
TECHNICAL NOTES

Ternary Representation of Trivariate Data¹

Peter M.A. Sloot² and Carl G. Figdor

The Netherlands Cancer Institute, Department of Biophysics, 1066 CX Amsterdam, The Netherlands

Received for publication February 26, 1988; accepted August 5, 1988

A new fast trivariate display technique based on the ratio of three measured variables and of relative cell number, obtained from flow cytometric measurements, is described.

Key terms: On-line display software; centrifugal elutriation; forward-, side-, and back-scattering

In recent years the detection equipment applied in flow cytometric (FCM) measurements has developed from relatively simple univariate techniques to highly advanced multivariate systems. As a result, direct qualitative identification of the various clusters in multidimensional variable space is required to interpret the FCM data. It was pointed out by Valet and others that one method of obtaining this information is to display a set of three variables simultaneously in a box (2,3,5,9). Other methods rely on complex clustering algorithms, e.g., the "ribbon method" (4), or on a combination of clustering algorithms and multivariate statistical analysis (8). However, when direct information concerning the memory content during a FCM experiment is required, three-dimensional (3D) box display techniques are generally applied.

The disadvantage of programs evaluating (contour drawn) 3D boxes is determined by the time-consuming programs on the one side, and the predetermined viewing point on the other hand. To exclude the possibility of hidden distributions, not visible in the isometric box, a combination of more than one viewpoint should be considered (there are six relevant permutations of the axes). To partly overcome these problems, we developed a new, very fast projection technique that provides additional and complementary information on the cell clusters studied.

COMPUTATION METHOD

To illustrate the principle of a ternary representation, an artificial data file is generated.

It is convenient to describe a 3D datafile as a mixture of trivariate Gaussian density distributions $p_i(\bar{r}|\phi_i)$ (8). It must be stressed, however, that the representation technique is not limited to this specific type of distribution.

$$p_i(\bar{r}|\phi_i) = \frac{1}{(2\pi)^{3/2} |\bar{\Sigma}_i|^{1/2}} \exp(-1/2 (\bar{\mu}_i - \bar{r}) \bar{\Sigma}_i^{-1} (\bar{\mu}_i - \bar{r})^T) \quad [1]$$

and $\bar{r} = (x, y, z) \in \mathbf{R}^3$. Here, $0 \leq x, y, z < 64$ represent, for instance, the channel numbers corresponding to the digitized detection signals (i.e., 6 bits per channel). The mixed distribution is parameterized by:

$$\phi_i = (\bar{\mu}_i, \bar{\Sigma}_i) \quad [2]$$

The vector $\bar{\mu}_i$ is the vector of the mean and the matrix $\bar{\Sigma}_i$ is the (co)variance matrix of the single distribution 'i':

$$\bar{\Sigma}_i = \begin{bmatrix} \sigma_{xx}^i & \sigma_{xy}^i & \sigma_{xz}^i \\ \sigma_{yx}^i & \sigma_{yy}^i & \sigma_{yz}^i \\ \sigma_{zx}^i & \sigma_{zy}^i & \sigma_{zz}^i \end{bmatrix} \quad [3]$$

σ_{pq}^i is the variance ($p = q$) or the covariance ($p \neq q$) term of the distribution.

The mixed trivariate data file $P(\bar{r}|\Phi)$, parameterized by

$$\Phi = (\alpha_0, \alpha_1, \dots, \alpha_{m-1}; \phi_0, \phi_1, \dots, \phi_{m-1}) \quad [4]$$

is described by

$$P(\bar{r}|\Phi) = \sum_{i=0}^{m-1} \alpha_i p_i(\bar{r}|\phi_i) \quad [5]$$

where the number and fraction of the component populations is represented by m and α_i , respectively. The parameter Φ completely describes a simulated trivariate Gaussian data file. Next, the new graphical representation is discussed and the properties of this projection

¹This research was supported by a grant from STW LGN 260353.

²Address reprint requests to P.M.A. Sloot, The Netherlands Cancer Institute, Division of Biophysics, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands.

technique are illustrated by means of a simulated data file (derived from Equation 5).

The ternary representation is a nonlinear projection from a three-dimensional variable space (x,y,z) to an equilateral triangle in two-dimensional (X,Y) space. The geometrical construction of this projection is shown in Figure 1. From an arbitrary point in the triangle, three lines, parallel to the sides of the triangle, are drawn. The length of these lines are denoted by Q_x , Q_y , and Q_z ($0 \leq Q_i \leq 1$; $i=x,y,z$). Q_x is defined by

$$Q_x = \frac{x}{x + y + z} \quad [6]$$

Q_y and Q_z are defined accordingly. Note that x , y , and z are the coordinates of a point in the original three-dimensional space. For instance, cells in a three-dimensional data file with channel number entries 14, 4, and 22 are mapped to a point P in the triangular space defined by the distances $Q_x = 0.35$, $Q_y = 0.1$, and $Q_z = 0.55$. In this way, each point in the (x,y,z) space is projected to a corresponding point in the triangle. Since we are interested in a computational scheme defining the complete projection to a two-dimensional (X,Y) plane, a general equation for a projected point P must be derived. From Figure 1, it is easily derived that the screen coordinates X and Y can be obtained from

$$X = 1 - 1/2 \cdot Q_z - Q_x \quad \text{and} \quad Y = Q_z \cdot \cos(30^\circ) \quad [7]$$

Equation 7 essentially describes (apart from a scaling factor) the complete production of a ternary plot from the coordinates $x,y,z \in \{0, 1, \dots, 63\}$, for a 6 bits resolution per variable. Finally, a threshold is defined that indicates the minimum number of cells that must be present for the computational scheme to evaluate a dot in the X,Y plane.

To illustrate the procedure, a simulated mixed trivariate distribution is calculated as an example, in accordance with Equations 1-5. The statistical variables used to produce this distribution are

$$\bar{\Sigma}_1 = \begin{bmatrix} 100 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}; \quad \bar{\mu}_1 = \begin{bmatrix} 40 \\ 20 \\ 20 \end{bmatrix}; \quad \alpha_1 = 2/3$$

$$\bar{\Sigma}_2 = \begin{bmatrix} 100 & 70 & 70 \\ 70 & 100 & 0 \\ 70 & 0 & 100 \end{bmatrix}; \quad \bar{\mu}_2 = \begin{bmatrix} 30 \\ 30 \\ 30 \end{bmatrix}; \quad \alpha_2 = 1/3$$

the result of which is shown in Figure 2. It can be demonstrated that the width of the distributions measured along the dashed lines P_x , P_y , and P_z indicates the ratio of the minors of the principal axes (data not shown). Distribution 1 in Figure 2, for instance, is elongated along P_x , indicating a significant broadening in the x direction in the original x,y,z domain. Accordingly, information on (co)variance and correlation between the three

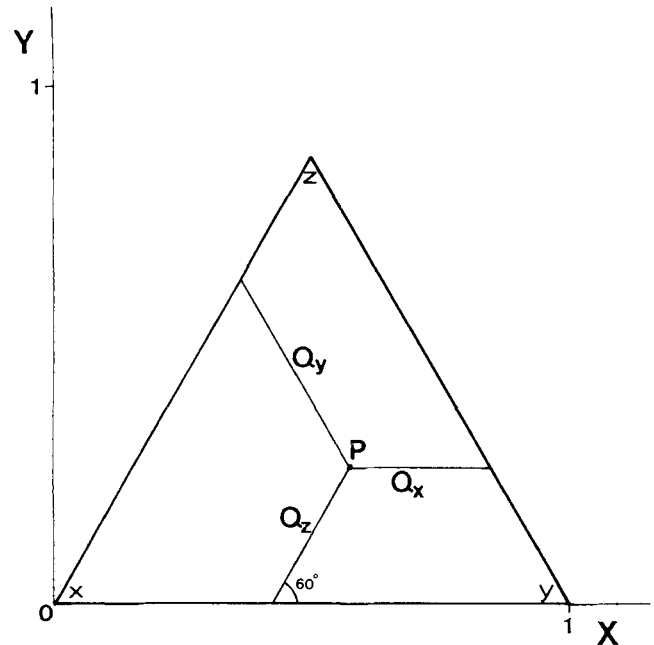


FIG. 1. Illustration of the ternary representation technique. The fractions in the original trivariate space are denoted by Q_x , Q_y , Q_z ; the domain axes, x , y , z , and the graphical axes, X , Y , are defined in the text.

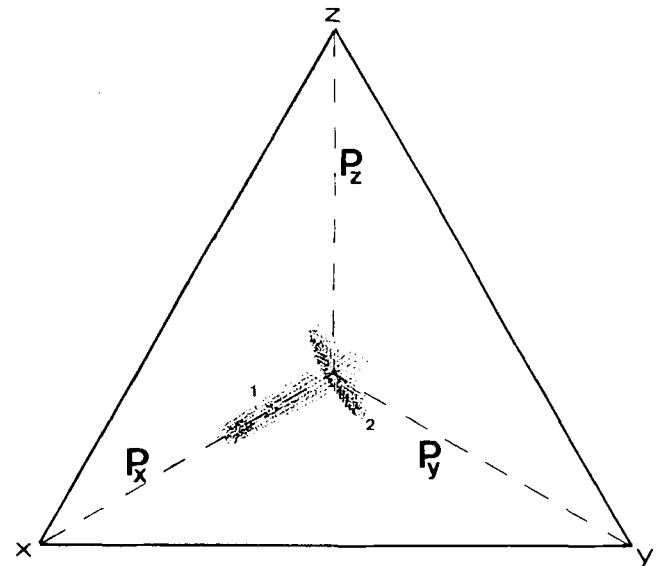


FIG. 2. Ternary representation of a simulated mixture of two trivariate normal distributions. The variables, x , y , z , and the directions, P_x, P_y, P_z , are explained in the text.

variables in x,y,z space is reflected by the width of the distributions along P_x , P_y , and P_z in triangular space.

APPLICATION OF TERNARY REPRESENTATIONS

In our laboratory we developed equipment to monitor on-line the centrifugal elutriation (CE) (1) of human

peripheral leukocytes, by means of a specially designed flow-cytometer and controlling system (7). The variables measured by this so-called computer-assisted centrifugal elutriation system (CACE, Patent applied for in the U.S.A.) are: forward-scattering (FS), side-scattering (SS), and back-scattering (BS) (6). The data obtainable from the trivariate light-scattering experiments can be displayed by means of the algorithm described above.

Figure 3 shows a ternary representation of a typical CE fraction. The fraction shown contains monocytes (distribution 1) and two populations of granulocytes, i.e., neutrophils (distribution 2) and eosinophils (distribution 3). In addition, Figure 4 shows a 3D box display of the same data, where the most favorable permutation of the three axes is applied.

DISCUSSION

In this study we describe the principle of an on-line graphical display technique which we call "ternary representation." The necessity of this complementary on-line representation technique is emphasized by means of a typical example. In contrast to Figure 4, Figure 3 shows unambiguously three completely separated distributions. Furthermore, the data in Figure 3 indicate *qualitatively* the relative intensities of the three scattering directions, as well as the scattering-angle dependent width (variance of the distributions). This is summarized in Table 1.

The computer programs to calculate the graphical representation are written in the language C on a micro-computer, based on the motorola MC 68000 CPU (Unicorn 10 Microproject, The Netherlands). After optimization of the display codes, the representation of

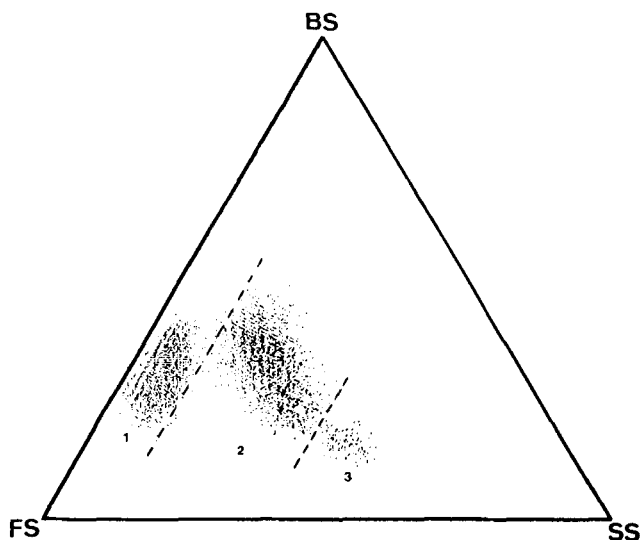


FIG. 3. Ternary representation of a mixture of human peripheral blood monocytes (distribution 1), neutrophilic and eosinophilic granulocytes (distributions 2 and 3, respectively), measured in a flow system. The three observables are forward-scattering (FS), side-scattering (SS), and back-scattering (BS).

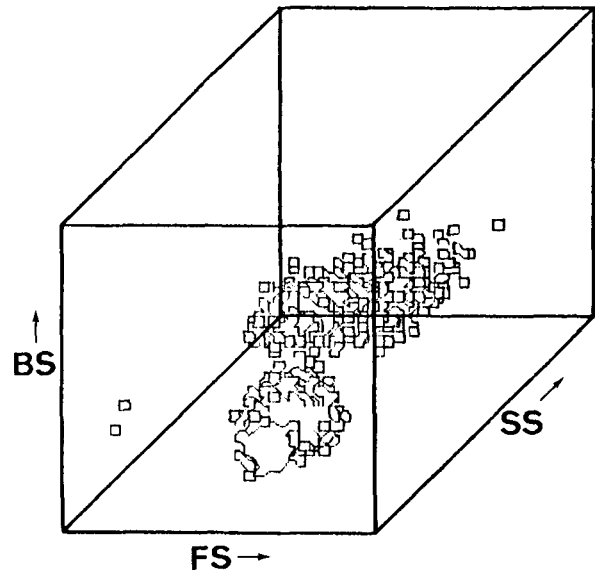


FIG. 4. A three-dimensional box-display of the blood cell mixture shown in Figure 3.

Table 1
Qualitative Interpretation of the Data Shown in Figure 3^a

Distribution 1 (Monocytes)	$FS \approx 2.0 \times BS \approx 8.0 \times SS$; $S_{FS} \approx S_{BS} > S_{SS}$
Distribution 2 (Neutrophils)	$FS \approx 1.5 \times BS \approx 2.0 \times SS$; $S_{FS} < S_{BS} < S_{SS}$
Distribution 3 (Eosinophils)	$FS \approx 2.5 \times BS \approx 0.8 \times SS$; $S_{FS} \approx S_{BS} \approx S_{SS}$

^aFS, SS, and BS are the relative scattering intensities for the forward-, side-, and back-scattering directions. S_i ($i = FS, SS, BS$) indicates the corresponding width of the distributions with respect to the three principal directions.

the ternary plot was obtained within seconds, whereas the 3D box plot required several minutes. This is mainly due to the fact that 3D box plots require a contour representation.

It is concluded that application of the fast ternary display technique for software processed trivariate graphical representations has several advantages compared to the conventional 3D box plot and is well suited for on-line purposes in FCM measurements. In our laboratory both on-line (real-time) and off-line ternary representations are applied routinely, for display of trivariate scattering patterns of human leukocytes (7,8).

LITERATURE CITED

- Figdor CG, Leemans JM, Bont WS, De Vries JE: Theory and practice of centrifugal elutriation. *Cell Biophys* 5:105-118, 1983.
- Kachel V, Meier H: Control unit for on-line handling of three-parameter flow cytometric data. In: *Flow Cytometry IV*, Laerum OD, Lindom T, Thorund E (eds). Universitetsforlaget, Bergen Oslo Tromsø, 1980, pp 120-124.

3. Kachel V, Schneider H: On-line three parameter data uptake, analysis, and display device for flow cytometry and other applications. *Cytometry* 7:25-40, 1986.
4. Murphy RF: Analysis of flow cytometric data using cluster analysis. *Cytometry* 6:302-309, 1985.
5. Ormerod MG, Payne AWR: Display of three-parametric data acquired by a flow cytometer. *Cytometry* 8:240-243, 1987.
6. Slood PMA, Figdor CG: Elastic light scattering from nucleated blood cells: Rapid numerical analysis. *Appl Optics* 25:3559-3565, 1986.
7. Slood PMA, Carels MJ, Tensen P, Figdor CG: Computer assisted centrifugal elutriation. I. Detection system and data acquisition equipment. *Computer Methods Programs Biomed* 24:179-188, 1987.
8. Slood PMA, van der Donk EML, Figdor CG: Computer assisted centrifugal elutriation. II. Multiparametric statistical analysis. *Computer Methods Programs Biomed* 27:37-46, 1988.
9. Valet G: Graphical representation of three-parameter flow cytometer histograms by a newly developed fortran IV computer program: In *Flow Cytometry IV*, Laerum OD, Lindmo T, Thorund E (eds). Universitetsforlaget, Bergen Oslo Tromsøe, 1980, pp 125-129.