sum\log w_j^{(y)}\}. \tag{21}$$

(c) (Regularization) However, there is additional problem that arises when using (21) to compare different seeds. If the seed ν converges to one of the y_j, then the corresponding log weight diverges, $\log w_j^{(y)} \to -\infty$. Thus for the multiplicative model and the additive-in-x model, the log-likelihood has a singular maximum at this limiting choice of seed.

The basic cause of the problem is that the polar decomposition of the error model breaks down when the radial value is close to 0. A simple way to resolve the problem is to "regularize" the log-likelihood by approximating $J^{-1}\sum\log w_j$, the log of the

geometric mean of the weights, by $\log(J^{-1}\sum w_j)$, the log of the arithmetic mean. This substitution yields the "regularized" scaled log-likelihood

$$l^{\text{scaled,reg}} = l^{\text{scaled,reg}}(\mu, \nu)$$
$$= -\frac{1}{2}\{2J + 2J\log(\text{RSS}/(2J)) - 2J\log(J^{-1}\sum w_j) + 2J\log(J^{-1}\sum w_j^{(y)})\}. \tag{22}$$

This approach works for all three models. However, note that for the additive-in-y model, the final two terms in (21) cancel one another out and it does not matter whether or not the scaled log-likelihood is regularized. In this case equations (21) and (22) are identical to one another.

(d) (Optimization) The value of $-l^{\text{scaled,reg}}(\mu, \nu)$ can then be minimized numerically over the four parameters $\Re(\mu), \Im(\mu), \Re(\nu), \Im(\nu)$, e.g. using the black-box minimizer `nlm` from R. Once μ and ν have been estimated, the regression parameters a_0, a_1, a_2 and the angular parameter ψ can be estimated using Sections 4.1 and 4.2 respectively. Call the resulting estimator the approximate maximum likelihood estimator (AMLE).

(e) (Interpretation) Estimates of the regression and variance parameters are reported using the definitions in Section 4.2. That is, they are not affected by the scaling of the response variables introduced in (21). Further, since the FEC model has eight parameters, an "unbiased" estimate of the error variance can be defined by

$$\hat{\sigma}^2 = \frac{\text{RSS}}{2J - 8} \tag{23}$$

where RSS is defined by (18).

(f) (Standard errors) Once the parameters have been estimated by the AMLE, it is important to include standard errors. These can be obtained as follows. Sections 4.1 and 4.2 discuss optimization over the regression parameters a_0, a_1, a_2 and the angular parameter ψ, respectively. If these parameters are left in the model, then a version of the approximation (22) is obtained for the log-likelihood, where RSS depends on these parameters, as well as the four parameters in μ and ν. Differentiating this version of (22) twice numerically at the AMLE and changing the sign yields the 8×8 approximate observed Fisher information matrix, I_{obs}, say. Inverting I_{obs} gives the approximate variance matrix, $\hat{\Sigma}$, say for the AMLE. In particular, the square roots of the diagonal elements give the standard errors.

Further, the 2×2 submatrix of $\hat{\Sigma}$ for the real and imaginary parts of the seed μ can be used to construct a confidence ellipse for μ (and similarly for ν).

5 Outline data

In general, an object in the plane can be represented either by a continuous outline or by a finite set of landmarks. So far the paper has focused on the landmark case.

Mathematically an outline can be represented as a continuous curve $\{f(u) \in \mathbb{C} : u \in I\}$, where the index variable u ranges through an interval $I = [a, b]$. The curve is either open if $f(a) \neq f(b)$, or closed if $f(a) = f(b)$. We note that the index variable is a convenient tool to describe the curve, but is not an essential part of the curve. In particular the curve can be re-parameterized by any monotone function $\Phi(u)$.

In general, given two curves, it is not possible to match a given point on the first curve to a particular point on the second curve. In Biology, matching points on the two objects are called *homologous* if they have the same biological interpretation. In the landmark case it has been assumed above that the landmarks are homologous. However,in the outline setting it is not assumed that any information about homology is available. Fortunately, under the growth model approximate matching can be carried out mathematically.

To proceed further, make the simplifying assumption that X and Y are "star-shaped" about their centers μ and ν respectively. A star-shaped curve can thus be written in polar coordinates with the angular part θ playing the role of the index variable j. That is, for the X outline, there is a radial function $r(\theta)$ such that we can write

$$x(\theta) = e^{i\theta} r(\theta), \quad \theta \in [\theta^{(0)}, \theta^{(1)}].$$

The curve is closed if $\theta^{(0)} = 0$, $\theta^{(1)} = 2\pi$ and $r(0) = r(2\pi)$. The name "star-shaped" arises because the ray from the center μ at angle θ, $\theta \in [\theta^{(0)}, \theta^{(1)}]$, intersects the X outline exactly once.

A similar representation is assumed to hold for the Y outline holds. with radial function $s(\theta)$ and, for a given value of θ, we declare $x(\theta)$ and $y(\theta)$ to be matched. Of course this matching procedure assumes the registration parameters are known or at least estimated.

The fitting of growth models to outline data is left to future work.

6 Three-dimensional version of the growth model

It is straightforward to define a three-dimensional version of the growth model. If a direction on the unit sphere in $\mathbb{R}^d$ is represented by a unit vector $\boldsymbol{u}$, say, and if the standard basis directions are represented by the unit vectors

$$\boldsymbol{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \boldsymbol{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \boldsymbol{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

then, with $\boldsymbol{e}_3$ denoting the polar direction, the growth function L can be written as

$$L(\boldsymbol{u}) = a_0 - b\boldsymbol{u}^T \boldsymbol{e}_3 = a_0 - bu_3.$$

Unfortunately, the elegance of complex arithmetic is no longer available to simplify the fitting procedure.

7   Numerical considerations

If the seeds μ and ν are known, the estimation of the remaining parameters is straight-forward and can be computed using standard linear regression and directional statistics algorithms. However, the estimation of the seed is more challenging. Here are some preliminary recommendations. The objective function is given by changing the sign of the regularized scaled log-likelihood (22).

(a) Start by doing a grid search for the seeds, looking for the smallest value of the objective function.

(b) Then use a black box optimizer to refine the estimate of the seeds. In this paper the `nlm` function in R has been used to minimize the objective function. The log-likelihood given in (22), and the parameter is the 4-dimensional pair of seeds.

In general there do not seem to be numerical problems using the regularized scaled log-likelihood, provided there is enough information in the data to estimate the seeds. However, as illustrated in the next section, there can be a high correlation between the estimates of the seeds μ and ν, making it difficult to estimate them individually.

8   Examples

8.1   Simulated data

To illustrate the behavior of the fitting algorithm, consider the following simulated data set. The x values at some initial time are given y

$$x_j = (1 + (j-1)/7)\exp(\pi i(j-1)/7), \ j = 1, \ldots 8,$$

so that the angular pars are equally-spaced on a semi-circle and the radial parts increase in an arithmetic progression from 1 to 2. The y values at some final time follow the FEC model with $\alpha = \beta = 0$, $a_0 = 1.2, b = 0.2$, with $CN(0, \sigma^2)$ noise, $\sigma = 0.1$.

The additive-in-y FEC model has been fitted, with the results plotted in Figure 2. In the figure the two seeds have been shifted to lie at the origin and the configurations have been rotated by the fitted angles $\hat{\alpha}$ and $\hat{\beta}$. Growth is smallest in the vertically upwards direction and largest in the vertically downwards direction. The black numbers closest to the origin correspond to the x configuration. The gray lines show the direction of growth. The red numbers on the gray lines show the fitted y landmarks under the growth model and the green numbers show the actual y configuration.

The estimated regression and angular parameters (with standard errors) are given by $\hat{a}_0 = 1.230 \ (0.022)$, $\hat{a}_1 = 0.163 \ (0.027)$, $\hat{a}_2 = -0.012 \ (0.013)$, $\hat{\psi} = -0.033 \ (0.031)$. Also, $\hat{b} = 0.163$. These estimates are broadly compatible with the true values, though a_1 and b are somewhat under-estimated. The estimated value of a_2 is is compatible with the true value $a_2 = 0$ (since $\alpha = 0$). Similarly, the estimated value of ψ is compatible with the true value $\psi = 0$.
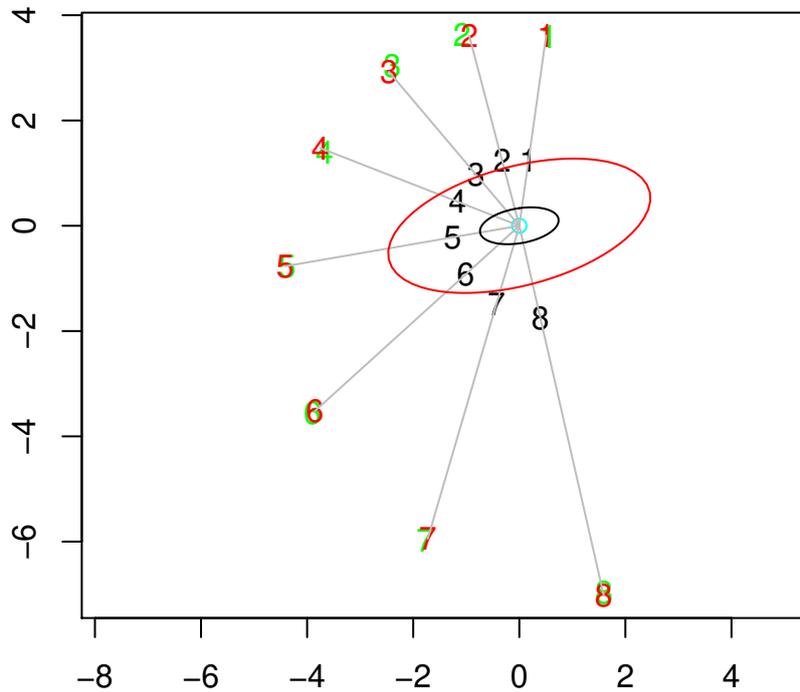
Figure 2: Additive-in-y FEC model fitted to a simulated data set between an initial time and a final time. The data have been shifted and rotated so that the initial and final seeds are located at the origin and the polar direction is vertical. The black numbers nearest the origin are the initial landmark locations. The red numbers at the end of the gray lines radiating from the origin are the fitted final landmarks under the model. The green numbers are the final landmarks. The black ellipse about the origin represents a 95% confidence region for the initial seed. The larger red ellipse is the corresponding region for the final seed.

Also plotted in Figure 2 are 95% confidence ellipses for μ (the inner black ellipse) and ν (the outer red ellipse). The seeds are not very tightly determined by the data, even though the noise standard deviation is small. Note the estimated seeds are pushed to the left from their true values. In particular, if the true regression parameters were used in the plot, then landmarks 1 and 8 for x would lie on a vertical line and the true seed for x would lie midway between them. The reason that the seeds are not very accurately determined seems to be due to the high canonical correlations between the estimates of μ and ν (0.981 and 0.980). See, e.g. Mardia et al. (1979, Ch 10) for a description of canonical correlation.

To save space only the fit from the additive-in-y model has been plotted. However, the figures for the multiplicative and the additive-in-x models are similar.

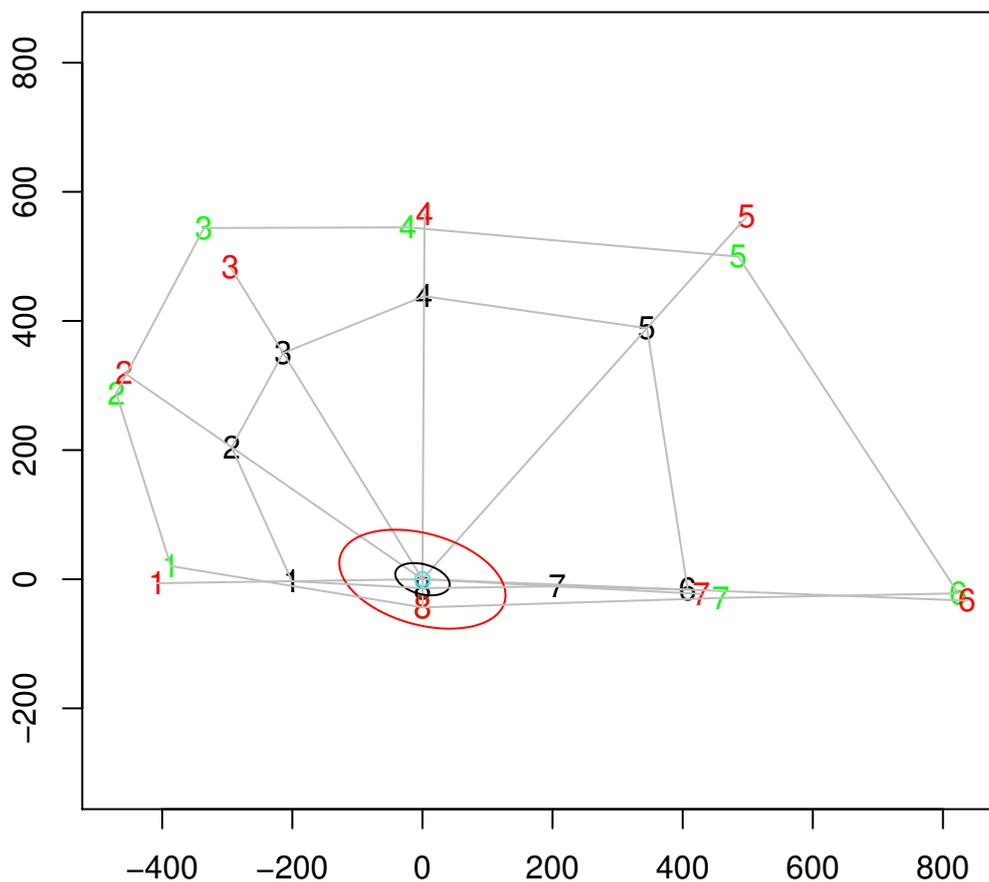**Multiplicative FEC model for rat data**

Figure 3: Multiplicative FEC model fitted to rat data between the initial time and the final time. The data have been shifted and rotated so that the initial and final seeds are located at the origin and the polar direction is vertical. The black numbers nearest the origin are the initial landmark locations. The red numbers at the end of the gray lines radiating from the origin are the fitted final landmarks under the model. The green numbers are the final landmarks. The initial landmarks have been joined by a gray polygon; similarly for the final landmarks. The black ellipse about the origin represents a 95% confidence region for the initial seed. The larger red ellipse is the corresponding region for the final seed.
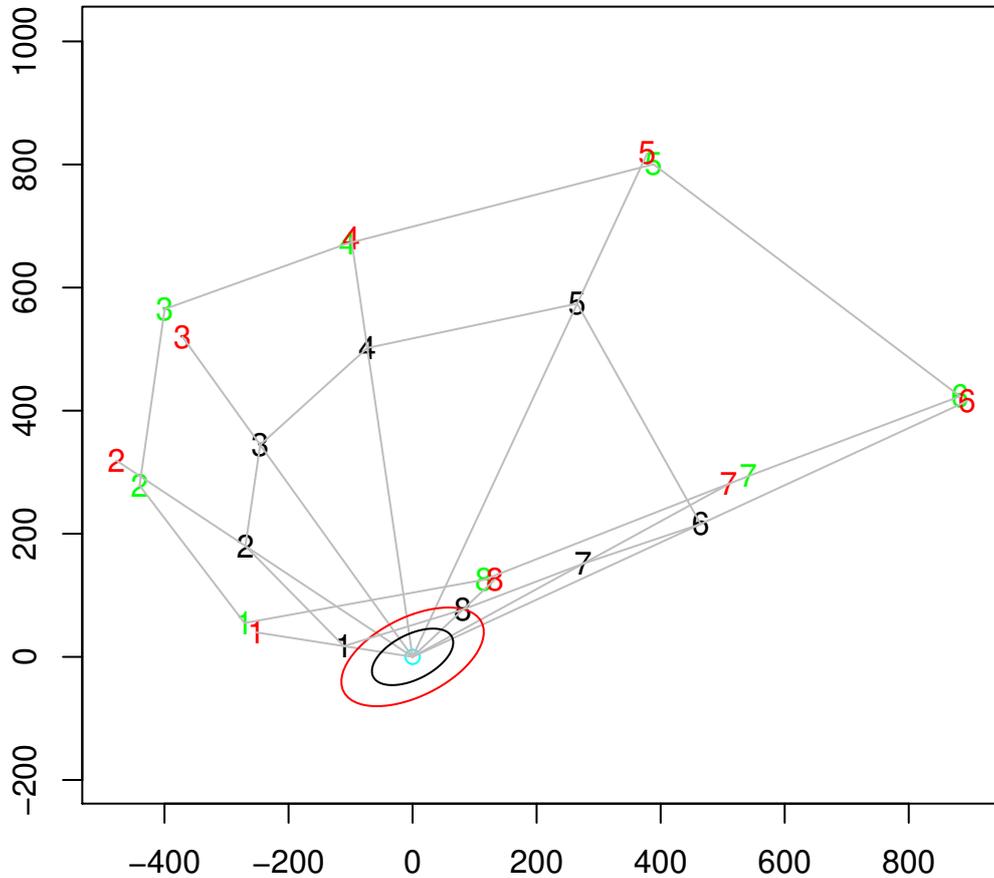
Additive in x FEC model for rat data

Figure 4: Additive-in-x FEC model fitted to rat data between the initial time and the final time. See caption to Figure 3 for more explanation.

8.2 Rat calvarial data

The data set considered here consists of the position of $J = 8$ biological landmarks from a two-dimensional midsagittal section of the calvarium, (the skull without the lower jaw) from 18 different rats at 8 different ages from birth (7 days old) to adulthood (150 days old). Many researchers have investigated craniofacial growth laws using this data set, e.g. Bookstein (1991, 2018); Dryden and Mardia (2016); Kenobi et al. (2010); Kent and Mardia (2002); Kent et al. (2001); Le and Kume (2000); Mardia et al. (2013); Moss et al. (1985, 1984, 1983); Starke et al. (2003).

A detailed description of the data is given in Bookstein (1991, Table 3.4.1) and Book-
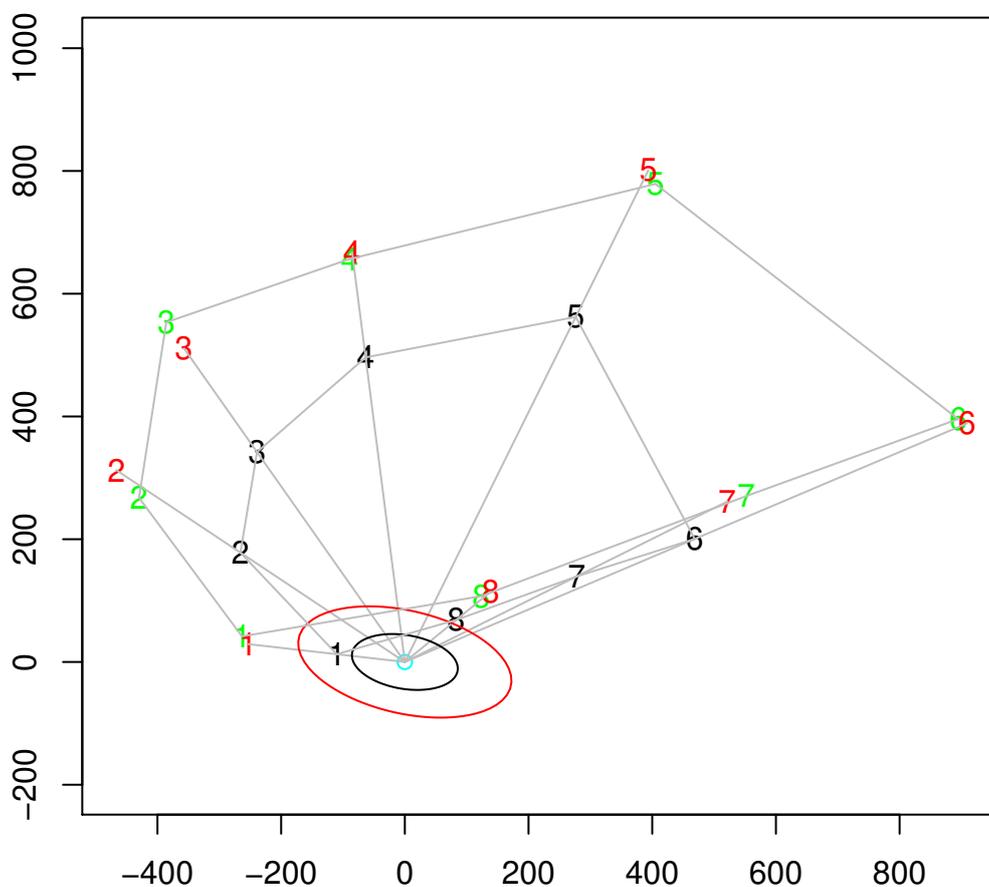
Additive in y FEC model for rat data

Figure 5: Additive-in-y FEC model fitted to rat data between the initial time and the final time. See caption to Figure 3 for more explanation.

stein (2018, p. 122). The data can be found in e.g. Dryden (2019). The landmarks, labelled 1-8, have the following definitions:

1: Bas, Basion	2: Opi, Opisthion
3: IPP, Interparietal Suture	4: Lam, Lambda
5: Brg, Bregma	6: SES, Spheno-ethmoid Synchondrosis
7: ISS, Intersphenoidal Suture	8: SOS, Spheno-occipital Synchondrosis

Landmark 1 lies at the back of the head and landmark 4 lies at the top of the head. The upper part of the jaw lies to the right of landmark 6. Note that landmarks 1,8,7,6 are

Table 1: The fitted regression parameters (with standard errors and p-values) for the multiplicative, additive-in-x, and additive-in-y FEC models. Also given is the maximized log-likelihood for each model.

Par	Mult	Add-x	Add-y
$\hat{a}_0$	0.702 (0.043)	0.916 (0.072)	0.903 (0.067)
$\hat{a}_1$	0.441 (0.055)	0.607 (0.068)	0.602 (0.067)
$\hat{a}_2$	0.071 (0.044)	-0.119 (0.060)	-0.106 (0.058)
$\hat{\psi}$	-0.041 (0.028)	-0.031 (0.024)	-0.037 (0.027)
log-lik	-61.85	-57.77	-58.21

nearly collinear.

For the purposes of this paper we ignore any differences between the individual rats and focus only on the changes in size and shape of the "average" configurations obtained by means of Generalised Procrustes analysis of the 18 configurations considered at at each of the 8 times. In addition, the data at just the initial and final times are used to fit the FEC model. All three variants of the models (multiplicative, additive-in-x, and additive-in-y) have been fitted. A number of features can be noted.

(a) In each of Figures 3–5, the standardized data have been plotted. In particular, the two seeds lie at the origin and the configurations have been rotated by the fitted parameters $\hat{\alpha}$ and $\hat{\beta}$. Note that the top of the head points roughly upwards in each figure, confirming this aspect of intuition about growth models. Growth is smallest in the vertically upwards direction and largest in the vertically downwards direction.

(b) The black numbers closest to the origin correspond to the initial configuration. The gray lines show the direction of growth. The red numbers on the gray lines show the fitted landmarks at the final time under the growth model and the green numbers show the final configuration.

(c) The estimated seeds for x and y lie just above landmark 8 for the multiplicative model, and below the line segment connecting landmarks 1 and 8 for both additive models. These locations (for rats) are broadly similar to one another, but they go against the intuition for human heads, where it is expected the seeds would lie nearer the top of the skull.

(d) At first sight the models appear to fit reasonably well. However, remember the model contains 8 parameters and the data contain $2J = 16$ degrees of freedom, where $J = 8$ is the number of landmarks. Hence there is considerable scope for overfitting.

(e) Although the three models are non-nested, it is still interesting to compare their log-likelihoods. The multiplicative model has the smallest log-likelihood and the

additive-in-x model the largest. In particular,

$$2(l_{\text{add-x}} - l_{\text{mult}}) = 8.16, \quad 2(l_{\text{add-x}} - l_{\text{add-y}}) = 0.88.$$

Hence the additive models are similar to one another, and both are considerably better than the multiplicative model. (Taking χ_1^2 as an approximate benchmark distribution, note that $8.16 > 3.84$, the upper 5% critical value of χ_1^2.)

(f) By construction, the maximum growth rate in each figure is greatest in the vertically downwards direction. For the multiplicative model there is only one landmark below the seed (landmark 8). Since this landmark is very close to the seed for both x and y, the data provide only limited confirmation of the model in the downwards direction. The situation is even more extreme for the two additive models; there are no landmarks below the seeds.

(g) The fitted regression parameters (with standard errors) are given in Table 1. All three models have a similar interpretation. First $\hat{a}_2$ is compatible with a population value of 0, so no rotation of the x-configuration is needed to fit the growth model. Since $\hat{a}_0$ is significantly different from 0, there is no supporting evidence for an exponential version of the CS model. It is also reasonably clear in each case that $\hat{a}_0$ is significantly different from both $\hat{a}_1$ and $\hat{b}$. Hence there is no supporting evidence for an exponential version of the RCS strain model. Finally, $\hat{\psi}$ is compatible with a population value of 0, so there is no need for a rotation of the y-configuration (more specifically, there is no need to rotate the x- and y-configurations differently from one another).

(h) Also plotted in Figures 3–5 are 95% confidence ellipses for μ (the inner black ellipse) and ν (the outer red ellipse). Hence the seeds are not very tightly determined by the data.

(i) Moss et al. (1983) fitted a similar growth model with an estimated seed between landmarks 7 and 8.

(j) In summary, at first sight the FEC growth model(s) seem to provide a plausible fit to the data. However, the argument for a biological interpretation of the seeds is not very convincing as the seeds lie near the boundary (multiplicative model) or even outside the convex hull of the landmarks (additive models). Further, it is important not to read too much into the fitted model. For data with a limited number of landmarks such as the rat data, there is a tendency to overfitting.

9   Conclusions

Mathematically, the FEC growth model proposed here is more elegant and tractable than earlier approaches. This paper has simplified one key aspect of the fitting process. If the seeds are known, then standard closed-form estimators can be used for the remaining

parameters in the model. Thus, issues relating to the estimation of the seeds can be separated from the estimation of the remaining variables.

If a single model is to be used, then the additive-in-y model has several appealing features. It is closest in character to standard regression models since the errors are defined on the same scale as the response variable. Also, it is simpler to describe than the other models since the regularization step in Section 4.3 is not needed. That is, it makes no difference to the log-likelihood whether or not the weights are regularized. Further, for the examples considered here, the the fit is as good as or better than the other two models.

Acknowledgments

References

Bookstein, F. L. (1981). Comment on "issues related to the prediction of craniofacial growth", *American Journal of Orthodontics* **79**: 442–448.

Bookstein, F. L. (1991). *Morphometric Tools for Landmark Data: Geometry and Biology*, Cambridge University Press, Cambridge.

Bookstein, F. L. (2018). *A Course in Morphometrics for Biologists: Geometry and Statistics for Studies of Organismal Form*, Cambridge University Press, Cambridge.

Dryden, I. L. (2019). `shapes` *package*, R Foundation for Statistical Computing, Vienna, Austria. Contributed package, Version 1.2.5.
URL: *http://www.R-project.org*

Dryden, I. L. and Mardia, K. V. (2016). *Statistical Shape Analysis with Applications in R*, Wiley, Chichester.

Grenander, U., Srivastava, A. and Saini, S. (2007). A pattern-theoretic characterization of biological growth, *IEEE Transactions on Medical Imaging* **26**: 648–59.

Kenobi, K., Dryden, I. L. and Le, H. (2010). Shape curves and geodesic modeling, *Biometrika* **97**: 567–584.

Kent, J. T. and Mardia, K. V. (2002). Modelling strategies for spatial-temporal data, *in* A. B. Lawson and D. G. T. Denison (eds), *Spatial Cluster Modelling*, Chapman and Hall/CRC, pp. 213–226.

Kent, J. T., Mardia, K. V., Morris, R. J. and Aykroyd, R. G. (2001). Functional models of growth for landmark data, *in* K. V. Mardia and R. G. Aykroyd (eds), *Proceedings in Functional and Spatial Data Analysis*, Leeds University Press, pp. 109–115.

Le, H. and Kume, A. (2000). Detection of shape changes in biological features, *Journal of Microscopy* **200**: 140–147.

Mardia, K. V., Fallaize, C. J., Barber, S., Jackson, R. M. and Theobald, D. L. (2013). Bayesian alignment of similarity shapes, *Annals of Applied Statistics* **7**: 989–1009.

Mardia, K. V. and Jupp, P. E. (2000). *Directional Statistics*, Wiley, Chichester.

Mardia, K. V., Kent, J. T. and Bibby, J. M. (1979). *Multivariate Analysis*, Academic Press, London.

Miyoshi, T. and Hyodo, M. (2006). Aging effects on face images by varying vertical feature placement and transforming face shape, *IEEE International Conference on Systems, Man and Cybernetics* **2**: 1548–1553.

Moss, M. L., Skalak, R., Patel, H., Sen, K., Moss-Salentijn, L., Shinozuka, M. and Vilmann, H. (1985). Finite element method modeling of cranifacial growth, *American Journal of Orthodontics* **87**: 453–474.

Moss, M. L., Skalak, R., Patel, H., Shinozuka, M., Moss-Salentijn, L. and Vilmann, H. (1984). An allometric network model of craniofacial growth, *International Journal of Orthodontia and Dentistry for Children* **85**: 316–332.

Moss, M. L., Vilmann, H. and Shalak, R. (1983). Statistical testing of an allometric centered model of craniofacial growth, *American Journal of Orthodontics* **83**: 5–18.

Portman, N. (2009). *The Modelling of Biological Growth: a Pattern Theoretic Approach*, PhD thesis, University of Waterloo.

Ramanathan, N. and Chellappa, R. (2006). Modeling age progression in young faces, *IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 1, New York, pp. 387–394.

Ramanathan, N., Chellappa, R. and Biswas, S. (2009). Computational methods for modeling facial aging: a survey, *Visual Languages and Computing* **20**: 131–144.

Rao, C. R. (1958). Some statistical methods for comparison of growth curves, *Biometrics* **14**: 1–17.

Rao, C. R. and Suryawanshi, S. (1996). Statistical analysis of shape of objects based on landmark data, *Proc. Natl. Acad. Sci. USA* **93**: 12132–12136.

Shaw, R., McIntyre, M. and Mace, W. (1974). The role of symmetry theory in event perception, *in* R. MacLeod and H. Pick, Jr (eds), *Studies in Perception: Essays in honor of J. J. Gibson*, Cornell University Press, Ithica, NY.

Starke, J., Rübel, J. and Lux, C. (2003). Modelling the dynamics of craniofacial growth, *Annals of Operations Research* **119**: 75–100.

Todd, J. T. and Mark, L. S. (1981a). Issues related to the prediction of craniofacial growth, *American Journal of Orthodontics* **79**: 63–80.

Todd, J. T. and Mark, L. S. (1981b). A reply to Dr. Bookstein, *American Journal of Orthodontics* **79**: 449–455.

Yamaguchi, M. and Oda, M. (1999). Does cardioidal strain change in real front-view facial images tend to change the perceived age?, *Electronics and Communications in Japan Part III, Fundamental Electronic Science* **82**: 39–48.