Mathl Comput. Modelling, Vol. 12, No. 9, pp. 1145–1154, 1989
Printed in Great Britain.

0895-7177/89 $3.00 + 0.00

ANALYSIS OF FLOW CYTOMETRIC ANEUPLOID DNA HISTOGRAMS: VALIDATION OF AN AUTOMATIC PROCEDURE AGAINST *AD HOC* EXPERIMENTAL DATA

C. BRUNI,[1] F. LAMPARIELLO,[2] S. LUCIDI[2] and C. MICHELI[3]

[1]Dipartimento di Informatica e Sistemistica, Università di Roma "La Sapienza", Via Eudossiana 18, 00184 Roma, Italia

[2]Istituto di Analisi dei Sistemi ed Informatica del CNR, Viale Manzoni 30, 00185 Roma, Italia

[3]Divisione di Fisica e Scienze Biomediche, ENEA-Casaccia, S.P. Anguillarese 301, 00060 Roma, Italia

(Received January 1989; accepted for publication February 1989)

Communicated by E. Y. Rodin

Abstract—In this paper we present an improved version of a method for the automatic analysis of flow cytometric DNA histograms from samples containing a mixture of two cell populations. The procedure is tested against two sets of *ad hoc* experimental data, obtained by mixing cultures of cell lines in different known proportions. The potentialities of the method are enlightened and discussed with regard to its capability of recovering the population percentages, the DNA index and the G_0/G_1, S, $G_2 + M$ phase fractions of each population. On the basis of the obtained results, the procedure appears to be a promising tool in the flow cytometric data analysis and, in particular, in problems of diagnosis and prognosis of tumor diseases.

1. INTRODUCTION

In the recent literature, many methods and algorithms have been proposed for the analysis of DNA flow cytometric data in the case of a single cell population (see, for example, Ref. [1] and references cited therein). On the contrary, only few results are available concerning the analysis of DNA histograms from samples containing two cell populations. This problem appears to be of great relevance in the study of human malignancies both for diagnosis and prognosis purposes. A practical graphical method has been proposed in Ref. [2]; a computer-based system can be found in Ref. [3]. Moreover, a cell analysis program, developed by P. S. Rabinovitch (Multicycle, Phoenix Flow Systems, San Diego, Calif.), is available for fitting complex tumor DNA histograms.

Recently, a mathematical model was proposed [4, 5] for describing the DNA content distribution of a sample containing a mixture of two cell populations and a procedure was presented for the fully automatic estimation of the unknown model parameters. The procedure was tested against both simulated data and a few real data from clinical material. The results there obtained were satisfactory, but further validations appeared to be appropriate, in particular with respect to the experimental data.

The aim of this paper is to perform a further analysis along this line. More specifically, we have carried out some *ad hoc* experiments by mixing cultures of two cell lines in different known proportions. Subsequently, we have applied to the corresponding flow cytometric histograms a new version of the previous method, improved both in the model and in the estimation procedure.

In the following sections, first we briefly describe the method and the modifications introduced in order to improve its efficiency from a computational point of view and the reliability of the obtained estimates. Thereafter we account for the experimental material and the measurement method, and we report the numerical results obtained by the revised procedure. Finally, we discuss and interpret these results in order to enlighten the potentialities of the method in flow cytometric data analysis.

2. MATHEMATICAL MODEL AND ESTIMATION PROCEDURE

In this section we briefly account for the method proposed in Refs [4, 5] and, more specifically, we describe both the changes performed on the models of DNA and fluorescence distributions and the improvements introduced in the parameter estimation procedure.

In the following, we denote by the indices $i = 1, 2$ the parameters of the lower and the higher DNA content populations, respectively.

The DNA content distribution $g^{(i)}(y)$ of each population consists of two concentrated masses $(G_1^{(i)}, G_2^{(i)})$, corresponding to the G_0/G_1 and $G_2 + M$ phases, located at $y_1^{(i)}$ and $2y_1^{(i)}$ values of DNA content respectively, and of a mass $(S^{(i)})$, corresponding to the S-phase, which is assumed to be piecewise constant distributed in the interval $[y_1^{(i)}, 2y_1^{(i)}]$.

In order to avoid the analytical complications deriving from a constraint on the G_0/G_1 peak locations, the condition $y_1^{(2)} - y_1^{(1)} > \epsilon > 0$, introduced in Refs [4, 5], has been removed. This allows the two populations to fully overlap, and in this case their distinction may be accomplished, in principle, only if they exhibit a sufficiently different behavior with respect to the flow cytometric analysis. However, this situation will not be considered in this study; the effective capability of the procedure to recover two fully overlapped populations out of their mixture requires further investigation.

A second change introduced here on the DNA content distribution model is related to the number of equal subintervals in $[y_1^{(i)}, 2y_1^{(i)}]$ with constant S-phase density: in particular these numbers, $n^{(i)}$, $i = 1, 2$, are allowed to be different for the two populations in order to make the model more flexible. They are chosen by the operator, in the set $\{1, 2, 4\}$, on the basis of the shape of the measured histogram.

As far as the fluorescence distribution model is concerned, we describe both the staining and instrumental dispersions by a Gaussian density function. More specifically, denoting by x the fluorescence intensity, we assume that the fluorescence density $p^{(i)}(x \mid y)$, for a cell of the ith population with DNA content y, is a Gaussian with mean y and variance

$$v^{(i)}(y) = \gamma^{(i)} y + \delta^2 y^2. \tag{1}$$

Motivations about this choice are given in Refs [4, 5].

In fluorescence histograms it is frequently observed that the location of the $G_2 + M$ peak is not exactly at twice that of the G_0/G_1 peak, as we have assumed in the DNA distribution model. This effect is probably due to an instrumental bias which causes the fluorescence intensity x not to be exactly proportional to the DNA content y, and it has been accounted for in our model by a suitable shift ρ on the fluorescence intensity axis. This corresponds to considering for $p^{(i)}(x \mid y)$ the Gaussian with mean $y + \rho$ and the same variance (1). Differently from Refs [4, 5] where this shift was manually performed on the measured histogram, the parameter ρ is now a continuous variable to be automatically estimated together with the other unknown parameters.

In this paper we do not consider any model for the debris background because the aim is here to validate our procedure against *ad hoc* experimental data which lack a significant debris component.

As is well-known, the *fluorescence distribution density* $f(x; \boldsymbol{\theta})$ can be mathematically expressed as follows:

$$f(x; \boldsymbol{\theta}) = \sum_{i=1}^{2} \int_{y_1^{(i)}}^{2y_1^{(i)}} p^{(i)}(x \mid y) g^{(i)}(y)\, dy, \tag{2}$$

where, on the basis of the previous assumptions, the vector $\boldsymbol{\theta}$ of unknown parameters to be estimated has the following $n^{(1)} + n^{(2)} + 10$ components:

$$\boldsymbol{\theta} = [y_1^{(1)} \gamma^{(1)} G_1^{(1)} A_1^{(1)} \ldots A_{n^{(1)}}^{(1)} G_2^{(1)} \mid y_1^{(2)} \gamma^{(2)} G_1^{(2)} A_1^{(2)} \ldots A_{n^{(2)}}^{(2)} G_2^{(2)} \mid \delta\, \rho]^{\mathrm{T}}. \tag{3}$$

In vector (3), $A_k^{(i)}$, $k = 1, \ldots, n^{(i)}$, denote the S-phase fractions on each subinterval of $[y_1^{(i)}, 2y_1^{(i)}]$ with uniform distribution, so that

$$S^{(i)} = \sum_{k=1}^{n^{(i)}} A_k^{(i)}.$$

Since the experimental histogram gives the number of cells in each of the channels in which the fluorescence axis is subdivided, in order to relate it to $f(x;\boldsymbol{\theta})$, it is necessary to discretize this latter continuous function, by integrating it over each channel. We denote by $[j_i, j_f]$ the prefixed channel interval over which the analysis will be effectively performed. This interval, which is obviously finite, has to be chosen so that tails of the fluorescence histogram due to effects not considered in the model (background and cell aggregations) are excluded. Therefore, a normalization to unit area of the same function $f(x;\boldsymbol{\theta})$ on the prefixed interval is also required. Thus, denoting by $H_j(\boldsymbol{\theta})$ the discretized and normalized fluorescence density:

$$H_j(\boldsymbol{\theta}) = \frac{\int_{j-1}^{j} f(x;\boldsymbol{\theta})\,dx}{\sum_{k=j_i}^{j_f} \int_{k-1}^{k} f(x;\boldsymbol{\theta})\,dx}, \qquad j = j_i, \ldots, j_f, \tag{4}$$

the observation equation is

$$\frac{N_j}{N} = H_j(\boldsymbol{\theta}) + R_j, \qquad j = j_i, \ldots, j_f, \tag{5}$$

where N_j is the cell number registered in the jth channel,

$$N = \sum_{j=j_i}^{j_f} N_j,$$

and R_j is an error term due to the finiteness of the cell sample.

In order to estimate the unknown parameter vector $\boldsymbol{\theta}$, we adopt here the *maximum likelihood approach* instead of the minimum χ^2 error index, which was used in Ref. [4]. This choice is justified by the well-known statistical properties of the achieved estimation [6].

The maximum likelihood estimation for $\boldsymbol{\theta}$ is the value which minimizes the function

$$L(\boldsymbol{\theta}) = -\log \prod_{j=j_i}^{j_f} H_j(\boldsymbol{\theta})^{N_j} = -\sum_{j=j_i}^{j_f} N_j \log H_j(\boldsymbol{\theta}). \tag{6}$$

Of course, it is necessary to take suitable constraints on $\boldsymbol{\theta}$ into account, which are a nonnegativity constraint of obvious physical meaning for all the components with the exception of ρ, and the normalization constraint:

$$G_1^{(1)} + \sum_{k=1}^{n^{(1)}} A_k^{(1)} + G_2^{(1)} + G_1^{(2)} + \sum_{k=1}^{n^{(2)}} A_k^{(2)} + G_2^{(2)} = 1. \tag{7}$$

By performing suitable changes of variables [5], the previous constrained minimization problem has been transformed into an unconstrained one.

3. MINIMIZATION ALGORITHM

The unconstrained minimization of L is performed by using essentially the same technique as employed in Refs [4, 5]. More specifically, an algorithm is designed which implements a version of Newton's method; at each step, the search direction in the parameter space is determined by using the analytical expressions for the first and the second order derivatives of L. As regards the stabilization technique, a nonmonotone line search strategy is adopted, which allows considerable computational savings both in the number of line searches and in the number of function evaluations. Further improvements in this direction have been achieved, by introducing in the minimization algorithm the following modifications.

In order to guarantee a satisfactory approximation in the S-phase reconstruction, the convolution integrals which appear in equation (2) are numerically computed, as in Refs [4, 5], by using a second order formula (Simpson's integration rule). On the contrary, we have verified that a relevant reduction of the computational effort, without sensible loss of precision, can be obtained by performing a first order interpolation to compute the integrals in equation (4).

By applying these integration procedures, the function $H_j(\boldsymbol{\theta})$ turns out to be expressed in terms of linear combinations of Gaussian density functions. The same property holds true for the first and the second order derivatives of $H_j(\boldsymbol{\theta})$ with respect to the components of $\boldsymbol{\theta}$, and therefore for those of L. A second computational advantage has been achieved by exploiting the fast decay of the Gaussian function: in the computation of the above-mentioned linear combinations, the contribution of each Gaussian is definitely ignored as soon as it assumes a value lower than a prefixed threshold (the value 10^{-4} has been used in the data processing reported here). The saving in computing time thus obtained, is particularly relevant in the evaluation of the derivatives.

Of course, the minimization algorithm ensures in practice convergence to a local minimum; in order to obtain a physically reliable final estimation and to reduce the processing time, the choice of the starting point deserves particular care. In Refs [4, 5], this choice was left to the experience and skillfulness of the operator. Here, a procedure has been implemented to perform this task automatically. On the basis of the measured histogram to be processed, the user should select, on the fluorescence intensity axis, four nonadjacent channel intervals corresponding to the location of the G_0/G_1 and $G_2 + M$ phases of the two populations as well as the modal channel value of the four peaks. By exploiting this information, the computing program automatically determines an initial guess for $\boldsymbol{\theta}$. Should any of the previous phases not be clearly identifiable, the operator can nevertheless make a tentative choice for them and check its acceptability by verifying the fitting of the computed histogram with respect to the experimental one. In order to further improve this choice, a fixed number of iterations may be performed by minimizing a χ^2 error index, before initiating the maximum likelihood estimation process.

4. TESTING THE ESTIMATION PROCEDURE AGAINST *AD HOC* EXPERIMENTAL DATA

In this section we describe two sets of experiments which have been specifically carried out with the aim of testing the capability of the revised procedure in the separation of two populations when applied to real data.

In the experiments, trout spleen, mouse spleen, mouse thymo and human thyroid cells were used. Cells were cultured in RPMI supplemented with 10% fetal calf serum, nonessential amino acids and antibiotics (100 UI/ml penicillin and 100 μg/ml streptomycin).

Cells were washed twice with PBS pH 7.4, appropriately diluted with isotonic liquid and counted using a Coulter counter. The suspensions were centrifuged at 1000 rpm for 10 min and the cell pellet was fixed in 70% ethanol and stored at 4°C.

The cells were taken out of the ethanol and prepared for flow cytometric analysis. In particular, we have considered two couples of the previous cell populations: the first one obtained by mixing trout and mouse spleen cells, the second one by using mouse thymocytes and human thyroid cells. We have measured them both separately and in mixtures of various known proportions.

Cells were treated for 10 min at room temperature with 0.5% pepsin–HCl solution (Serva, Heidelberg, F.R.G.), pH 1.8. Those relative to the first experiment were stained with DAPI (4′6′-diamidino′-2-phenylindolo) [7], to a final concentration of 5 μg/ml in Tris buffer containing 40 mM $MgCl_2$ and 10^6 cells. In the second experiment, cells were stained with a dye solution of ethidium bromide (Serva, Heidelberg, F.R.G.), mithramycin (Pfizer Inc., New York, U.S.A.) and $MgCl_2$, in 0.1 M Tris–HCl buffer, pH 7.5 [8] to a final concentration of 5 μg/ml ethidium bromide, 12.5 μg/ml mithramycin, $7.4 \cdot 10^{-3}$ M $MgCl_2$ and 10^6 cells. Samples were filtered through a 70 μm dia nylon mesh and kept at 4°C for 15 min before flow analysis.

Cells stained with DAPI were measured using a Partec Pas II flow cytometer (Partec AG, Arlesheim, Switzerland). Excitation wavelengths around 350 nm were selected by two filters: BG38 and UG1. A TK 420 dichroic mirror was used to separate the incident light from the emission light. The fluorescence light was filtered by a GG 435 step filter. In the second experiment, an ICP flow cytometer (ORTO) was used. The excitation light was selected by a BG12 filter (around 405 nm). The fluorescence light, separated by a TK 455 dichroic mirror, was filtered by a RG 610 step filter.

About 10^4 cells were accumulated for each histogram.

As far as the data processing is concerned, the minimization algorithm has been coded in FORTRAN 77 in double-precision arithmetic. The computational results have been obtained by using a VAX-11/780 under the VMS operating system.

In Tables 1–4, we report the results corresponding to the two sets of experiments. In particular, the first two columns show the known proportions of the two populations in the measured mixture and the estimated proportions, respectively. The subsequent three columns report the estimated cell cycle phase percentages; in parentheses we have indicated the corresponding percentages of the three phases within each of the two populations. The last column indicates the estimated value of the DNA index, i.e. the ratio of the estimated locations of the two G_0/G_1 peaks. Tables 1 and 2 refer to the experiment with trout ($i = 1$) and mouse ($i = 2$) spleen cells, whereas Tables 3 and 4

Table 1. Experiment with trout (Tr) and mouse (Mo) spleen cells ($n^{(1)} = n^{(2)} = 1$)

Sample prop. (%)		Estim. prop. (%)	G_1	S	G_2	$y_1^{(2)}/y_1^{(1)}$
Tr	0	0	0 (0)	0 (0)	0 (0)	
Mo	100	100	83.35 (83.35)	4.37 (4.37)	12.28 (12.28)	
Tr	20	20.62	13.06 (63.33)	7.14 (34.62)	0.42 (2.05)	1.31
Mo	80	79.38	67.59 (85.14)	4.00 (5.04)	7.79 (9.82)	
Tr	25	20.91	15.02 (71.83)	5.13 (24.52)	0.76 (3.65)	1.31
Mo	75	79.09	65.27 (82.53)	5.17 (6.54)	8.65 (10.93)	
Tr	33.4	29.78	21.25 (71.35)	7.18 (24.11)	1.35 (4.54)	1.30
Mo	66.6	70.22	57.31 (81.60)	5.09 (7.26)	7.82 (11.14)	
Tr	50	43.45	33.31 (76.67)	7.60 (17.49)	2.54 (5.84)	1.31
Mo	50	56.55	44.89 (79.37)	6.53 (11.56)	5.13 (9.07)	
Tr	66.6	63.13	52.98 (83.92)	6.05 (9.58)	4.10 (6.50)	1.30
Mo	33.4	36.87	27.13 (73.58)	6.89 (18.68)	2.85 (7.74)	
Tr	75	74.72	58.63 (78.46)	11.81 (15.82)	4.28 (5.72)	1.30
Mo	25	25.28	15.95 (63.08)	7.83 (30.98)	1.50 (5.94)	
Tr	100	99.69	82.23 (82.48)	9.92 (9.95)	7.54 (7.56)	
Mo	0	0.31	0.31 (100)	0 (0)	0 (0)	

Table 2. Experiment with trout (Tr) and mouse (Mo) spleen cells ($n^{(1)} = n^{(2)} = 2$)

Sample prop. (%)		Estim. prop. (%)	G_1	S	G_2	$y_1^{(2)}/y_1^{(1)}$
Tr	0	0	0 (0)	0 (0)	0 (0)	
Mo	100	100	83.13 (83.13)	4.44 (4.44)	12.43 (12.43)	
Tr	20	20.99	12.08 (57.54)	8.26 (39.37)	0.65 (3.09)	1.31
Mo	80	79.01	65.52 (82.92)	5.96 (7.55)	7.53 (9.53)	
Tr	25	21.04	13.94 (66.29)	6.21 (29.55)	0.88 (4.16)	1.32
Mo	75	78.96	63.53 (80.45)	7.07 (8.96)	8.36 (10.59)	
Tr	33.4	33.01	20.11 (60.91)	11.71 (35.49)	1.19 (3.60)	1.31
Mo	66.6	66.99	55.88 (83.41)	3.68 (5.50)	7.43 (11.09)	
Tr	50	41.90	32.62 (77.85)	6.37 (15.21)	2.91 (6.94)	1.31
Mo	50	58.10	43.55 (74.94)	9.57 (16.48)	4.98 (8.58)	
Tr	66.6	67.70	51.64 (76.28)	12.21 (18.04)	3.85 (5.68)	1.30
Mo	33.4	32.30	25.16 (77.91)	4.49 (13.90)	2.65 (8.19)	
Tr	75	74.61	58.13 (77.90)	11.94 (16.01)	4.54 (6.09)	1.30
Mo	25	25.39	15.18 (59.76)	8.79 (34.63)	1.42 (5.61)	
Tr	100	99.64	82.28 (82.58)	9.85 (9.88)	7.51 (7.54)	
Mo	0	0.36	0.36 (100)	0 (0)	0 (0)	

Table 3. Experiment with mouse thymocytes (Tm) and human thyroid cells (Ty) ($n^{(1)} = n^{(2)} = 1$)

Sample prop. (%)		Estim. prop. (%)	G_1	S	G_2	$y_1^{(2)}/y_1^{(1)}$
Tm	0	1.27	1.27 (100)	0 (0)	0 (0)	
Ty	100	98.73	80.98 (82.02)	6.31 (6.39)	11.44 (11.59)	
Tm	20	19.97	17.83 (89.26)	0 (0)	2.14 (10.74)	1.25
Ty	80	80.03	70.27 (87.79)	3.49 (4.38)	6.27 (7.83)	
Tm	25	23.72	19.66 (82.89)	1.91 (8.06)	2.15 (9.05)	1.27
Ty	75	76.28	65.10 (85.35)	4.13 (5.41)	7.05 (9.24)	
Tm	50	62.23	42.90 (68.93)	11.91 (19.14)	7.42 (11.93)	1.25
Ty	50	37.77	28.69 (75.96)	4.07 (10.78)	5.01 (13.26)	
Tm	80	80.17	62.55 (78.01)	4.38 (5.47)	13.24 (16.52)	1.27
Ty	20	19.83	12.53 (63.20)	5.41 (27.27)	1.89 (9.53)	
Tm	100	100	81.64 (81.64)	6.41 (6.41)	11.95 (11.95)	
Ty	0	0	0 (0)	0 (0)	0 (0)	

Table 4. Experiment with mouse thymocytes (Tm) and human thyroid cells (Ty) ($n^{(1)} = n^{(2)} = 2$)

Sample prop. (%)		Estim. prop. (%)	G_1	S	G_2	$y_1^{(2)}/y_1^{(1)}$
Tm	0	1.34	1.34 (100)	0 (0)	0 (0)	
Ty	100	98.66	80.85 (81.95)	6.25 (6.33)	11.56 (11.72)	
Tm	20	18.79	17.77 (94.60)	0 (0)	1.02 (5.40)	1.25
Ty	80	81.21	70.43 (86.72)	5.41 (6.66)	5.37 (6.62)	
Tm	25	27.15	18.93 (69.74)	6.41 (23.60)	1.81 (6.66)	1.28
Ty	75	72.85	63.59 (87.29)	2.38 (3.27)	6.88 (9.44)	
Tm	50	59.87	43.28 (72.29)	11.29 (18.86)	5.30 (8.85)	1.25
Ty	50	40.13	32.40 (80.73)	2.76 (6.87)	4.97 (12.40)	
Tm	80	82.16	62.57 (76.16)	7.19 (8.75)	12.40 (15.09)	1.27
Ty	20	17.84	13.26 (74.32)	2.74 (15.36)	1.84 (10.32)	
Tm	100	100	81.80 (81.80)	6.67 (6.67)	11.53 (11.53)	
Ty	0	0	0 (0)	0 (0)	0 (0)	

refer to the one with mouse thymocytes ($i = 1$) and human thyroid cells ($i = 2$). Tables 1 and 3 report the results obtained by means of the model with a single S-phase compartment ($n^{(1)} = n^{(2)} = 1$); Tables 2 and 4 refer to the case of two S-phase compartments for each population ($n^{(1)} = n^{(2)} = 2$).

In order to show the fitting property of our method, in Figs 1–8 some measured histograms (•——•) are plotted for the two experiments, along with the corresponding final estimated distributions (——). All figures refer to the choice $n^{(1)} = n^{(2)} = 1$.

5. DISCUSSION

A first argument of discussion refers to the capability of the procedure to recover out of a given mixture the proportions of the two populations. In this regard, as far as the first experiment is concerned, we have plotted in Fig. 9 the estimated percentages of trout spleen cells (second columns of Tables 1 and 2) vs the known percentages of the same cell population in the sample (first columns of Tables 1 and 2). In the same figure both the results corresponding to the choices $n^{(1)} = n^{(2)} = 1$ and $n^{(1)} = n^{(2)} = 2$ are represented. Obviously, the percentages of the second population (mouse spleen cells) are represented in the same figure as the complements of the percentages of the first one. Similarly, the population percentages relative to the second experiment are plotted in Fig. 10.

The absolute mean deviations between the estimated and the known percentages for the first experiment are 2.37 in the case $n^{(1)} = n^{(2)} = 1$ and 1.91 in the case $n^{(1)} = n^{(2)} = 2$, with maximum deviations of 6.55 and 8.1, respectively. For the second experiment, the mean deviations are 2.5 in the case $n^{(1)} = n^{(2)} = 1$ and 2.79 in the case $n^{(1)} = n^{(2)} = 2$, with maximum deviations of 12.23 and 9.87, respectively. From these results it appears that the procedure is able to properly estimate the proportions of the two populations. Furthermore, from Figs 9 and 10, we note that the deviations are less significant when the sample proportions are unbalanced, whereas the worst cases occur always in correspondence to sample proportions of 50%. We remark that the procedure has been applied also to the samples containing the single four populations with quite satisfactory results regarding the capability of the algorithm to exclude in these cases the presence of a second population.

The above fact may be explained to some extent by interpreting the flow cytometric processing of the cells of a given sample as repeated trials and by considering as a success the belonging of each cell to one of the two populations. By assuming these trials to be independent and by taking into account the uncertainty due to the sample finiteness, the success frequency turns out to be proportional to a random variable with a binomial distribution [6]. Denoting by p the known fraction of one of the two populations, as is well-known, the variance of this random variable is proportional to $p(1-p)$. Therefore, the estimate of the fraction p is expected to be more reliable the more the percentages are unbalanced (p near to 0 or to 1).

A second point concerns the possibility of estimating the value of the DNA index. From the final columns of Tables 1–4 it appears that the procedure yields an estimate of this index virtually without any fluctuation, both with respect to the proportions in the sample and to the choice of the parameters $n^{(1)}$ and $n^{(2)}$. The stability of this estimate can be related to the fact, as already

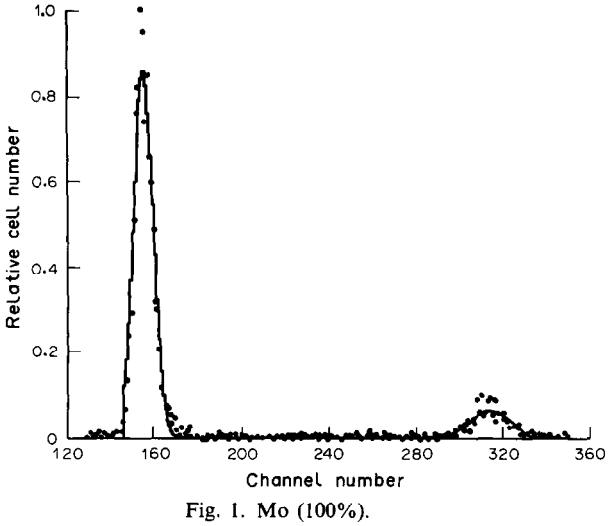

Fig. 1. Mo (100%).

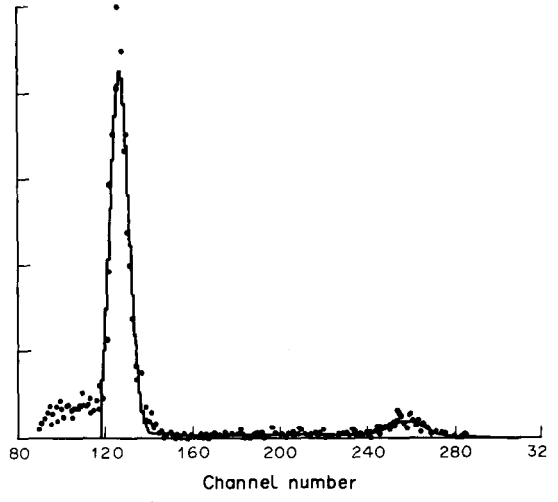

Fig. 2. Tr (100%).

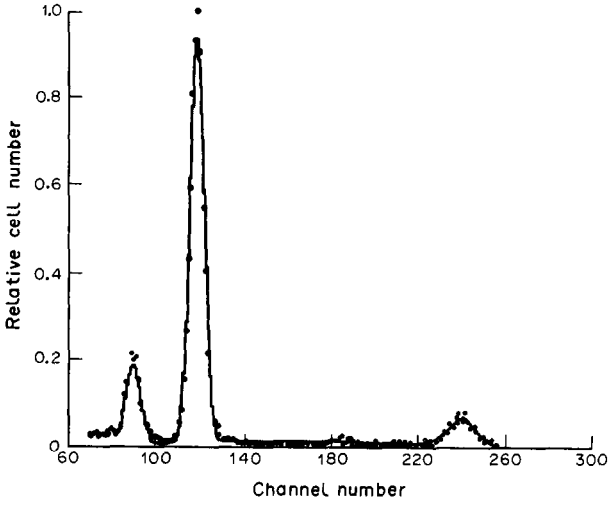

Fig. 3. Tr (25%)–Mo (75%).

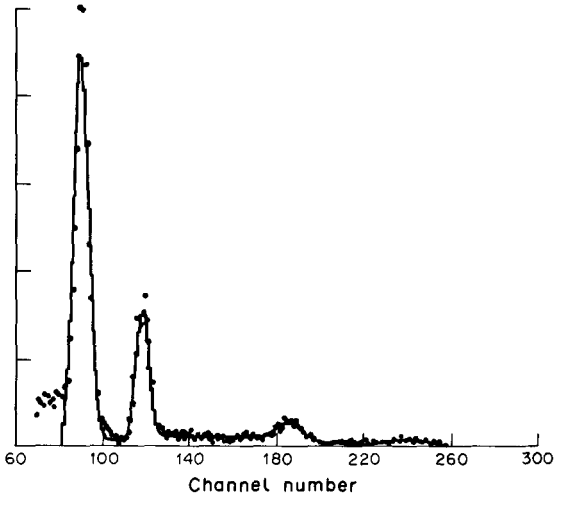

Fig. 4. Tr (75%)–Mo (25%).

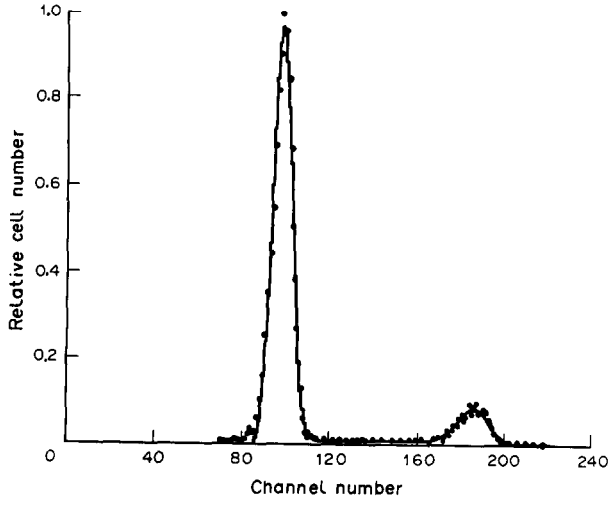

Fig. 5. Ty (100%).

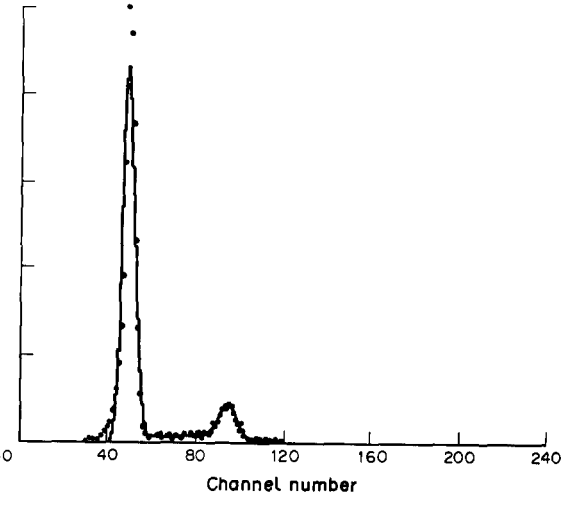

Fig. 6. Tm (100%).

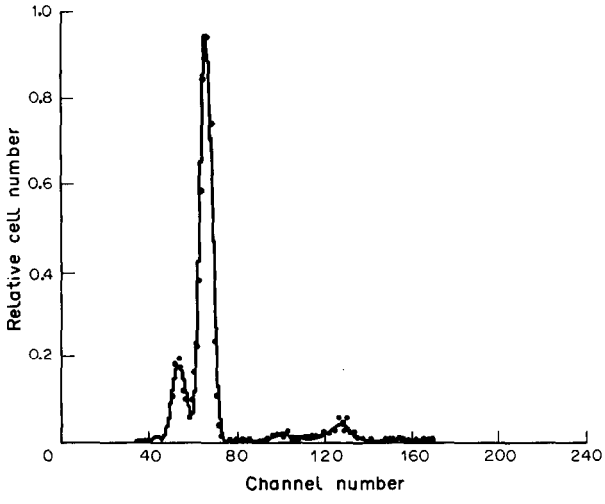

Fig. 7. Tm (20%)–Ty (80%).

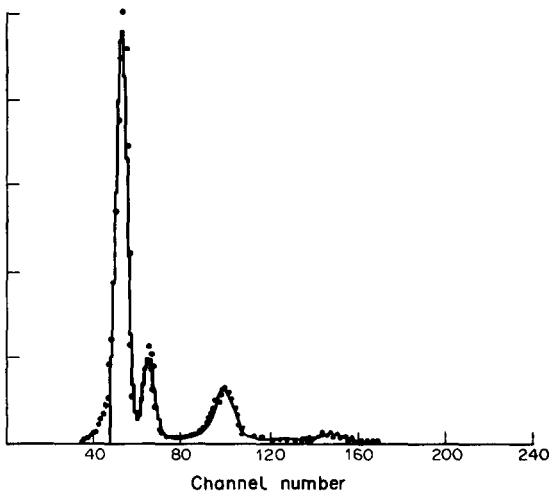

Fig. 8. Tm (80%)–Ty (20%).

pointed out, that our procedure is able to automatically compensate for a possible instrumental bias on the fluorescence intensity.

A third item of interest is related to the problem of estimating the phase fractions of each population from the overall histogram. In Fig. 11, with reference to the first experiment, we have plotted the estimated percentages of the G_0/G_1, S, G_2+M phases, for both trout and mouse spleen cells, separately for the cases $n^{(1)}=n^{(2)}=1$ and $n^{(1)}=n^{(2)}=2$ (third, fourth and fifth columns of Tables 1 and 2). The same kind of representation is given in Fig. 12 for the second experiment.

From these figures it appears that, in any case, the phase fraction estimates are sufficiently stable, and therefore, probably reliable, when the corresponding population constitutes the greater part of the sample. On the contrary, the phase fraction estimates of the lower percentage population turn out to be less significant.

The obtained results do not exhibit meaningful differences when changing the value of the parameters $n^{(1)}$ and $n^{(2)}$, so they do not allow us to give any specific indications about their choice. This fact could be caused by the almost uniform S-phase shape of the considered cell lines. Further investigations on this point are needed.

Finally, we observe that the above conclusions are in agreement with those previously formulated in Ref. [4], with reference to the analysis of simulated data.

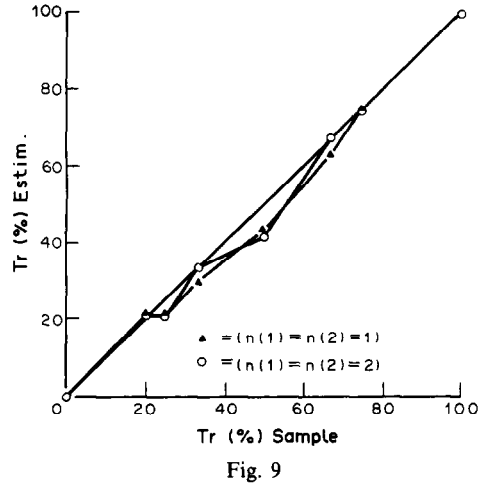

Fig. 9

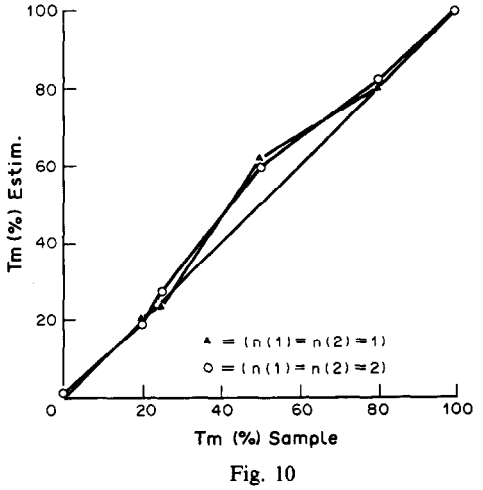

Fig. 10

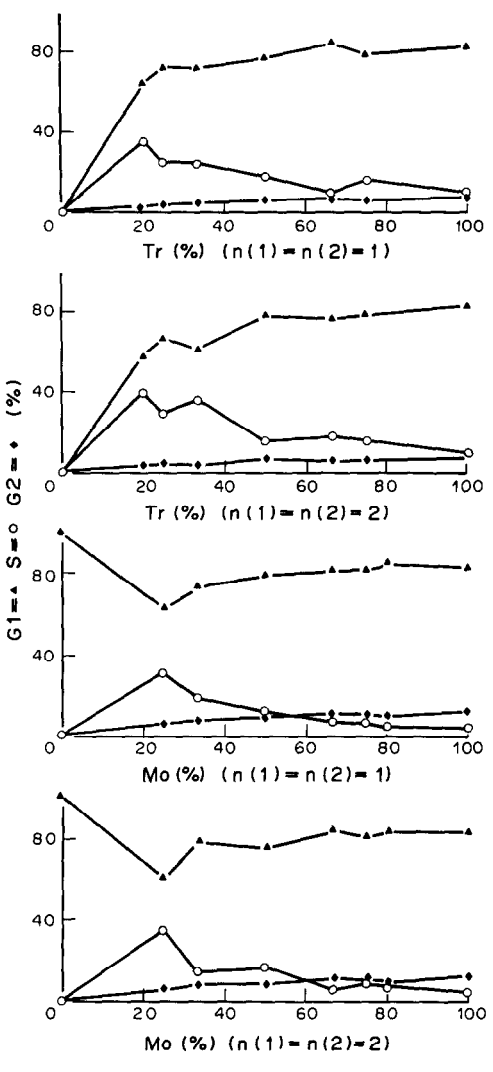

Fig. 11

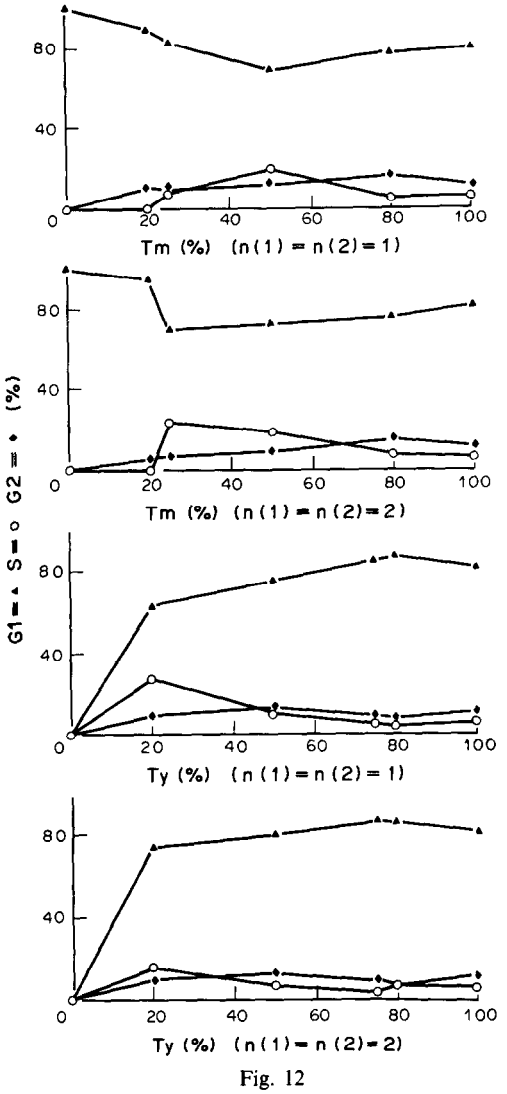

Fig. 12

6. CONCLUDING REMARKS

Two relevant features of the revised method for aneuploid DNA histogram analysis presented in this paper are (1) the introduction of an automatic choice of the initial estimate of the unknown model parameters and (2) the considerable reduction of the computational cost inherent in the optimization process. The first point is important with respect to the requirement of obtaining physically reliable estimates, while the second is particularly interesting with respect to the possibility of avoiding the use of a mainframe computer (some encouraging applications have been performed by employing an advanced technology PC). A further important feature of the revised model is the introduction of a shift parameter on the fluorescence intensity, allowing an automatic compensation of a possible instrumental bias.

From the tests performed it appears that the procedure is able to recover the DNA index value, as well as the percentage of each population in the mixture. We have also observed that the more the sample proportions are unbalanced, the more this last capability increases. This fact confirms the interest, already stressed in Ref. [4], about the potential of the method in problems of the early diagnosis of cancer diseases (the possibility of detecting and quantitatively evaluating an abnormal stemline during an early development stage).

The procedure appears also to be a useful tool for estimating the cell cycle phase fractions of each of the two mixed populations. In particular, since the S-phase percentage may account for

the proliferative activity in several tumor types, the possibility of quantitatively evaluating this parameter can help in prognostic problems. However, we have noted that the phase fraction identification turns out to be less reliable when the overall proportion of the corresponding population is small.

Further investigations would be useful with respect to either the previous considerations or to the other questions raised, like that of the choice of the number of S-phase compartments. It would be also of interest to analyze the behavior of the estimation procedure when the DNA index takes values near to 1 or 2. Finally, we remark that it is necessary to include in the mathematical description of the fluorescence distribution and adequate background model, in order to successfully apply our method of analysis to histograms from solid tumor samples. A function accounting for the actual cell fragment distribution has been proposed recently [9].

Acknowledgements—This work was partially supported by Progetto Finalizzato Oncologia, CNR, under Grant Nos 83.00847 and 56825.

The authors gratefully acknowledge the Laboratorio di Dosimetria e Biofisica, ENEA-Casaccia, and the Institut für Strahlenbiologie, Münster Universität, where the measurements were performed.

REFERENCES

1. H. Baisch, H.-P. Beck, I. J. Christensen, N. R. Hartmann, J. Fried, P. N. Dean, J. W. Gray, J. H. Jett, D. A. Johnston, R. A. White, C. Nicolini, S. Zeitz and J. V. Watson, A comparison of mathematical methods for the analysis of DNA histograms obtained by flow cytometry. *Cell Tiss. Kinet.* **15,** 235–249 (1982).
2. P. S. Ritch, S. E. Shackney, W. H. Schuette, T. L. Talbot and C. A. Smith, A practical graphical method for estimating the fraction of cells in S in DNA histograms from clinical tumor samples containing aneuploid cell populations. *Cytometry* **4,** 66–74 (1983).
3. L. G. Dressler, L. Seamer, M. A. Owens, G. M. Clark and W. McGuire, Evaluation of a modeling system for S-phase estimation in breast cancer by flow cytometry. *Cancer Res.* **47,** 5294–5302 (1987).
4. C. Bruni, L. Capurso, G. Koch, M. Koch, F. Lampariello, S. Lucidi and L. Teodori, Automatic analysis of flow cytometrically determined DNA distributions in the presence of abnormal stemlines. *Mathl Modelling* **7,** 1325–1338 (1986).
5. F. Lampariello and S. Lucidi, Analysis of DNA distributions from flow cytometry by means of an optimization procedure. In *System Modelling and Optimization* (Edited by A. Prékopa, J. Szelezsán and B. Strazicky); *Lecture Notes in Control and Information Sciences*, Vol. 84, pp. 478–487. Springer, New York (1986).
6. H. Cramer, *Mathematical Methods of Statistics*. Princeton Univ. Press, N.J. (1963).
7. W. Göhde, J. Schumann and J. Zante, The use of DAPI in pulse cytophotometry. In *Proc. Third Int. Symp. Pulse Cytophotometry* (Edited by D. Lutz), pp. 229–238. European Press, Ghent (1978).
8. J. Zante, J. Schumann, W. Göhde and U. Hacker, DNA-fluorometry of mammalian sperm. *Histochemistry* **54,** 1–7 (1977).
9. F. Lampariello and G. Del Bino, Automatic parameter estimation of flow cytometric DNA distributions in the study of tumor cell kinetics. In *Identification and System Parameter Estimation (8th IFAC/IFORS Symp., Beijing, China)* (Edited by H.-F. Chen), pp. 1767–1772. Pergamon Press, Oxford (1988).