Comparing Spectral Unmixing Locally and in the Cloud

Jo-Anne Crofts1, Christopher Hall2, Hanan Ibrahim2, Sam Thompson2, Philip S Hobson3, Peter Nobes1, Steven Lim3, Rachael V Walker2, Tony Burpee1

1= Applied Cytometry, Matrix Business Centre, Nobel Way, Dinnington, Sheffield S25 3QB. 2= Flow Cytometry Facility, Babraham Institute, Cambridge, CB22 3AT
3= Flow Cytometry Science Technology Platform, The Francis Crick Institute, 1 Midland Road, London, NW1 1AT

AppliedCytometry
Pioneering Software Simplicity

INTRODUCTION

The aim of this study was to test the capability of spectral unmixing in the cloud and thus provide an opportunity for SRLs to unlock the full spectral potential of their existing cytometers. Spectral unmixing in the cloud would enable SRLs to remove autofluorescence and extend their panel designs without purchasing a dedicated spectral instrument, and without having to become an R software engineer.

In the study by Chris Hall et.al., the flowUnmix algorithm successfully converted detector data into fluorochrome data. It was used on several instruments for an objective comparison of performance. The flowUnmix algorithm in the study was implemented locally in R, a programming language and software environment commonly used for statistical computing and graphics.

This study sought to establish the data equivalence between flowUnmix algorithm implementations using Novocyte Penteon data as follows: running locally on R, running locally in an Open Container Initiative, "OCI", compliant container and running an OCI container in CytoSwarm, a cloud-based processing engine.

By establishing data equivalence, this study would allow cloud-based unmixing to be realized.

ABSTRACT

With the revelation and demonstration that spectral deconvolution (the conversion of detector data into fluorochrome data) can be achieved on FCS files from traditional flow cytometers ([1] https://doi.org/10.1101/2022.12.21.521417 Submitted to Cytometry Part A), it follows that there was a need for flow cytometrist's to have access to the relevant algorithms that produce this deconvoluted data, especially for the cytometers that are not designed for full spectral analysis.

Our aim was to compare the methodology and results of the following to be able to show data equivalence:

1) The deconvolution algorithm used in the Christopher Hall et.al. study (run using an R based program locally)

2) The same deconvolution algorithm but containerised (to freeze any library codes) to re-run the raw data locally

3) The same deconvolution algorithm (containerised) but run through the cloud-based algorithm processing engine

The raw data from Chris Hall *et al.* study [1] was processed using an 'in house' local and cloud-based algorithm engine, here, the files were scheduled to be run through the containerised algorithm. Output files were then analysed and compared to confirm data equivalence, and methodology used to demonstrate ease of use compared to an R programming platform.

We found that not only was data equivalence achieved but improved by the containerisation of the algorithm, which stops the app library codes in the background updating- which in turn could cause small changes within the algorithm every time it is called upon to use. The containerisation ensures truly repeatable results whether processing locally or in the cloud.

The scheduling characteristics of the 'in house' local and cloud-based algorithm engine made it possible to perform spectral deconvolution without the need for knowledge of R language or programming, and cloud-based algorithm processing free's up computer usage time for other tasks.

Overall, a cloud-based Spectral deconvolution system simplifies access to this technology opening the opportunity for SRLs to use spectral technology benefits such as Autofluorescence removal and simpler panel design.

METHODS AND MATERIALS

In this study we have used the Novocyte Penteon raw data files from the Chris Hall et.al. [1] study (which are available on the flow repository website). For sample preparation and data acquisition details from this study, please see the bioRxiv preprint linkhttps://doi.org/10.1101/2022.12.21.521417. This study seeks to determine if data that is process locally in R, locally in an Open Container Initiative, "OCI", compliant instance or in the cloud in CytoSwarm, which uses OCI compliant instances, is equivalent.

The Kolmogorov-Smirnov (KS) statistic is a nonparametric test that is used to determine if two samples come from the same distribution. The KS test is often used in statistical analysis, such as in comparing the performance of different models as is the case here. The K-S probability ranges from 0 to 1, with values closer to 1 indicating a higher likelihood that the two datasets are drawn from the same distribution, and values closer to 0 indicating a lower likelihood. A K-S probability of 1.0 indicates that the distributions are identical.

To establish data equivalence between the different instances of the flowUnmix algorithm we used the following method:

FIGURE 1

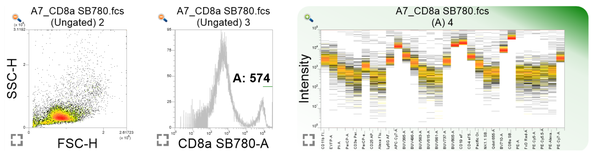


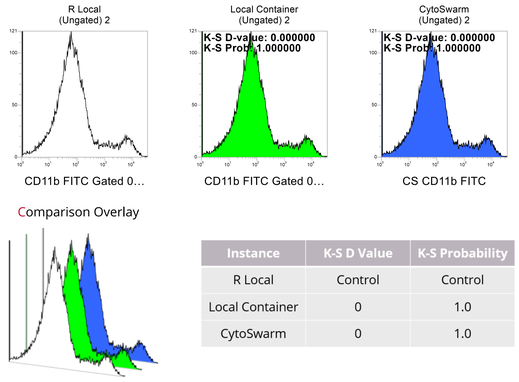
Figure 1 provides an example of the positive gating workspace used for the single stained positive controls, showing CD8a SB780 single colour raw data FCS file.

1. We used the VenturiOne software to gate and save (as FCS files) the positive population for each single stained control, ensuring that we selected a minimum of 500 bright positive events creating a set of positively gated single stain controls.
2. Each implementation of the flowUnmix algorithm was provided with the same full set of positively gated single stain controls, which were used to create the unmix matrix, together with the files which were to be unmixed, this process was performed in triplicate.
3. For each implementation of the flowUnmix algorithm the triplicate outputs were then compared using Kolmogorov-Smirnov statistics in VenturiOne to establish within-implementation data equivalence.
4. The output of the 3 types of implementation of the flowUnmix algorithm were then compared using Kolmogorov-Smirnov statistics in VenturiOne to establish between-implementation data equivalence.

RESULTS

The Kolmogorov-Smirnov (K-S) probability is a statistical measure that ranges from 0 to 1, with values closer to 1 indicating a higher likelihood that the two datasets under comparison are drawn from the same distribution, whereas values closer to 0 suggest a lower likelihood. A K-S probability of 1.0 indicates that the distributions are identical.

FIGURE 2



Instance	K-S D Value	K-S Probability
R Local	Control	Control
Local Container	0	1.0
CytoSwarm	0	1.0

Figure 2 depicts the resultant CD11b FITC histogram obtained from the triplicate-tested, fully-stained, and unmixed datasets for each of the three distinct implementations; namely R local, Local container, and CytoSwarm cloud-based engine.

In the present study, we observed that the K-S D-value was consistently zero when compared to the control in all instances. Furthermore, the K-S probability value was 1.0, indicating that regardless of the technology employed, such as local R, local container, or cloud-based CytoSwarm, the results were indistinguishable.

DISCUSSION

The availability of cloud-based unmixing can be considered a significant advantage for laboratories and Shared Resource Laboratories (SRLs), as it has been demonstrated to be equivalent to using the R language locally, without requiring any programming skills in R. Our research findings indicate that containerizing the algorithm has not only achieved data equivalence but has also prevented any changes in R library components. As a result, long-term stability in unmixing can be achieved.

The reuse of existing conventional cytometers in a spectral role enhances the environmental sustainability of each laboratory. By enabling all users to perform spectral unmixing, a busy SRL can transition from being the primary actor to the primary advisor, saving time in an already busy department. The proposed approach's flexibility and options could prove invaluable to SRLs dealing with data from various flow cytometry instruments.

Furthermore, CytoSwarm, the cloud-based algorithm engine, optimizes computational efficiency by allowing users to use their computer usage time for other tasks.

CONCLUSION

The present study provides empirical evidence of data equivalence through the utilization of the flowUnmix algorithm using three different processing methods.

The implementation of a containerised approach ensures that consistent and identical results can be obtained every time the methodology is applied, offering a valuable tool for researchers to enhance the reliability and accuracy of their analytical outcomes. This is particularly advantageous on a cloud-based platform, which offers high-speed processing and global accessibility, allowing researchers from all corners of the world to benefit from this methodology.

Based on these findings, we postulate that the combination of conventional flow cytometers and innovative cloud-based technologies represents a valuable resource for laboratories and SRLs seeking to optimize their data analysis capabilities. The integration of such platforms offers a unique opportunity to extract additional value from existing resources, thereby enhancing the efficiency and productivity of research. Ultimately, this approach can lead to a more comprehensive understanding of complex biological systems and accelerate the pace of scientific discovery.

ACKNOWLEDGEMENTS

Thanks to Chris Hall *et al* for allowing us to use the raw Penteon data from their study [1]
Thanks to Agilent for use of the Penteon.

REFERENCES

1.Christopher Hall , Hanan Ibrahim , Sam Thompson , Philip S Hobson , Jo-Anne Crofts , Peter Nobes , Steven Lim , Tony Burpee , Rachael V Walker. Back to the Future- Unleashing your cytometer's spectral potential https://doi.org/10.1101/2022.12.21.521417 Submitted to Cytometry Part A

SCAN QR CODE

MADE POSSIBLE WITH VENTURIONE CytoSwarm

FIND OUT MORE