Cytometry
PART A
Journal of the International Society for Advancement of Cytometry

Distributions of Autofluorescence After Compensation: Be Panglossian, Fret Not

Mario Roederer*

Flow Cytometry Core and ImmunoTechnology Section, Vaccine Research Center, NIAID, NIH, Bethesda, Maryland

Received 20 July 2015; Revised 2 November 2015; Accepted 31 December 2015

Grant sponsor: Intramural Research Program of the Vaccine Research Center, NIAID, NIH.

*Correspondence to: Mario Roederer, Vaccine Research Center, NIH, 40 Convent Dr., Room 5509, Bethesda, MD 20892-3015. E-mail: Roederer@nih.gov

Published online 9 February 2016 in Wiley Online Library (wileyonlinelibrary.com)

DOI: 10.1002/cyto.a.22820

Abstract

In this technical note, I describe an in silico model of multiparameter fluorescence measurements that takes into account intrinsic cellular autofluorescence and stained dye fluorescence distributions, fluorescence spectrum spillovers, and photon counting statistics. Using this model, it is easy to manipulate spectral variables as well as error terms to understand the impact of each on the distributions of estimated cell-associated dyes (e.g., conjugated monoclonal antibodies in immunophenotyping). An application of this model was to understand the genesis of "unusual" autofluorescence distributions that occasionally happen in multi-color immunophenotyping. These unusual distributions show striking correlated patterns (diagonals) for graphs of certain pairs of parameters. Here I show that these arise from combinations of spillover-spreading from unviewed parameters. While disconcerting to researchers taught to look for diagonals in distributions as heralding improper compensation, these distributions are in fact appropriate. In general, one can ignore the characteristics of cell distributions within the same limits as background (e.g., as proscribed by a fluorescence-minus-one, or FMO, control), as there is essentially no information content in that region.

Key terms

instrumentation; controls; data analysis

WE have occasionally noticed unusual distributions of autofluorescence in multicolor immunophenotyping panels (Fig. 1). Specifically, when displayed with certain pairs of parameters, the cells in the autofluorescence region show a correlated distribution (i.e., appearing on a diagonal). Because the appearance of a diagonal distribution in compensated plots has long been used as an indicator of potential improper compensation, the appearance of these distributions raised concerns that there may be problems with the panel, the instrument, or the compensation.

To address this, and other compensation-related questions, I developed an in silico model of fluorescence measurements by flow cytometry. This model takes as input parameters the noise distribution in the measurements, the fluorescence spillover between measurement detectors, and the proportions and fluorescence distributions of one or more subsets of cells. Using this model, I reproduced the unexpected distributions seen in some of our panels, and thereby can explain their origin.

The process of compensation is a transformation of measurement space into fluorescence space (in other words, transforming measurement values to estimates of cell-associated fluorescence). Measurements from any single detector comprise contributions from multiple fluorescences because of spectral spillover. With standard linear algebra, these are transformed into values that estimate the individual "pure" fluorescences (1). As previously described, this process also convolves measurement error from each of the individual detectors into the final values, leading to "spillover spreading error" (2). It is worth reiterating that this error is not introduced nor exacerbated by compensation: the error arises primarily from photon counting statistics (3), and is present in the measurement prior to any transformations.

ISAC
International Society for Advancement of Cytometry

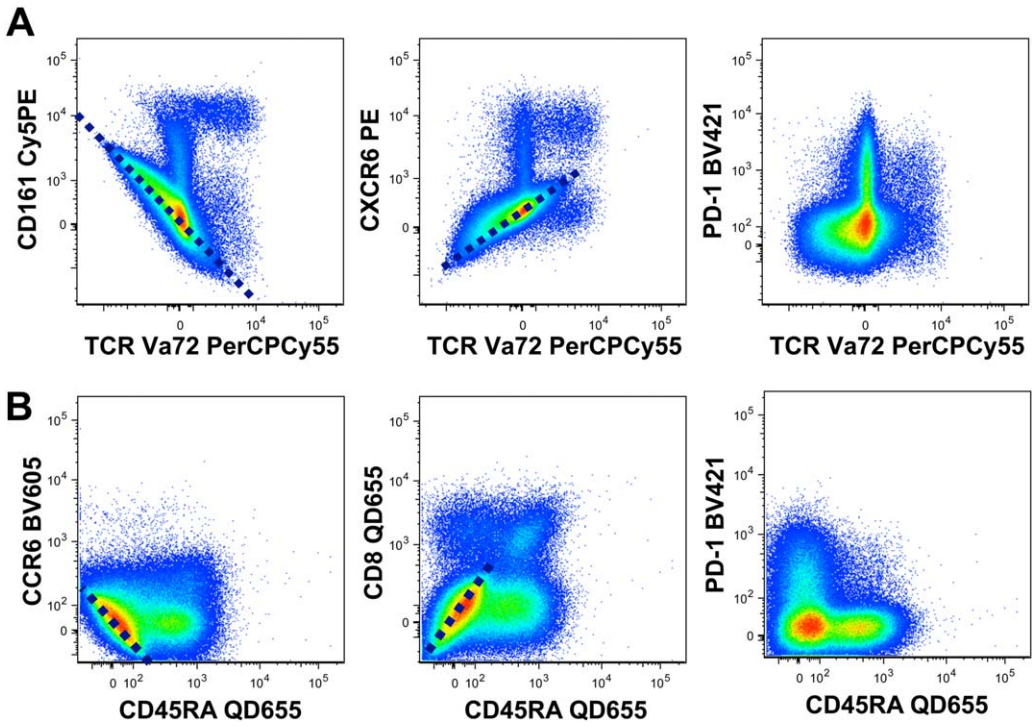

Figure 1. **A**: Peripheral blood mononuclear cells were stained with a 16-color immunophenotyping panel, and compensated using standard methods. Shown are three graphs from the same sample, plotting a PerCP-Cy5.5 reagent against a Cy5PE, PE, or BV421 reagent. All of these reagents stain a small minority of cells. Substantial spillover occurs amongst the first three reagents, but not with BV421. The distribution of autofluorescence (background) amongst the vast majority of cells not staining with these reagents shows an unexpected negative (left) or positive (middle) correlation, as indicated by the dashed line. The first two graphs are the only pairs of parameter combinations (out of 120) showing such an effect in this panel. **B**: Similar to Panel A, but using a different, 13-color panel. In this panel, four of 78 pairwise combinations exhibited obvious correlated background distributions.

The linear algebra used for spillover compensation performs a linear transformation. This effectively takes the correlated ("diagonal") distributions of uncompensated data and transforms them into uncorrelated ("cloudlike") distributions in the compensated plots. It should therefore come as no surprise that if the original distribution in the uncompensated plot is not correlated (i.e., cloudlike), then this same transformation may lead to an unexpected correlated (diagonal) distribution.

Autofluorescence will typically have an uncorrelated distribution in the uncompensated measurement space, particularly for small cells such as lymphocytes. This is because, at such low signal intensity, measurement errors dominate; noise will be random and thus not correlated between channels.

To understand the impact of measurement noise and spillover spreading on the distributions of fluorescence measurements after compensation, I developed an in silico model of flow cytometry detection (see Methods for details). Figure 2 illustrates an analysis using this model. Nine subsets of cells were modeled that have different staining patterns for two detectors. As shown in Figure 2A, the nine have been constructed to be nonoverlapping so that the impact of the model variables can be visualized. In this panel, a small amount of measurement noise is modeled (i.e., relatively high photon counts), leading to a mild amount of spillover spreading, as seen by the increased variance in the "*Z*" parameter for cells with higher "*X*" fluorescence. In this model, the variation of intrinsic autofluorescence is small, so the distributions of cells are very tight.

Figure 2B illustrates the impact of fivefold higher relative measurement noise (i.e., reducing the photon counts by a factor of 25). The amount of spillover spreading is dramatically increased, as is evident by the increased spread in the "*Z*" dimension. As well, it can be seen that the distributions of the events attained a negative correlation between the parameters "*Z*" and "*X*"—and this is much more obvious at the low "*X*" intensity populations. Note that for this example, there is no spillover into parameter "*Y*" and the distributions in the "*Y*" channel are uniform as expected.

Figures 2C and 2D illustrates the impact of spill over into the "*Y*" parameter. Depending on whether the spillover comes from parameter "*Z*" (Fig. 2C) or "*X*" (Fig. 2D), the shape of the distributions in the "*Z* vs. *Y*" bivariate plots is strikingly different. In both cases, there is apparently highly correlated distribution between parameters "*Z*" and "*Y*", either negative or positive. It should be stressed, however, that these apparent correlations exist only within the background region in the "*Z*" parameter (i.e., occur in the area where negative events are found).

Figure 3A illustrates in more detail why these correlations arise; in this figure, only the five populations having positive "*Y*" fluorescence are modeled. Here, the critical observation is

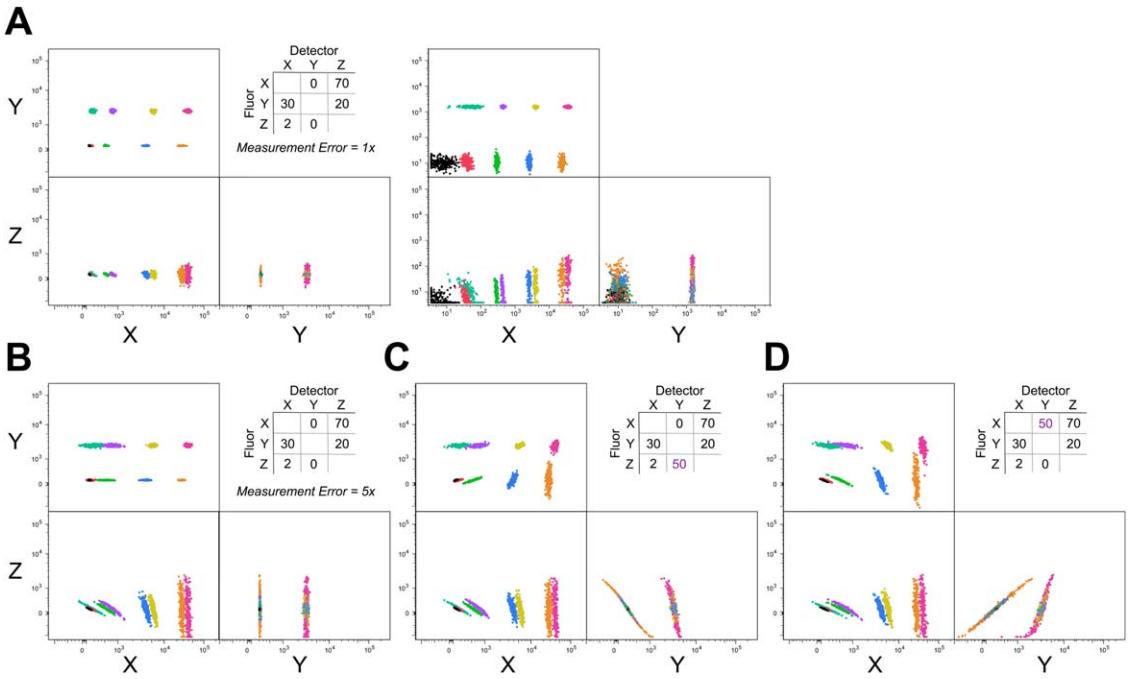

Figure 2. In silico modeling of flow cytometry data. In the these graphs, eight different subsets are modeled, each shown with a different color. The data comprises of three measurement parameters, "*X*," "*Y*," and "*Z*". Four of the subsets are positive for *Y*, eight of the subsets express different levels of *X*. Each panel shows three bivariate plots comparing every parameter against every other parameter. In addition each panel shows the spillover matrix for the three detectors. All data are shown properly compensated. **A**: A small amount of measurement error is modeled and shown using Logicle scaling (left) or logarithmic scaling (right). **B**: Compared to A, relative measurement error is increased fivefold, resulting in more spillover spreading in the compensated distributions. **C, D**: Compared to B, additional spectral spillover is introduced into detector *Y* from either detector *Z* (C) or *X* (D), resulting in the opposite background correlations.

in the uncompensated panel. The populations of cells already have a significant spreading in the "*X*" dimension because of the spillover spreading error from the "*Y*" parameter. As compensation is applied partially or fully, the events at the high-end of each population are pushed further down than the events at the low-end, leading to an apparent negative correlation post compensation.

As shown in Figures 3B and 3C, these unusual distributions are occurring within the autofluorescence or background region within the "*Z*" parameter. Since all of these populations are modeled with no "*Z*" fluorescence, they represent the fluorescence minus one (FMO) controls. Thus, the area occupied by these populations, illustrated by the purple-shaded region, is in the negative distribution for events.

This serves as a reinforcement that everything below the FMO boundary is negative and carries essentially no information. The distribution of events below the FMO boundary should not, on their own, raise concerns that there may be problems with the instrument or compensation settings.

A question arises as to what happens when highly autofluorescent cells are studied. Typically, in such a setting, the autofluorescence distribution is not randomly distributed in the uncompensated measurement space, but is correlated. The in silico model allows for the specification of such an autofluorescence distribution. In such a case, it is still possible to encounter unusually correlated distributions in the autofluorescence region after compensation. This is primarily because the autofluorescence emission spectrum is distinct from the fluorescence dyes that are being measured; thus, the spillover matrix used to compensate stained cells will not "properly" correct the correlation between the autofluorescence measurements. Hence, the autofluorescence distribution achieves an under- or over-compensated appearance in the compensated, fluorescence dye space. Similar to the situation above, such distributions should not, on their own, raise concerns about compensation settings or instrument settings.

Finally, in some situations it may be useful to use "autofluorescence compensation" to improve sensitivity by correcting for autofluorescence measured an otherwise unused channel (4). In this case, the resulting corrected values will now be dominated by noise and therefore subject to the same potential distributions as described in Figures 2 and 3.

In summary, the intricacies of background measurements (i.e., of cells unstained in a particular parameter) can lead to unusual or unexpected distributions following spectral unmixing (compensation). Fortunately, these occur relatively rarely, but when they do occur they should not, on their own, raise undue concern.

Methods

The *in silico* model was implemented in JMP (version 11, SAS Institute, Cary, NC). Scripts are available by request to

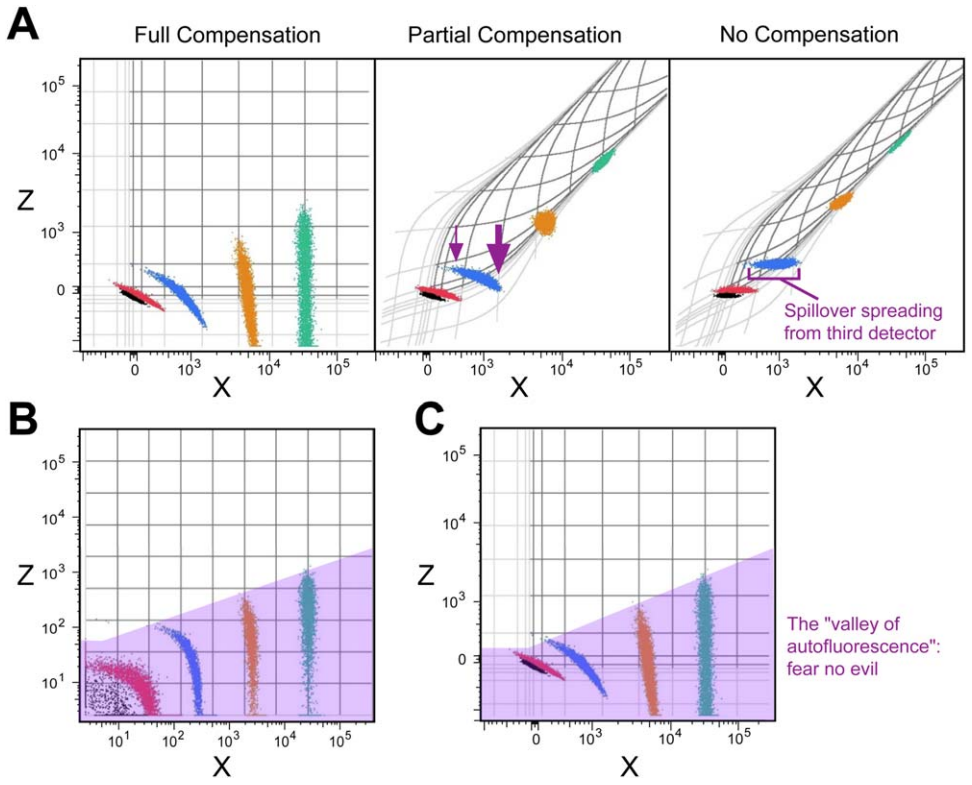

Figure 3. Similar modeling of distributions as in Figure 2, but with five populations expressing varying levels of *X*, and all expressing *Y* (not shown, but see Fig. 2). Grid lines in these figures are drawn along equal fluorescence values and are thus rectilinear in the fully compensated plot. Light gridlines correspond to values below zero. The grid lines help visualize the transformation between the uncompensated measurement space and the compensated fluorescence space. **A**: The distributions are shown either fully compensated (left), uncompensated (right), or with compensation settings at half of correct (middle). Note that the unviewed fluorescence in parameter *Y* introduces spillover spreading in parameter *X*, resulting in an uncorrelated spreading of the events when viewing parameters *X* and *Z*, in the uncompensated plot. Consequently, as compensation is increased, the events in this single population with higher levels of *X* will be pushed further down in the *Z* parameter (arrows, middle plot). This accounts for the correlated distribution appearing in the compensated plots. **B**,**C**: The same data as in A, shown fully compensated and with logarithmic (B) or Logicle (C) scaling. The shaded regions indicate the autofluorescence and background region as defined by fluorescence minus one (FMO) controls. It is generally safe to ignore unusual distributions in this shaded region—as long as the central tendency (as described by the median of the florescence distribution) is not biased away from the control distribution, which might indicate under- or over-compensation.

the author. The script requires the following as inputs: number of parameters (colors); for each parameter: a relative error factor and the spillover coefficient into other parameters; number of types of autofluorescence; for each autofluorescence: the mean and (log normal) CV of the autofluorescence distribution for every parameter; the number of subsets to model; for each subset: the number of events, the autofluorescence type to use, and the mean and (log normal) CV of the fluorescence distribution for every parameter.

The script creates a new data table containing one entry for every cell specified. For every parameter of each cell, it creates a value for the intrinsic autofluorescence and the cell-associated dye fluorescence, using the specified mean and CV as a basis for random number generation. The sum of all parameters' intrinsic dye fluorescence values multiplied by the respective spillover coefficients is added to the intrinsic autofluorescence to compute the "true" signal. "Measured" signal is computed from this by introducing a Poisson-based randomization to simulate photon counting errors (i.e., a random value proportional to the square root of the "true" signal; using the detector's error factor as the constant of proportionality). Finally, these "measured" signals undergo compensation (using the inverse of the spillover matrix) to generate the compensated values. These are stored both as log-transformed and as logicle-transformed (5) in the spreadsheet.

The script has the options to generate grid lines, where grids are equidistant in either measurement space or in dye space (the latter is shown in Figs. 3B and 3C). A separate script is available to perform successive partial to full compensations, saving a separate graphic for each step (e.g., that can be concatenated into a movie).

Acknowledgments

I thank Andrew Ishizuka and Bob Seder for the sample data in Figure 1A, Maggie Beddall for sample data in Figure 1B, Pratip Chattopadhyay and Steve Perfetto for discussions and comments, and all members of my laboratory for continued support, discussions, and enthusiasm.

Literature Cited

1. Roederer M. Compensation. In: Robinson JP, Darzynkiewicz Z, Dean PN, Dressler LG, Rabinovitch PS, Stewart CC, Tanke HJ, Wheeless LL, editors. Current Protocols in Cytometry. New York: Wiley; 1999.
2. Nguyen R, Perfetto S, Mahnke YD, Chattopadhyay P, Roederer M. Quantifying spillover spreading for comparing instrument performance and aiding in multicolor panel design. Cytometry A 2013;83A:306–315.
3. Roederer M. Spectral compensation for flow cytometry: Visualization artifacts, limitations, and caveats. Cytometry 2001;45:194–205.
4. Roederer M, Murphy RF. Cell-by-cell autofluorescence correction for low signal-to-noise systems: application to epidermal growth factor endocytosis by 3T3 fibroblasts. Cytometry 1986;7:558–565.
5. Parks DR, Roederer M, Moore WA. A new "Logicle" display method avoids deceptive effects of logarithmic scaling for low signals and compensated data. Cytometry A 2006;69A:541–551.