A statistical mixture model for estimating the proportion of unreduced pollen grains in perennial ryegrass (Lolium perenne L.) via the size of pollen grains

Jansen, R.C.; Nijs, A.P.M. den

Published in:
Euphytica

DOI:
10.1007/BF00023761

Euphytica **70**: 205–215, 1993.

A statistical mixture model for estimating the proportion of unreduced pollen grains in perennial ryegrass (*Lolium perenne* L.) via the size of pollen grains

R.C. Jansen[1] & A.P.M. Den Nijs[2]
[1] *Department of Population Biology &* [2] *Department of Arable and Forage Crops, Centre for Plant Breeding and Reproduction Research (CPRO-DLO), P.O. Box 16, 6700 AA Wageningen, The Netherlands*

Received 24 May 1993; accepted 27 August 1993

Key words: Box-Cox transformation, maximum likelihood, mixtures of normal distributions, perennial ryegrass, *Lolium perenne,* ploidy breeding, diploid (unreduced) pollen grains

Summary

The size of pollen grains is commonly used to indicate the ploidy level of pollen grains. In this paper observations of the diameter of pollen grains are evaluated from one diploid accession of perennial ryegrass (*Lolium perenne* L.), which was expected to produce diploid (unreduced) pollen grains in addition to haploid pollen grains. The considerable overlap of the diameter distributions of haploid and diploid pollen grains severely hampers the accurate estimation of the proportion of diploid pollen grains. To overcome this problem we develop in this paper a statistical normal mixture model and we describe a method to test for the production of diploid pollen grains from a diploid parent, and to estimate the proportion of diploid pollen grains.

Introduction

Crosses between plants of different ploidy may yield new and sexually originated polyploid plants and can therefore be of great value for the breeding of tetraploid perennial ryegrass (*Lolium perenne* L.). Perennial ryegrass plants of exceptional genotypes may produce unreduced diploid pollen grains in addition to haploid pollen grains. Plants of such genotypes are therefore of particular interest for breeding polyploids by interploidy crosses.

The size of pollen grains is related to chromosome number (Stanley & Linskens, 1974). Conventional methods to indicate the ploidy level are often based on measurements of the diameter or volume of the pollen grain. Crude estimates of proportions of diploid pollen grains are commonly provided by applying simple thresholding techniques (cf. Den Nijs & Stephenson, 1988; Eijlander, 1988; Van Tuyl et al., 1989), or by using sets of control data of haploid and diploid pollen grains (cf. Veronesi et al., 1988). In this paper normal mixture models (McLachlan & Basford, 1988) are described to assess the production of diploid pollen grains more accurately. It will be assumed that the observed distribution of the size of pollen grains derives from a mixture of underlying normal distributions, the components of the mixture representing haploid and diploid pollen grains. For each sample the normal mixture model contains a separate parameter for the proportion of diploid pollen grains. This makes it possible to study differences in the production of diploid pollen grains between samples. The model also contains parameters for the mean sizes of haploid and diploid pollen grains as well as parameters for shifts of mean sizes between samples. This makes it pos-

sible to correct the observed pollen sizes for variation between samples. A simple algorithm to find the maximum likelihood estimates has been described by Jansen (1993a). Calculations can be easily carried out in the statistical package GENSTAT (GENSTAT 5 Committee, 1987; Jansen, 1993b).

The method relies on the assumptions that the size of haploid pollen grains is normally distributed, and that the size of diploid pollen grains is normally distributed with the same variance. It is not a priori clear whether these assumptions hold for diameter (which was measured in our experiment) or for volume (which may be derived from the diameter). Our aim is to find the measure for the size of pollen grains for which these normality assumptions hold best.

Materials and methods

Data set

In preceding experiments large numbers of accessions were screened for the production of diploid pollen grains (Den Nijs & Stephenson, 1988). A plant showing an appreciable fraction of extra large pollen grains was cloned, vernalized and grown in climatized glasshouses. A control accession producing only haploid pollen grains was treated similarly. Pollen grain samples from several plants per clone were collected during the two weeks flowering period (in total 14 samples for the first accession and 16 samples for the control accession). Pollen grains were stained with carmine-acetic acid and diameters of well-stained pollen grains were measured by using a microscope with an ocular micrometer (1 unit = 2.6 μm). Table 1 shows the data of the accession that was expected to produce a mixture of haploid and diploid pollen grains. Table 2 shows the data of the control accession which was expected to produce only haploid pollen grains. Samples 1–5 of Table 1 were collected successively in time at 15° C, samples 6–10 at 20° C, and samples 11–14 at 25° C. Samples 1–4 of Table 2 were collected successively in time at 15° C, samples 5–11 at 20° C, and samples

Table 1. Observed frequencies of diameters and internal volumes of pollen grains in 14 samples of one accession which was expected to produce diploid in addition to haploid pollen grains. Diameters were measured, internal volumes were derived from the diameters (see text).

Sample	(a) Diameter (b) Internal volume/100															Total
(a)	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	
(b)	2	3	4	5	7	9	12	14	18	21	26	31	36	42	49	
1			3	24	58	45	21	7	6	2						166
2		1	7	19	39	41	32	14	9	2	2					166
3			4	13	34	58	45	26	12	3	1					196
4			1	5	5	9	18	29	29	14	10	8	4	4	3	139
5		2	1	1	3	12	27	41	32	16	11	3	2			151
6	1	3	32	107	73	15	8	7								246
7		2	6	19	70	63	45	11	11	2	1					230
8				17	37	32	44	30	26	15	10	7	5	1		224
9		2	13	29	42	42	37	51	23	7	2	1				249
10			7	17	42	54	78	48	8	2						256
11	1	6	6	23	48	52	36	14	12	8	2					208
12	2	8	26	46	56	38	24	17	2	1						220
13		3	13	32	58	32	37	18	6	2						201
14			1	1	4	16	34	24	17	16	12	9	4	2		140
Total	4	27	120	353	569	509	486	337	193	90	51	28	15	7	3	2792

(a) Diameter: 1 unit = 2.6 μm.
(b) Internal volume/100: 1 unit = $(2.6\ \mu m)^3/100$.

12–16 at 25° C. The flowering period was earliest at 25° C and latest at 15° C.

Statistical background

The data in Table 1 are from an accession which was expected to produce a mixture of haploid and diploid pollen grains. Let p_{hj} and p_{dj} denote the expected proportions of haploid and diploid pollen grains in sample j, respectively (j = 1. . .14). Similarly for each sample, we define parameters for the underlying normal distributions, i.e. for the mean sizes of haploid and diploid pollen grains and their variance. Let μ_{hj} and μ_{dj} denote the mean sizes of haploid and diploid pollen grains in sample j, respectively, and let σ_j^2 denote the variance in sample j. Observations are blocked into samples. It is assumed that

$$\mu_{hj} = \mu_h + \beta_j,$$
$$\mu_{dj} = \mu_d + \beta_j$$

in which β_j denotes the effect of sample j. Let y_{ij} denote the i-th outcome in sample j. Since observations are grouped, only a lower and an upper bound are observed, i.e. y_{ij} is replaced by (y_{lij}, y_{uij}) with $-\infty < y_{lij} < y_{uij} < \infty$. The probability that a haploid pollen grain has a size between y_{lij} and y_{uij} is

$$P_{hj}(y_{lij}, y_{uij}) = F\left(\frac{y_{uij} - \mu_{hj}}{\sigma_j}\right) - F\left(\frac{y_{lij} - \mu_{hj}}{\sigma_j}\right),$$

where F(•) is the probability density function of the standard normal distribution. Similarly, the probability that a diploid pollen grain has a size between y_{lij} and y_{uij} is

$$P_{dj}(y_{lij}, y_{uij}) = F\left(\frac{y_{uij} - \mu_{dj}}{\sigma_j}\right) - F\left(\frac{y_{lij} - \mu_{dj}}{\sigma_j}\right).$$

The corresponding mixture probability $P(y_{lij}, y_{uij})$ that a pollen grain with unknown ploidy has a size between y_{lij} and y_{uij} is

Table 2. Observed frequencies of diameters and internal volumes of pollen grains in 16 samples of one accession which produces only haploid pollen grains. Diameters were measured, internal volumes were derived from the diameters (see text).

Sample	(a) Diameter (b) Internal volume/100										Total
(a)	8	9	10	11	12	13	14	15	16	17	
(b)	1	2	3	4	5	7	9	12	14	18	
1			2	38	79	100	31	7			257
2			2	16	55	92	31	6	4		206
3				1	26	83	90	40	2		242
4			4	38	95	92	18	4			265
5			22	66	123	92	18	4			325
6		1	3	25	83	117	38	6			273
7		1	9	34	75	65	13	6			203
8			21	49	51	25	9	7			162
9				2	33	85	60	15			195
10			1	16	64	87	42	4	1		215
11			2	24	102	64	5				197
12			4	13	88	64	23	5			197
13	1	15	54	85	88	55	16	2			316
14	1	7	33	82	89	48	16	4	1		281
15		1	7	26	82	98	64	36	6	1	321
16		1	6	28	71	57	27	9	1		200
Total	2	26	170	543	1204	1224	515	154	16	1	3855

(a) Diameter: 1 unit = 2.6 μm.
(b) Internal volume/100: 1 unit = $(2.6\ \mu m)^3/100$.

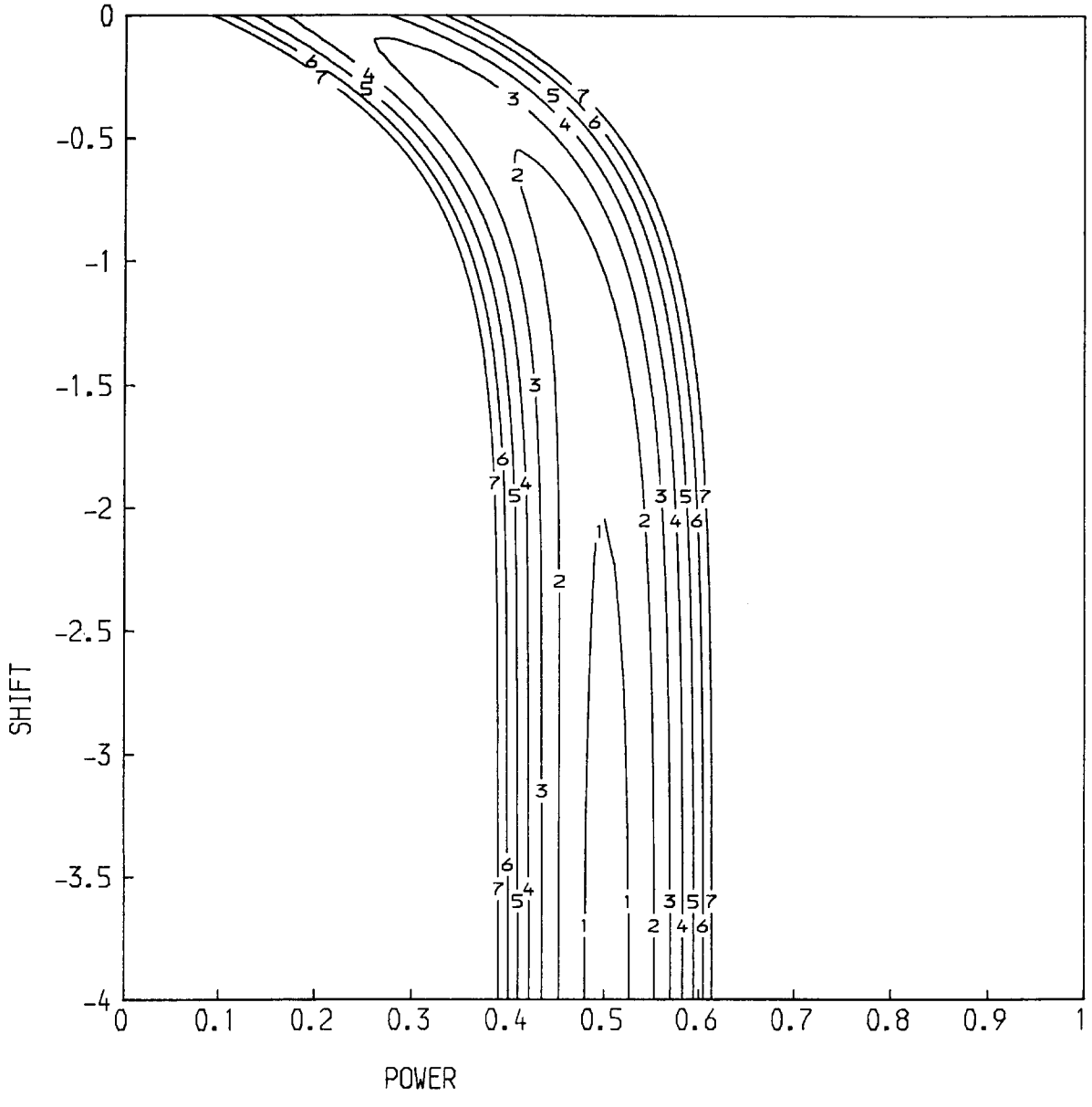

Fig. 1. Contour plot of the log-likelihood profile of the power (λ_2) and shift (λ_3) parameters. All values of (λ_2,λ_3) on the same contour have the same likelihood. The likelihood is at its maximum at ($\hat{\lambda}_2$,$\hat{\lambda}_3$) = ($^1/_2$, −4). The seven contours show the shape of the log-likelihood surface close to its maximum, the other contours came too close to each other for plotting. Differences between contours correspond to differences of one unit in log-likelihood, so that the approximate 95 per cent confidence region is enclosed by the third contour. It shows that the optimum transformation will be to subtract the sample minimum from internal volume and then take the square root of the shifted internal volume.

$$P(y_{lij},y_{uij}) = p_{hj}P_{hj}(y_{lij},y_{uij}) + p_{dj}P_{dj}(y_{lij},y_{uij}).$$

A simple method for the maximum likelihood estimation of the parameters has been described by Jansen (1993a). Calculations can be easily carried out in GENSTAT (GENSTAT 5 Committee, 1987; Jansen, 1993b).

To assess the production of diploid pollen grains the likelihood of the normal mixture model is compared to the likelihood of the normal (no-mixture) model. Unfortunately, distributional properties of

the difference in the log-likelihoods are not known when testing the number of components in a mixture. McLachlan & Basford (1988) described a procedure to simulate its distribution. According to this procedure, the production of diploid pollen is significant at the level of 100 (1-α) per cent if the observed difference in log-likelihoods exceeds the upper 100 (1-α) per cent in the simulations. For instance, the confidence level of the test is 90 per cent when it is based on nine simulations and 95 per cent when it is based on 19 simulations.

Apart from the above statistical modelling, the choice of an appropriate measure for the size of pollen grains should be made. The aim is to find a measure for which the normality assumption holds best. Three measures will be considered first: diameter, internal volume and external volume. Thereafter, we discuss data transformations to obtain a possibly better fitting model.

Since fertile pollen grains of perennial ryegrass are almost spherical, external volume v_{ex} can be derived from the diameter d by

$$v_{ex} = \frac{4}{3}\pi\left(\frac{d}{2}\right)^3.$$

Pollen grains of perennial ryegrass have uniformly thick cell walls (Hayward & Manthriratna, 1972). The thickness of the cell wall is close to unity on the scale used, and approximately independent of the size of pollen grains. Therefore, the internal volume v_{in} is given by

$$v_{in} = \frac{4}{3}\pi\left(\frac{d}{2} - 1\right)^3.$$

Since external volume and internal volume are closely related, we restrict ourselves temporarily to diameter and internal volume of the pollen grains.

We used the procedure of Box & Cox (1964) to find a transformation of internal volume, such that the normality assumption holds best on the transformed scale. The Box-Cox transformation (with shift parameter λ_1 and power parameter λ_2) of internal volume equals $(v_{in} - \lambda_1)^{\lambda_2}$ for $\lambda_2 > 0$, and $\log(v_{in} - \lambda_1)$ for $\lambda_2 = 0$. Both internal volume and diameter are embedded in this transformation framework, since for $\lambda_1 = 0$ and $\lambda_2 = 1$ the transformation equals identity, while for $\lambda_1 = 0$ and $\lambda_2 = {}^1/_3$ the transformation gives (a constant multiple of) the diameter. In any sample the shift λ_1 should not exceed the sample minimum $v_{min,j}$ of the internal volume. Therefore we introduce λ_3 ($-\infty < \lambda_3 \leq 0$) and use

$$\lambda_{1,j} = v_{min,j}(1 - 10^{\lambda_3})$$

as the shift in sample j (Atkinson, 1985), so that for each sample a separate shift $\lambda_{1,j}$ is used, which can take values between 0 ($\lambda_3 \to -\infty$) and the sample minimum $v_{min,j}$ ($\lambda_3 = 0$). For reasons of computer accuracy only the values of λ_3 with $-4 \leq \lambda_3 \leq 0$ are considered.

For fixed (λ_2,λ_3) we can find the log-likelihood $L_{max}(\lambda_2,\lambda_3)$ maximized over the parameters of the normal distribution. To compare the log-likelihood for various values of (λ_2,λ_3), the maximum likelihood $L_{max}(\lambda_2,\lambda_3)$ can be plotted over a range of plausible values. An approximate 95 per cent confidence region for (λ_1,λ_2) is found from those values for which

$$L_{max}(\hat{\lambda}_2,\hat{\lambda}_3) - L_{max}(\lambda_2,\lambda_3) \leq +\frac{1}{2}\chi^2_{2,0.95} \approx 3,$$

in which $(\hat{\lambda}_1,\hat{\lambda}_2)$ are the values of (λ_1,λ_2) where the likelihood is at its maximum (Box & Cox, 1964).

To show the effect of the change from external volume to internal volume by the correction for cell wall thickness, a plot of the likelihood as a function of the correction constant can be drawn using the optimum transformation.

Results

Statistical analysis

Firstly, the results of the analysis of the control accession (Table 2) are presented. The main goals of this analysis were the choice of an appropriate measure for the size of pollen grains and the choice of the data transformation, such that the normality assumption holds best on the transformed scale. Figure 1 shows a contour plot of the likelihood $L_{max}(\lambda_2,\lambda_3)$. Only seven contours were plotted, the

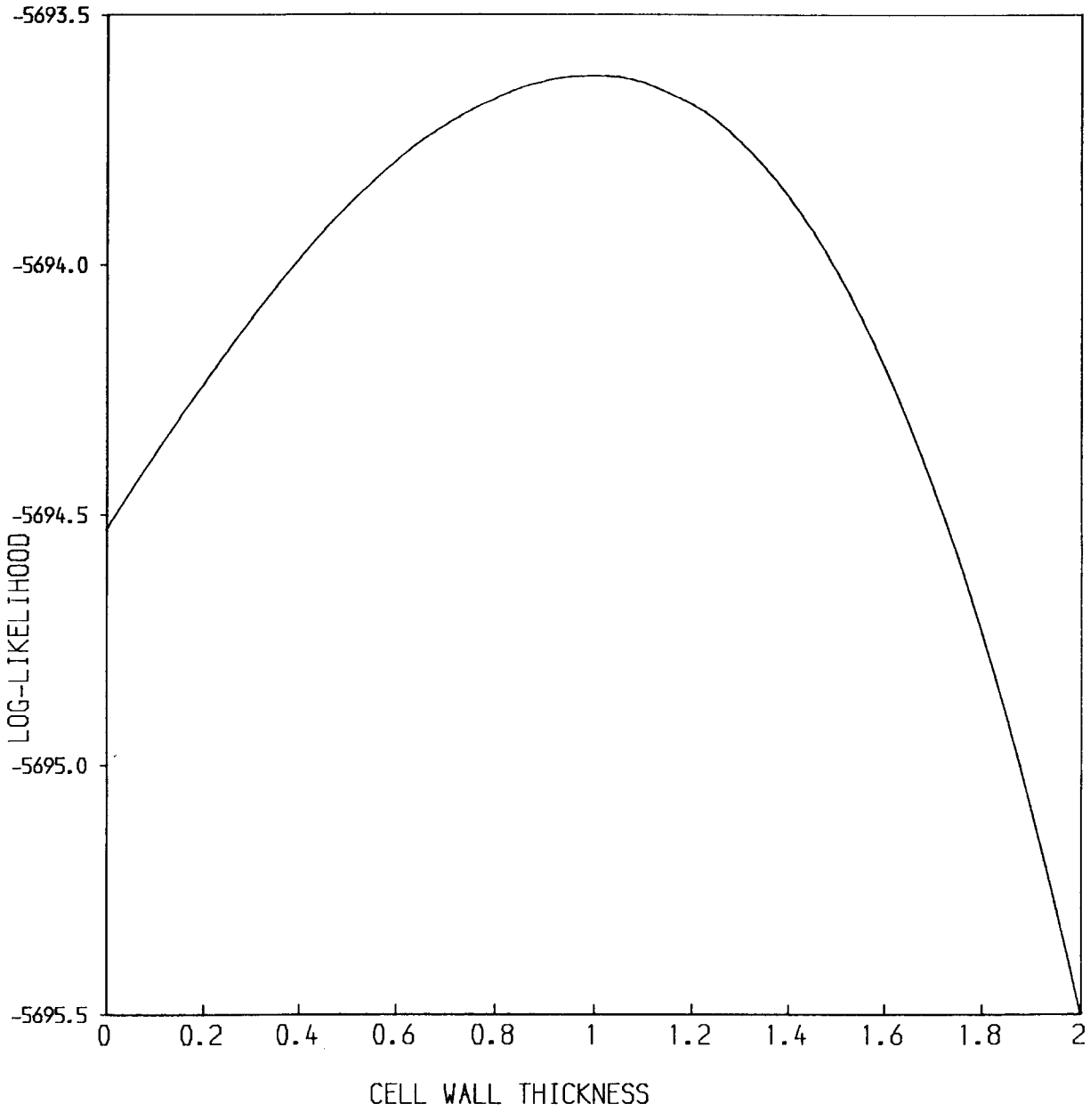

Fig. 2. Log-likelihood profile of the cell wall thickness. The likelihood is at its maximum at about unity. It shows that the normality assumption fits slightly better to transformed internal volumes than to transformed external volumes (internal volume is the volume of the pollen grain minus the cell wall; the cell wall thickness is known be constant and about equal to unity).

other contours were too close together for plotting. The likelihood is at its maximum at $(\hat{\lambda}_2,\hat{\lambda}_3) = (^1/_2, -4)$. Thus, the optimum transformation will be to subtract the sample minimum from the internal volume and then take the square root of the shifted internal volume. The approximate 95 per cent confidence region is enclosed by the third contour. The confidence region firmly excludes $(\lambda_2,\lambda_3) = (1,0)$. This implies that the fit of the normal model to the internal volume can be significantly improved by prior data transformation. The confidence region just excludes $(\lambda_2,\lambda_3) = (^1/_3,0)$. Thus, the fit of the normal model to the diameter can also be significantly improved by prior data transformation. Figure 2

shows the log-likelihood as a function of the correction for thickness of the cell wall by using the optimum transformation. As mentioned above, the thickness of the cell wall is close to unity. The likelihood is also maximal at about unity. This demonstrates that the normality assumption fits slightly better to the transformed internal volume than to the transformed external volume.

We now turn to the analysis of the data of the accession that was expected to produce diploid in addition to haploid pollen grains (Table 1). To obtain as normally distributed data as possible, we used the shifted square root transformation of the internal volume which proved best in the analysis of the data of the control accession. In order to assess the production of diploid pollen grains, the distribution of the difference between the log-likelihood for the model with production of diploid pollen grains and that for the model without production of diploid pollen grains was simulated under the null hypothesis of no production of diploid pollen grains. The observed difference (117.6) highly exceeded the largest simulated difference (13.3). The test was based on 19 simulations and leads us to reject the null hypothesis of no production of diploid pollen grains at a confidence level of 95 per cent. A second test was carried out to assess the production of still another pollen grain type. The observed difference (6.5) slightly exceeded the largest simulated difference (5.8). This test was based on only 9 simulations, since it involved much computation. The test leads to rejection of the null hypothesis of no production of a third pollen grain type at the confidence level of 90 per cent. Estimated proportions and mean internal volumes of haploid, diploid and (possibly) tetraploid pollen grains are shown in Tables 3 and 4. Transformation very much improved the fit of the normal mixture model to the internal volume. The normal mixture model fits slightly better to the transformed internal volume than to the diameter. Especially the estimates of the proportion of diploid pollen grains based on internal volume without transformation highly differed from the estimates presented in Table 3. This makes clear that one should carefully choose an appropriate measure for the size of pollen grains.

Table 3. Proportions of diploid and (possibly) tetraploid pollen grains as estimated with the mixture model assuming three types of pollen grains. The proportion of haploid pollen grains is equal to 1 minus the proportions of diploid and tetraploid pollen grains.

Sample	Diploid	Tetraploid
1	0.11 ± 0.04	0
2	0.10 ± 0.05	0
3	0.02 ± 0.04	0
4	0.70 ± 0.11	0.16 ± 0.05
5	0.90 ± 0.05	0.04 ± 0.03
6	0.07 ± 0.02	0
7	0.08 ± 0.03	0
8	0.47 ± 0.19	0.13 ± 0.05
9	0.53 ± 0.05	0.01 ± 0.01
10	0.87 ± 0.04	0
11	0.14 ± 0.05	0
12	0.01 ± 0.03	0
13	0	0
14	0.70 ± 0.11	0.28 ± 0.05

Biological implications

The data in Table 1 are from an accession which was expected to produce diploid pollen grains in addition to haploid pollen grains. This expectation is also supported by a comparison of Table 1 and Table 2: the data in Table 1 clearly cover a wider range of values than those in Table 2. Note that the internal

Table 4. Mean internal volumes of haploid, diploid and (possibly) tetraploid pollen grains as estimated with the mixture model assuming three types of pollen grains (± standard error of mean).

Sample	Haploid	Diploid	Tetraploid
1	785 ± 3	1506 ± 9	–
2	865 ± 5	1689 ± 11	–
3	1002 ± 4	1843 ± 10	–
4	878 ± 6	1653 ± 12	3235 ± 16
5	782 ± 11	1561 ± 18	3148 ± 26
6	581 ± 2	1288 ± 8	–
7	850 ± 5	1666 ± 10	–
8	792 ± 1	1432 ± 7	2831 ± 7
9	678 ± 2	1397 ± 8	2903 ± 9
10	553 ± 8	1107 ± 8	–
11	830 ± 19	1680 ± 30	–
12	729 ± 1	1522 ± 8	–
13	842 ± 5	–	–
14	666 ± 3	1308 ± 8	2711 ± 10

1 unit = $(2.6\ \mu m)^3$.

volume of the largest pollen grains in Table 1 is almost 25 times higher than the internal volume of the smallest pollen grains, i.e. the internal volume of pollen grains varies enormously. The large variation between samples is also a striking feature of the data in Table 1. Some samples cover a much wider range than others. For instance, the 166 observations in sample 1 of Table 1 take values from 11 to 18 on the original measuring-scale, whereas the 139 observations in sample 4 take values from 11 to 23. Furthermore, the modes of the frequency distributions of the individual samples are clearly shifted. For example, the mode of sample 5 in Table 1 equals 16 on the original measuring-scale, whereas the mode of sample 6 equals only 12. The corresponding internal volume of the pollen grain in sample 5 is more than twice the corresponding internal volume in sample 6. This large variation could hardly be explained by regular variation in the mean size of haploid and diploid pollen grains. In contrast with this, the sample modes in Table 2 are either 12 or 13.

The statistical evaluation of the data in Table 1 clearly demonstrates the production of at least two types of pollen grain. The production of a third type of pollen grain is also indicated (at least in three of the 14 samples), the corresponding test statistic being significant (at least) at the 90 per cent confidence level. We present and discuss below the results of the model assuming a mixture of three types of pollen grain. The estimates of the three corresponding proportions and the mean internal volumes are presented in Tables 3 and 4. Figure 3 shows histograms for some samples with the fitted component distributions and the mixture distributions superimposed (the fitted distributions were backtransformed to the original scale). The estimated proportions of diploid pollen grains vary between 0 (sample 13) and 0.90 (sample 5, Table 3). This indicates that the production of diploid pollen grains is strongly influenced by environmental factors which may vary from one sample to another. However, no clear effect of temperature on the production of unreduced pollen grains is found. At each temperature samples were collected successively in time (samples 1–5 were collected at 15° C, samples 6–10 at 20° C and samples 11–14 at 25° C). The samples obtained at the end of the flowering period show the larger proportions of unreduced pollen grains (Table 3). This suggests a relation between plant age and production of unreduced pollen grains.

Discussion

In this paper we present a statistical analysis for the assessment of ploidy via the size of pollen grains. Although the statistical methods were developed for the analysis of pollen grain data from perennial ryegrass, the methods and their results probably have a wider applicability.

Our statistical analysis concerns the data obtained from two accessions of perennial ryegrass and leads to the following remarks. First, there is a large variation in the proportion of diploid pollen grains between samples. An effect of plant age on the production of unreduced pollen grains is suggested. Screening for accessions with a good production of diploid pollen grains is often based on a single sample, but our results show that conclusions based on a single sample may be quite misleading.

Second, our analysis indicates the production of haploid, diploid and possibly tetraploid pollen grains. However, further cytological evidence is required to definitely assess the production of tetraploid pollen grains. In any case, our statistical method makes it possible to test for the presence of two or more types in the pollen grain mixture and to estimate their proportions.

Third, the analysis of diameters and that of volumes lead to different results. A different goodness-of-fit of the normal (mixture) models to the data is found; in addition rather different estimates of the proportions of diploid pollen grains in the samples are found. Therefore, one should carefully choose an appropriate measure for the size of pollen grains. It was demonstrated that the normal (mixture) models fitted best to the transformed internal volume. This measure is obtained by subtracting the sample minimum from the internal volume and then taking the square root of the shifted internal volume. In the case of non-spherical pollen grains the diameter is unlikely to be a good measure for the size of pollen grains. Pollen grains of for in-

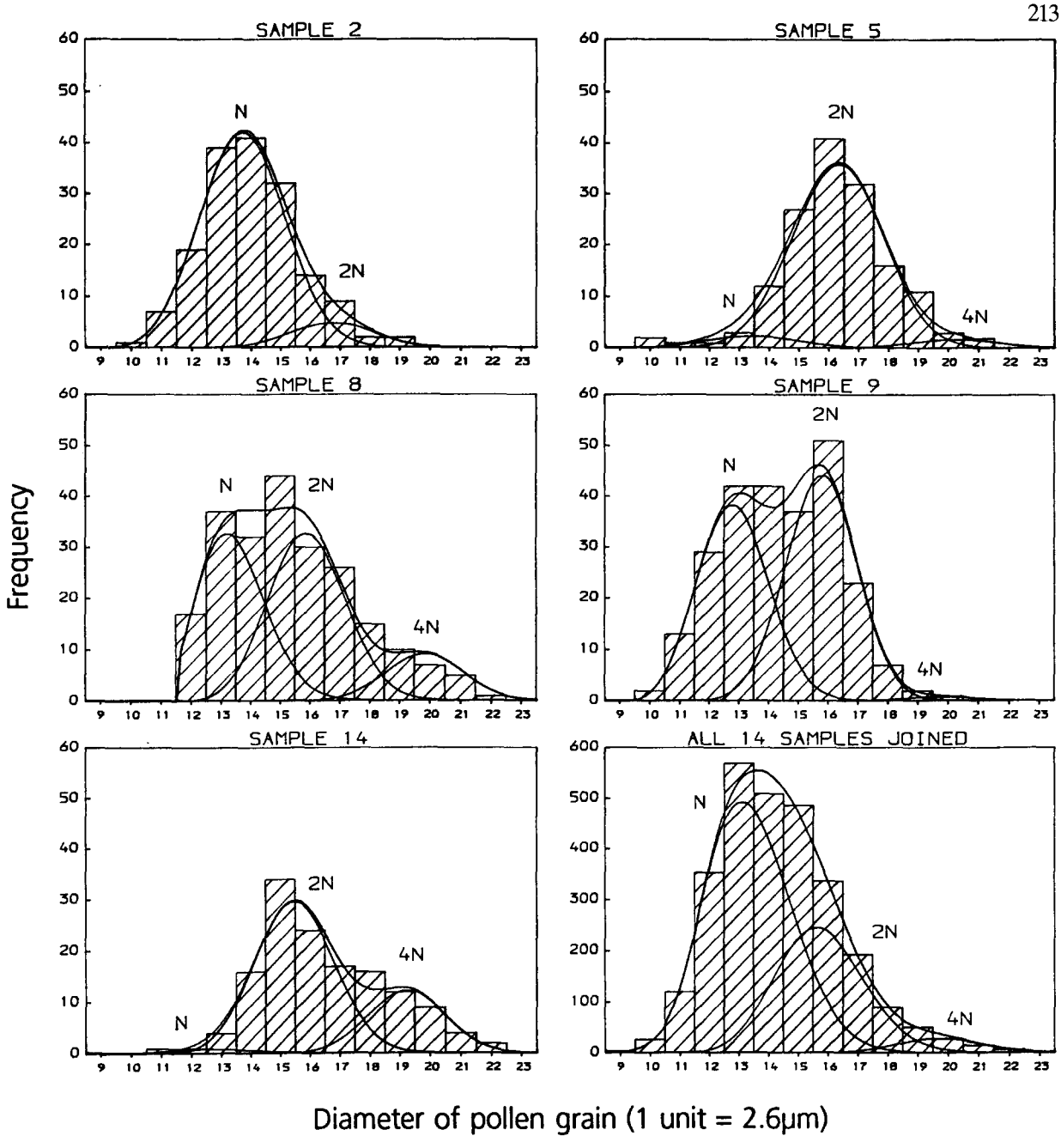

Fig. 3. Some histograms of the sample data and graphs of the fitted diameter distributions for haploid (N), diploid (2N) and possibly tetraploid (4N) pollen grains and a graph of their fitted mixture distributions (see also Tables 1, 3 and 4). Sample 2 contains a large proportion of haploid pollen grains. Sample 5 contains a large proportion of diploid pollen grains. Sample 8 is clearly a mixture of haploid, diploid and (possibly) tetraploid pollen grains. Sample 9 is a mixture of predominantly haploid and diploid pollen grains. Sample 14 is a mixture of predominantly diploid and (possibly) tetraploid pollen grains.

stance lily, alfalfa and radish are non-spherical, so that in these species volume measurements are most natural. But also in these cases the normality assumption will probably hold best for the shifted square root of the volume.

To our knowledge there is no proof of any mathematical relation between the ploidy and the size of pollen grains, though it is sometimes supposed that the volume of diploid pollen grains is twice the volume of haploid ones (cf. Jacobsen, 1978; Sala et al.,

1989). In our data the mean internal pollen volumes of the three types of pollen grain are remarkably close to the ratio 1:2:4 (Table 4). We suggest that the third type of pollen grain, of which the internal volume is four times the internal volume of haploid pollen grains, represents tetraploid pollen grains. The production of tetraploid pollen grains in perennial ryegrass has never been reported before, but is not unlikely, since Van Wagenvoort et al. (1992) observed diploid, tetraploid and even octoploid cells in meiosis I in a cytological study of a different accession of perennial ryegrass. Omara (1976) observed the production of pollen mother cells with 2 up to 58 chromosomes in perennial ryegrass. Therefore, mixtures of even more than three types of pollen grain may possibly occur. However, reliable detection of other types of pollen grain by our method would probably require much larger data sets.

We also had the data of a control accession which was known to produce only haploid pollen grains (Table 2). The analysis of these data provided important information on the size distribution of haploid pollen grains and on data transformation. Such information could hardly be obtained from the mixture data (Table 1), since the component distributions are poorly separated. For the analysis of Table 1 we used the results of the analysis of Table 2. However, no precise information is available on the size distribution of diploid pollen grains (nor on that of the possibly tetraploid pollen grains). We assumed, for example, that the three types of pollen grain have equal variance (within the same sample and after data transformation). We also assumed that the size differences between the three types of pollen grain are constant over the 14 samples (again, on the transformed scale). These assumptions cannot be checked properly but do seem to be reasonable. At least, they make a robust parameter estimation possible by linking the component distributions; the number of observations per sample as well as the number of diameter classes per sample are too small to fit more extended models. Besides, without these assumptions the number of parameters per sample would readily exceed the multinomial degrees of freedom per sample (i.e. the number of diameter classes minus 1).

Flow cytometry has been proposed as an alternative method for estimating the proportion of diploid pollen grains in pollen grain samples (van Tuyl et al., 1989). This method produces DNA histograms of individually released nuclei from the pollen grain mixture. However, the distributions of both types of nuclei may also overlap and similar problems arise as dealt with in the present paper (Jansen 1993a). Furthermore, analysis of the DNA histograms is often complicated by noise due to cell debris.

Acknowledgements

The authors would like to thank G.J.O. Jansen and W.E. van de Weg for critical reading of the manuscript.

References

Atkinson, A.C., 1985. Plots, transformations and regression. Oxford, Clarendon Press.

Box, G.E.P. & D.R. Cox, 1964. An analysis of transformations (with discussion). J.R. Statist. Soc. B, 26: 211–246.

Den Nijs, A.P.M. & A.G. Stephenson, 1988. Potential of unreduced pollen for breeding tetraploid perennial ryegrass. In: M. Cresti, P. Gori, E. Pacini (Eds). Sexual reproduction in higher plants. Proceedings of the tenth international symposium on the sexual reproduction in higher plants. Springer Verlag, Berlin.

Eijlander, R., 1988. Manipulation of the 2n gametes frequencies in *Solanum* pollen. Euphytica S: 45–50.

GENSTAT 5 Committee, 1987. GENSTAT 5 reference manual. Clarendon Press, Oxford.

Hayward, M.D. & M.A.P.P. Manthriratna, 1972. Pollen development and variation in genus *Lolium*. Z. Pflanzenzüchtg. 67: 131–144. Verlag Paul Parey, Berlin.

Jacobsen, E., 1978. Die Chromosomverdopplung in der Züchtung dihaploider Kartoffeln. These. Rheinischen Friedrich-Wilhelms-Universität, Bonn.

Jansen, R.C., 1993a. Maximum likelihood in a generalized linear finite mixture model by using the EM algorithm. Biometrics 49: 227–231.

Jansen, R.C., 1993b. Maximum likelihood in a finite mixture model by exploiting the GLM facilities of GENSTAT. Genstat Newsletter (in press).

McLachlan, G.J. & K.E. Basford, 1988. Mixture models. Inference and application to clustering. Marcel Dekker, New York.

Omara, M.K., 1976. Cytomixis in *Lolium perenne*. Chromosoma 55: 267–271.

Sala, C.A., E.L. Camadro, M.T. Salaberry & A.O. Mendiburu,

1989. Cytological mechanism of 2n pollen formation and unilateral sexual polyploidization in *Lolium*. Euphytica 43: 1–6.

Stanley, R.G. & H.F. Linskens, 1985. Pollen. Biology, biochemistry, management. Springer Verlag, Berlin.

Van Tuyl, J.M., J.N. De Vries, R.J. Bino & T.A.M. Kwakkenbos, 1989. Identification of 2n pollen producing interspecific hybrids of *Lilium* using flow cytometry. Cytologia 54: 737–745.

Van Wagenvoort, M. & A.P.M. Den Nijs, 1992. Implications of 2n pollen for breeding tetraploid perennial ryegrass. In: A. Mariani, S. Tavoletti (Eds). Proceedings of European Workshop "Gametes with somatic chromosome number in the evolution and breeding of polysomic species. Achievements and perspectives". Porziuncola, Assisi (Italy).

Veronesi, F., A. Mariani & S. Tavoletti, 1988. Screening for 2n gamete producers in diploid species of genus *Medicago*. Genet. Agr. 42: 187–200.