



HAL
open science

Generalized unmixing model for multispectral flow cytometry utilizing nonsquare compensation matrices

David Novo, Gérald Grégori, Bartek Rajwa

► **To cite this version:**

David Novo, Gérald Grégori, Bartek Rajwa. Generalized unmixing model for multispectral flow cytometry utilizing nonsquare compensation matrices. *Cytometry Part A*, 2013, 83A (5), pp.508-520. 10.1002/cyto.a.22272 . hal-00807657

HAL Id: hal-00807657

<https://hal.science/hal-00807657>

Submitted on 20 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Published in final edited form as:

Cytometry A. 2013 May ; 83(5): 508–520. doi:10.1002/cyto.a.22272.

Generalized Unmixing Model for Multispectral Flow Cytometry Utilizing Nonsquare Compensation Matrices

David Novo¹, Gérald Grégori², and Bartek Rajwa^{3,*}

¹De Novo Software, 3250 Wilshire Blvd. Suite 803 Los Angeles, CA 90010

²Mediterranean Institute of Oceanography, Aix-Marseille Université, UMR CNRS 7294/IRD 235, France

³Bindley Bioscience Center, Purdue University, 1203 W. State Street, West Lafayette, Indiana 47907

Abstract

Multispectral and hyperspectral flow cytometry (FC) instruments allow measurement of fluorescence or Raman spectra from single cells in flow. As with conventional FC, spectral overlap results in the measured signal in any given detector being a mixture of signals from multiple labels present in the analyzed cells. In contrast to traditional polychromatic FC, these devices utilize a number of detectors (or channels in multispectral detector arrays) that is larger than the number of labels, and no particular detector is a priori dedicated to the measurement of any particular label. This data-acquisition modality requires a rigorous study and understanding of signal formation as well as unmixing procedures that are employed to estimate labels abundance. The simplest extension of the traditional compensation procedure to multispectral data sets is equivalent to an ordinary least-square (LS) solution for estimating abundance of labels in individual cells. This process is identical to the technique employed for unmixing spectral data in various imaging fields. The present study shows that multispectral FC data violate key assumptions of the LS process, and use of the LS method may lead to unmixing artifacts, such as population distortion (spreading) and the presence of negative values in biomarker abundances. Various alternative unmixing techniques were investigated, including relative-error minimization and variance-stabilization transformations. The most promising results were obtained by performing unmixing using Poisson regression with an identity-link function within a generalized linear model framework. This formulation accounts for the presence of Poisson noise in the model of signal formation and subsequently leads to superior unmixing results, particularly for dim fluorescent populations. The proposed Poisson unmixing technique is demonstrated using simulated 8-channel, 2-fluorochrome data and real 32-channel, 6-fluorochrome data. The quality of unmixing is assessed by computing absolute and relative errors, as well as by calculating the symmetrized Kullback–Leibler divergence between known and approximated populations. These results are applicable to any flow-based system with more detectors than labels where Poisson

noise is the dominant contributor to the overall system noise and highlight the fact that explicit incorporation of appropriate noise models is the key to accurately estimating the true label abundance on the cells.

Keywords

spectral unmixing; compensation; Poisson regression; multispectral and hyperspectral flow cytometry

Introduction

Classes of bioparticles are often defined by the type and quantity of biomarkers present in each analyzed particle. Flow cytometry (FC) typically quantifies the presence of these biomarkers by tagging them with fluorescent molecules. However, the raw FC measurements do not directly yield the biomarker quantity or label concentration; instead, they provide values that are proportional to the number of photons measured by the individual photodetectors.

The optical pathway of FC instruments is arranged in an attempt to separate signals from different fluorochromes by routing them into dedicated detectors; however, owing to spectral overlap and imperfect filters, a complete separation is almost never possible. Therefore, the fluorescence emitted by every fluorochrome may be simultaneously collected by more than one detector (in extreme cases, all the detectors). This process can be mathematically represented as a linear mixing of signals and is a subject of study in various fields of science ranging from chemometrics to imaging and remote sensing (1–6).

Let \mathbf{r} denote the vector of observations of length L (the number of detectors employed in the FC system), \mathbf{M} an $L \times p$ spectral-signature matrix (p being the number of labels used in an experiment), \mathbf{a} the vector of length p of abundances in which a_i represents abundance (amount) of the i th label in the measured object, and \mathbf{e} a vector of length L which denotes noise. Therefore, the phenomenon of “spectral spillover” that leads to signal mixing may be represented using a basic linear spectral mixture equation:

$$\mathbf{r} = \mathbf{M}\mathbf{a} + \mathbf{e}. \quad (1)$$

The linear-mixture model assumes that multiple signals measured from every particle can be expressed as a linear combination of spectral signatures with appropriate abundances a_1, a_2, \dots, a_n . The cytometry literature usually refers to these values (however incorrectly) as “compensated fluorescence.”

In traditional polychromatic FC, the number of detectors employed is equal to the number of labeled markers; thus, in order to find the abundances (or values linearly correlated with abundances), the unmixing operation can readily be performed by multiplying the measured data vectors (or raw fluorescence observations) by the inverse of the spectral-signature matrix (also called the mixing matrix):

$$\hat{\alpha} = \mathbf{rM}^{-1}, \quad (2)$$

where $\hat{\alpha}$ is the unmixed approximation of α . Although the mixing matrices are a priori unknown, they can be easily approximated by employing single-stained controls and normalizing the resultant spectra. This process leading to the recovery of abundances is known as FC compensation and is described extensively in the FC literature (7,8).

However, the number of detectors employed in an FC experiment does not have to be limited to the number of fluorochromes and may be significantly larger. This type of optical arrangement is characteristic of an emerging class of spectral FC systems, which attempt to measure an approximation of the full spectrum emitted by every analyzed bioparticle. The measurements produced by a spectral system may represent fluorescence, Raman, or surface-enhanced Raman scattering characteristics (9–12).

An attempt to recover abundances from spectral measurements leads to a mixing model with matrices that are not square, resulting in an overdetermined system of equations. This is seemingly a trivial problem, as the standard compensation approach can easily be extended by using the pseudoinverse of an overdetermined mixing matrix in a process known as ordinary least-square (OLS) minimization.

Although overdetermined unmixing is a new issue for FC analysis, it is often used in various imaging techniques ranging from microscopy to remote sensing (2,3). These techniques usually rely on OLS to find the optimal vector of abundances. However, the OLS method is valid only if the noise in Eq. (1) is Gaussian and has equal variance irrespective of the signal level. Therefore, it is legitimate to inquire whether this widely accepted approach is appropriate for spectral FC and other techniques based on fluorescence.

In this report, we will demonstrate that, owing to the physics of signal formation in cytometry, the OLS solution is biased and does not provide a correct estimation of abundances for spectral FC systems. Therefore, it should not be employed for fluorescence-, Raman-, or surface-enhanced Raman scattering-based cytometry. We will also propose and discuss alternative approaches: an approximation based on minimization of percentage error using weighted least squares (WLS), a technique explicitly addressing the distribution of the fluorescence signal and employing a generalized linear model (GLM), and a simplified solution using a variance-stabilization transformation commonly employed in image denoising.

The reported data represent simulations and real multispectral measurements obtained using a 32-channel experimental system designed at Purdue University (12). The goal of the simulations is to demonstrate the known and proposed approaches in a simple and straightforward fashion without reference to any particular biological application. This simulation also allows us to validate unmixing algorithms by comparing abundances known a priori to the estimated values after unmixing.

In the case of the real experimental data, we are able to compare the unmixed abundances to abundances obtained by measuring control samples. Additionally, the changes in

distribution of estimated intensities introduced by different unmixing methodologies demonstrate their impact on the estimation of fluorochrome concentration and on the relative position of biological populations in the feature space.

Materials and Methods

Multispectral Measurement of Human Lymphocytes

Multispectral detection system—The example multispectral data were collected using a prototype of a multispectral 32-channel fluorescence system that has already been extensively described elsewhere (12). Briefly, the fluorescence emission is collected at 90° to the laser direction using an A10766 compact spectrometer (Hamamatsu, Japan). The spectrometer unit includes a polychromator that according to the manufacturer's data has a grating groove density of 600 g/mm, a spectral range from 200 to 900 nm, a focal length of approximately 100 mm, and an F value of 3.3. The dispersed signal is projected onto a Hamamatsu 7260-01 32-channel multianode PMT linear array photodetector (Hamamatsu). The detailed specifications of the linear-array PMT employed were provided in the Supporting Information of a recent report published in *Cytometry Part A* (12).

Flow cytometer—The FC fluidics used for the multispectral detection was based on a customized FC500 flow cytometer (Beckman Coulter, Miami, FL). The system was equipped with two air-cooled lasers: a uniphase argon ion, 488 nm, 20 mW output, and a Coherent red solid-state diode, 635 nm, 25 mW.

The flow cell consists of a 150- × 450- μ m rectangular-channel BioSense enhanced quartz optics, mounted with vertical (upward) flow path. The liquid sheath was distilled water filtered through a 0.2- μ m filter. The pressure applied to the sheath tank was kept constant at 28 psi (~191 kPa). The lymphocyte concentrations in the various samples analyzed were in the range of 1,000–3,500/ μ L. The sample flow rate was maintained at approximately 1,000 events per second. In order to handle the modified forward-scatter detection subsystem and the 32 channels of fluorescence recorded using the multi-anode PMT device, the data-control electronics of the FC500 instrument was upgraded to the system designed for the Beckman-Coulter Gallios flow cytometer. The Beckman-Coulter electronics was also used to control the flow cytometer fluidics.

Flow cytometry data acquisition—Control of the modified FC500 acquisition (start and stop acquisition, clean, rinse, etc.) as well as light-scatter data collection was performed by the CXP software package (Beckman Coulter). Acquisition of the fluorescence data was implemented using the custom-built *Cytospec* package developed by Valery Patsekina and J. Paul Robinson (Purdue University), which recorded all values simultaneously for each single particle analyzed. All the output files were saved into a custom binary format and subsequently converted into comma-separated values (.csv).

Monoclonal antibodies—CD45-FITC, CD4-PE, CD8-ECD, and CD3-Cy5 human monoclonal antibodies (catalog number 6607013) were obtained from Beckman Coulter. CD19-PECy7-labeled human antibody (catalog number 25-0199) was obtained from eBioscience (San Diego, CA).

Blood collection and sample preparation—Venous blood was collected under human-use protocol 0506002740 by a standard venipuncture procedure using 7-mL EDTA Venoject tubes. A 100- μ L aliquot of whole blood was taken from the venous sample, mixed with 10 μ L antibody solution, and incubated for 10 min at room temperature. All samples were prepped on a standard Q-prep using the 35-second cycle and the ImmunoPrep reagent system (Beckman Coulter).

Simulations and data analysis—All data processing, as well as the simulations, was performed using R language for statistical computing (13). The simulated FC data sets were generated using a hierarchical stochastic process. In the first step, a “true abundance” for each fluorochrome was simulated for each cell using a random-number generator (RNG) sampling from either normal, truncated-normal, or log-normal distributions with defined means and coefficients of variation (CVs). Cell- by-cell photon noise due to the stochastic nature of photon emission was then simulated using a Poisson RNG with mean parameter equal to the simulated abundance. This resulted in a photon emission vector (PEV) of integer values of length p describing the number of emitted photons for each label. The PEV was then mixed by multiplying it by the multispectral spillover matrix, which was normalized columnwise to 1. This resulted in a mixed-PEV (MPEV) of values of length L . In the third step, simulating the generation of photoelectrons and measurement of the detected signals [detection vector (DV)], a gamma RNG was used to produce a vector of random numbers distributed according the gamma distribution, with shape parameter equal to the corresponding value in the MPEV plus one, and scale parameter equal to one. The utilized gamma distribution was selected as a generalization of the Poisson distribution for real numbers (see Supporting Information Materials for details).

All simulations were conducted using an eight-anode PMT array. Therefore, the spillover matrix contained eight-point approximations of spectra of two simulated fluorochromes (see Supporting Information Fig. S1). The spectra were assumed to be Gaussians with varying standard deviations. The example demonstrated in this report used spectra with full width at half maximum approximately equal to 2.76 detector channels ($\sigma = 1.175$), and maxima at detector channels 3 and 6, respectively.

It is important to note that the simulations were performed assuming idealized conditions; therefore, no electronic noise of the FC instrumentation was simulated. As can be seen by inspecting Eq. (1), mixing (and hence unmixing) occurs on an individual cell basis, and the population from which the cells (intensity vectors \mathbf{r}) arise does not affect the unmixing process. Thus, the techniques described in this article are relevant to a broad range of sample types.

Results

The Gaussian Model of Spectral Unmixing

Figure 1 shows a simulation of signal distribution obtained from three cell populations generated without applying a mixing matrix (ground truth). Therefore, the plots represent the DV values (abundance with noise present) for $\mathbf{M} = \mathbf{1}$ (identity matrix). The same cell populations were used to generate DV outputs from the eight-anode PMT (see Supporting

Information Fig. 1 for the mixing matrix \mathbf{M}). These simulated PMT outputs were subsequently unmixed according to the various algorithms described below.

The basic spectral-mixing model as expressed by Eq. (1) is nonidentifiable, meaning that there is no unique solution unless additional information, particularly the noise model, is specified (5). In remote sensing, it is common to state explicitly that \mathbf{e} represents additive Gaussian noise with an expected value of zero and covariance matrix $\sigma^2\mathbf{1}$. Following these assumptions, in the case for which the number of detectors is larger than the number of labels, spectral unmixing can be performed by solving an LS problem (2,14):

$$\min_{\alpha \in \mathbb{R}} \{(\mathbf{r} - \mathbf{M}\alpha)^T (\mathbf{r} - \mathbf{M}\alpha)\} \quad (3)$$

Assuming no additional constraints, the OLS approximation value of α can be obtained by the closed-form equation

$$\hat{\alpha} = (\mathbf{M}^T \mathbf{M})^{-1} \mathbf{M}^T \mathbf{r}. \quad (4)$$

Unmixing the simulated data using OLS results in some cells with negative abundances, and the estimated low-intensity populations have a distorted, characteristically “spread” shape compared to the true α (compare Figs. 1 and 2A). Not only were the low-intensity populations pushed toward zero, they were affected by this bias more than the high-intensity population (double positive) as well. The distortion of the low-intensity population is illustrated in Figure 3B and measured by symmetrized Kullback-Leibler divergence (SKLD) between the distribution of the true abundances and the distribution recovered by OLS (15). Since the negative abundance values have no physical interpretation, the result obtained using OLS is obviously problematic. If the vector $\hat{\alpha}$ is assumed to be proportional to biomarker abundances, the negative results would suggest negative photon emission, negative concentration, and consequently negative quantity of biomarkers. This result violates the physical constraints of the system described by Eq. (1). When OLS was applied to mixed data in which the noise arising from emission and detection processes was not present, the recovered abundances did not exhibit any of the spreading seen in Figure 2A (data not shown).

Unmixing of the experimental data set yields similar results, with the problem especially acute for weak signals. Large portions of abundances representing autofluorescence (AF) are pushed below zero (Fig. 5A). Again, this result violates the physical constraints and cannot possibly be correct. The unstained controls show measurable and obviously positive values of AF intensity, contradicting the computed estimation (Fig. 6).

We will next apply more sophisticated unmixing approaches traditionally employed in the field of imaging in an attempt to obtain $\hat{\alpha}$ vectors that do not violate the physical constraints of the model.

The Non-Negative Least-Squares Unmixing

In order to avoid the problem of obtaining negative abundances after unmixing, many imaging applications explicitly employ a non-negativity constraint in the model. An additional constraint is also included, stating that the unmixed abundances must sum to 100% of the mixed input signal (per the law of conservation of energy). The constrained formulation of the problem leads to the following model:

$$\min_{\alpha \in \Delta} \left\{ (\mathbf{r} - \mathbf{M}\alpha)^T (\mathbf{r} - \mathbf{M}\alpha) \right\} \quad \text{s.t.} \quad \Delta = \left\{ \alpha \mid \left(\sum_{i=1}^P \alpha_i = \sum_{i=1}^P r_i \right) \wedge (\alpha_i \geq 0) \right\}. \quad (5)$$

where α_i and r_i are the elements of α and \mathbf{r} vectors.

Unlike the unconstrained model in Eq. (3), Eq. (5) does not have a closed-form solution. Therefore, the vectors α must be found numerically, for instance, by employing the traditional Lawson-Hanson algorithm or a newer approach by Bro and De Jong (16,17).

As we can see in Figures 2B and 5B, negative abundance values were indeed eliminated, and the results returned are physically feasible. However, the populations recovered by non-negative least-squares (NNLS) seem to be “clipped,” and the data points (events) pile up on the axes. In addition, the artificial spreading of the dim populations demonstrates that this result is not a good approximation of the true populations seen in Figure 1. Even though the solution provided by Eq. (5) is not the same as the ad hoc approach of simply solving Eq. (3) and setting the negative results to zero, it yields results quite similar to that method.

The Alternative Poisson Noise Model

A detailed understanding of the assumptions employed in the formulation of the linear mixing model is the key to obtaining more accurate unmixing results. As stated above, the LS formulation proposed in Eqs. (3) and (5) explicitly assumes that the errors are represented by additive white Gaussian noise with zero mean and a variance that does not change with the signal intensity (homoskedasticity). However, since FC observation of single cells in flow involves detection of photons emitted by the fluorescent molecules on the surface of or inside the cells, the errors cannot be distributed this way. The noise variance is not identical for every measured cell, and it is not identical for high- and low-intensity signals (18,19).

This is a consequence of the fact that the process of photon emission and detection involves Poisson processes. Photons are emitted by fluorochromes at random time intervals and the distribution of their arrival at the photocathode is closely approximated by a Poisson distribution (20). However, even if for the purpose of our model an assumption is made that the variance in the photon emission is zero and that the photons arriving at the photocathode are equally spaced in time, the number of emitted photoelectrons is not constant, as the probability of photoelectron emission is also governed by a Poisson process (20–22).

The final measured signal is proportional to the number of photoelectrons generated on the last photocathode. In the idealized case in which no additional noise sources are present and the detector offers 100% efficiency, the model of FC observation could be expressed as

$$\mathbf{r} \sim \text{Poisson}(\mathbf{M}\boldsymbol{\alpha}). \quad (6)$$

The important consequence of the model shown in Eq. (6) is heteroskedasticity of the data, that is, the fact that the expected variance increases with the increased signal, in contrast to the stable variance assumptions described previously. Therefore, unmixing approaches that specifically incorporate the Poisson nature of signal and noise should arrive at better results for FC data, since the unmixing model then more accurately reflects the underlying physics of signal formation.

Percentage Error Estimation via WLS

A common approach, owing to its mathematical simplicity, is to continue with the assumption that the observations include a normally distributed noise component with a variance that grows with the signal intensity. Essentially, it mimics the Poisson behavior of signal using Gaussians and results in the fact that measurements with lower variance have proportionally more influence on the estimate of abundances than they would under LS (23).

These requirements can be met by utilizing a special case of WLS and minimizing the mean absolute percentage error (MAPE) instead of the squared error as with OLS. Therefore, the percentage error may be defined as (observed value – predicted value)/observed, which we will express as

$$E_p = \mathbf{j}^T |(\mathbf{r} - \mathbf{M}\boldsymbol{\alpha}) \circ \bar{\mathbf{r}}|, \quad (7)$$

where \mathbf{j} is an $L \times 1$ sum vector of 1, and \mathbf{j}^T is its transpose (the sum vector is used to find the sum of the elements of the computed vector), $\boldsymbol{\alpha}$ is a vector of abundances and \mathbf{r} is the actual FC measurement of a cell, $\bar{\mathbf{r}}$ is a vector of reciprocal values to the elements of vector \mathbf{r} (Hadamard inversion), and \circ denotes element-wise multiplication (Hadamard product).

The closed-form solution to Eq. (7) allows us to compute $\hat{\boldsymbol{\alpha}}$ directly (see Supporting Information Materials for details):

$$\hat{\boldsymbol{\alpha}} = (\mathbf{M}^T \mathbf{W}^2 \mathbf{M})^{-1} \mathbf{M}^T \mathbf{W}^2 \mathbf{r}. \quad (8)$$

It is easy to observe that Eq. (3) is a special case of Eq. (8) in which $\mathbf{W} = \mathbf{1}$ (identity matrix). As intended, the weights are inversely proportional to the signal, providing us with a simple solution that recognizes the increase of variance (uncertainty) with the increase of signal.

The results shown in Figures 2C and 3C demonstrate that the MAPE approach improved the abundance estimations for the simulated data. However, the cellular data seem to be distorted and some populations were pushed below zero (Fig. 5C). Owing to the presence of weights, the observed signals cannot contain zeros. Therefore, the vectors \mathbf{r} must be shifted by a small value before unmixing. Since the shift impacts the shape and location of the unmixed abundance distribution, we choose the shift value empirically in order to minimize SKLD between the controls and the unmixed distributions.

GLM Formulation of the Unmixing Process

The OLS regression can be considered as a specific case of a more general theory of regression mathematics known as GLMs. The theory of GLM is well described in Refs. 24 and 25, and a detailed discussion of this topic is beyond the scope of this article. The GLM approach fits the data by maximizing the log-likelihood and can be used with response variables that have distributions other than Gaussian and/or are not homoskedastic. Thus, the GLM approach allows solution of Eq. (1) for cases in which \mathbf{e} is not normally distributed, as shown in Eq. (6).

In the idealized case, the distribution of noise \mathbf{e} can be approximated by a Poisson distribution. However, the detectors used in FC instruments do not report photon counts directly but convert light into analog electronic signals (even though this information is subsequently digitized). Therefore, the raw detector output is better represented by real rather than natural numbers. Faithful modeling of true continuous distributions of analog signals produced by a PMT is a difficult topic beyond the scope of this article, as the Poisson model ceases to be appropriate if the secondary emission statistics is taken under careful consideration (22). However, if we assume a noiseless and uniform secondary emission process in which gain does not vary for different photoelectrons, we may approximate the FC data using just a simple continuous generalization of a Poisson distribution, which expresses it as a special case of the well-understood gamma distribution (see Supporting Information Materials for details).

Incidentally, this framework allows us to use a gamma RNG to produce real (floating point) rather than integer-based data directly during *in silico* FC measurements, as mentioned in the Materials and Methods.

Furthermore, it can be demonstrated that the log-likelihood function for this specific gamma distribution is the same as that for the Poisson distribution (see Supporting Information Materials). Therefore, in order to find the solution to Eq. (3), we followed the approach to the identity-link Poisson regression suggested by Venables and Ripley (26) and minimized the deviance function \mathcal{D} that assesses the goodness of fit by comparing the log-likelihood under the saturated model (i.e., the model in which the number of parameters is equal to the number of observations) to the log-likelihood under the proposed Poisson model:

$$\mathcal{D} = 2\mathbf{j}^T \left(\mathbf{r} \circ \log \left(\mathbf{r} \circ \overline{\mathbf{M}\boldsymbol{\alpha}} \right) - (\mathbf{r} - \mathbf{M}\boldsymbol{\alpha}) \right), \quad (9)$$

The $\log(\cdot)$ notation denotes element-wise logarithm. $\overline{\mathbf{M}\boldsymbol{\alpha}}$ is a vector of reciprocal values to the elements of vector $\mathbf{M}\boldsymbol{\alpha}$ (Hadamard inversion) and \circ denotes element-wise multiplication (Hadamard product).

In contrast to the OLS approach, the minimization of deviance in the Poisson regression problem has no general closed-form solution. Therefore, the vector $\hat{\boldsymbol{\alpha}}$ is found using optimization methods. Furthermore, a sum-to-one equivalent constraint can be added as a soft penalty, providing the complete unmixing model:

$$\hat{\alpha} = \arg \min_{\alpha} \left\{ 2\mathbf{j}^T (\mathbf{r} \circ \log(\mathbf{r} \circ \overline{\mathbf{M}}\alpha)) - (\mathbf{r} - \mathbf{M}\alpha) + \lambda (\|\mathbf{r}\|_1 - \|\alpha\|_1) \right\} \text{ s.t. } \mathbf{M}\alpha > \mathbf{0}. \quad (10)$$

The additional penalty parameter λ in Eq. (10) allows us to control the level of certainty in the model. This parameter can be set to 0 or to some very low value if the accuracy (or completeness) of \mathbf{M} is suspect. In other words, in the experimental setting in which not all the fluorochromes present are known, we cannot expect that the entire signal is unmixed utilizing only the spectra describing the known fluorochromes.

Simulated data unmixed using Poisson GLM is shown in Figure 2D. It is evident that the clusters look very similar to the distribution of true abundances shown in Figure 1.

In order to quantitatively assess the ability of the various unmixing algorithms to recover the original simulated abundances, we calculated the root mean square error (RMSE) and mean normalized error (MNE) to measure the differences between the simulated true abundances (Fig. 1) and the unmixed estimations (Fig. 2). The RMSE values were similar for OLS and Poisson unmixing (Figs. 4A–4C). However, the MNE values improved when Poisson unmixing was used, demonstrating that the relative error can be minimized without significantly affecting the absolute values (Figs. 4D–4F). The SKLD between the Poisson-unmixed estimation of the negative populations and the true abundance is minimal (Fig. 3D), as both distributions almost completely overlap (Fig. 3E).

The effect of unmixing under experimental conditions was assessed by examining the AF profile of unstained controls. We compared it using SKLD to the estimated AF determined by the various unmixing algorithms (Fig. 6). The SKLD measure indicated that the AF recovered using Poisson unmixing was the most similar to the actual AF (Fig. 6D). The Poisson algorithm, which properly unmixed the low abundance (dim) signal, did not have a significant effect on the abundance estimation of the bright CD45 signal (Fig. 7).

Unmixing via Variance-Stabilizing Transformation

The field of imaging often uses variance stabilization as a first step in image denoising operations. Therefore, it is interesting to establish whether unmixing of measurements that are known to be Poisson distributed can be performed using an OLS minimization following the transformation of the mixing model into something approximately Gaussian. The transformation proposed by Bar-Lev and Enis, or Anscombe and Freeman–Tukey transformations belonging to a wider class of variance-stabilization functions described by Bar-Lev and Enis, can be used for this purpose (27,28).

The Bar-Lev/Enis transformation (BET) is defined as

$$\begin{aligned} \mathcal{A}_{a,b}(x) &= (x+2a-b)(x+a)^{-\frac{1}{2}}, \\ \mathcal{A}_{a,b,c}(x) &= \mathcal{A}_{a,b}(x) + (x+c)^{-\frac{1}{2}}. \end{aligned}$$

The transformation has been shown to exhibit optimal variance-stabilizing performance for a Poisson distribution for

$$a = \frac{3}{8} + \frac{3^{-\frac{1}{2}}}{2}, \quad b = \frac{3}{8} + \frac{3^{\frac{1}{2}}}{4}.$$

Therefore, the compensation process can be expressed as an LS minimization of transformed values:

$$\hat{\alpha} = \arg \min_{\alpha} \|\mathcal{A}_{a,b}(\mathbf{r}) - \mathcal{A}_{a,b}(\mathbf{M}\alpha)\|_2^2 \quad \text{s.t.} \quad \alpha \geq \mathbf{0}.$$

The results demonstrate that the simulated data set was correctly compensated following the mixing model correction (Supporting Information Fig. S3). However, the real 32-channel data unmixing did not provide a major improvement over NNLS results (Supporting Information Fig. S4).

Discussion

The presented work proposes a new approach to the problem of unmixing in multispectral FC. We extend the well-established concept of unmixing as used in other fields such as remote sensing, spectral imaging, and chemometrics and modify it for use with cytometry systems that utilize a number of detectors larger than the number of labels (9–12).

First, we used simulations to evaluate various unmixing algorithms. This allowed us to compare the unmixed abundances with the known input values that were used for the simulations. Understandably, each unmixing algorithm is able to recover the simulated abundance in the absence of noise, as in that case the linear mixing process is a completely reversible operation [see Eq. (1) with $\mathbf{e} = 0$]. However, once realistic Poisson noise was introduced, the unmixing algorithms yielded different results (Fig. 2). This key observation, that noise alone is sufficient to reproduce the “spreading” artifacts that are widely seen with traditional compensation (29), led to the hypothesis that a more refined treatment of the noise in the unmixing model should result in improved recovery of the actual abundances.

The abundances estimated by the OLS and NNLS algorithms were the most dissimilar to the simulated input data (Figs. 2A and 2B and 3). The reason is that both algorithms are based on the implicit assumption that the data are homo-skedastic (2). The problem of unmixing is mathematically equivalent to multiple regression, and these two algorithms arrive at a solution by minimizing the ℓ^2 norm between the actual observation and the regression result. However, the simulated values are contaminated with Poisson noise, resulting in a variance that depends on the magnitude of the observed intensities (therefore, the data are heteroskedastic). The optimal way of minimizing the ℓ^2 norm is to fit the high values as closely as possible, while the precise estimate of the low (dim) values remains relatively unimportant since their absolute contribution to the ℓ^2 norm is negligible. This results in highly unrealistic estimates of low-abundance populations. Similar results were obtained for cellular data acquired with a spectral flow cytometer (Fig. 5). The data unmixed with OLS showed large numbers of events with negative abundance values, particularly for dim signals (as in the case of AF) (Figs. 5A and 6). Even after unmixing, some populations

demonstrated artifactual correlations (“diagonal” orientation), which are often considered an indication of “undercompensated” data. Data unmixed with NNLS (Fig. 5B) naturally did not exhibit negative values, but very dim populations seemed abruptly truncated at zero. These results demonstrate that LS minimization is not appropriate for overdetermined FC data with Poisson noise present.

It is important to appreciate the relevance of negative values from the observed FC signals as compared with the unmixed label abundances. Since the FC measurements are usually not performed on calibrated scales or on an absolute scale of true photon counts, the observations from FC instruments may indeed span from negative to positive values. These negative values do not necessarily signify the presence of any particular noise type. They can also be safely rescaled before being used in any unmixing model. However, a valid unmixing model should avoid negative values for the unmixed label abundances, which should be proportional to actual photon emission, hence positive. However, as the *in silico* experiment demonstrated, even if all the observed values from the FC instrument are strictly positive the resultant unmixing may produce negative values for the label abundances if heteroskedasticity is not properly addressed. In the context of a linear mixture with mixing matrix columnwise normalized to one [as in Eq. (1)], these negative abundances have no physical interpretation and indicate an inappropriate unmixing model.

Based on these observations, we hypothesized that minimization algorithms that weighted the magnitude of the error in relation to the size of the observed signal should result in superior performance. The first tested alternative involved MAPE unmixing, which indeed provided a significant improvement over OLS in the simulated data set (Fig. 2C). The MAPE algorithm “stabilizes” the variance by effectively weighting each observation by the square root of the observation. This minimizes the relative instead of the absolute error. Since MAPE can be performed using a closed-form solution, it might be recommended as a quick and computationally inexpensive strategy to cope with the bias of the OLS method. However, MAPE requires that all observations be >0 . In theory, if measured on an absolute (or relative but calibrated) scale all measurements from an FC instrument should be positive since even in the absence of a fluorescence signal the observation should include additive noise (owing to stray light, intrinsic dark-current noise, offset signal, etc). Yet, in practice, the tested multispectral data sets contain a number of zero values owing to the discriminator threshold level. The common ad hoc solution of shifting these values from zero to some small constant for the calculation of the weights may distort the relationship between “bright” populations, as shown in Figure 5C.

The more sophisticated alternative treatments of zero were developed in the field of compositional analysis (30). There are also alternative formulations of percentage error that are not constrained to having observations >0 ; however, these formulations do not have a closed-form solution and require an iterative algorithm for minimization (25). We did not explore this further; instead, we reasoned that, if an iterative approach was required, it was preferable to use the more rigorous technique discussed further below.

The second tested alternative involved using our a priori knowledge of the physics of fluorescence signal formation and finding appropriate solutions utilizing our understanding

of the noise characteristics of the system. The proposed model approximates the process of fluorescence signal detection using a PMT employing a special case of gamma distribution that extends the Poisson distribution to the continuous domain. This model not only allowed us to offer GLM-based signal estimation but also was directly applicable to *in silico* experiments and simulations, surpassing the limitations of a traditional discrete Poisson model.

An FC signal is primarily the result of a cascade of random processes resulting in an overdispersed Poisson distribution (21). However, the output from a PMT is an analog signal (as opposed to integer values, which would be obtained using the detectors in photon-counting mode) (31). Thus, *in silico* simulations of signal formation, as well as unmixing equations, must take this fact into consideration. Simplistic simulations utilizing Poisson RNGs would produce only integer values for the number of photons emitted by the fluorochromes and for the photoelectrons generated on each photo-cathode. Spectral mixing would indeed generate real numbers but only as an artifact of matrix multiplication. Yet, these values would be subsequently truncated to integers to simulate photon detection at the photocathode via another Poisson process. The application of a gamma function described in the Supporting Information Materials to model the continuous yet Poisson nature of detected signals solves this problem and justifies the use of Poisson GLM as the basis for unmixing algorithms.

The explicit application of Poisson regression led to superior results for both simulated and real data (Figs. 2D, 3, 5D, 6, and 7). Since we know the input values in the case of the simulated data, the simplest direct metrics of unmixing quality are RMSE and MNE, showing the difference between simulated input data and the unmixed estimations. For the reported *in silico* experiments, the RMSE values are similar for OLS and Poisson unmixing, as expected (Figs. 4A–4C). However, the MNE values indeed improved when a Poisson model (allowing for heteroskedasticity), rather than a homoskedastic Gaussian, was applied, demonstrating that a relative error can be minimized without significantly affecting the absolute values (Figs. 4D–4F). The visual examination of the scatter plots alone reveals the large impact of the proper noise model. The low-abundance populations are not spread below zero, and the variance within populations is not artificially extended (Fig. 2D). The high abundance (double-positive) population produced after unmixing looks almost indistinguishable from the one obtained using OLS. This is because the OLS and Poisson unmixing return virtually identical results when all the values in the vector \mathbf{r} are similar (resulting in homogeneity of variance).

The unmixing of real 32-channel data was also improved when a Poisson model was used (Fig. 5D). As with the simulations, the effect of unmixing on low-intensity populations demonstrates the dramatic impact of the proper noise model. The OLS and NNLS algorithms resulted in a completely distorted AF distribution compared to the known control (Fig. 6). The Poisson-estimated AF abundance does not contain negative values and its distribution is most similar to the controls, as shown by SKLD (Fig. 6D). Although the application of the Poisson model does not lead to perfect results owing to a slight positive bias, it produces a good-quality approximation of the control AF, surpassing the estimation produced by Eq. (4), which yields values that fall below zero. These negative values

indirectly affect the shape and location of other populations. We also demonstrated that the high-intensity population of CD45¹ cells was also estimated correctly (Fig. 7).

A key component of GLM-based unmixing is the so called “link function” (25). The canonical link for Poisson GLM is the logarithm function. However, multispectral FC requires the use of the identity link since the observed Poisson-distributed fluorescence signals are linearly dependent on the abundances. The use of an identity-link function is not common for a Poisson regression but has been investigated (4).

A common general approach to regression of heteroskedastic data is an iterative version of WLS—iteratively reweighted least squares. This algorithm uses the Newton-Raphson technique to approximate the maximum likelihood of a GLM (25). However, in the case of identity-link Poisson, the simple iteratively reweighted least squares is not guaranteed to converge (32). Therefore, we do not consider this method to provide any advantage over direct numerical minimization of deviance (26). The detailed analysis of various alternative approaches to stable computation of ML estimates in identity-link Poisson regression is beyond the scope of this article.

As in the case of MAPE, the observed signal should not be zero, since the deviance residual is undefined for such values. However, the deviance residuals for observations close to zero approach zero. This is not the case with MAPE, as the small \mathbf{r} leads to increasing values of error. If any element of $\mathbf{M}\boldsymbol{\alpha}$ is zero, the deviance function \mathcal{D} also breaks down. Therefore, one simple solution is to add a small constant to \mathbf{r} and disallow $\boldsymbol{\alpha}$ for which there are zeros in $\mathbf{M}\boldsymbol{\alpha}$. Consequently, the obtained result may demonstrate a small positive bias (Figs. 5D and 6). Again, a more sophisticated treatment of zero values is possible using the methodology developed within the field of compositional analysis (30).

The Poisson deviance can be seen as an estimate of the KLD between two spectral vectors (33). Therefore, the described unmixing model can also be discussed within the context of generalized spectral unmixing proposed in the field of remote sensing. In this framework, it has also been shown that minimizing a symmetrized version of divergence results in superior results in terms of accuracy of spectral matching compared with minimizing the ℓ^2 norm of the error (34). Therefore, our result independently confirms conclusions reached by the remote-sensing community.

Finally, as the third alternative to OLS, we utilized a variance-stabilizing transformation to unmix the data. Similar solutions are routinely used in the field of microarray analysis and imaging (35,36). The use of BET yielded more realistic results when applied to simulated data (Supporting Information Fig. S3); however, when applied to experimental data, it resulted in data that were similar to those produced by NNLS (Supporting Information Fig. S4). We can only speculate that the lack of robustness caused the BET to perform worse than Poisson-based unmixing approaches. Although our signal-formation model is based on known physics of fluorescence, it is still a simplification. As indicated in the previous sections, a more laborious model might include overdispersion owing to nonideal PMT characteristics as well as the Gaussian noise component added by the system electronics. Yet, the BET is optimized for a simple Poisson distribution. The work in the field of

microarray data processing suggests that more appropriate variance-stabilization transforms can be found empirically using custom-built functions based on power transformation (36).

Conclusions

The research reported herein was based on the simple rationale that the goal of unmixing is to gain knowledge regarding the contribution of different fluorochromes to the total measured signal regardless of the signal intensity. The visualization approach commonly used in FC involving scatter plots, as well as the traditional FC terminology describing samples as “positive” or “negative,” suggests that practitioners are indeed interested in minimizing the error of estimation for low-abundance signals (negative population) just as much as for high-abundance signals (positive population) when both are present in the mixture. The reason is that classification of cells into negative, dim, and positive categories is often important in both research and clinical settings. If the estimated abundances are proportional to the quantity of fluorochrome present in bioparticles, this proportionality should always hold, no matter what the measured signal intensity range. In addition, any quantitative estimates of biological phenomena from underlying abundance of fluorescence markers are obviously confounded in the presence of unmixed negative abundances, especially when those observations arise primarily as a result of improper treatment of noise in the system.

The use of more detectors than labels (an overdetermined case) provides a unique opportunity from a data-analysis point of view. Since the noise in each detector is independent of the noise in all the other detectors, applying appropriate regression techniques allows us to arrive at a “best-fit” value for each label that is superior to the value that would have been determined if the number of detectors equaled the number of labels. The reported results show that Poisson unmixing provides the most accurate representation of the true underlying signal, and we expect that with the proliferation of spectral FC instruments the presented framework will be commonly employed for unmixing of spectral data sets.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Our special thanks to J. Paul Robinson (Purdue University) for allowing us to access the prototype of the multispectral flow cytometry system and to Dr. Tom Goldstein (Rice University) for assistance with mathematical notation and implementation of GLM.

Grant sponsors: National Institute of Biomedical Imaging and Bioengineering (grant number 1R21EB015707), PhRMA Foundation (2012 Research Starter Grant in Informatics), and De Novo Software Inc.

LITERATURE CITED

1. Settle JJ. On the relationship between spectral unmixing and subspace projection. *IEEE Trans Geosci Rem Sens.* 1996; 34:1045–1046.
2. Keshava N, Mustard JF. Spectral unmixing. *IEEE Signal Process Mag.* 2002; 19:44–57.

3. Zimmermann, T. Microscopy techniques. Vol. 95. *Advances in Biochemical Engineering/Biotechnology*. Springer Verlag; Berlin/Heidelberg; 2005. Spectral imaging and linear unmixing in light microscopy.; p. 245-265. Available from: <http://dx.doi.org/10.1007/b102216>. [July 12, 2012]
4. Palkki RD, Lanterman AD. Chemical mixture estimation under a Poisson Raman spectroscopy model. *Opt Eng*. 2010; 49:113601–113611.
5. Wolbers M, Stahel W. Linear unmixing of multivariate observations: A structural model. *J Am Stat Assoc*. 2005; 100:1328–1342.
6. Akerjord M-A, Christophersen N. Assessing mixing models within a common framework. *Environ Sci Technol*. 1996; 30:2105–2112.
7. Bagwell CB, Adams EG. Fluorescence spectral overlap compensation for any number of flow cytometry parameters. *Ann N Y Acad Sci*. 1993; 677:167–184. [PubMed: 8494206]
8. Roederer, M. Compensation in flow cytometry.. In: Robinson, JP.; Darzynkiewicz, Z.; Dobrucki, J.; Hyun, WC.; Nolan, JP.; Orfao, A.; Rabinovitch, PS., editors. *Current Protocols in Cytometry*. Vol. 1.14. John Wiley & Sons; Hoboken, NJ: 2002. Available from: <http://www.currentprotocols.com/protocol/cy0114>. [July 14, 2012]
9. Goddard G, Martin JC, Naivar M, Goodwin PM, Graves SW, Habbersett R, Nolan JP, Jett JH. Single particle high resolution spectral analysis flow cytometry. *Cytometry Part A*. 2006; 69A:842–851.
10. Watson DA, Brown LO, Gaskill DF, Naivar M, Graves SW, Doorn SK, Nolan JP. A flow cytometer for the measurement of Raman spectra. *Cytometry Part A*. 2008; 73A:119–128.
11. Nolan JP, Sebba DS. Surface-enhanced Raman scattering (SERS) cytometry. *Methods Cell Biol*. 2011; 102:515–532. [PubMed: 21704852]
12. Grégori G, Patsekín V, Rajwa B, Jones J, Ragheb K, Holdman C, Robinson JP. Hyper-spectral cytometry at the single-cell level using a 32-channel photodetector. *Cytometry Part A*. 2012; 81A: 35–44.
13. Ihaka R, Gentleman R. R: A language for data analysis and graphics. *J Comput Graph Stat*. 1996; 5:299–314.
14. Settle JJ, Drake NA. Linear mixing and the estimation of ground cover proportions. *Int J Rem Sens*. 1993; 14:1159–1177.
15. Kullback S, Leibler RA. On information and sufficiency. *Ann Math Stat*. 1951; 22:79–86.
16. Lawson, CL.; Hanson, RJ. *Solving Least Squares Problems*. Society for Industrial and Applied Mathematics; Philadelphia, PA: 1995. p. 350
17. Bro R, De Jong S. A fast non-negativity-constrained least squares algorithm. *J Chemometr*. 1997; 11:393–401.
18. McCutcheon MJ, Miller RG. Fluorescence intensity resolution in flow systems. *J Histochem Cytochem*. 1979; 27:246–249. [PubMed: 374582]
19. Ubezio P, Andreoni A. Linearity and noise sources in flow cytometry. *Cytometry*. 1985; 6:109–115. [PubMed: 3979215]
20. Coates PB. Photomultiplier noise statistics. *J Phys Appl Phys*. 1972; 5:915–930.
21. Prescott JR. A statistical model for photomultiplier single-electron statistics. *Nucl Instrum Meth*. 1966; 39:173–179.
22. Lachs G. The statistics for the detection of light by nonideal photomultipliers. *IEEE J Quant Electr*. 1974; 10:590–596.
23. Tofallis C. Least squares percentage regression. *J Mod Appl Stat Meth*. 2009; 7:526–534.
24. McCullagh, P.; Nelder, JA. *Generalized linear models*. 2nd ed.. Chapman and Hall/CRC; Boca Raton, FL: 1989. p. 532
25. Fox, J. *Applied Regression Analysis and Generalized Linear Models*. 2nd ed.. Sage Publications, Inc; Thousand Oaks, CA: 2008. p. 688
26. Venables, WN.; Ripley, BD. *Modern applied statistics with S*. 4th ed.. Springer; Berlin: 2002. p. 512
27. Bar-Lev SK, Enis P. On the classical choice of variance stabilizing transformations and an application for a Poisson variate. *Biometrika*. 1988; 75:803–804.

28. Anscombe FJ. The transformation of Poisson, binomial and negative-binomial data. *Biometrika*. 1948; 35:246–254.
29. Roederer M. Spectral compensation for flow cytometry: Visualization artifacts, limitations, and caveats. *Cytometry*. 2001; 45:194–205. [PubMed: 11746088]
30. Martín-Fernández, JA.; Palarea-Albaladejo, J.; Olea, RA. Dealing with zeros.. In: Pawlowsky-Glahn, V.; Buccianti, A., editors. *Compositional Data Analysis*. John Wiley & Sons Ltd; New York, NY: 2011. p. 43-58. Available from: <http://onlinelibrary.wiley.-com/doi/10.1002/9781119976462.ch4/summary>. [September 2, 2012]
31. Snow C. “Kit.” Flow cytometer electronics. *Cytometry Part A*. 2004; 57A:63–69.
32. Marschner IC. Stable computation of maximum likelihood estimates in identity link Poisson regression. *J Comput Graph Stat*. 2010; 19:666–683.
33. Hastie T. A closer look at the deviance. *Am Stat*. 1987; 41:16–20.
34. Chang, C-I. Spectral information divergence for hyperspectral image analysis.. *Geoscience and Remote Sensing Symposium, 1999; IGARSS '99 Proceedings. IEEE 1999 International*; New York. IEEE; 1999. p. p509-511.
35. Huber W, Von Heydebreck A, Sultmann H, Poustka A, Vingron M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics*. 2002; 18:S96–S104. [PubMed: 12169536]
36. Lin SM, Du P, Huber W, Kibbe WA. Model-based variance-stabilizing transformation for Illumina microarray data. *Nucleic Acids Res*. 2008; 36:e11. [PubMed: 18178591]

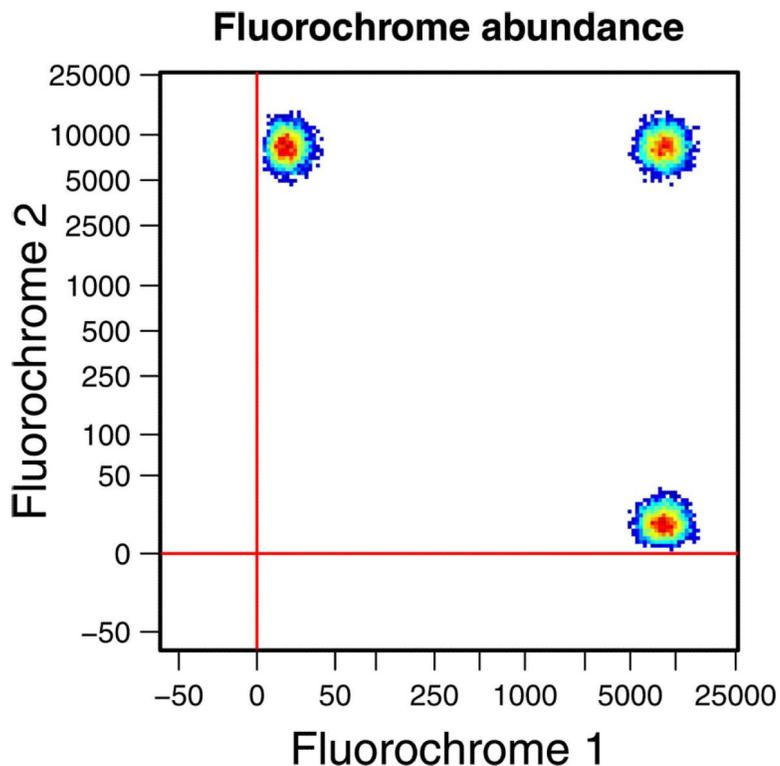


Figure 1.

Density plot representing simulated abundances of two fluorochromes; 15,000 cells were simulated in this in silico experiment. The abundances were drawn from log-normal distributions. The negative populations have logmean of $\log(10)$ and $CV = 0.2$ ($\sigma = 0.1$), the positive population has logmean of $\ln(10000)$ and $CV = 0.2$. The simulated data set includes emission shot noise and detection noise but no spectral overlap (i.e., $\mathbf{M} = 1$). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

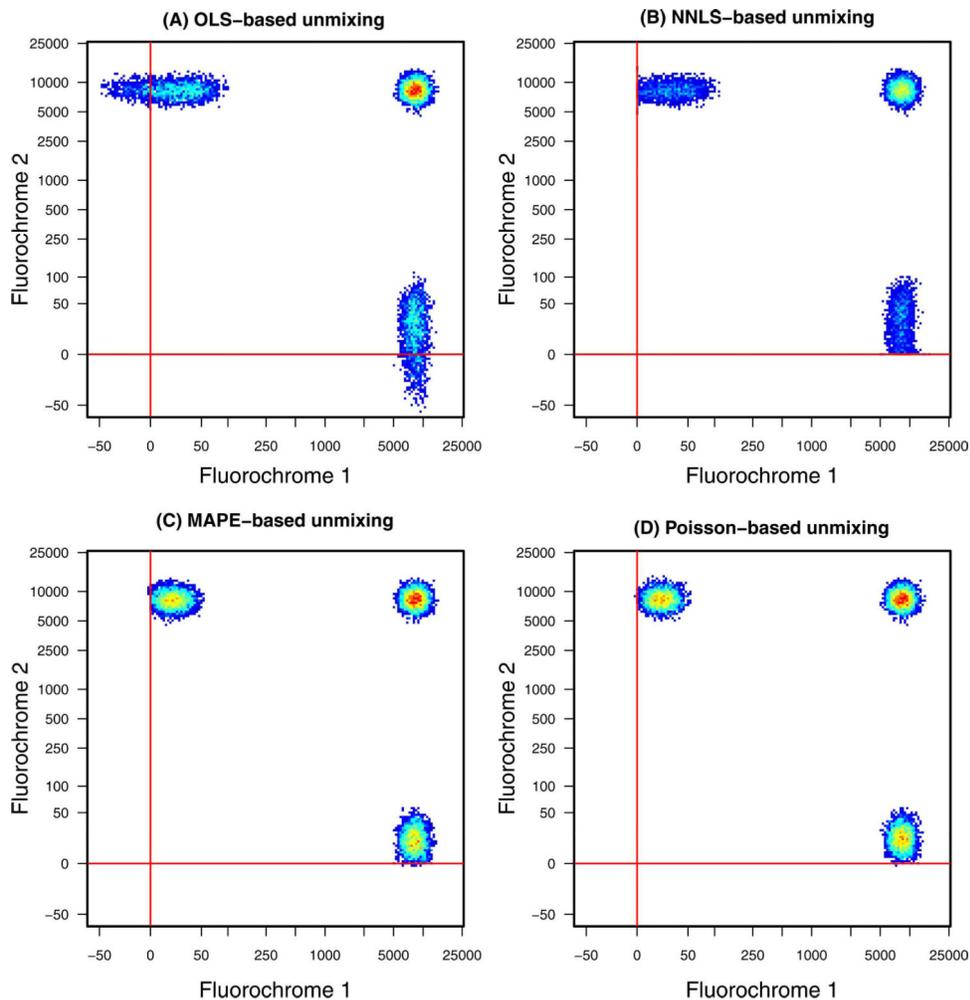


Figure 2. The results of spectral unmixing. For the purpose of visualization, the data have been transformed using a generalized log transformation (see Supporting Information Materials for details, and Fig. S1 for the mixing matrix M): OLS-based unmixing (A), NNLS-based unmixing (B), MAPE-based unmixing (C), and Poisson-based unmixing (D). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

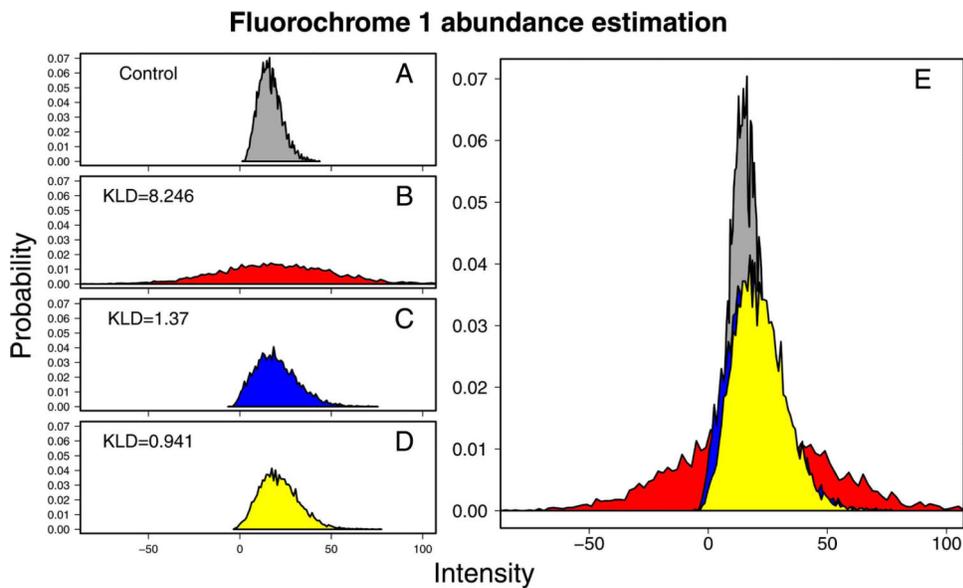


Figure 3. Comparison of unmixing results for a simulated low-intensity (negative) population. Plots **A**—**D** illustrate the shape of distributions: the distribution of original signal measured in the absence of spectral mixing (**A**), OLS-based unmixing (**B**), MAPE-based unmixing (**C**), and Poisson-based unmixing (**D**). Plot **E** shows overlap of the unmixed distributions and the original abundance. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

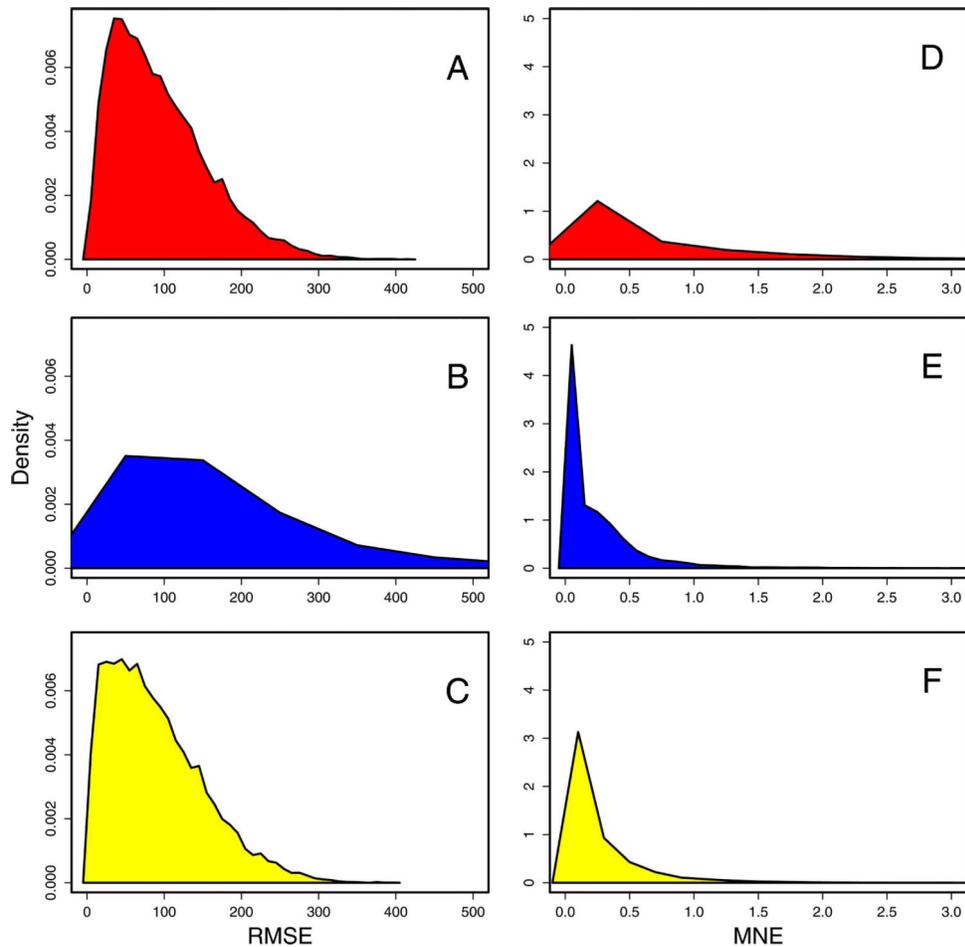
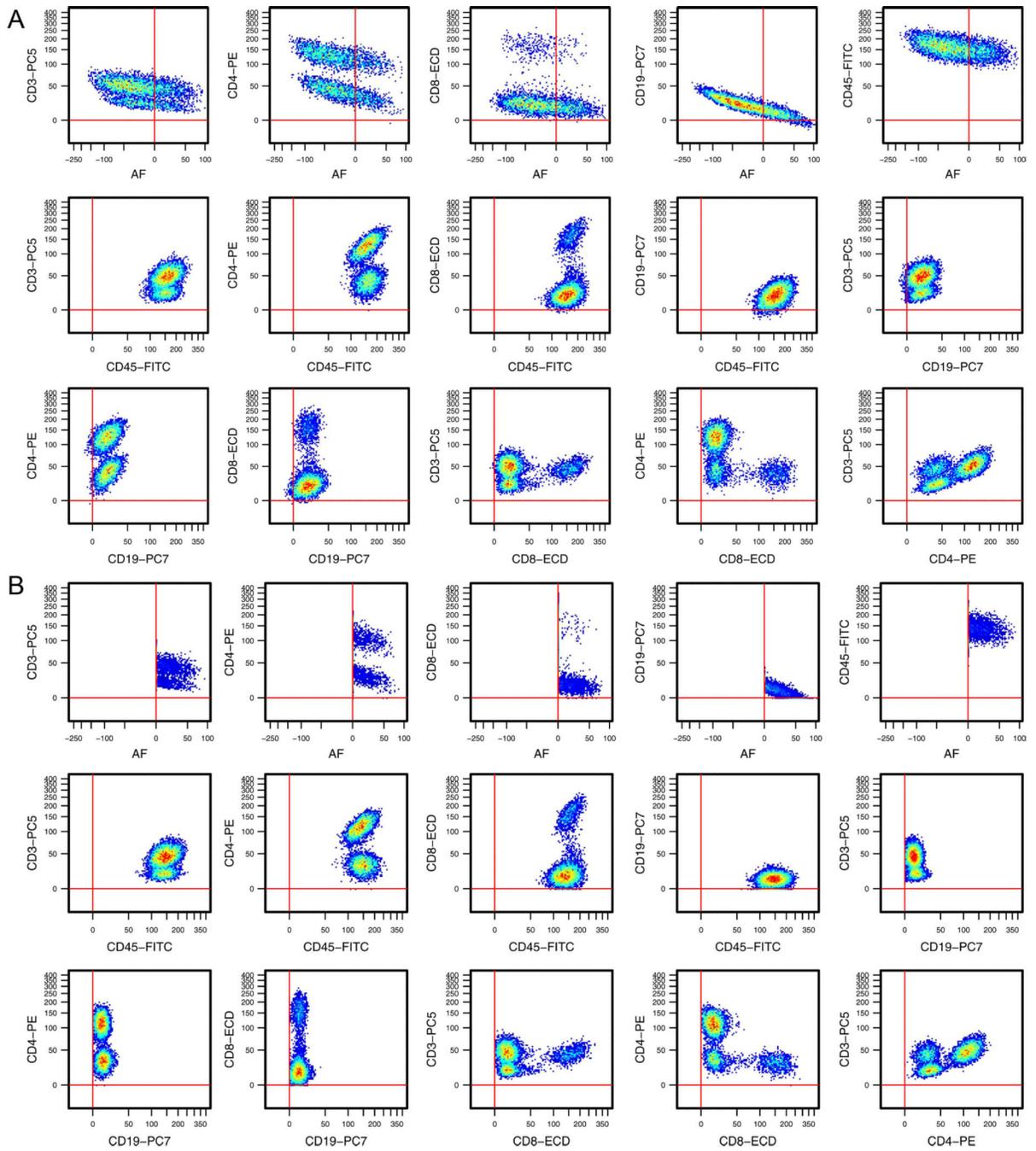


Figure 4. Distributions of RMSE and MNE for simulated flow-cytometry data. The distribution of RMSE is similar for OLS-based (A) and Poisson-based unmixing (C), and different than error distribution for MAPE unmixing (B). The distribution of relative (normalized) error is broader for the OLS method (D) and much narrower for MAPE (E) and Poisson-based (F) techniques. The median MNE for OLS is 0.3, and the 75th percentile is 0.87. In contrast, the median relative error for Poisson-based unmixing is only 0.1, with 75th percentile of 0.32. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]



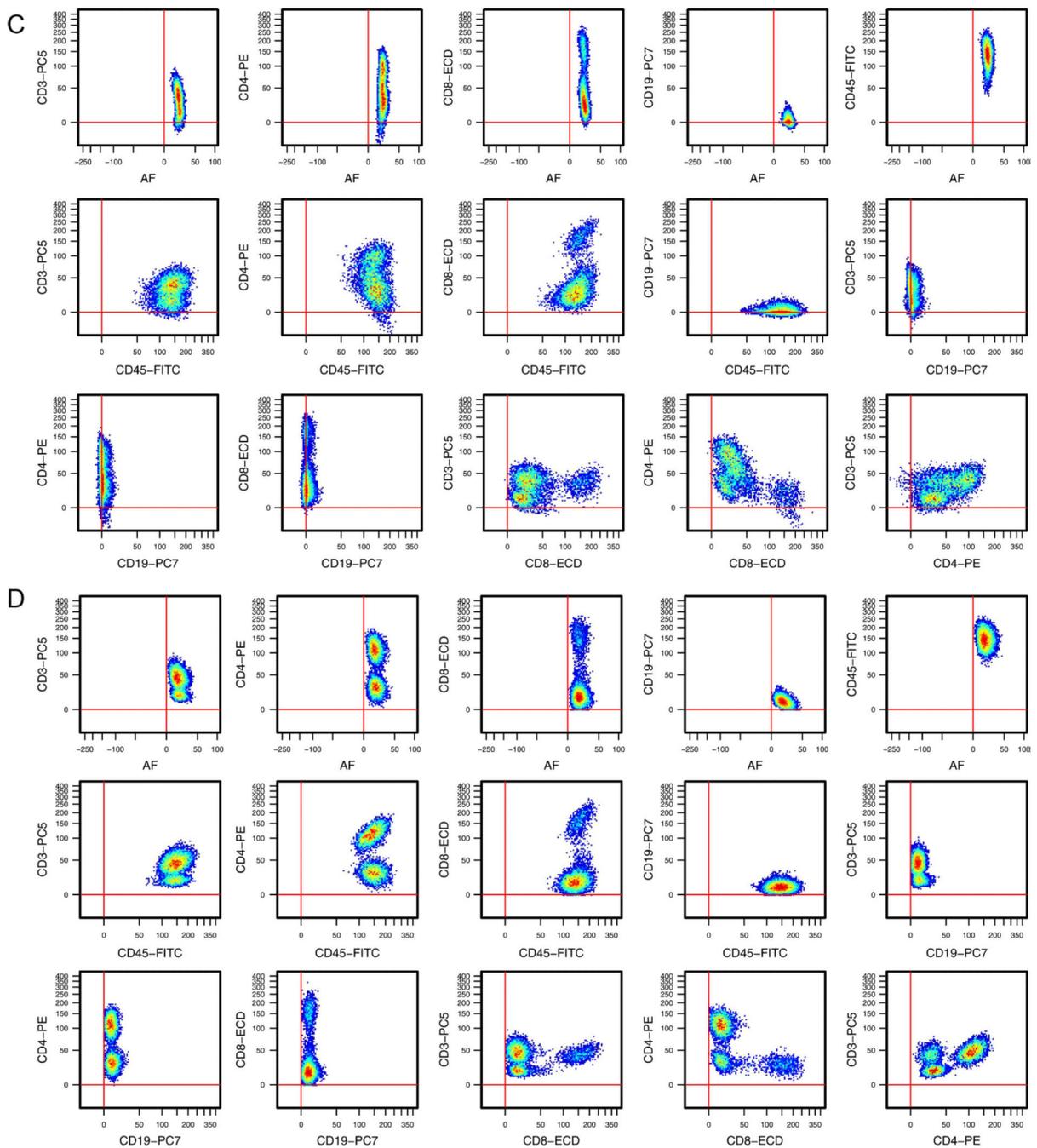


Figure 5. Unmixing of a 32-channel multispectral data set. OLS-based unmixing (A), NNLS unmixing (B), MAPE-based unmixing (C), and Poisson-based unmixing (D). The data were transformed using glog function with $m = 0.02$ (see Supporting Information Materials, and Fig. S2 for the estimated mixing matrix \mathbf{M}). Note that the density plots illustrating autofluorescence (the top row) are zoomed in on the low-intensity region. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

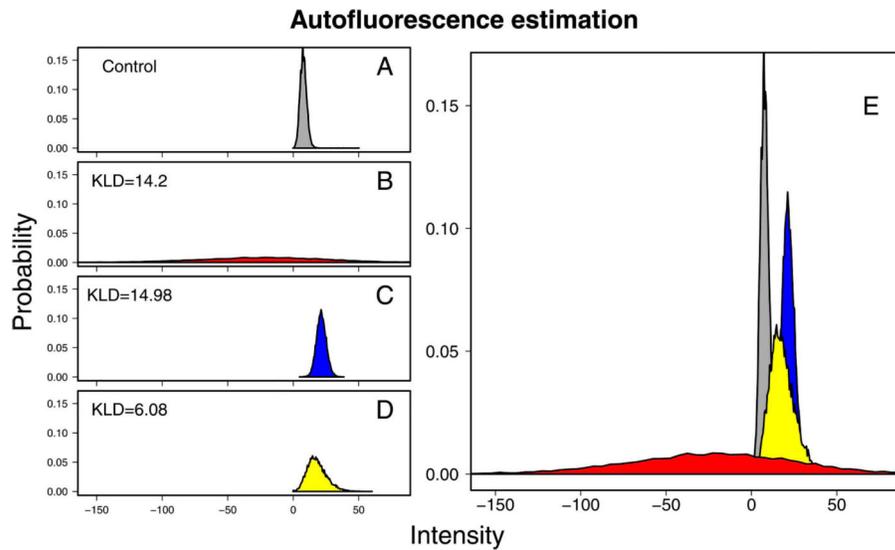


Figure 6. Comparison of the true autofluorescence (**A**) with the estimated autofluorescence recovered using OLS- (**B**), MAPE- (**C**) and Poisson-based (**D**) unmixing. Plot E illustrates the overlap between the recovered distributions and the known control. The difference between the unmixed values and the true autofluorescence is provided in plots B—D as a SKLD. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

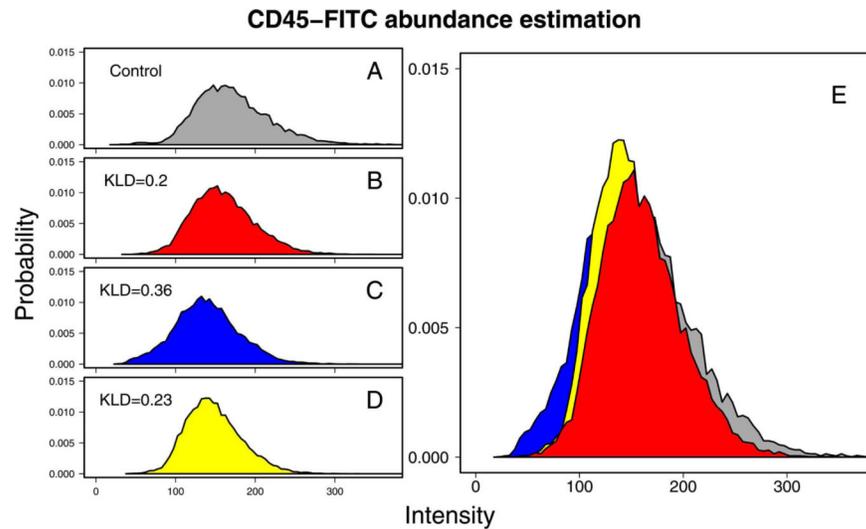


Figure 7. Comparison of the true measured CD45-FITC signal (A) with the estimated abundance recovered using OLS- (B), MAPE- (C) and Poisson-based (D) unmixing. Plot E illustrates the overlap between the recovered distributions and the known control. The difference between the unmixed values and the control is provided in plots B–D as a SKLD. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]