Cytometry 8:545–551 (1987)

Spectral Analysis of Flow Cytometric Data: Design of a Special-Purpose Low-Pass Digital Filter

Peter M.A. Sloot, Pieter Tensen, and Carl G. Figdor

Divisions of Biophysics and Immunology, The Netherlands Cancer Institute, Antoni van Leeuwenhoek Huis, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

Received for publication January 22, 1987; accepted June 9, 1987

Spectral decomposition of flow cytometric datafiles of arbitrary dimension reveal information of both the signal and the noise components that constitute the histograms. This spectral information is used to construct a low-pass digital filter, which removes the high-frequency noise from the actual data. It is shown that this procedure guarantees nontrivial smoothing of the flow cytometric data in accordance with the local experimental situation. As a consequence optimal reconstruction of the signal is possible, which facilitates unambiguous interpretation of the data files and mathematical estimation of the statistical parameters.

Key terms: Fourier transform, noise rejection

Interpretation of flow cytometric (FCM) data is hampered by the presence of noise superimposed on the signal of the studied phenomenon (e.g., fluorescence, light scatter). Both deterministic and stochastic noise may contribute to the histograms. Deterministic noise is present as a consequence of (known) unavoidable nonsystematic instrumental errors (17,21,22), whereas stochastic noise may arise from a statistically insufficient number of cells (15,16). The main purpose of analyzing FCM data is the classification and detection of homogeneous (sub)populations. This can be accomplished by means of statistical parametric and nonparametric analysis which may be extended by iterative algorithms (1,10,11,13,14,24).

The reliability of histogram analysis and the convergence time of iterative parametric procedures are strongly correlated to the estimation accuracy of the initial parameters (e.g., mean and [co]variance) (10,14). Furthermore, the statistical significance of nonparametric analysis such as randomization algorithms is affected by the signal-to-noise ratio (1,11,13,24). Flow cytometry is a technique whereby fluorescence and/or elastic light scatter of biological entities, such as cells, are studied. The response of a population of cells to the incident light deviates from a one-channel histogram as a result of instrumentional resolution and biological spread (i.e., slight differences in size, chromophore uptake, and optical density). This homogeneous broadening (of a single population) extends over several channel numbers and may be regarded as one period of a basically low-frequency signal. The actual signal, however, may be corrupted by (for instance) additive white noise (which is composed of frequency components with no preference to any special frequency). Therefore, it is assumed that the Fourier coefficients of the intrinsic signal decrease with increasing frequency, whereas the Fourier coefficients of the noise remain approximately constant (7). The validity of this assumption is investigated by means of spectral analysis of FCM data. As a consequence the Fourier tranform (FT) of the FCM data must contain information to discriminate between signal and noise.

In the present study we describe a procedure to specify the significant signal part of FCM data. This information facilitates the design of a digital filter with characteristics tuned to the specific local experimental situation, which allows unambiguous interpretation of FCM histograms.

MATERIALS AND METHODS

Spectral Analysis

FCM histograms may be considered as a discrete representation of a (continuous) detector-signal distribution, which is digitized into an N-channel histogram by means of analog/digital conversion. This raw distribution is converted to a histogram with a fixed class width and a fixed number of classes (e.g., 0–5 V is mapped to 0–63 channels).

The frequency content of the histogram (represented by the function f[x]) can be studied by means of spectral decomposition. This is established by calculating the

This work was supported by Stichting Technische Wetenschappen (STW), grant No. LGN 22053.

discrete Fourier transform (DFT) $F(\omega)$ from the N-channel histogram f[x] (9,20):

$$F(\omega): = \sum_{x=0}^{N-1} f(x) \exp(-i\pi x\omega) \quad (1)$$

where $F(\omega)$ is composed of an imaginary part (Im[$F(\omega)$] and a real part (Re [$F(\omega)$]). The normalized frequency $0 \leqq \omega \leqq 1$ is defined by $\omega = f\,T$; T is the interval between each of the N data points; and $0 \leqq f \leqq 2\pi$ (12).

Since we are especially interested in the relative amplitude of the Fourier coefficients, the output is converted to polar coordinates:

$$\begin{aligned} |\,A(\omega)\,| &= \{\mathrm{Re}[F(\omega)]^2 + \mathrm{Im}\,[F(\omega)]^2\}^{1/2} \\ \Phi(\omega) &= \tan^{-1}\{\mathrm{Im}\,[F(\omega)]\}/\{\mathrm{Re}\,[F(\omega)]\} \end{aligned} \quad (2)$$

with $|A(\omega)|$ the absolute value of the magnitude of the Fourier coefficients and $\phi(\omega)$ the corresponding phase. To speed up the computational procedures a widely used algorithm derived by Cooley and Tukey (3), the fast Fourier transform (FFT), is used.

The interpretation of the spectral analysis is straightforward: Assume that the relevant information in f(x) can be described by a signal function s(x) containing only the low-frequency components, and a function n(x) with a random distribution of frequency components representing the noise that needs to be rejected:

$$f(x) = s(x) + n(x) \quad (3)$$

Since the FT is a linear operation, the Fourier transformation of equation 3 is constructed of a monotonically decreasing function $S(\omega)$ and a more-or-less constant part $N(\omega)$, where $S(\omega)$ and $N(\omega)$ result from the FT of s(x) and n(x), respectively. From this Fourier spectrum the cutoff frequency, ω_c, above which no relevant information is expected, can be deduced. This is accomplished by means of a linear regression fit of the right part ($\omega \geqq 0.5$) of the normalized spectrum and extrapolation to the intersection with the monotonically decreasing left part of the spectrum. This procedure guarantees that no significant signal information is lost.

The Low-Pass Filter

Once the cut off frequency ω_c is estimated, the high-frequency part of the spectrum ($\omega > \omega_c$) can be removed by convolution of the histogram function f(x) with a function h(x), which is only sensitive to the low-frequency part s(x) ($\omega \leqq \omega_c$) of f(x) (6).

Here we illustrate some asymptotic properties of the functions in the frequency domain:

$$s(x) = [s(x) + n(x)] * h(x)$$

Fourier transformation results in

$$S(\omega) = [S(\omega) + N(\omega)] \cdot H(\omega) \quad (4)$$

where the convolution theorem (20) is applied; $H(\omega)$ (the transfer function) is the FT of h(x) and * denotes the convolution.

Rearranging equation 4 results in

$$H(\omega) = \frac{S(\omega)/N(\omega)}{S(\omega)/N(\omega) + 1} \quad (5)$$

from which two regions (see Fig. 1) can be deduced:

(i) $|\,N(\omega)\,| \gg |\,S(\omega)\,|$:
The irrelevant high frequency part where $|\,H(\omega)\,| \to 0$
(ii) $|\,N(\omega)\,| \ll |\,S(\omega)\,|$:
The relevant low frequency part where $|\,H(\omega)\,| \to 1$

The convolution of the histogram function with an impulse response having a frequency characteristic comparable to that shown in Figure 1 results in rejection of the high-frequency components of the data and therefore results in an unambiguous smoothed histogram. Similar low-pass filter techniques are common in signal and time series analysis (2,8).

The Window Definition

For computation purposes, a finite number of impulse response terms is used in the estimation of $H(\omega)$. Therefore, a certain overshoot in the frequency domain of $H(\omega)$, near the discontinuity ω_c, will occur [Gibbs phe-

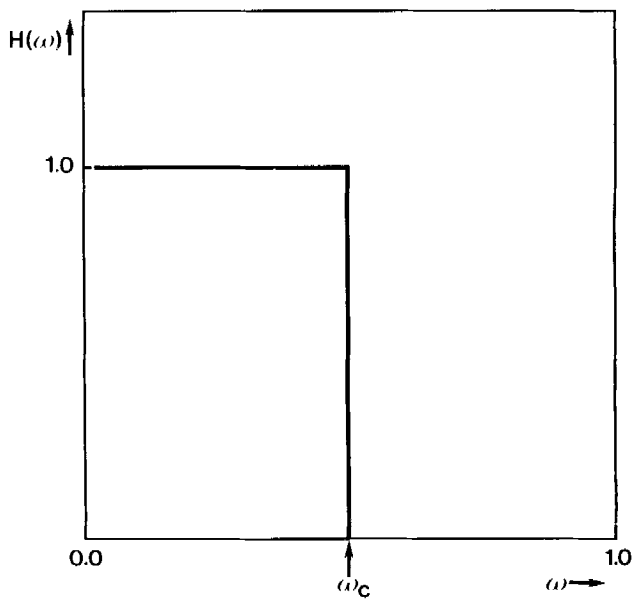

FIG. 1. The Fourier transform of the magnitude of an ideal impulse response function vs. the normalized frequency. ω_c: the cutoff frequency.

nomenon (6)]. This error can be reduced by means of a modifying window w(n). Several windows for special purposes have been proposed in literature 1,2,23). In this study, the kernel of a very simple Lanczos window is applied (2), in order to make the implementation on a microcomputer more feasible:

$$w(n): = \left[\frac{\sin\left(\frac{2n\pi}{Np-1}\right)^m}{\frac{2n\pi}{Np-1}} \right] \quad (6)$$

where m depends on the span of filter (2·Np). The optimal m for a certain number of smoothing points Np can be derived from the effective transfer function $H_{eff}(\omega)$:

$$H_{eff}(\omega) = FT(h(n)\cdot w(n)) \quad (7)$$

Several filter parameters have to be taken into account: The transition band width (δ) which defines the steepness of the discontinuity in Figure 1, and the passband tolerance (ϵ) which is defined by the amount of overshoot in the regions on both sides of ω_c. The values of both parameters should be minimal for an optimal filter. If m in equation 6 increases, ϵ decreases and δ increases. However, if m decreases, ϵ increases and δ decreases. Apparently, for a certain number of smoothing points, NP, an optimal value for m can be calculated. From equations 4 and 6 the filtered histogram f*(x) is calculated by means of a discrete convolution of the modified symmetric impulse response function b(n) with the original data f(x):

$$f^*(x) = \sum_{n=0}^{Np} b(n)\,[f(x-n) + f(x+n)]$$

where

$$b(n) = \frac{\sin \omega_c n\pi}{n\pi} \cdot w(n) \quad (8)$$

The first term of b(n) results from the inverse Fourier transform of the ideal transfer function $H(\omega)$ (with an ideal rectangular frequency response characteristic—see Fig. 1), whereas the second term is defined by equation 6. The span of the filter depends on the required accuracy and the maximum allowable computation time. Calculations demonstrated that half times the number of channels (N) is an acceptable value for Np. Equation 8 implies that once Np, ω_c, and m are defined, b(n) can be calculated and the filtered histogram results from a number of simple multiplications. Furthermore, an adequate definition of Np, ω_c, and m allows noise reduction of the original FCM data in accordance with the local system performance and the current type of experiment. Finally, equation 8 implies a simple extrapolation to a filter algorithm for arbitrary dimensions, e.g., two-dimensional (bivariate) distributions.

The computations were carried out on a computer configuration based on an MC68000 microprocessor (Motorola) in the language C. The final program was implemented in BASIC on a microcomputer interfaced to a FACS IV flow cytometer.

RESULTS

In this section the characteristics and the advantages of the special-purpose low-pass filter in relation to one-dimensional simulated FCM data are investigated. In addition the results are extrapolated to higher dimensions and a multiparametric filter algorithm is described. Furthermore, the fundamental assumption is studied that FCM data are constructed of a low-frequency signal part, whereas the noise is confined to a high-frequency part.

First, the FT of the product of the impulse response function with the window function (equation 7) is calculated for $\omega_c = 0.5$, Np = 32 (= 1/2N) at different values of m. The results, shown in Figure 2, suggest an optimal window for $1 < m < 3$. Since no complex values of w(n) are permitted (m is an integer), further study is required to estimate the best value of m. In accordance with the fundamental assumption, as mentioned above, an artificial histogram is constructed: The amplitude of the numerically generated white noise resembles a Gaussian distribution with a mean value at zero amplitude (no noise present) and a standard deviation (S) that fulfills the condition that 99.7% of the white noise has

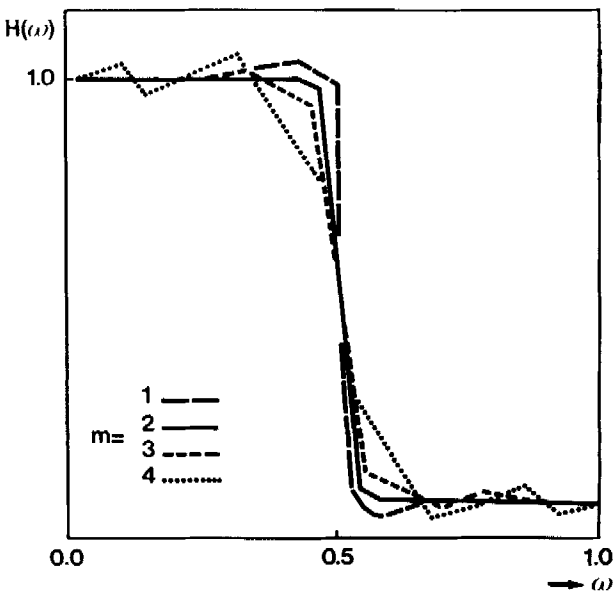

FIG. 2. The influence of the window parameter m on the quality of the filter. Np = 32, $\omega_c = 0.5$, $H(\omega)$ = absolute value of the magnitude of the transfer function.

an amplitude between 0 and 10% of the top value (TV) of the signal:

$$3S = TV/10 \quad (9)$$

By means of a pseudo-random generator (4), 64 elements were chosen to construct a packet of superimposed white noise.

The normalized signal was presented by a Gaussian distribution with a standard deviation of ten channels and a mean value of 31:

$$s(x) = \frac{1}{(2\pi)^{1/2} \cdot 10} \exp -\tfrac{1}{2} [(x-31)/10]^2 \quad (10)$$

The result, obtained with the particular noise in this sample, is shown in Figure 3a–c. From these distributions the FFTs were calculated. Figure 3d and f show the estimated cut-off frequency for the signal with and without additional noise, respectively. All tested noise and data files showed the same relation between the "real" and the estimated ω_c namely, that the latter has a slightly higher value. This is a consequence of the large number of frequency components of the noise, relative to the number of frequency components of the signal, near ω_c. Therefore a small but negligible amount of the noise is still present, whereas the filter procedure assures that no signal information is lost.

Since the information derived from Figure 3d is normally not available, the filter algorithm (equation 8) with $\omega_c = 0.19$ (Fig. 3f), Np = 32, and m= 0, 1, 2, 3, 4, 5 is applied to the simulated data file (Fig. 3c). The influence of the window parameter m on the filter procedure is studied by comparing the filtered histogram to the "original" signal (Fig. 3a) by means of a chi-square test. The typical result shown in Figure 4 indicates an optimal value for m = 2. Various packets of noise were chosen and added to the same and other signal distributions. The results were comparable to those shown in Figures 3 and 4 (data not shown). Therefore the value m = 2 was applied to construct the window. The result of the filtered histogram with m= 2 is shown in Figure 5. In order to investigate the influence of the applied filter to the Gaussian shape of the signal function (Fig. 3a), the smoothed histogram was fit to a normal distribution. No significant changes in shape could be detected. This is in line with the general phenomenon that no amplitude distortion is expected as long as the cutoff frequency (ω_c) is chosen beyond the signal region.

After establishment of the characteristics of this filter procedure and the window parameters, the algorithm was extended to higher dimensions, implemented on a microcomputer, and tested with several experimental (FCM) histograms.

In Figure 6a, a masked plot of a bivariate scatter histogram (64 × 64 channels) of a mixed population of human peripheral blood lymphocytes and monocytes, isolated as described by Figdor et al. (5), and measured with a FACS IV, is shown as an example. Although the total number of cells in this histogram is 50,000, the maximum number of cells/channel is only 170 (in the lymphocyte population), due to the broad monocyte dis-

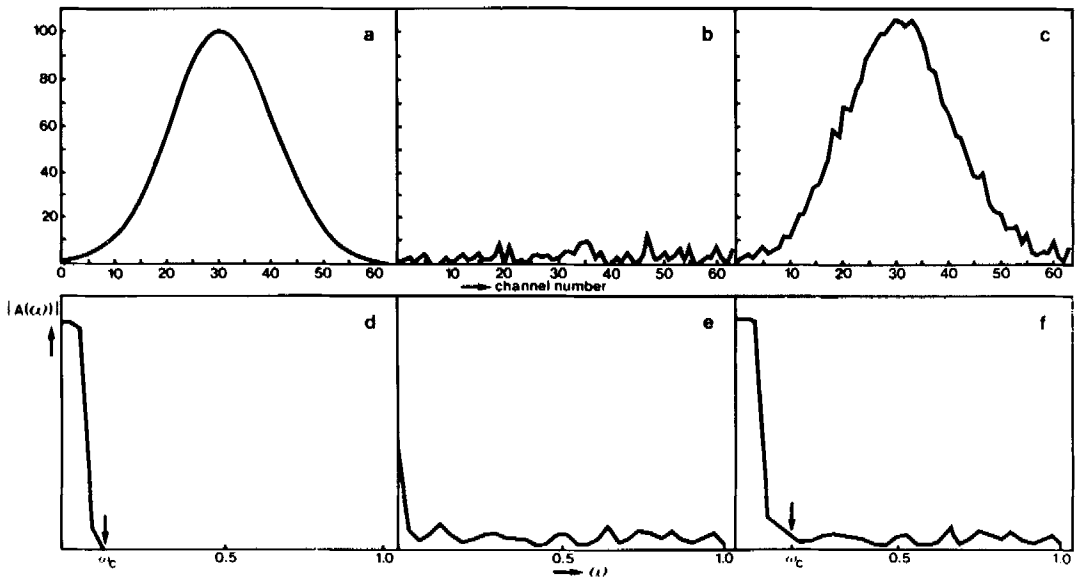

FIG. 3. A construction of a typical FCM data file, and the Fourier transform (FT) of the artificial FCM data file. ω: normalized frequency. $|A(\omega)|$: polar representation of the FT coefficients. The arrows indicate the estimated cutoff frequency ω_c. a) The normalized signal (s[x]) vs. arbitrary units. b) The random chosen white noise (n[x]) vs. arbitrary units. c) The complete data set (f(x)) vs. arbitrary units. d) FT of the normalized signal (S(ω)) vs. ω. e) FT of the random chosen noise N(ω)) vs. ω. f) FT of the complete data set (F(ω)) vs. ω.

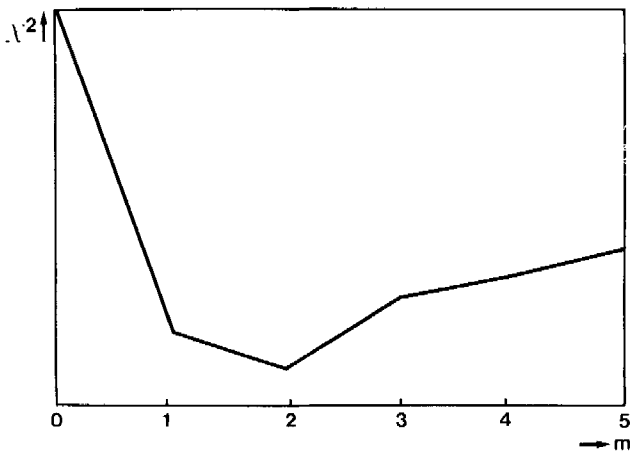

FIG. 4. Relative chi-square (χ^2) vs. the window parameter m. Calculated by comparing the filtered histogram to the original signal (Fig. 3a).

tribution. Therefore, high variance of histogram values is present in the data file, which indicates the necessity of spectral analysis. The result of two-dimensional Fourier transform is shown in Figure 6b. These data show that there are indeed two distinct regions of interest: one containing the signal information and one containing the noise information. The frequency spectrum of other data files showed a close resemblance to the data shown in Figure 6b. Estimation of the cutoff frequency in two orthogonal directions resulted in a mean value of ω_x in the forward-scatter (FS) direction of approximately 0.20 ± 0.02 ($= \omega_{cx}$), whereas for the side-scatter (SS) direction a value of 0.29 ± 0.03 ($= \omega_{cy}$) was calculated. Similar results were obtained with other FS vs. SS data files. It was concluded therefore that routine application

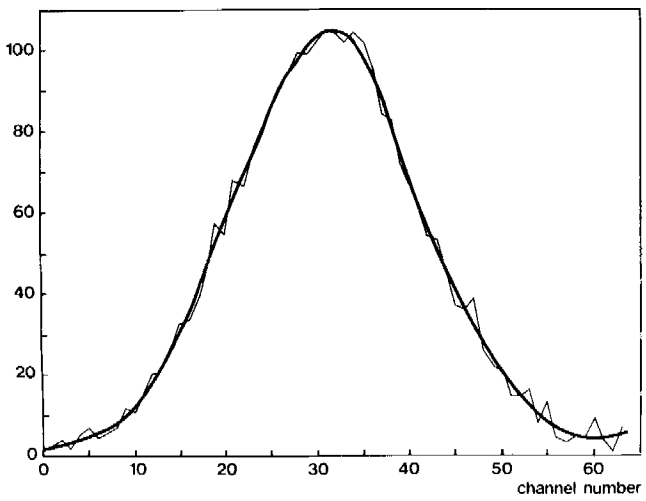

FIG. 5. The filtered artificial histogram (heavy line) plotted together with the original data from Figure 3c (thin line): m = 2, Np = 32 and $\omega_c = 0.19$.

of the filter procedure with $\omega_{cx} = 0.20$ and $\omega_{cy} = 0.29$ is allowed. Consequently, no precalculation of the cutoff frequencies by means of an FT is necessary, and direct convolution can be applied.

To smooth this bivariate histogram a neglible correlation between the FS intensities and the SS intensities is assumed (18). This implies that the FCM histogram can be separated. Therefore a cascade of two one-dimensional filters is applied: one operating on the rows and one operating on the columns (2). The data in Figure 6c show the filtered histogram calculated in accordance with the filter procedure described by equation 8. Preliminary statistical parametric analysis of the smoothed bivariate histogram shows a close resemblance to morphological characterization, by means of May-Grünwald Giemsa staining of cytocentrifuge preparations of the original sample (5). This indicates that no significant shape changes or amplitude deformation is introduced by the filtering procedure, as was expected from theoretical considerations in the previous section.

DISCUSSION

In this study a mathematical technique is presented to reject noise from FCM data files of arbitrary dimension. It is demonstrated that the fundamental assumption, that FCM data may be regarded as functions with fast-fluctuating noise components superimposed on slowly varying signals, is valid. One exception, which might lead to misinterpretation, occurs when the logarithmic noise distribution at the origin (due to a nonzero trigger threshold level on the master signal) is smoothed to look like a real subpopulation (Fig. 6). In routine application, however, this artifact will not obscure the interpretation of the complete data set.

When relatively high-frequency signals are studied, a sharp discrimination between noise and signal becomes more difficult and phase reversals may occur in the filter procedure (12), thus leading to errors in the smoothed data file. However, in the data files we have studied so far, no high-frequency components of the signal could be detected, and therefore it seems fair to conclude that spectral analysis results in a unique estimation of both the signal and the noise part of the data. Moreover, spectral analysis showed only minor changes in the cutoff frequencies of different data files. Consequently, information of the cutoff frequency is available, and application of a low pass digital filter becomes feasible. In addition, spectral analysis allows experimental minimalization of the noise present. Accordingly it can be used to study the local system performance or interpret the physical parameters of the experiment.

The results presented here indicate that the application of the proposed filter procedure facilitates unambiguous smoothing of FCM data in accordance with the frequency content of the histograms. Furthermore, a straightforward generalization to higher dimension is allowed, provided that the correlation between the various detection parameters is small enough to allow the application of a cascade of one-dimensional filters. From

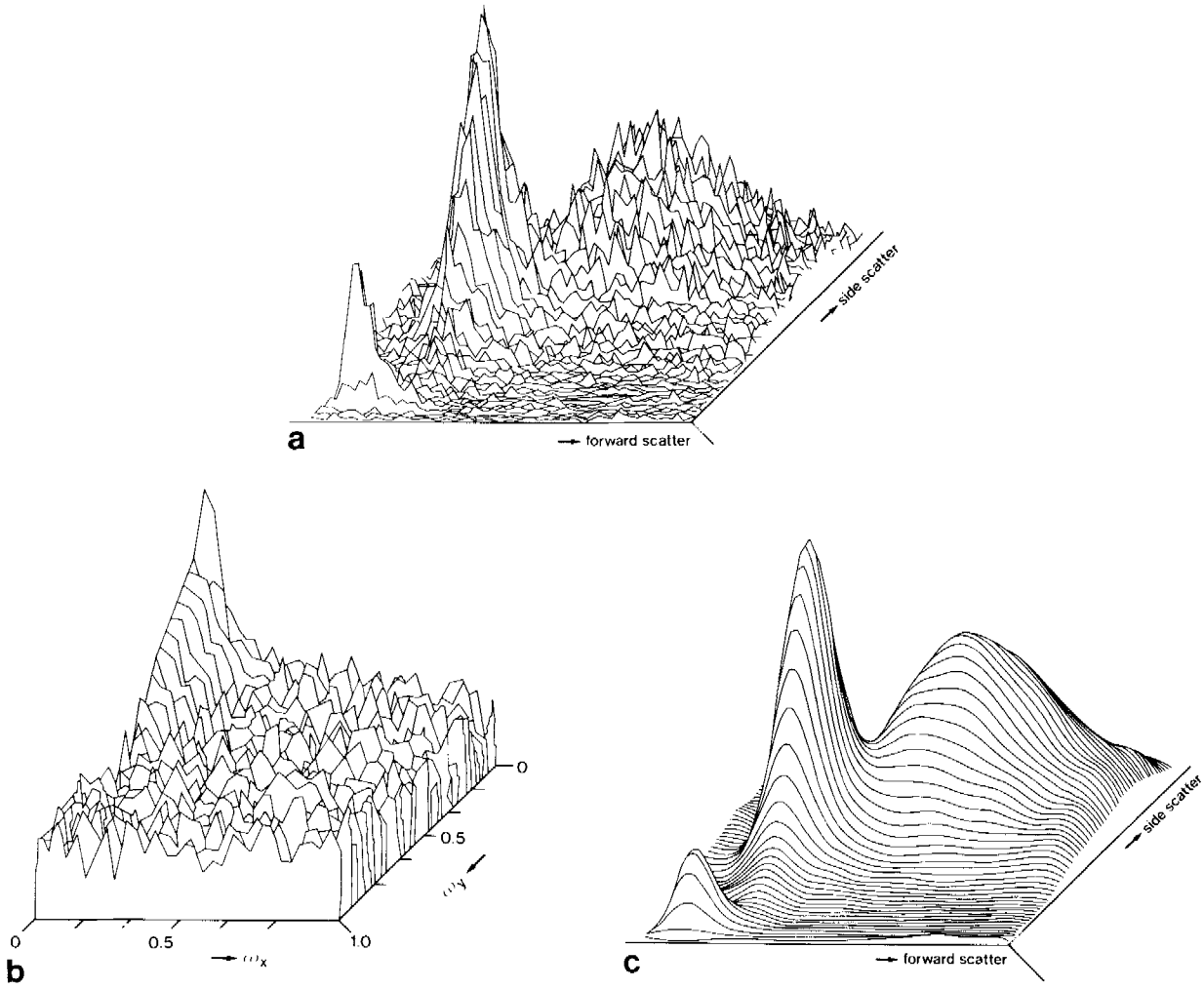

FIG. 6. Illustration of the described filter procedure for multiparameter histograms. a) Bivariate histogram of a mixed population of human lymphocytes and monocytes. b) Spectral decomposition of the histogram. c) Modified histogram after noise rejection.

previous theoretical calculations it is concluded that only a small correlation is present between FS and SS for nucleated blood cells, such as peripheral blood lymphocytes, whereas other combinations of parameters result in even smaller correlations (18). Therefore, application of a cascade of two one-dimensional filters instead of a more complex and time-consuming multidimensional filter procedure is justified.

In the method presented here, ω_c, Np, and m can be chosen in dependence of the histograms at hand, whereas the individual coefficients b(n) cannot be modified. Consequently, adaptation of the histogram data is limited. On the other hand, however, few parameter methods allow simpler and faster handling of the filter algorithm. The complete process of (off-line) filtering of a bivariate data file, with predetermined cutoff frequencies, takes only a few minutes if the algorithm is written in BASIC and implemented on a micro-computer. Implementation of the algorithm (written in the language C) on locally-designed light-scatter equipment indicates that the procedure is also applicable to other-than-commercially available flow cytometers (19).

The sources of the various programs are available from the authors upon request.

ACKNOWLEDGMENTS

The authors thank Dr. W.S. Bont for critical reading of the manuscript, E.H. Philippus for technical assistance, and Marie Anne van Halem for secretarial help.

LITERATURE CITED

1. Bagwell CB, Hudson JL, Irvin GL: Nonparametric flow cytometric analysis. J Histochem Cytochem 27:293–296, 1979.
2. Cappellini V, Constantinides AG, Emilliani P: Digital Filters and Their Applications. Academic Press, New York, 1978.
3. Cooley JW, Tukey J: An algorithm for machine calculation of complex Fourier Series. Math Comp 19:297–309, 1965.
4. Ehrhardt JC: Generation of pseudorandom numbers. Med. Phys. 13:240–241,1986.

5. Figdor CG, Van Es WL, Leemans JMM, Bont WS: A centrifugal elutriation system of separating small numbers of cells: J Immunol Methods 68:73–87, 1984.
6. Gaskill JD: Linear Systems, Fourier Transforms and Optics. John Wiley, New York, 1978.
7. Inonye T, Harper T, Rasmussen NC: Nucl. Instrum. Methods 67: 125–132, 1969.
8. Johnson AT: Microcomputer program for the design of digital filters. *Comput Methods Programs Biomed* 21:203–210, 1985.
9. Ledoux CA: The Fast Fourier Transform as a Signal Processing Technique. NUSC Technical Document 6148, 1983.
10. Mann RC, Hand RE Jr, Braslawasky GR: Parametric analysis of histograms measured in flow cytometry. Cytometry 4:75–82, 1983.
11. Mann RC, Hand RE Jr: The randomization test applied to flow cytometric histograms. Comput Programs Biomed 17:95–100, 1983
12. Marchand P. Marmet L: Binomial smoothing filter: a way to avoid some pitfalls of least-squares polynomial smoothing. Rev Sci Instrum 54:1034–1041, 1983.
13. Recchia M, Rocchetti M: The simulated randomization test. Comput Programs Biomed 15:111–116, 1982.
14. Redner AR, Walker HF: Mixture densities, maximum likelihood and the EM algorithm. SIAM Rev 26:195–293, 1984.
15. Schuette WH, Shackney SE, MacCollum MA, Smith CA: A count-dependent filter for smoothing flow cytometric histograms. Cytometry 5:487–493, 1984.
16. Schuette WH, Shackney SE, Marti GE: Count-dependent filter for smoothing bivariate FCM histograms. Cytometry 7:274–280, 1986.
17. Shapiro HM, Feinstein DM, Kirsch AS, Christenson L: Multistation multiparameter flow cytometry: Some influences of instrumentational factors on system performance. Cytometry 4:11–19, 1983.
18. Sloot PMA, Figdor CG: Elastic light scatter from nucleated blood cells: Rapid numerical analysis. Appl Optics 25:3559–3565, 1986.
19. Sloot PMA, Carels MJ, Tensen P, Figdor CG: Computer-assisted centrifugal elutriation. I: Detection system and data acquisition equipment. Comput Methods Programs Biomed 24:179–188, 1987.
20. Spiegel MR: Fourier Analysis. Schaum's Outline Series. McGraw-Hill, New York, 1974.
21. Steinkamp JA: Flow cytometry. Rev Sci Instrum 55:1357–1400, 1984.
22. Ubezio P, Andreoni A: Linearity and noise sources in flow cytometry. Cytometry 6:109–115, 1985.
23. Woo HW, Kim Y, Tompkins WJ: Development and applications of an interactive digital filter design program. Comput Methods Programs Biomed 21:11–22, 1985.
24. Young IT: Proof without prejudice: Use of the Kolmogorov-Smirnov test for the analysis of histograms from flow systems and other sources. J Histochem Cytochem 25:935–941, 1977.