

Review of: "Quantitative Single-Molecule Imaging with Statistical Machine Learning"

Mohamadreza Fazel¹

¹ Arizona State University

Potential competing interests: The author(s) declared that no potential competing interests exist.

In the past two decades, single molecule localization microscopy (SMLM) has become a standard tool in unraveling details of sub-cellular environments. This technique employs independent and sparse blinking of single molecules to localize them with precisions below the diffraction limit of light. The list of resulting localizations can next be employed to either reconstruct high resolution images of cellular organelles or to gain quantitative insight into sub-cellular environments, such as counting single molecules.

This paper is part of a series of papers that leverage tools from statistical analysis to count single molecules within clusters given a list of localizations. In the previous works, the authors assumed that the average number of localizations per molecule, λ , can be precalibrated and given as an input to the algorithm. This paper, however, employs expectation maximization (EM) technique to remove this constraint and learns λ as well as molecular counts. The authors also use AIC method for model selection to pick the best model between monomer, dimer and trimer that fits data in the presence and absence of missed labels.

There are multiple concerns associated to this study discussed in the following:

- 1) It is known that the maximum likelihood estimation techniques, including EM, tend to over-interpret the data. This is manifest in the results presented in figure 1 where for monomer and dimer cases the algorithm still accepts non-zero ratios of trimers (π_3). The authors also proposed the AIC method to select the best model. However, this method as well fails under certain conditions, for example panel IV in figure 1.
- 2) While intuitively more accurate predictions are expected with increasing λ , errors associated to the fraction π_3 increase with larger λ . This observation is not a general trend in the counting problem, but it is probably due to the tendency of the proposed algorithm to over-interpret the data. In principle, larger λ values lead to more localizations, which, in turn, will be interpreted as higher molecule counts and will result in overestimation of π_3 . As such, the proposed algorithm counterintuitively yields higher errors for larger λ .
- 3) The manuscript includes a thorough discussion with several figures regarding the estimation of monomer, dimer and trimer fractions (π_1 , π_2 , π_3) under different circumstances. However, the authors have not explored the algorithm's robustness in estimating λ despite this being one of their main goals.
- 4) There are several published counting algorithms; however, the authors have not compared their results

to any of the existing algorithms.