Mixture analysis of asymmetry: modelling directional asymmetry, antisymmetry and heterogeneity in fluctuating asymmetry

Abstract
The occurrence of different forms of asymmetry complicates the analysis and interpretation of patterns in asymmetry. Furthermore, between-individual heterogeneity in developmental stability (DS) and thus fluctuating asymmetry (FA), is required to find relationships between DS and other factors. Separating directional asymmetry (DA) and antisymmetry (AS) from real FA and understanding between-individual heterogeneity in FA is therefore crucial in the analysis and interpretation of patterns in asymmetry. In this paper we introduce and explore mixture analysis to (i) identify FA, DA and AS from the distribution of the signed asymmetry, and (ii) to model and quantify between-individual heterogeneity in developmental stability and FA. In addition, we expand mixtures to the estimation of the proportion of variation in the unsigned FA that can be attributed to between-individual heterogeneity in the presumed underlying developmental stability (the so-called hypothetical repeatability). Finally, we construct weighted normal probability plots to investigate the assumption of underlying normality of the different components. We specifically show that (i) model selection based on the likelihood ratio test has the potential to yield models that incorporate nearly all heterogeneity in FA; (ii) mixtures appear to be a powerful and sensitive statistical technique to identify the different forms of asymmetry; (iii) restricted measurement accuracy and the occurrence of many zero observations results in an overestimation of the hypothetical repeatability on the basis of the model parameters; and (iv) as judged from the high correlation coefficients of the normal probability plots, the underlying normality assumption appears to hold for the empirical data we analysed. In conclusion, mixtures provide a useful statistical tool to study patterns in asymmetry.

Keywords
Antisymmetry, directional asymmetry, EM algorithm, fluctuating asymmetry, hypothetical repeatability, mixtures, normality, statistics.

INTRODUCTION
Fluctuating asymmetry (FA, i.e. random deviations from perfect symmetry, Van Valen 1962) has become a popular measure of developmental stability (DS, i.e. the ability of an individual to buffer its development against random perturbations, Palmer & Strobeck 1992). The measurement and analysis of asymmetry must follow a number of specific guidelines for three reasons.

First, degrees of asymmetry are often small and may become confounded with measurement error (Lundström 1960; Greene 1984; Palmer & Strobeck 1986). Therefore, repeated measurements and mixed model analysis are generally applied to separate real asymmetry from measurement error (Palmer & Strobeck 1986; Palmer 1994; Mørila & Björklund 1995; Van Dongen et al. 1999a).

Second, because the estimation of individual DS by single trait asymmetry is an attempt to estimate a variance with two datapoints (i.e. left and right trait value), the correlation between individual asymmetry and the presumed underlying DS is weak (Whitlock 1996; Houle 1997). As a result, estimates of heritability of FA and between-trait correlations in FA underestimate patterns in DS. Whitlock (1996; see also Van Dongen 1998a; Whitlock 1998; Gangestad &
Thornhill 1999) proposed a method, based on the hypothetical repeatability \( R \) of individual asymmetry, to convert patterns in FA into patterns in the presumed underlying DS. \( R \) estimates the proportion of the total variation in the unsigned FA that results from between-individual heterogeneity in the underlying DS (Whitlock 1996).

Other forms of asymmetry, which do not necessarily reflect DS, may be present as well (Palmer & Strobeck 1992; Graham et al. 1993; Rowe et al. 1997). There is at present no consensus whether these other forms can provide important information about underlying patterns of stability. The distribution of the signed asymmetry (i.e. left–right trait value) is in the case of real FA assumed to be reflected by a normal or a combination of several normal distributions with zero means. An admixture of several normal distributions with zero mean results in a positive kurtosis (leptokurtosis) but does not influence the mean (Palmer & Strobeck 1992; Van Dongen 1998a, b). Directional asymmetry (DA), where the degree of asymmetry is biased in one direction, is assumed to be normally distributed with a mean differing from zero. The distribution of antisymmetry (AS) is typically bimodal, thus showing a lack of symmetrical individuals relative to FA (Fig. 1, see also Palmer & Strobeck 1992; Rowe et al. 1997). As the occurrence of different forms of asymmetry has far-reaching implications for both the analysis (Palmer & Strobeck 1986, 1992; Palmer 1994; Whitlock 1996, 1998; Van Dongen 1998a, b, 1999) and interpretation (Palmer & Strobeck 1992; Graham et al. 1993) of asymmetry data, their identification and separation are an important aspect in the analysis of FA and DS.

In spite of its relevance, there are only a few specific tests that allow the three forms of asymmetry to be distinguished. DA can be detected by mixed model analysis (Palmer & Strobeck 1986; Van Dongen et al. 1999a), whereas a negative kurtosis (platykurtosis) is indicative of AS (e.g. Palmer & Strobeck 1986). While these tests are relatively powerful when only one type of asymmetry is present, they have limited power and relevance when different types co-occur (Palmer & Strobeck 1992). The distribution of the signed asymmetry becomes skewed when FA and DA are combined, possibly affecting the performance of mixed ANOVA and regression models that are used to detect DA (Palmer & Strobeck 1986; Van Dongen et al. 1999a) and assume normality. A blend of FA and AS may result in both leptokurtic or platykurtic distributions with zero mean (Palmer & Strobeck 1992; Van Dongen 1998b). Therefore, a mix of different levels of FA and of FA and AS cannot always be separated on the basis of the kurtosis only (Palmer & Strobeck 1992; Van Dongen 1998b). Hence, there is a need for a unifying statistical technique that allows the identification of admixtures of the different forms of asymmetry.

The three types of asymmetry and admixtures among them can be represented as combinations of several normal distributions with different means and/or variances (Palmer & Strobeck 1992). Mixture analysis is a statistical technique that allows modelling of combinations of several normal distributions (Böhning et al. 1992). In this paper we explore the applicability of mixtures in the analysis of patterns in asymmetry. In the next section we introduce a suite of models that allow testing of specific hypotheses (with the likelihood ratio test) enabling us to separate the different forms of asymmetry. In addition, mixture models will be used to quantify between-individual heterogeneity in FA and DS. Based on the parameter estimates of the latter we introduce a new estimate of the hypothetical repeatability (further called model-based \( R \), see Whitlock 1996, 1998; Van Dongen 1998a; and Gangestad & Thornhill 1999 for other procedures that will be denoted as database estimates). Weighed normal probability plots are constructed to verify the assumption of normality. We then analyse both empirical and simulated datasets to explore possibilities and difficulties in the use of mixtures. First, we evaluate the efficiency of mixtures to model heterogeneity in FA and thus in the underlying DS. For three datasets (gamma1, gamma2 and winter moth, see below) we select models on the basis of the likelihood ratio test (below). Ideally, the selected model should reflect real FA only and should describe (nearly) all variation in DS (as expressed by the model-based \( R \) present in the data (as expressed by the databased \( R \), estimated following Van Dongen 1998a). Secondly, we investigate the effect of restricted measurement accuracy on the performance of mixtures. Asymme-

**Figure 1** Density plots of real fluctuating asymmetry (FA), directional asymmetry (DA) and antisymmetry (AS). Distributional characteristics [i.e. normal distributions with given mean \( \mu \) and variance \( \sigma^2 \): \( N(\mu, \sigma^2) \)] are indicated.

©1999 Blackwell Science Ltd/CNRS
try of a trait is continuous except for meristic traits, which require particular methods (Palmer 1994). However, due to technical limitations like available measurement tools and time constrains, measurement accuracy renders asymmetry discrete. As a result, datasets often contain many zero observations from nearly, but often not perfectly, symmetrical individuals. To what extent this affects the performance of mixtures and possible solutions to the problem are explored in two datasets (gamma 1b and White-starred robin). Finally, we illustrate the sensitivity of mixtures to detect DA and AS with three datasets (Indian meal moth, rare antsymmetry and barn swallow, see below). In a final section we investigate the shape of the likelihood surface of mixtures with two simulations.

MATERIALS AND METHODS

Mixtures and the maximization of the likelihood by the EM algorithm

A random variable $X$ is said to follow a mixture of $k$ different normal distributions (further called components) if:

$$X \sim \sum_{j=1}^{k} q_j N(\mu_j, \sigma_j^2)$$

where $\mu_j$ and $\sigma_j^2$ represent the mean and variance of the $j$th component and $q_j$ represents the proportion of the total population following this distribution with $q_1 + q_2 + \ldots + q_k = 1$ (Böhning et al. 1992). The corresponding density function is given by:

$$f(x; \theta) = \sum_{j=1}^{k} q_j \phi(x; \theta_j)$$

where the unknown parameter vector $\theta$ consists of $k$ component parameters $\theta_1, \ldots, \theta_k$ with $\theta_j = (\mu_j, \sigma_j^2)$ and $k$ proportions $q_1, \ldots, q_k$. $\phi(x; \theta_j)$ denotes the normal density function of component $j$:

$$\phi(x; \theta_j) = \frac{1}{\sigma_j \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x-\mu_j}{\sigma_j}\right)^2}$$

The log-likelihood of a $k$ component mixture can then be written as:

$$l(P; x) = \sum_{i=1}^{n} \ln \left( \sum_{j=1}^{k} q_j \phi(x_i; \theta_j) \right)$$

This likelihood function can be maximized by the EM algorithm, a technique to maximize the likelihood of an incomplete dataset (Dempster et al. 1977). The analysis of mixtures can be fitted in this framework by observing that each datapoint $i$ has $k$ unknown probabilities $p_{ij}$ reflecting its chance to be generated from the $j$th component. The EM algorithm then follows two steps. In the E-step, these individual probabilities are determined from the current parameter estimates (in iteration $t$) and from the data as:

$$p_{ij}^{(t+1)} = \frac{q_j^{(t)} \phi(x_i; \theta_j^{(t)})}{\sum_{j=1}^{k} q_j^{(t)} \phi(x_i; \theta_j^{(t)})}$$

In the M-step, these new posterior probabilities $p_{ij}$ and the data are used to obtain improved estimates of the different parameters. In the most general model, the mean values and variances of each component are given by:

$$\mu_j^{(t+1)} = \frac{\sum_{i=1}^{n} p_{ij}^{(t+1)} x_i}{\sum_{i=1}^{n} p_{ij}^{(t+1)}}$$

$$\sigma_j^2^{(t+1)} = \frac{\sum_{i=1}^{n} p_{ij}^{(t+1)} (x_i - \mu_j^{(t+1)})^2}{\sum_{i=1}^{n} p_{ij}^{(t+1)}}$$

The E- and M-steps are then iteratively repeated until the likelihood converges to a maximum. Upon convergence the posterior probabilities can be used to classify an observation into a particular component (Böhning et al. 1992) and to construct a weighted normal probability plot (see below).

Models of asymmetry and model selection

The particular nature of the three forms of asymmetry limits the number of parameter settings that need to be evaluated. FA is modelled as a normal distribution with zero mean, DA as a normal distribution with nonzero mean, and AS as a bimodal distribution consisting of two normal distributions with nonzero means of equal magnitude and opposite sign, equal variance and equal density (Palmer & Strobeck 1992). Thus, FA, DA and AS are modelled by components with one, two and two parameters, respectively. In addition $k-1$ proportions are estimated in a $k$-component mixture. The three different forms of asymmetry can be formally denoted as (see also Table 1 and Fig. 1):

FA: $X \sim N(0, \sigma^2)$
DA: $X \sim N(\mu, \sigma^2)$
AS: $X \sim 0.5 \times N(-\mu, \sigma^2) + 0.5 \times N(\mu, \sigma^2)$

The multicomponent mixture models considered in this paper are described in Table 1.

Model selection can be based on the likelihood ratio test. In some cases this test does not fulfill the usual regularity conditions needed for asymptotic chi-squared distribution theory. Let us illustrate this. Suppose we want to test if a two-component FA model should be selected over a single component FA model. The two component model can be written as:

$$q_1 N(\mu_1, \sigma_1^2) + q_2 N(\mu_2, \sigma_2^2)$$

with $\mu_1 = \mu_2 = 0$
The null hypothesis can be formulated in two different ways: $q_2 = 0$ or $\sigma^2 = \sigma^2$. However, in the first case $\sigma^2$ is undefined whereas in the latter case $\sigma_1$ (and by definition also $\sigma_2$) is undefined. In addition, mixture models in which one component reflects AS are not nested within mixture models of FA or DA components, again invalidating the classical likelihood ratio theory. Asymptotic chi-squared distribution theory does hold for the null hypothesis $\mu_1 = \mu_2$ or $\mu_1 = \mu_2 = 0$. When regularity conditions are not satisfied the distribution of the likelihood ratio test statistic is obtained from Monte Carlo simulations. Data are repeatedly (1000 times) generated under the conditions of the null hypothesis and the null and the alternative model are fitted to the data. The proportion of samples for which the likelihood ratio test statistic was larger than the observed value was used as $P$-value (further denoted as $p_{sim}$).

To illustrate that the classical likelihood ratio test may be misleading when regularity conditions are not satisfied, we simulate the distribution of the likelihood ratio test statistic under the null hypothesis for three different test situations. Simulations are based on the gamma2 dataset (see below) and we refer to the results section for more details.

HYPOTHETICAL REPEATABILITY

An important parameter in the study of patterns in real FA is the hypothetical repeatability $R$ (Whitlock 1996). This parameter estimates the proportion of the total variation in the unsigned asymmetry (i.e. absolute value of the signed asymmetry) that results from between-individual heterogeneity in DS (Whitlock 1996). Traditionally, estimation is based on distributional characteristics of signed and/or unsigned asymmetry (Whitlock 1998; Van Dongen 1998a; Gangestad & Thornhill 1999). The parameter estimates of mixtures directly estimate heterogeneity in FA, and thus in the presumed underlying DS.

Therefore, if the selected model reflects real FA only, parameter estimates can be used to estimate $R$ as well. Deriving computational formule relies on the close relationship between parameters of the normal (like the signed asymmetry) and the half-normal distribution (like the unsigned asymmetry) (Whitlock 1996). Following Whitlock (1996) and Van Dongen (1998a) it can be shown that the total variance in the unsigned asymmetry can be decomposed into two components: one reflecting sampling variance ($\sigma^2_s$), the other between-individual heterogeneity in DS ($\sigma^2_{DS}$), which are given by:

$$\sigma^2_s = \frac{\pi - 2}{\pi} \sigma^2_{sign FA}$$

and

$$\sigma^2_{DS} = \sum_{j=1}^{k} q_j (\mu_{unsigned FA_j} - \bar{\mu}_{unsigned FA})^2$$

where

$$\sigma^2_{sign FA} = \sum_{j=1}^{k} q_j \sigma^2_{j}, \mu_{unsigned FA_j} = 2 \sqrt{\sigma^2_j / (2 \pi)}$$

and

$$\bar{\mu}_{unsigned FA} = \sum_{j=1}^{k} q_j \bar{\mu}_{unsigned FA_j}$$

The hypothetical repeatability is then given by the proportion of the total variance in the unsigned FA reflecting heterogeneity in the underlying DS:

$$R = \frac{\sigma^2_{DS}}{\sigma^2_s + \sigma^2_{DS}}$$

Model assumptions

It is generally assumed that the different forms of asymmetry can be described by one or more normal distributions (Palmer & Strobeck 1992) and that between-individual heterogeneity in DS results in a blend of
normal distributions with different variances and zero mean (Whitlock 1996). In order to verify this assumption we construct weighted normal probability plots (NP plot) using the posterior probabilities as weight factors. These plots contrast the observed values against the expected values under the assumption of normality and under the parameter estimates of the selected mixture model. If this assumption is correct, this plot should form a straight line (Neter et al. 1990; Sokal & Rohlf 1995). The expected value of observation $i$ is obtained as:

$$E_i = \sum_{j=1}^{k} p_j \left[ \frac{z}{\sqrt{\sigma_j^2}} \left( \frac{r - 0.375}{n + 0.25} \right) + \mu_j \right]$$

where $r$ is the rank order of the $i$th observation and $z(\lambda)$ denotes the $(\lambda) \times 100$th percentile of the standard normal distribution (see also Neter et al. 1990). We construct NP plots to allow visual inspection of the shape of the association and we compute the coefficient of correlation. As a rule of thumb, values above 0.95 can be considered as indicative for approximate normality (Neter et al. 1990).

Datasets

**Gamma1 data:** We simulated a dataset of 1000 asymmetry values from normal distributions with mean zero and varying variance. The distribution of the variance followed a gamma distribution with $\alpha = 2$ and $\beta = 1$. In this way we modelled between-individual heterogeneity in DS at a continuous scale.

**Gamma1b data:** This dataset was derived from Gamma1 after rounding the data to the nearest integer.

**Gamma2 data:** To evaluate the performance of mixtures with relative small samples we analysed the first 200 asymmetry values of the Gamma1 dataset.

**Winter moth data (WM):** We analysed asymmetry in the length of the front tibia from 715 winter moths (*Operopherta brimuta*). Measurements were performed under a microscope and are described in Van Dongen et al. (1999b, c). We used the signed asymmetry by averaging over the two within-subject repeats. Mixed regression analysis revealed that measurement error (ME) was relatively small (Van Dongen et al. 1999b).

**White-starred Robin data (WR):** We analysed tarsus asymmetry of 435 individuals caught in the Taita Hills, Kenya. Left and right side were measured twice with callipers and averaged for our analysis. ME was small (Lens et al. 1999; Lens & Van Dongen 1999) and 39% of the observations equaled zero.

**Indian meal moth data (IMM):** The asymmetry in front tibia lengths of 414 Indian meal moths (*Plodia interpunctella*) was measured as for the winter moth and part of the data are reported in Van Dongen et al. (1999c). Averages over the two within-subject repeats were used.

**Rare antisymmetry data (RA):** To illustrate the sensitivity of mixtures to detect antisymmetry we simulated a dataset of 1000 observations of which 90\% exhibited real FA (mean and variance equal to zero and one, respectively) and 10\% were antisymmetric (means of the bimodal distribution equal to $-3$ and 3 and variance equal to 1).

**Barn swallow data (BS):** There has been much recent debate on whether the negative relationship between the unsigned individual FA and trait size of signalling traits reflects condition-dependent FA or AS (see Rowe et al. 1997 favouring AS and Møller 1999 favouring FA). Rowe et al. (1997) have argued that a proportion of barn swallows showed antisymmetry in their tail feathers, while Møller (1999), based on a much larger dataset, did not find indications for the presence of AS. We analyse tail asymmetry data from 1469 barn swallows (*Hirundo rustica*) (see Møller 1999 for details) kindly provided to us by A.P. Møller. We analyse both sexes together but separate analyses gave similar results.

**Performance of the EM-algorithm**

It is well known that the surface of the likelihood of mixture models may exhibit several local optima. As a result, the EM-algorithm may fail to converge to the maximum likelihood value (i.e. global optimum). To circumvent this problem, it has been suggested to try several starting values, to let the EM-algorithm run for a long time and then to select the highest local optimum. There is, however, no general rule to select starting values and to terminate the iteration procedure (Böhning et al. 1992). In order to evaluate the importance of this problem in the context of modelling asymmetry we explore the likelihood surface for the joint estimation of different parameters in two simulated datasets. The first exhibits two different levels of real FA ($\mu_1 = \mu_2 = 0$, $\sigma_1^2 = 1$ and $\sigma_2^2 = 3$) with equal probability ($q_1 = q_2 = 0.5$). The second represents a mixture of real FA and antisymmetry (RA dataset, see above). The likelihood surface was constructed for the joint estimation of two parameters keeping the others fixed and equal to the parameter values of the simulation model. We examined the joint estimation of the two variances and the two proportions for the real FA dataset and the joint estimation of the mean and proportion and the mean and variance of the antisymmetry component in the RA dataset.

**RESULTS**

**Analysis of the datasets**

(a) Model selection and the performance of mixtures in describing heterogeneity in FA

**Gamma1:** A four-component mixture was selected. Its
likelihood was higher compared to the three-component model (LR = 8.3, $p_{s} = 0.004$), whereas an additional fifth component did not significantly increase the likelihood (LR = 0.01, $p_{s} = 0.6$). The mean values did not differ from zero (LR = 2.2, d.f. = 4, $P = 0.7$). Variances (proportions) of the four components equaled 0.20 (17%), 1.67 (52%), 15.3 (29%) and 71.3 (2%). The correlation coefficient of the NP plot equaled 0.996. The hypothetical repeatability calculated from the data and the model parameter estimates both equaled 0.40. The four-component FA mixture therefore reflected all observed heterogeneity in FA and observed and expected distributions matched very closely.

**Gamma2:** A three-component mixture was selected. Its likelihood was higher compared to the two-component model (LR = 6.0, $p_{s} = 0.01$), whereas an additional fourth component did not significantly increase the likelihood of the model (LR = 2.6, $p_{s} = 0.07$). The mean values did not differ significantly from zero (LR = 2.4, d.f. = 3, $P = 0.5$). The variances (proportions) equaled 0.16 (28%), 5.3 (61%) and 23.7 (11%). The correlation coefficient of the NP plot equaled 0.995. The hypothetical repeatabilities estimated from the data ($= 0.340$) and the model parameter estimates ($= 0.335$) were of a comparable magnitude.

We determined the distribution of the likelihood ratio test statistic for three different hypotheses by Monte Carlo simulations. The null hypothesis was in each case the three-component FA mixture. The alternative hypotheses were (i) a three-component FA-AS, (ii) a four component FA and (iii) a three-component DA mixture. In the first two cases, standard regularity conditions are not met. A traditional likelihood ratio would have been performed with one and two degrees of freedom, respectively. However, the distribution of the likelihood ratio test statistic under the null hypothesis did not follow these $\chi^2$ distributions (Fig. 2). For the third alternative hypothesis, where the standard regularity conditions were satisfied, the likelihood ratio test statistic followed the expected $\chi^2$ distribution with three degrees of freedom (Fig. 2). This result suggests that the traditional likelihood ratio test is conservative when regularity conditions are not satisfied (see Stram & Lee 1994 for a similar result in the context of longitudinal models).

**Winter moth:** A three-component mixture of FA best described the distribution of the signed asymmetry. Mean values did not differ from zero (LR = 6.14, d.f. = 3, $P = 0.1$), and an additional fourth FA component did not explain significantly more variation in the signed asymmetry (LR = 1.4, $p_{s} = 0.1$). However, two FA components were insufficient (LR = 38.2, $p_{s} < 0.0001$). There was no indication of asymmetry (LR = 0.0, $p_{s} = 0.8$). The correlation coefficient of the NP plot equaled 0.97. Nevertheless, at the tails of the distribution it did not closely follow the straight line (Fig. 3). Thus, the WM data appear to reflect real FA. The model-based hypothetical repeatability of 0.47 approximated the database value of 0.48, indicating that model selection on the basis of the likelihood ratio test leads to models that embody nearly all heterogeneity in DS.

(b) **Effect of restricted measurement accuracy**

**Gamma1b:** Rounding the observations of the Gamma1 dataset to the nearest integer simulates restricted measurement accuracy and the occurrence of many zero observations. This has two important consequences. First, the databased hypothetical repeatability was slightly increased towards 0.41. Secondly, modelling multicomponent mixtures of real FA led to models containing one component with zero variance, inflating the likelihood towards infinity. Limiting the lower bound of the variance to a very small value (e.g. $10^{-8}$) can solve this technical problem. In this way the zero observations are modelled separately in a component with zero mean and near zero variance. This approach led to the selection of a four component mixture (i.e. zero observations + three FA components). Mean values did not differ significantly from zero (LR = 0.06, d.f. = 3, $P = 0.9$). An additional
component did not significantly increase the likelihood ($LR = 0.02$, $p_{\text{sim}} > 0.6$) while a three component model had a significantly lower likelihood ($LR = 4.3$, $p_{\text{sim}} = 0.03$). Variances and proportions (within brackets) equaled $10^{-8}$ (30%), 3.5 (46%), 18.5 (22%) and 79 (2%). The correlation coefficient of the NP plot equalled 0.997. The model-based value of $R$ equalled 0.445, hereby overestimating $R$ of the underlying model (i.e. 0.40, see above). The reason for this is that the zero observations are all assumed to show nearly perfect DS (i.e. variance of FA $\approx 0$), whereas this is a underestimate of their real DS and thus increases heterogeneity in DS.

White-stared Robin: Limiting the variances of the mixture components to $10^{-8}$ resulted in the selection of a three-component model (zero observations + two FA components). Mean values did not differ from zero ($LR = 3$, d.f. = 2, $P = 0.2$). An additional fourth component did not significantly increase the likelihood ($LR = 0.01$, $p_{\text{sim}} > 0.5$), while a two component model had a significantly lower likelihood ($LR = 128$, $p_{\text{sim}} < 0.0001$). Variances and proportions (within brackets) equaled $10^{-8}$ (39%), 0.008 (42%) and 0.18 (19%). The correlation coefficient of the NP plot equalled 0.98. As before, the model-based value of $R$ (= 0.522) was higher than the databased value (= 0.507), thereby overestimating between-individual heterogeneity in DS.

(c) Sensitivity of mixtures in detecting DA and AS

Indian meal moth: A two component mixture was selected. Mean values differed significantly from zero ($LR = 7.12$, d.f. = 2, $P = 0.02$). For both components the signed asymmetry was on average negative [mixture model: $0.81 \times N(-0.006, 0.003) + 0.19 \times N(-0.004, 0.0001)]$. An additional third component did not improve the fit significantly ($LR = 2.8$, $p_{\text{sim}} = 0.44$), but a single component was insufficient to describe the data ($LR = 20.2$, $p_{\text{sim}} < 0.001$). There was no indication of antisymmetry ($LR < 0$). The NP plot formed a nearly straight line ($r = 0.99$) indicating approximate normality of the individual components.

Rare antisymmetry: An FA-AS mixture was selected over the two-component FA model. The additional parameter required to model AS (i.e. one mean value, Table 1) significantly increased the likelihood of the model ($LR = 4.9$, $p_{\text{sim}} = 0.02$). An additional third FA component did not significantly increase the likelihood of the model ($LR = 0.1$, $p_{\text{sim}} = 0.6$). Predicted and observed densities of the FA-AS mixture matched closely and the correlation coefficient of the NP plot equalled 0.99.

Barn swallow: This dataset contained many zero observations. This resulted in the convergence of the EM-algorithm towards a zero-variance model for one of the components in the mixture. As before, we avoided this problem by limiting the lower bound of the variances to $10^{-8}$. A four-component FA-AS mixture was selected. The additional parameter to model AS (i.e. one mean value, Table 1) increased the likelihood significantly relative to a four component FA mixture ($LR = 11.1$, $p_{\text{sim}} = 0.03$). There was no directional component apart from the AS (i.e. mean values of the three other components did not differ significantly from zero: $LR = 3$, d.f. = 2, $P = 0.2$). Adding a fifth FA component did not significantly increase the likelihood ($LR = 0.1$, $p_{\text{sim}} > 0.5$). A three component FA-AS mixture had a significantly lower likelihood ($LR = 340$, $P < 0.0001$). The correlation coefficient of the NP plot for the four component FA-AS mixture equalled 0.98. Figure 4
relatively high variance, relative to the global optimum (Fig. 5). This explains why during the estimation of antisymmetry for the datasets above, the EM-algorithm failed to detect any antisymmetry unless the initial mean was chosen relatively large. Otherwise, the EM-algorithm converged to a model with near zero mean and relatively high variance of the AS component, hereby reflecting FA instead of AS. Thus, in order to detect antisymmetry, the selection of appropriate starting values appears important. This problem becomes more critical with decreasing proportion of individuals that exhibit antisymmetry (data not shown).

**DISCUSSION**

Because the identification of different forms of asymmetry (fluctuating asymmetry, directional asymmetry and antisymmetry) is a central topic in the study of patterns in asymmetry, there is a need for sensitive and powerful statistical methods to differentiate the three types of asymmetry. Furthermore, real FA is often composed of several components with different variances (Whitlock 1996; Gangestad & Thornhill 1999; Van Dongen 1999). The hypothetical repeatability (Whitlock 1996) is an important parameter estimating the proportion of variation that is due to this heterogeneity. However, there is no method to describe this presumed heterogeneity. We show that mixtures may provide a very useful tool to

![Figure 4](image-url)  
**Figure 4** (a) Observed distribution of the signed asymmetry (histogram) and predicted density distribution (solid line) of the four component FA-AS mixture for the barn swallow dataset. Plot (b) shows a blow up illustrating the very rare antisymmetry. Mean and variance of the different mixture components as well as their proportions are indicated in the top figure.

illustrates that only four observations accounted for the AS component in the data. This is also supported by the parameter estimates since 0.2% (0.2 x 1469 = 4) of the data follow the bimodal distribution. To what extent this small proportion of antisymmetry is biologically relevant is a discussion beyond the scope of this paper.

**Performance of the EM-algorithm**

Estimating parameters under real FA resulted in a smooth likelihood surface without any local optima for the joint estimation of the two variances and of the two proportions (data not shown). For the rare antisymmetry dataset, the joint estimation of the mean and proportion of the AS component showed no local optima in the likelihood surface (data not shown). However, joint estimation of mean and variance of the AS component tended to exhibit a weak local optimum near a mean equal to zero and a

![Figure 5](image-url)  
**Figure 5** Likelihood surface for the joint estimation of the mean and variance of the antisymmetry component in a FA-AS two-component mixture of the rare antisymmetry data. All other variables of the model were kept fixed and equal to the simulation model parameters. The surface exhibited a weak local optimum indicated by the arrow.
characterize different types of asymmetry in the distribution of the signed asymmetry. In addition, it appears a valuable technique in describing between-individual heterogeneity in developmental stability and in estimating the hypothetical repeatability. Construction of weighed normal probability plots allows examining the underlying assumption of normality of the different components. Analyses of eight different datasets specifically demonstrated the following. (i) Model selection of different mixtures of real FA on the basis of the likelihood ratio test (with Monte Carlo simulation-based significance levels) leads to a number of components that pick up (nearly) all heterogeneity in DS. Construction of weighed normal probability plots further confirmed a close match between the observed and estimated distribution. It is therefore likely that the parameter estimates of the multicomponent mixture models provide a good approximate estimate of the between individual heterogeneity in FA and the presumed underlying DS. (ii) A lack of accurate measurement and the presence of many zero observations (apparent perfect symmetry) leads to convergence towards mixtures that contain a component with zero variance and thus infinite likelihood. This technical problem can be solved by limiting the minimal magnitude of the variances to a very small value (e.g. $10^{-8}$). However, this leads to an overestimation of the hypothetical repeatability as estimated from the mixture model parameter estimates. (iii) Mixtures appear to be a very sensitive tool to detect directional asymmetry and/or antisymmetry. In the RA and the BST datasets we showed significant antisymmetry for a small proportion of the total sample (10% and 0.2%, respectively). For the IMM dataset, we found a significant directional component that remained undetected by mixed model analysis (Van Dongen et al. 1999c and unpublished results). (iv) The underlying assumption of normality appeared to hold for the four empirical datasets analysed.

It is, however, important to note that our analyses are based on relatively large datasets (N ranging between 200 and 1500). Further research is needed to evaluate the performance of mixtures in the analysis of asymmetry. In particular, an estimation of the statistical power of the likelihood ratio test is indispensable to rate the usefulness of mixtures in the asymmetry context. Whereas one can intuitively assume that statistical power to detect different forms of asymmetry will sharply decrease with smaller sample sizes, the use of mixtures for model based estimation of the hypothetical repeatability might be very suited when sample sizes are relatively small. Recently, Van Dongen (1999) showed that $R$ cannot be estimated accurately with small sample sizes (< 80) and that it is biased towards smaller values. This bias resulted from a long tail of negative values of $R$, which are biologically meaningless. Yet, replacing the negative values by zero would lead to an upward bias. Using model-based estimation of $R$ will always render values larger than or equal to zero, and might be less subject to sampling variation. In addition, the likelihood ratio test based on Monte Carlo simulations provides an objective criterion to decide if the heterogeneity in DS (i.e. the positive value of $R$) is statistically significant. Further study is needed to evaluate the use of mixtures in this context.

ACKNOWLEDGEMENTS

SVD and LL are Postdoctoral Fellows with the Fund for Scientific Research Flanders (FWO Vlaanderen). We thank Erik Matthysen and two anonymous referees for their comments on an earlier version of the manuscript. The programs used to construct the mixtures evaluated in this paper can be obtained from SVD at HTTP://biowww.uia.ac.be/ulsvd/ongen/index.html.

REFERENCES


©1999 Blackwell Science Ltd/CNRS


**BIOSKETCH**

Stefan Van Dongen has studied insect–plant interactions in a fragmented landscape. Recently he became interested in the use of fluctuating asymmetry and an estimate of developmental stability, and how they relate to stress at both the individual and population level.

Editor, A. Sasaki

Manuscript received 5 July 1999
First decision made 13 August 1999
Manuscript accepted 6 September 1999