BECKMAN COULTER
Life Sciences

Unlocking Insights: The Vital Role of Unmixing Algorithms in Spectral Flow Cytometry

Introduction

Spectral flow cytometry offers significant improvements in performance and multiplexing capabilities beyond conventional flow cytometry. Unlike conventional flow cytometry, which collects only a discrete portion of the emission spectrum using single filters per fluorochrome, spectral flow cytometry captures the full spectrum for all fluorochromes on each cell using multiple detectors. Therefore, spectral flow cytometry requires more complex methods than compensation to distinguish fluorochromes. This is necessary because the instrument must distinguish between multiple fluorescent profiles across the entire visible light spectrum rather than from a few distinct channels. The process of deconvoluting fluorochrome emission spectra across an array of detectors in spectral flow cytometry is referred to as *spectral unmixing*. Unmixing requires single-stained reference controls, as well as noise-reducing mathematical algorithms. This article will give a brief explanation of the ideas behind compensation and unmixing, highlighting their differences, and discuss the current unmixing algorithms being used.

The Difference between Compensation and Spectral Unmixing: Understanding their Role in Flow Cytometry

In conventional flow cytometry, each fluorochrome is measured using a single primary detector, and the peak emission spectra are collected using a bandpass filter. As each fluorochrome emits light at different wavelengths when excited by a laser, the emission of one fluorochrome overlaps the optimal detection channel of a different fluorochrome, known as spectral spillover. The use of multiple fluorescent dyes to label different cellular components or markers brings in the challenge of spectral spillover, causing a single detector to pick up signals from multiple dyes. To address such spillover, we apply a mathematical tool called compensation that removes the signal of any given fluorochrome from all detectors except the primary detector dedicated to measuring that specific fluorochrome **(Figure 1)**. Compensation works by generating a compensation matrix derived from the spillover calculated using single-color compensation controls. Single-color controls are essential for determining how much each fluorochrome spills over into other channels. This compensation matrix is then applied to the experimental data to subtract the overlapping fluorescence and isolate the signal from each fluorochrome.

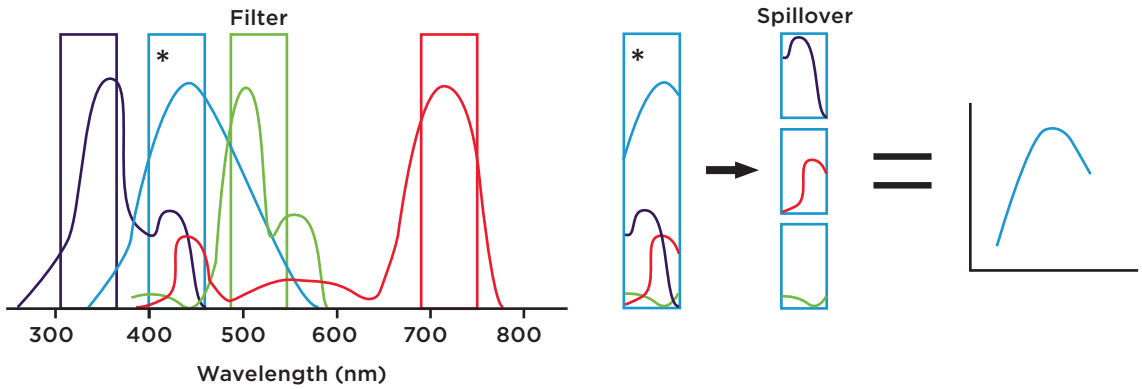

Figure 1. Compensation in Conventional Flow Cytometry: In conventional flow cytometry, each detector uses a bandpass filter to capture the peak emission of specific fluorophores. Any spillover from other fluorophores is corrected using a mathematical process called a "compensation" matrix.

*Denotes peak emission.

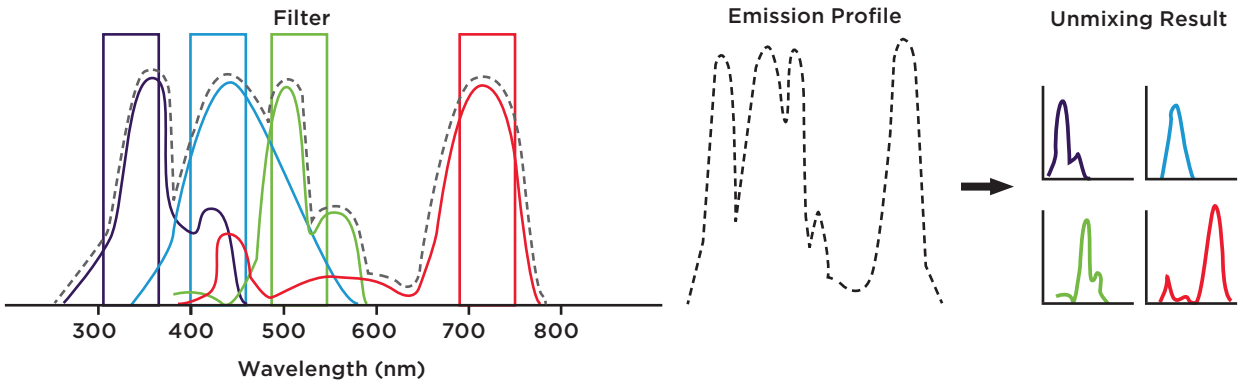

Figure 2. Unmixing in Spectral Flow Cytometry: Spectral flow cytometers use a detector array to capture the full spectral signature of all fluorophores. These combined signals create a single complex waveform, which is then separated into individual fluorophore signatures using an unmixing algorithm.

In contrast to conventional flow cytometry, spectral flow cytometry uses multiple detectors to capture the full spectrum emission of each fluorochrome across multiple lasers used in the system, to create a detailed spectral signature. This enables use of a larger number of fluorochromes, even those with overlapping emission spectra, thereby enabling more complex and detailed analyses of cell populations.

Thus, traditional compensation methods used for discrete wavelength channels in conventional flow cytometry are insufficient for the continuous spectra captured in spectral flow cytometry. Therefore, spectral unmixing is employed, allowing for precise analysis of complex, multicolor experiments **(Figure 2)**.

Spectral unmixing deconvolutes various fluorophore signatures based on their unique spectral characteristics. While not mathematically identical to conventional compensation, the overall principle is the same, as it can separate the overlapping spectra using the premeasured single-stained or unstained samples as the basic pure spectra. Thus, unmixing relies on robust single-stain controls that identically match the spectral signatures in the experimental sample. Through this approach, fluorochromes with similar emission but different spectral signatures can be distinguished from each other, allowing them to be used together in a panel.

Unmixing algorithms commonly rely on statistical methods such as ordinary least squares and weighted least squares. Before discussing the available options for spectral unmixing algorithms, it's important to briefly understand the common mathematical principles these algorithms follow. For a more in-depth dive into mathematics, please refer to the references at the end of this application note.

The Mathematics of Unmixing

In conventional flow cytometry, the compensation model comprises a full rank equation of system where number of fluorochromes matches detector. The abundance of an arbitrary number of fluorochromes across the same number of detectors in such square matrices is calculated, using a basic linear equation:

$$\mathbf{r} = \mathbf{M}_{avg}\,\mathbf{a} + \mathbf{e}$$

$$\mathbf{a} = \mathbf{r}\,\mathbf{M}_{avg}^{-1} - \mathbf{e}$$

- **r** is observed raw detector values from an arbitrary cell.
- $\mathbf{M}_{avg}$ is the average spectral signature matrix calculated for each detector.
- **a** is true abundance of each fluorochrome on that cell. a is calculated as the difference between the M_{avg} and the actual emission of each fluorochrome from an individual cell (M) from the sample.
- **e** denotes an unknown noise matrix.
- $\mathbf{M}_{avg}^{-1}$ (often called compensation matrix) is reversible because of square matrices.

As spectral flow cytometers utilize a larger number of detectors than fluorochromes, it involves an overdetermined system, therefore M_{avg}, is rectangular and is not irreversible. Most unmixing algorithms use the mathematics below to "unmix" the data and calculate the approximate abundance of each fluorophore:

$$\mathbf{a = (M_{avg}^{T} M_{avg})^{-1} - M_{avg}^{T} r.}$$

- M_{avg}^{T} - is the transpose of M_{Avg}, because M_{avg} is not reversible.

Most of the unmixing algorithm aims to calculate the approximate (a) that satisfies the equation, $r= M_{avg}$ a+e.

The Effect of Noise on Calculated Abundances

Two key sources of noise can exist in flow cytometer measurements. The one most familiar to users is *instrument noise*, which arises from electronics, stray light, etc. Modern commercial flow cytometers are designed so well that this source of noise is usually negligible in many experiments, except when measuring small or dim particles. The second source is the variation between the actual emission (M) from an individual cell in the sample and the average emission (M_{avg}) used in the unmixing calculations. Each particle's M cannot be the same as M_{avg} because photon emission follows Poisson statistics, making it impossible for them to be equal. The brief exposure of cells to the laser can lead to low photon emission and significant deviation in the emission profile of individual fluorochromes from the average. This type of noise is typically the main cause of the observed spread in unmixed data.

Different Unmixing Algorithms Used in Various Spectral Flow Cytometers

The spectral unmixing algorithm is crucial because it enables the accurate separation and quantification of overlapping fluorochromes, thus enabling precise identification and analysis of multiple markers within a single sample. The following are various algorithms for spectral unmixing:

1. Least-square method (LSM) for spectral unmixing

- This is the simplest and most common method used for estimating the abundance **(a)** of different fluorochromes in individual cells within multispectral datasets.
- It finds the best function match for the data by minimizing the sum of the squares of the errors/noise between the observed data and the actual data.
- LSM assumes that the errors or noise are homoscedastic, which means that the mean of the noise is zero/constant, and the variance of the noise is the same in each detector, irrespective of signal level.
- However, in practical scenarios, detectors with more signal have more noise than those with less signal. Hence, over a large population of cells, low-abundance fluorochromes with dim signals have a narrow distribution and decreased variance. The mismatch can cause population distortion (spreading).

2. Weighted least-square (WLSM) algorithm for spectral unmixing

- This is an extension of traditional LSM, but it gives more importance (or "weight") to some errors/noises over others.
- In WLSM regression, weights are assigned to each observation based on the variance of the error term, enabling more accurate modeling of heteroscedastic data.
- The weights are inversely proportional to the error variance. Data points with lower variability or higher reliability are given more weight in estimating abundances.
- In WLSM regression, data points with higher weights have a stronger influence on the fitted line. This improves accuracy, especially in data with varying variability.
- **Note:** Commercially available WLSM algorithm allows users to manually input the weight or let the software perform multiple iterations to find the optimal setting for minimizing signal spreading. It is important to note that while the multiple iterations of WLSM may yield better results in reducing spreading, this process can be significantly time-consuming, making it unsuitable for live unmixing. The speed of this iterative method is comparable to that of the Poisson method.

3. Poisson algorithm for spectral unmixing

- The Poisson distribution models the probability of a certain number of events occurring in a fixed interval, assuming the events happen at a constant rate and independently of each other.
- This unmixing approach incorporates the Poisson nature of signal and noise. Photons are emitted by fluorochromes at random time intervals, and the distribution of their arrival at the detector is closely approximated by a Poisson distribution.
- This algorithm assumes that the errors or noise are heteroscedastic.
- In a Poisson distribution, the expected value (which sometimes matches the mean) and variance are equal, meaning the average number of events and the variability around this average are both defined by the same parameter.
- Because this algorithm more accurately reflects the underlying physics of signal formation, it can help in distinguishing true signals from background noise, which is particularly useful where the signal-to-noise ratio is low.

4. Poisson hybrid

The Poisson regression is too slow to handle millions of events unmixing. To address this issue, a combination of Poisson and LSM was introduced to expedite the unmixing process while achieving better performance. The hybrid takes LSM's result as the initial estimated parameter and then does the regression by iterative weighted least square (IRLS). This approach incorporates the Poisson nature of signal and noise, leading to improved efficiency and accuracy. Therefore, Poisson hybrid provides a better result than LSM within an acceptable time.

Summary of available unmixing algorithms

Features	LSM	WLSM	Poisson	Poisson Hybrid
Assumption	Assumes constant variance (homoscedasticity) of errors	Allows for varying variance (heteroscedasticity) of errors	Counts follow a Poisson process. Allows for varying variance (heteroscedasticity) of errors	Counts follow a Poisson process. Allows for varying variance (heteroscedasticity) of errors
Advantages	Simple, easy to implement, provides unbiased estimates when assumptions are met	Accurate with low data spreading for varying data	Accurate with low data spreading	Accurate with low data spreading
Disadvantages	Sensitive to outliers, assumes constant variance of errors	Using correct weights based on error variance is crucial; incorrect weights can cause biased results	Assume mean equals variance (limitation to over-dispersed data)	Assume mean equals variance (limitation to over-dispersed data)
Computational Complexity	Low	Moderate	High	Mid-high
Speed of Algorithm for Unmixing	Fast	Moderate	Slow	Moderate

Conclusion

Spectral flow cytometry enables the analysis of many cellular features in one experiment using multiple fluorochromes. Unlike traditional methods, it employs spectral unmixing—a mathematical technique that separates signals from different fluorochromes, producing an unmixing matrix that estimates the abundance of each fluorochrome on a cell. To achieve accurate results, it's important to have a good reference control and an understanding of the unmixing algorithm used by the software, as the accuracy of the results depends on how well the algorithm reflects the real physical processes being studied (as explained in the summary).

References

1. Novo D. A comparison of spectral unmixing to conventional compensation for the calculation of fluorochrome abundances from flow cytometric data. Cytometry. 2022;101(11):885–91.

2. Novo D, Grégori G, et al. Generalized unmixing model for multispectral flow cytometry utilizing nonsquare compensation matrices. Cytometry Part A, 2013 83(5): 508–520.

3. Sharma S, Boyer J, et al. A practitioner's view of spectral flow cytometry. Nature Methods. 2024 May; 21(5) 740-743.

4. Xian-guang Fan, Yu-liang Zhi, et al. An effective spectral unmixing algorithm for flow cytometry based on GA and least squares. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2022, 264:120254.

For Beckman Coulter's worldwide office locations and phone numbers, please visit Contact Us at beckman.com

2024-GBL-EN-106619-v2