

## A CONTINUOUS PARAMETRIC SHAPE MODEL

ASGER HOBOLTH, JAN PEDERSEN AND EVA B. VEDEL JENSEN

*Department of Mathematical Sciences, University of Aarhus,  
Ny Munkegade, 8000 Aarhus C, Denmark*

(Received March 16, 2001; revised July 31, 2002)

**Abstract.** In this paper we propose a flexible continuous parametric shape model for star-shaped planar objects. The model is based on a polar Fourier expansion of the normalized radius-vector function. The expected phase amplitudes are modelled by a simple regression with parameters having nice geometric interpretations. The suggested *generalized  $p$ -order model* is an extension of first- and second-order Gaussian shape models, and in particular the Gaussian assumption is relaxed. The statistical analysis is straightforward, as demonstrated by an application concerning shape discrimination of two cell nuclei populations.

*Key words and phrases:* Cancer diagnostics, featureless objects, Fourier descriptors, radius-vector function, shape, star-shaped objects.

### 1. Introduction

Recently, shape modelling of featureless objects has attracted much attention in the statistical literature. The Gaussian model with cyclic invariance properties, described by Grenander and Miller (1994), has played a predominant role. One line of research is concerned with the application of the Gaussian model as a prior in Bayesian object recognition. Such an application is discussed in Grenander and Miller (1994). The group around Håvard Rue has also contributed significantly to this research, cf. e.g. Rue and Syversveen (1998) and Rue and Hurn (1999). In Hansen *et al.* (2002) a similar Bayesian analysis is performed where also the time aspect is taken into account.

Another line of research treats likelihood analysis of the Gaussian model and is useful for describing rather than finding the objects. A very important contribution is the paper by Kent *et al.* (2000) where the model is used for the standardized edge transformation vector, see also Kent *et al.* (1996). (The standardized edge transformation vector only contains shape information.) In particular, the eigendecomposition of the circulant covariance matrix is described. In the follow-up paper Hobolth *et al.* (2002a) the corresponding theory is developed for the standardized vertex transformation vector. Likelihood analysis has also been considered in Hurn *et al.* (2001).

In Hobolth and Jensen (2000) a continuous approach is used, which may have a general appeal because the model and its parameters do not relate to a particular number of vertices. Furthermore it appears natural to represent the boundary of an object continuously. The continuous counterpart of the standardized vertex transformation vector is the normalized residual process, as introduced in Hobolth and Jensen (2000). Continuous models are also mentioned in Kent *et al.* (2000) and Hobolth *et al.* (2002a).

In this paper we represent the shape of a random planar star-shaped object in

terms of the normalized radius-vector function  $R = (R(t))_{t \in [0,1]}$  and suggest a flexible continuous statistical model for  $R$ . The model relies on a polar Fourier expansion of the normalized radius-vector function

$$R(t) = 1 + 2\sqrt{c_1} \cos(2\pi(t - d_1)) + 2 \sum_{s=2}^{\infty} \sqrt{C_s} \cos(2\pi s(t - D_s)), \quad t \in [0, 1].$$

We show that the first phase amplitude  $c_1$  and the first phase angle  $d_1$  play special roles as parameters of asymmetry and discuss in detail how the remaining random phase amplitudes  $C_s$  and phase angles  $D_s$  characterise the shape of the random object. Under the proposed *generalized  $p$ -order model* the expected phase amplitudes  $\lambda_s = E(C_s)$  satisfy the simple regression equation

$$\lambda_s^{-1} = \alpha + \beta(s^{2p} - 2^{2p}), \quad s \geq 2,$$

where  $\alpha > 0$ ,  $\beta > 0$  and  $p > 1/2$ . We show that  $p$  determines the smoothness of the object boundary while the parameters  $\alpha$  and  $\beta$  determine the ‘global’ and ‘local’ shape, respectively. Thus, an important new feature of the proposed model is that the statistical parameters relate directly to the geometry of the object and not to, say, the covariance function of  $R$ . The phase angles  $D_s$  are assumed to be uniformly distributed, but there is no restriction on the distribution of the phase amplitudes  $C_s$ . Exponentially distributed amplitudes correspond to a Gaussian model and generalized gamma distributed amplitudes offer a simple extension which allows for both heavier and lighter tails than the exponential ones. The generalized  $p$ -order model is an extension of the models used in Kent *et al.* (2000) and Hobolth *et al.* (2002a) where exponentially distributed phase amplitudes are used and  $p = 1$  or 2. Hobolth *et al.* (2002b) in a follow-up paper consider an application with  $p = 2.5$ .

Fourier expansion of the radius-vector function has been used in many applications and a statistical introduction can be found in Stoyan and Stoyan ((1994), pp. 80–88). We refer the reader to Lestrel (1997) for a review of biological applications and Loncaric (1998) for a survey of the engineering literature. Common to the usual approaches is that statistical models are not formulated, but instead parameters of global structure and roughness are defined directly from the Fourier coefficients. Stoyan and Stoyan ((1994), p. 83) also define shape parameters in a non-statistical context.

In Section 2 we review some well-known properties of the radius-vector function and analyse the geometry of the Fourier coefficients. This analysis is the basis for the construction of the generalized  $p$ -order model proposed in Section 3. Statistical inference is discussed in Section 4, and the model is applied to a data set of normal mantle cell nuclei and cell nuclei from a mantle lymphoma in Section 5. The analysis shows that cell nuclei from the mantle lymphoma are more ‘irregular’ than normal cells (significantly different  $\beta$ -values in the two groups). Shape discrimination of the two cell nuclei populations was the original motivation for studying the continuous shape model. Finally Section 6 contains ideas for future research.

## 2. Geometry of the radius-vector function

Let  $K$  be a compact subset of  $\mathbf{R}^2$  and suppose  $K$  is star-shaped with respect to  $z \in K$ , i.e. the intersection between every line through  $z$  and  $K$  is a line segment. We will describe  $K$  in terms of its radius-vector function  $(r_K(t; z))_{t \in [0,1]}$  where  $r_K(t; z)$  is

the distance from  $z$  to the boundary of  $K$  along a ray with angle  $2\pi t$  relative to a fixed axis.

The radius-vector function is invariant under translation and rotation. To be more specific, let  $z_0 \in \mathbf{R}^2$  and

$$A = \begin{pmatrix} \cos 2\pi t_0 & -\sin 2\pi t_0 \\ \sin 2\pi t_0 & \cos 2\pi t_0 \end{pmatrix}, \quad t_0 \in [0, 1].$$

Then,

$$r_{AK+z_0}(t; Az + z_0) = r_K((t - t_0) \bmod 1; z), \quad t \in [0, 1].$$

A scaling transformation yields

$$r_{\alpha K}(t; \alpha z) = \alpha r_K(t; z), \quad t \in [0, 1], \quad \alpha > 0.$$

The shape of  $K$  is thus, up to shifts in  $t$ , represented by the normalized radius-vector function

$$\tilde{r}_K(t; z) = \frac{r_K(t; z)}{\int_0^1 r_K(u; z) du}, \quad t \in [0, 1].$$

Note that the normalized radius-vector function is a continuous analogue of a standardized vertex transformation vector, cf. Hobolth *et al.* (2002a). Below we simply write  $r(\cdot)$  for  $\tilde{r}_K(\cdot; z)$  in cases where it causes no confusion.

A detailed shape description of  $K$  can be obtained from a Fourier series expansion of the normalized radius-vector function  $r$ ,

$$r(t) = 1 + \sqrt{2} \sum_{s=1}^{\infty} a_s \cos(2\pi st) + \sqrt{2} \sum_{s=1}^{\infty} b_s \sin(2\pi st), \quad t \in [0, 1],$$

where the Fourier coefficients are

$$(2.1) \quad a_s = \sqrt{2} \int_0^1 r(t) \cos(2\pi st) dt, \quad b_s = \sqrt{2} \int_0^1 r(t) \sin(2\pi st) dt, \quad s \geq 1.$$

The Fourier coefficient at phase 0 is 1 because of the normalization of the radius-vector function. Letting

$$a_s = \sqrt{2c_s} \cos(2\pi sd_s), \quad b_s = \sqrt{2c_s} \sin(2\pi sd_s), \quad s \geq 1,$$

we obtain the polar form

$$(2.2) \quad r(t) = 1 + 2 \sum_{s=1}^{\infty} \sqrt{c_s} \cos(2\pi s(t - d_s)), \quad t \in [0, 1],$$

where  $c_s = (a_s^2 + b_s^2)/2 \geq 0$  and  $d_s \in [0, \frac{1}{s}[$ ,  $s \geq 1$ . (If  $a_s = b_s = 0$ , let  $d_s = 0$ .) The coefficient  $c_s$  is called the  $s$ -th phase amplitude and  $d_s$  the  $s$ -th phase angle. It is immediate from (2.2) that  $c_s$  is rotation invariant since it remains unchanged under shifts in  $t$ .

Writing  $z = (z_1, z_2)$  the boundary of  $K$  can be represented as

$$(f_1(t), f_2(t)) = (z_1, z_2) + qr(t)(\cos(2\pi t), \sin(2\pi t)), \quad t \in [0, 1],$$

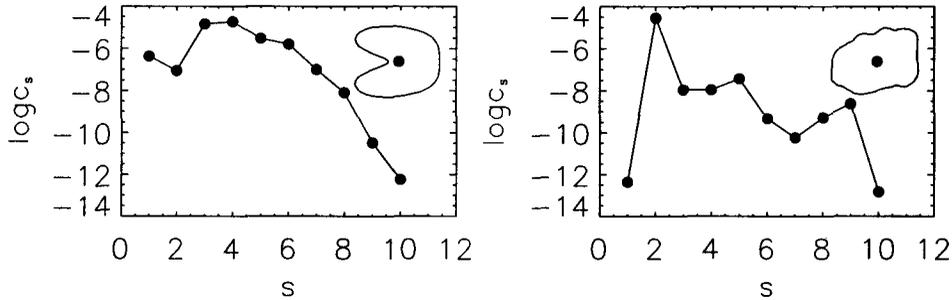


Fig. 1. The values of the phase amplitudes  $c_s$  are shown as a function of  $s$  for an asymmetric object (left) and a fairly symmetric object (right). An object is called symmetric if  $r(t) = r(t + 1/2)$  for all  $t \in [0, 1/2[$ .

where  $q$  is the integral of the radius-vector function. Combined with (2.1) it follows that

$$(a_1, b_1) = \frac{\sqrt{2}}{q} \left( \int_0^1 [f_1(t) - z_1] dt, \int_0^1 [f_2(t) - z_2] dt \right).$$

Thus, if  $K$  is symmetric (with respect to  $z$ ) then  $a_1 = b_1 = c_1 = 0$ . Here symmetry means 2-fold symmetry, cf. (2.4) below with  $s = 2$ . Conversely, a high value of  $c_1$  indicates a high degree of asymmetry relative to  $z$ , cf. Fig. 1. In the Appendix we show that the Fourier coefficients  $a_1$  and  $b_1$  can also be expressed as integrals on the interior of  $K$ .

To analyse the geometry of the higher order phase amplitudes let us consider an object where all but the  $s$ -th phase amplitude are zero such that

$$(2.3) \quad r(t) = 1 + 2\sqrt{c_s} \cos(2\pi s(t - d_s)).$$

For such an object  $z$  is the centre of mass, see the Appendix, and furthermore  $r(t)$  possesses  $s$ -fold symmetry in the sense that

$$(2.4) \quad r(t) = r\left(t + \frac{1}{s}\right) = \dots = r\left(t + \frac{s-1}{s}\right), \quad t \in \left[0, \frac{1}{s}\right].$$

In Fig. 2 we have plotted objects with radius-vector function of the form (2.3), corresponding to different values of  $s$ ,  $d_s = 0$  and varying values of  $c_s$ . In Fig. 3 we have illustrated how the  $s$ -fold symmetric objects contribute for small  $s$  to the ‘global’ shape of a given object  $K$  and for large  $s$  to the ‘local’ shape.

To sum up, we can interpret  $c_s$ ,  $s \geq 1$ , as shape parameters. For  $s = 1$ ,  $c_s$  is an asymmetry parameter. For  $s \geq 2$  small,  $c_s$  determines the ‘global’ shape of  $K$  while for  $s$  large  $c_s$  affects the ‘roughness’ of the boundary of  $K$ . Up to a shift in  $t$ ,  $d_s$ ,  $s \geq 1$ , are also shape parameters. For  $s \geq 2$ , they determine the relative orientation of the  $s$ -fold symmetric objects associated with  $K$ . In Zahn and Roskies (1972) the geometric interpretation of a Fourier expansion of the tangent-angle function is studied in a similar way.

Let us conclude this section by discussing how  $z$  can be chosen. In some applications  $z$  is ‘given by nature’. An important example comes from local stereology where  $K$  is actually a planar section through a biological cell, passing through the nucleus or nucleolus of the cell, cf. Jensen ((1998), Chapter 7). In other cases  $z$  is defined from  $K$ ,

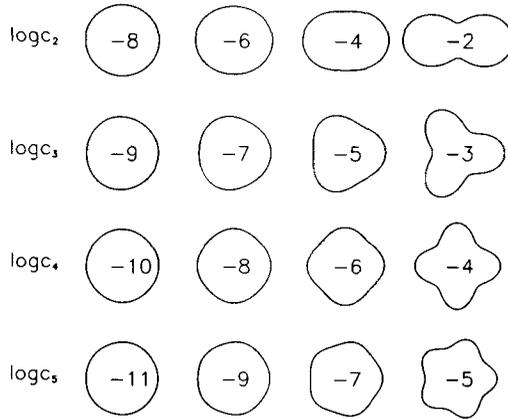


Fig. 2. Objects with radius-vector function of the form (2.3) with  $d_s = 0$ . In each row the value of  $s$  is constant ( $s = 2, 3, 4, 5$ ). The value of  $\log c_s$  is indicated in the interior of the object.

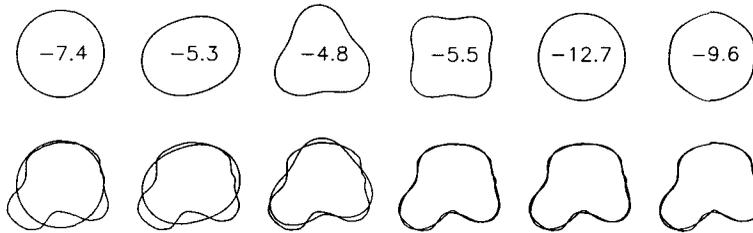


Fig. 3. The upper row shows the objects of the form (2.3) for  $s = 1, \dots, 6$  (left to right) associated with the object  $K$  shown in the lower row. The values of  $\log c_s$  are indicated in their interior. In the lower row, the reconstruction of  $K$  from the first  $s$  Fourier coefficients is also shown.

typically as the centre of mass, cf. Loncaric (1998) and Hobolth *et al.* (2002a). In the latter paper it is used that with  $z$  equal to the centre of mass the first phase amplitude of  $r_K(\cdot; z)$  is approximately zero when  $K$  is a small deformation of a circle. In the Appendix it is shown that the centre of mass of  $K$  can in fact be characterized by the property that the first phase amplitude of  $r_K(\cdot; z)^3$  is zero.

### 3. The generalized $p$ -order model

In this section we introduce the parametric statistical model for the normalized radius-vector function  $(R(t))_{t \in [0,1]}$ . The starting point is the polar expansion (2.2) of the normalized radius-vector function. As argued in the previous section the first phase angle  $d_1$  and phase amplitude  $c_1$  play special roles as asymmetry parameters. In this paper we treat  $c_1$  and  $d_1$  as non-random nuisance parameters. The expansion of the normalized radius-vector function in polar form therefore becomes

$$(3.1) \quad R(t) = 1 + 2\sqrt{c_1} \cos(2\pi(t - d_1)) + 2 \sum_{s=2}^{\infty} \sqrt{C_s} \cos(2\pi s(t - D_s)), \quad t \in [0, 1].$$

The remaining phase amplitudes  $C_s$  and angles  $D_s$ ,  $s \geq 2$ , should be modelled by distributions on  $\mathbf{R}_+$  and  $[0, 1/s]$ , respectively.

The expansion (3.1) makes it possible to construct a variety of shape models. A *generalized  $p$ -order model* is a parametric model satisfying

$$(3.2) \quad C_s \sim \lambda_s Z_s, \quad D_s \sim U[0, 1/s], \quad s \geq 2,$$

where the error variables  $Z_s$  have mean 1 and  $U[0, 1/s]$  is the uniform distribution on the indicated interval. Furthermore,  $C_s, D_s, s \geq 2$ , are all independent and the expected phase amplitudes  $\lambda_s = E(C_s)$  decrease as

$$\lambda_s^{-1} = \alpha_0 + \beta s^{2p}, \quad s \geq 2.$$

The parameters satisfy  $\alpha_0 > -\beta 2^{2p}$  and  $\beta > 0$  such that  $\lambda_s > 0$  for all  $s \geq 2$ . We further assume  $p > 1/2$ , which implies that  $R$  has finite variance, as discussed below.

In order to facilitate a geometric interpretation of the regression parameters we use the reparametrization

$$(3.3) \quad \lambda_s^{-1} = \alpha + \beta(s^{2p} - 2^{2p}), \quad s \geq 2,$$

where  $\alpha > 0, \beta > 0, p > 1/2$ . The parameter  $\alpha$  determines the ‘global’ shape of the object. If  $\alpha$  is high, objects of circular shape are expected while a low value corresponds to an elongated or, in the extreme, a ‘peanut-shell’ shape. The reason is that under (3.3),  $\alpha$  determines the expected phase amplitudes  $\lambda_s = E(C_s)$  for small  $s$  and  $C_s$  governs the global shape for small  $s$ , cf. Section 2. As discussed below  $p$  determines the smoothness of the boundary of  $K$ . For fixed  $p$  the parameter  $\beta$  determines the ‘local’ shape of the object since it controls the behaviour of  $\lambda_s$  when  $s$  is high. Precisely, as  $s \rightarrow \infty$ , we have that  $(\log s, \log \lambda_s^{-1})$  behaves as a line with slope  $2p$  and intercept  $\log \beta$ . For small values of  $\beta$  rather irregular objects are expected while high values yield regular objects.

The random phase angles  $D_s$  determine the relative orientation of the  $s$ -fold symmetric objects associated with  $K$ , cf. Section 2. The uniform distribution on the angles implies that these objects do not have a ‘preferred orientation’. A generalized  $p$ -order model is therefore expected to be appropriate for describing a population of objects which does not have a predominant non-circular shape. The shape variability of  $K$  is influenced by the variation of the error variables  $Z_s$ .

In Fig. 4 simulations from the model (3.1)–(3.3) with exponentially distributed error variables,  $p = 2$  and  $c_1 = 0$  are shown. The values of  $\alpha$  and  $\beta$  are typical for the objects studied in the data section. It is seen that in the corner corresponding to high values of  $\alpha$  and  $\beta$  the simulated objects are smooth and ‘circle’-like, while in the opposite corner the simulated objects are irregular.

To study the distribution of the radius-vector function let

$$R_1(t) = 2 \sum_{s=2}^{\infty} \sqrt{C_s} \cos(2\pi s(t - D_s)), \quad t \in [0, 1],$$

contain all the random Fourier terms of  $R(t)$ . Using (3.2) and independence of the phase angles and amplitudes it follows that  $R_1$  is a stationary process with covariance function

$$(3.4) \quad \sigma(t) = \text{cov}(R_1(t), R_1(0)) = 2 \sum_{s=2}^{\infty} \lambda_s \cos(2\pi st), \quad t \in [0, 1].$$

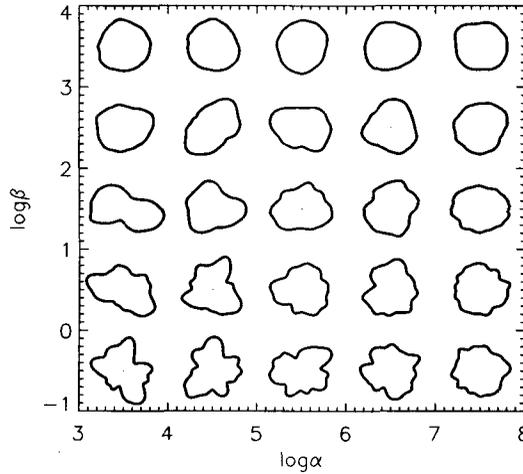


Fig. 4. Simulated objects under the second-order model with  $c_1 = 0$ , exponentially distributed error variables and the indicated values of  $\alpha$  and  $\beta$ .

The process  $R_1$  has zero Fourier coefficients at phases 0 and 1. Similar constraints were used by Hobolth *et al.* (2002a) and Kent *et al.* (2000) in a discrete time model. Properties such as continuity and differentiability of  $R_1$  (and hence also of  $R$ ) are determined by the parameter  $p$  as follows from Cramér and Leadbetter ((1967), Section 4.2 and 4.3).

Equation (3.4) gives the relation between the expected amplitudes and the covariance function. As an alternative to parametric specification of the  $\lambda_s$ s as in (3.3) one may suggest a simple parametric form of the covariance function  $\sigma$ , cf. e.g. Rue and Syversveen (1998). Since the amplitudes relate to the random geometry of the object we believe it is more natural to specify directly a parametric model for the expected amplitudes. Furthermore, the constraints on  $R_1$  are easier to handle and interpret in the spectral domain.

In the shape literature a random object is often modelled by a multivariate normal distribution with a circulant covariance matrix or by a stationary Gaussian process in continuous time, cf. Grenander and Miller (1994), Hobolth *et al.* (2002a), Rue and Hurn (1999), Hobolth and Jensen (2000), Kent *et al.* (2000). We now show that a Gaussian model is obtained by letting the error variables  $Z_s$  be exponentially distributed. This model will therefore be called the Gaussian  $p$ -order model. Using (2.2) and (3.2) it follows that if  $Z_s$  is exponentially distributed then

$$R_1(t) = \sqrt{2} \sum_{s=2}^{\infty} A_s \cos(2\pi st) + \sqrt{2} \sum_{s=2}^{\infty} B_s \sin(2\pi st), \quad t \in [0, 1],$$

where  $A_s, B_s, s \geq 2$ , are all mutually independent and  $A_s \sim B_s \sim N(0, \lambda_s)$ . This representation shows that  $R_1$  is a stationary Gaussian process. By (3.3) and Rogers and Williams ((1994), Theorem I.25.10), it follows that for the Gaussian  $p$ -order model the sample paths of  $R_1$ , and hence also of  $R$ , are  $k - 1$  times continuously differentiable where  $k$  is the integer satisfying  $p \in ]k - 1/2, k + 1/2]$ . In particular, if  $p$  is an integer then  $p = k$ . In the Gaussian first-order model, the sample paths of  $R$  are continuous and in the Gaussian second-order model the sample paths are continuously differentiable. The first- and second-order Gaussian models have been studied in the literature (most often

without the constraint  $\lambda_0 = \lambda_1 = 0$ ). In particular, these models appear as limits of discrete time first- and second-order Markov models, cf. e.g. Grenander ((1993), pp. 476 and 484).

#### 4. Statistical inference

If only a digitized version of the object is available, the radius-vector function cannot be determined accurately. To avoid this obstacle a low-pass filter, cf. Bloomfield (1976), can be used. The idea is to determine the parameter estimates from the low frequency Fourier coefficients only since they are robust to digitization effects. For the Gaussian  $p$ -order model, the analysis is particularly simple. Recall that in this case the phase angles  $D_s$  are uniform in  $[0, 1/s]$  and the phase amplitudes  $C_s$  are independent exponentially distributed with mean  $\lambda_s$ . In particular the distribution of the phase angles does not depend on unknown parameters. Using the first  $S$  phase amplitudes the likelihood function becomes

$$(4.1) \quad L(\lambda_s; c_s) = \prod_{s=2}^S \lambda_s^{-1} e^{-\lambda_s^{-1} c_s}.$$

Defining the expected amplitudes by (3.3) the maximum likelihood estimates for  $(\alpha, \beta, p)$  can be found by standard numerical methods. A likelihood function of the same form is considered in Hobolth *et al.* (2002a) and Kent *et al.* (2000).

If the normalized radius-vector function is only known at the data points  $t = 0, 1/n, \dots, (n-1)/n$ , the phase amplitudes  $c_s = (a_s^2 + b_s^2)/2$  can be approximated by discrete versions of the integrals (2.1). The specific value of  $n$  is not important, just as long as it is reasonably high. That is, different values of  $n$  give approximately the same value of  $c_s$ .

#### 5. Data analysis

The data set consists of 50 normal mantle cell nuclei and 50 cell nuclei from a mantle lymphoma (tumour in the mantle zone of a lymph node), cf. Fig. 5. The nuclei from each of the groups are sampled from a microscopic section among those with sectioned boundary in focus, using a semi-automatic segmentation procedure. The normalized radius-vector function  $r(t)$  is determined at  $t = 0, 1/n, \dots, (n-1)/n$  by tracing rays from the centre of mass to the boundary. Unless otherwise stated we use  $n = 100$ . The nuclei are rather homogeneous in size (about 15  $\mu\text{m}$  in diameter), so the normalization factor is almost the same for all the nuclei.

##### 5.1 Analysing each nuclear profile individually

First, each nuclear profile is analysed individually using the likelihood function (4.1). The cut-off value  $S$  is important. If  $S$  is too small we are not using important shape information, but on the other hand if  $S$  is too large the results are influenced by digitization effects, see Fig. 3. Unless otherwise stated we use  $S = 15$ .

For each object we determine the estimates of  $(\alpha, \beta, p)$  as explained in Section 4. In both samples the estimates of  $p$  are close to 2 for all nuclei. For the normal sample the average is 2.07 with a standard deviation of 0.21 while for the lymphoma sample the average is 2.02 and the standard deviation 0.28. Therefore we fix  $p = 2$  and consider the Gaussian second-order model only.

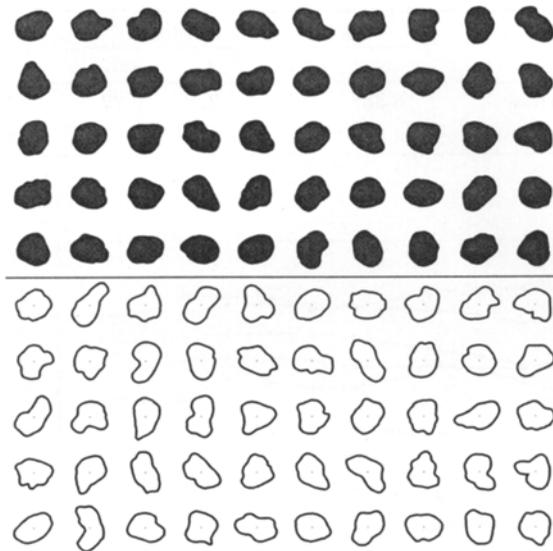


Fig. 5. The 50 normal mantle cell nuclei (upper panel) and the 50 cell nuclei from a mantle lymphoma (lower panel).

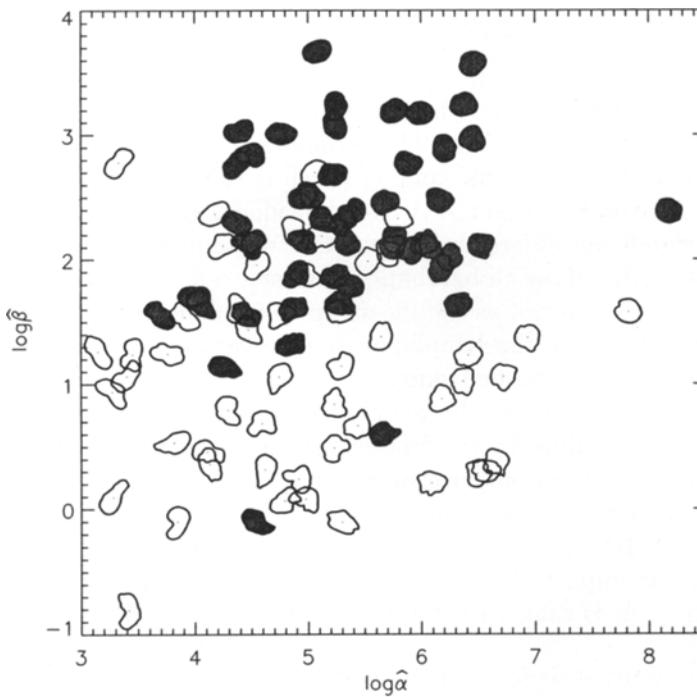
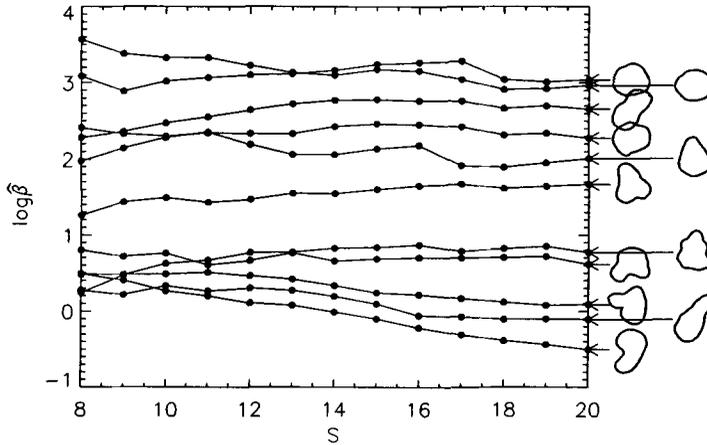


Fig. 6. The estimates of  $(\alpha, \beta)$  under the Gaussian second-order model. The hatched nuclei are from the normal mantle cells while the white nuclei are from cells in the mantle lymphoma.

Table 1. The average, standard deviation and correlation of  $(\log \hat{\alpha}, \log \hat{\beta})$  for each sample.

	$\log \hat{\alpha}$		$\log \hat{\beta}$		corr.
	av.	s.d.	av.	s.d.	
normal	5.35	0.84	2.26	0.72	0.27
lymphoma	4.94	1.11	1.09	0.81	0.03

Fig. 7. The estimates of  $\beta$  as a function of  $S$  for 11 nuclei.

The estimates of  $(\alpha, \beta)$  under the second-order model are shown for each nucleus in Fig. 6 and summarized in Table 1. The estimates of the local shape parameter  $\beta$  are on average lowest in the lymphoma sample, which is to be expected from the geometric interpretation of  $\beta$  given in Section 3. A  $t$ -test for identical  $\beta$ s, based on the distribution of  $\log \hat{\beta}$ , shows a significant difference between the two samples ( $p$ -value less than 0.05%). On average the estimates of the global shape parameter  $\alpha$  are also lowest in the lymphoma sample, but the difference is not as significant ( $p$ -value close to 5%). Furthermore we see that the estimates of  $\alpha$  from the lymphoma sample vary over a somewhat larger range than the estimates from the normal sample.

We also investigated how the choice of cut-off value  $S$  influences the analysis. Since the estimate of  $\alpha$  is determined by the first few amplitudes the estimate of this parameter only changes slightly when  $S$  is larger than 8, say. From Fig. 7 it is seen that the estimate of  $\beta$  does change with  $S$ , but the changes are rather small.

The number of data points  $n$  should be high compared to  $S$ , but otherwise the specific choice is less important. In Fig. 8 we see that for  $S = 15$  the estimates are stable, and the analysis is robust to the specific choice of  $n \geq 50$ .

## 5.2 Analysing the profiles under an iid-assumption

We now investigate whether the profiles within each of the groups can be regarded as independent and identically distributed realizations of a Gaussian  $p$ -order model. Let the indices  $(i, j)$  denote the  $j$ -th nucleus ( $j = 1, \dots, N = 50$ ) in the normal sample ( $i = 1$ ) or the lymphoma sample ( $i = 2$ ) and let  $c_{sij}$  be the corresponding phase amplitudes of the

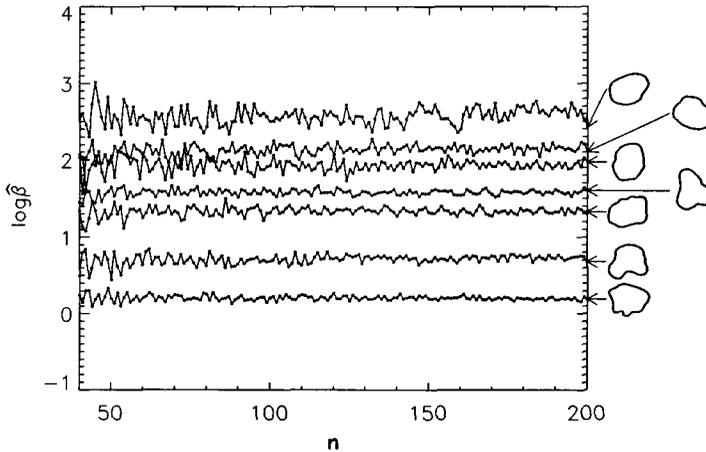


Fig. 8. The estimates of  $\beta$  as a function of  $n$  for 7 nuclei.

normalized radius-vector function. If we let  $\text{Exp}(\lambda)$  denote the exponential distribution with mean  $\lambda$  we want to investigate whether

$$(5.1) \quad C_{sij} \sim \text{Exp}(\lambda_{si}), \quad j = 1, \dots, N,$$

for each  $s = 2, \dots, S$  and  $i = 1, 2$ .

We now examine (5.1) by considering the more general model

$$C_{sij} \sim \Gamma(\gamma_{si}, \rho_{si}, \delta_{si}), \quad j = 1, \dots, N,$$

where  $\Gamma(\gamma, \rho, \delta)$  denotes the generalized gamma distribution with density

$$f(y) = \frac{\delta y^{\delta\gamma-1}}{\Gamma(\gamma)\rho^{\delta\gamma}} \exp\left(-\left(\frac{y}{\rho}\right)^\delta\right), \quad y > 0.$$

Here,  $\gamma, \delta > 0$  are shape parameters while  $\rho > 0$  is a scale parameter. The ordinary gamma distribution is obtained for  $\delta = 1$ , the Weibull distribution for  $\gamma = 1$  and the exponential distribution corresponds to  $\delta = \gamma = 1$ .

The class of generalized gamma distributions is in fact rather flexible. When  $\delta < 1$  ( $> 1$ ) the tails are heavier (lighter) than the exponential tails. When  $\delta\gamma \leq 1$  the density  $f(y)$  is strictly decreasing in  $y$ . Moreover  $\lim_{y \rightarrow 0} f(y)$  exists and is finite if and only if  $\delta\gamma \geq 1$ . When  $\delta\gamma > 1$  the density has a mode.

Plots of the empirical survival functions of  $c_{sij}$  for fixed  $s$  and  $i$  show that the distributions of the phase amplitudes have somewhat heavier tails than expected under (5.1) (the estimated values of  $\delta$  were less than 1). However in each sample the tendency is only significant for a few high values of  $s$ , and thus it is reasonable to consider exponentially distributed error variables, at least for low frequencies. The same conclusion is obtained by testing (5.1) using Bartlett tests.

Assuming the phase amplitudes  $c_{sij}$  are  $\text{Exp}(\lambda_{si})$ -distributed, the next step in the analysis is to fit a  $p$ -order model within each group,

$$(5.2) \quad \lambda_{si}^{-1} = \alpha_i + \beta_i(s^{2p_i} - 2^{2p_i}), \quad s = 2, \dots, S, \quad i = 1, 2.$$

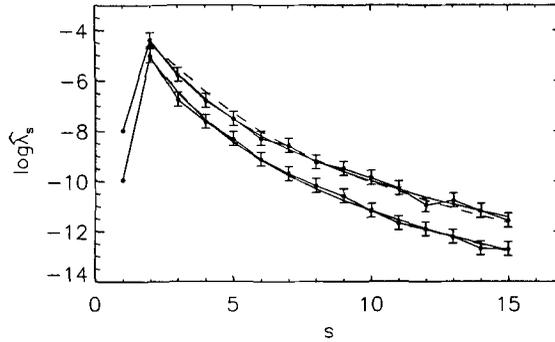


Fig. 9. The estimated regression  $\hat{\lambda}_s = [\hat{\alpha} + \hat{\beta}(s^{2\hat{p}} - 2^{2\hat{p}})]^{-1}$  in the Gaussian  $p$ -order model (solid) and the estimated regression under the Gaussian second-order model (dashed) is shown together with the average phase amplitudes as a function of  $s$  for the normal sample (lower curve) and the lymphoma sample (upper curve). The error bars are 95% confidence limits.

Table 2. The estimates and approximate confidence intervals, standard errors and correlation of  $(\log \hat{\alpha}, \log \hat{\beta})$ .

	$\log \hat{\alpha}$			$\log \hat{\beta}$			corr.
	est.	conf. int.	s.e.	est.	conf. int.	s.e.	
normal	5.08	4.81-5.35	0.14	1.97	1.89-2.05	0.04	-0.10
lymphoma	4.52	4.27-4.77	0.13	0.82	0.74-0.90	0.04	-0.13

The likelihood function is given by

$$L(\lambda_{si}; c_{sij}) = \prod_{j=1}^N \prod_{s=2}^S \lambda_{si}^{-1} \exp(-\lambda_{si}^{-1} c_{sij}) = \left\{ \prod_{s=2}^S \lambda_{si}^{-1} \exp(-\lambda_{si}^{-1} \bar{c}_{si.}) \right\}^N,$$

where  $\bar{c}_{si.} = \sum_{j=1}^N c_{sij} / N$  is the average of the amplitudes within the  $i$ -th group at phase  $s$  and  $\lambda_{si}$  is given by (5.2). As expected the estimated value of  $p$  is close to 2 in both samples (2.0 in the normal and 1.8 in the lymphoma sample), and again we consider the second-order model. The estimated regression lines are shown in Fig. 9 and in Table 2 the estimates and approximate standard errors and correlation coefficients based on the observed information are summarized. As in the previous subsection we observe a significant difference between the two samples in the value of  $\beta$ . The difference in  $\alpha$  is not as significant.

### 5.3 Simulations from the Gaussian second-order model

In the Gaussian second-order model truncated at  $S = 15$  we have

$$C_s \sim \text{Exp}(\lambda_s), \quad s = 2, \dots, 15, \quad \text{independent,}$$

with

$$(5.3) \quad \lambda_s^{-1} = \alpha + \beta(s^4 - 2^4).$$

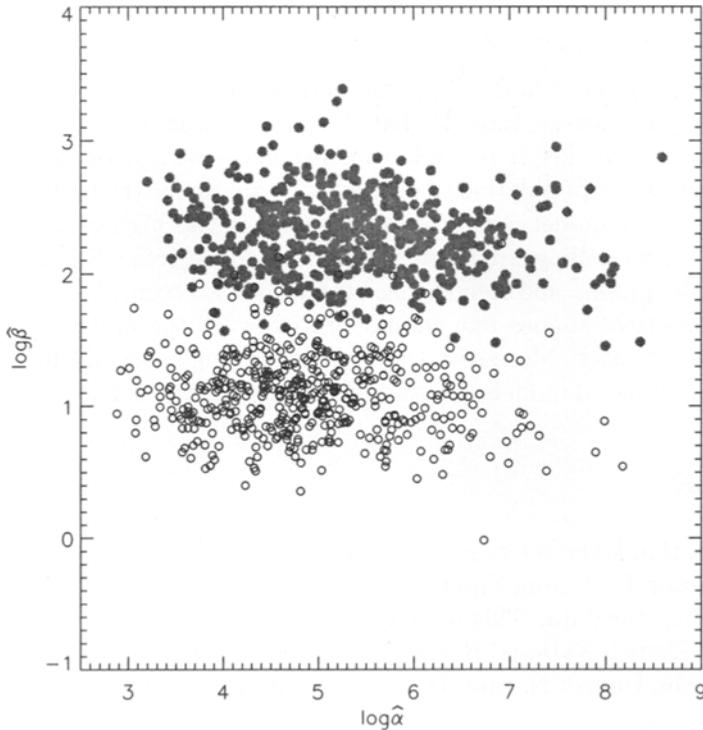


Fig. 10. Simulated distribution of  $(\hat{\alpha}, \hat{\beta})$  under the normal second-order model is shown for the normal sample ( $\bullet$ ) and the lymphoma sample ( $\circ$ ).

Table 3. The average, standard deviation and correlation of  $(\log \hat{\alpha}, \log \hat{\beta})$  for each sample.

	$\log \hat{\alpha}$		$\log \hat{\beta}$		corr.
	av.	s.d.	av.	s.d.	
normal	5.36	1.01	2.28	0.30	-0.14
lymphoma	4.89	0.95	1.10	0.30	0.03

In order to investigate the model more carefully we conducted the following simulation study. For each sample we calculated  $\lambda_s$  according to (5.3) with  $(\alpha, \beta)$  replaced by the average estimated value from Table 1 and simulated  $C_s \sim \text{Exp}(\lambda_s)$ ,  $s = 2, \dots, 15$ . From the values of  $C_s$  we calculated the maximum likelihood estimates of  $\alpha$  and  $\beta$ . This procedure was repeated 500 times for each of the samples and the result is shown in Fig. 10 and summarized in Table 3.

When we compare Figs. 6 and 10 it is seen that the variation in  $\log \hat{\alpha}$  is almost the same for the observed and simulated data for both groups. The variation range of  $\log \hat{\beta}$  is smaller in the simulation study than in the samples. One explanation is that the local shape variability in the data is somewhat higher than predicted from the Gaussian model, i.e. the assumption that the error variables are exponential distributed is not appropriate at high phases. Another reason could be that a well located ‘blob’ results in many high phase amplitudes.

## 6. Perspectives

The generalized  $p$ -order model is useful for describing a population of approximately circular objects. Let us stress, though, that the Fourier expansion (3.1) of the normalized radius-vector function makes it possible to construct a variety of shape models. If, for instance, one considers a population of objects with dominant triangular shapes, it would be natural to apply a model where on average  $C_3$  is the highest amplitude. A more challenging task is modelling approximately elliptical objects. Ellipses have vanishing amplitudes at odd phases and decreasing amplitudes at even phases. Thus, to model approximately elliptical shapes one should probably let the odd and even amplitudes decrease at different rates. Moreover, the even phase angles should have approximately the same values. Elliptical models are studied in Hobolth and Jensen (2000) and Hobolth *et al.* (2002b).

### Acknowledgements

The authors thank the referees for valuable comments which helped to improve the paper and Professor Flemming Sørensen, Institute of Pathology, University of Aarhus, for kindly providing the data. This work was supported in part by MaPhySto, funded by a grant from the Danish National Research Foundation. The second author is supported by a grant from the Danish Natural Science Research Council.

### Appendix

#### *Characterization of asymmetry and centre of mass*

Let  $x = (x_1, x_2)$  denote a generic point in  $\mathbf{R}^2$  and let  $\|x\| = (x_1^2 + x_2^2)^{1/2}$ .

**PROPOSITION.** *Let  $z = (z_1, z_2)$  be an interior point of a compact subset  $K$  of  $\mathbf{R}^2$ . Let  $K$  be star-shaped with respect to  $z$  and let the radius-vector function  $r_K(t; z)$  be continuously differentiable in  $t$ .*

(i) *We have*

$$(A.1) \quad \int_K \frac{x_1 - z_1}{\|x - z\|^2} dx_1 dx_2 = 2\pi \int_0^1 r_K(t; z) \cos(2\pi t) dt$$

$$(A.2) \quad \int_K \frac{x_2 - z_2}{\|x - z\|^2} dx_1 dx_2 = 2\pi \int_0^1 r_K(t; z) \sin(2\pi t) dt.$$

(ii) *If  $z$  is the centre of mass of  $K$  then*

$$(A.3) \quad \int_0^1 r_K(t; z)^3 \cos(2\pi t) dt = \int_0^1 r_K(t; z)^3 \sin(2\pi t) dt = 0.$$

*Conversely, if  $z$  is such that (A.3) is satisfied then  $z$  is the centre of mass of  $K$ .*

(iii) *Let  $r_K(t; z) = 1 + 2\sqrt{c_s} \cos(2\pi s(t - d_s))$ , where  $s \geq 2$ ,  $0 \leq c_s \leq 1/4$  and  $d_s \in [0, \frac{1}{s}]$ . Then  $z$  is the centre of mass of  $K$ .*

**PROOF.** Let  $F : [0, 1]^2 \rightarrow \mathbf{R}^2$  be defined by

$$F(v, t) = (z_1, z_2) + vr_K(t; z)(\cos(2\pi t), \sin(2\pi t)).$$

Then  $F$  is onto  $K$  and  $|\det(F'(v, t))| = 2\pi v r_K(t; z)^2$ . In order to prove (A.1) note that if  $x = (x_1, x_2) \in K$  is such that  $x = F(v, t)$  then

$$\frac{x_1 - z_1}{\|x - z\|^2} = \frac{\cos(2\pi t)}{v r_K(t; z)},$$

and from the transformation theorem we get

$$\int_K \frac{x_1 - z_1}{\|x - z\|^2} dx_1 dx_2 = 2\pi \int_0^1 r_K(t; z) \cos(2\pi t) dt.$$

The result (A.2) is proved similarly.

The same kind of arguments show that

$$(A.4) \quad \left( \int_K (x_1 - z_1) dx_1 dx_2, \int_K (x_2 - z_2) dx_1 dx_2 \right) \\ = \frac{2\pi}{3} \left( \int_0^1 r_K(t; z)^3 \cos(2\pi t) dt, \int_0^1 r_K(t; z)^3 \sin(2\pi t) dt \right).$$

The left-hand side is zero if and only if  $z$  is the centre of mass of  $K$ . Therefore (ii) is an immediate consequence of (A.4).

To prove (iii) one has to show that  $r_K(t; z) = 1 + 2\sqrt{c_s} \cos(2\pi s(t - d_s))$  satisfies the condition (A.3). This follows from elementary calculations, and is left to the reader.

## REFERENCES

- Bloomfield, P. (1976). *Fourier Analysis of Time Series: An Introduction*, Wiley, New York.
- Cramér, H. and Leadbetter, M. R. (1967). *Stationary and Related Stochastic Processes*, Wiley, New York.
- Grenander, U. (1993). *General Pattern Theory*, Oxford University Press, Oxford.
- Grenander, U. and Miller, M. I. (1994). Representations of knowledge in complex systems (with discussion), *J. Roy. Statist. Soc. Ser. B*, **56**, 549–603.
- Hansen, M. B., Møller, J. and Tøgersen, F. Aa. (2002). Bayesian contour detection in a time series of ultra-sound images through dynamic deformable template models, *Biostatistics*, **3**, 213–228.
- Hobolth, A. and Jensen, E. B. V. (2000). Modelling stochastic changes in curve shape, with an application to cancer diagnostics, *Advances in Applied Probability*, **32**, 344–362.
- Hobolth, A., Kent, J. T. and Dryden, I. L. (2002a). On the relation between edge and vertex modelling in shape analysis, *Scand. J. Statist.*, **29**, 355–374.
- Hobolth, A., Pedersen, J. and Jensen, E. B. V. (2002b). A deformable template model, with special reference to elliptical templates, *J. Math. Imaging Vision*, **17**, 131–137.
- Hurn, M. A., Steinsland, I. and Rue, H. (2001). Parameter estimation for a deformable template model, *Statist. Comput.*, **11**, 337–346.
- Jensen, E. B. V. (1998). *Local Stereology*, World Scientific, Singapore.
- Kent, J. T., Mardia, K. V. and Walder, A. N. (1996). Conditional cyclic Markov random fields, *Advances in Applied Probability*, **28**, 1–12.
- Kent, J. T., Dryden, I. L. and Anderson, C. R. (2000). Using circulant symmetry to model featureless objects, *Biometrika*, **87**, 527–544.
- Lestrel, P. E. (ed.) (1997). *Fourier Descriptors and Their Applications in Biology*, Cambridge University Press, New York.
- Loncaric, S. (1998). A survey of shape analysis techniques, *Pattern Recognition*, **31**, 983–1001.
- Rogers, L. C. G. and Williams, D. (1994). *Diffusions, Markov Processes, and Martingales. Volume 1: Foundations*, 2nd ed., Wiley, Chichester.

- Rue, H. and Hurn, M. A. (1999). Bayesian object recognition, *Biometrika*, **86**, 649–660.
- Rue, H. and Syversveen, A. R. (1998). Bayesian object recognition with Baddeley's delta loss, *Advances in Applied Probability*, **30**, 64–84.
- Stoyan, D. and Stoyan, H. (1994). *Fractals, Random Shapes and Point Fields. Methods of Geometrical Statistics*, Wiley, Chichester.
- Zahn, C. T. and Roskies, R. Z. (1972). Fourier descriptors for plane closed curves, *IEEE Trans. Comput.*, **C-21**, 269–281.