A framework for evaluating the performance of SMLM cluster analysis algorithms

Daniel J. Nieves[1], Jeremy A. Pike[1], Florian Levet[2,3], Juliette Griffié[3], Daniel Sage[3], Edward A.K. Cohen[4], Jean-Baptiste Sibarita[2], Mike Heilemann[5] and Dylan M. Owen[1]

[1] *Centre of Membrane Proteins and Receptors (COMPARE), University of Birmingham, UK.*
[2] *Université de Bordeaux, CNRS, Interdisciplinary Institute for Neuroscience, France.*
[3] *Université de Bordeaux, CNRS, INSERM, Bordeaux Imaging Center, France.*
[4] *Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland.*
[5] *Department of Mathematics, Imperial College London, UK.*
[6] *Institute of Physical and Theoretical Chemistry, Goethe-University Frankfurt, Germany.*

Single molecule localisation microscopy (SMLM) generates data in the form of Cartesian coordinates of localised fluorophores. Cluster analysis is an attractive route for extracting biologically meaningful information from such data and has been widely applied. Despite the range of developed cluster analysis algorithms, there exists no consensus framework for the evaluation of their performance. Here, we use a systematic approach based on two metrics, the Adjusted Rand Index (ARI; **Fig.1a**) and Intersection over Union (IoU; **Fig.1b**), to score the success of clustering algorithms in diverse simulated clustering scenarios mimicking experimental data. We demonstrate the framework using three analysis algorithms: DBSCAN, ToMATo and KDE, show how to deduce optimal analysis parameters and how they are affected by fluorophore multiple blinking. We propose that these standard conditions and metrics become the basis for future analysis algorithm development and evaluation.

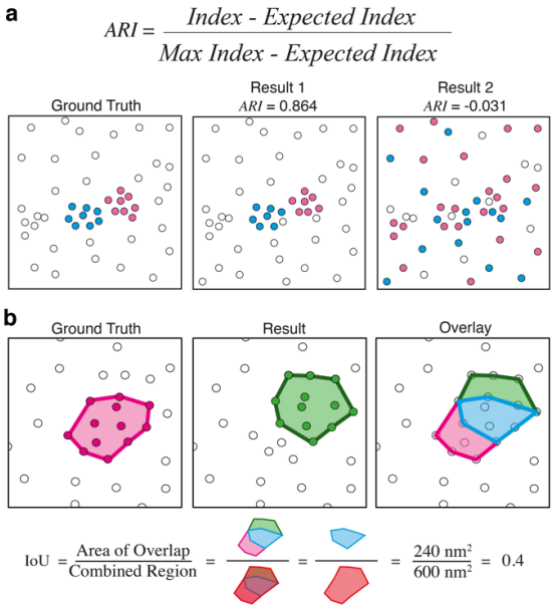

Fig.1 - a) ARI (theoretical range -1 to 1) compares the ground truth clustering to that of the results. Good agreement (Result 1) gives high scores, and bad agreement (Result 2) gives low or negative scores. **b**) IoU scoring uses polygons to compare the clustering result to the ground truth. Overlapping and combined areas are used to calculate the agreement. IoU has a range from 0-1, with good agreement giving high values, and poor agreement low values.