Correction of multiple-blinking artifacts in photoactivated localization microscopy

Jensen, Louis G.; Hoh, Tjun Yee; Williamson, David J.; Griffié, Juliette; Sage, Daniel; Rubin-Delanchy, Patrick; Owen, Dylan

DOI:
10.1038/s41592-022-01463-w

Correction of multiple-blinking artefacts in photoactivated localization microscopy

Louis G Jensen[1], Tjun Yee Hoh[2], David J Williamson[3], Juliette Griffié[4], Daniel Sage[5], Patrick Rubin-Delanchy[2]* and Dylan M Owen[6]*

Affiliations
[1]Department of Mathematics, Aarhus University, Aarhus, Denmark
[2]Institute for Statistical Science, School of Mathematics, University of Bristol, Bristol, UK
[3]Randall Centre for Cell and Molecular Biophysics, King's College London, London, UK
[4]Laboratory of Experimental Biophysics, Institute of Physics, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
[5]Biomedical Imaging Group, School of Engineering, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
[6]Institute of Immunology and Immunotherapy, School of Mathematics and Centre of Membrane Proteins and Receptors (COMPARE), University of Birmingham, Birmingham, UK.
*Contributed equally
Correspondence to:

louis.gammelgaard@gmail.com; patrick.rubin-delanchy@bristol.ac.uk; d.owen@bham.ac.uk;

Abstract
Photoactivated localization microscopy (PALM) produces an array of localizationordinates by means of photoactivatable fluorescent proteins. However, observations are subject to fluorophore multiple-blinking and each protein is included in the dataset an unknown number of times at different positions, due to localization error. This causes artificial clustering to be observed in the data. We present a "model-based correction" (MBC) workflow using calibration-free estimation of blinking dynamics and model-based clustering to produce a corrected set of localization coordinates representing the true underlying fluorophore locations with enhanced localization precision, outperforming the state-of-the art. The corrected data can be reliably tested for spatial randomness or analyzed by other clustering approaches, and descriptors such as the absolute number of fluorophores per cluster are now quantifiable, which we validate with simulated data and experimental data with known ground truth. Using MBC, we confirm that the adaptor protein, LAT, is clustered at the T cell immunological synapse.

Introduction
Single molecule localization microscopy (SMLM) methods, such as PALM, circumvent the diffraction limit of light by separating fluorophore detections in time through stochastic activation and photobleaching, and then localizing the resulting sparse distribution of point spread functions[1]. The resulting point-pattern is a purported realization of the underlying ground truth positions of the fluorophores, but is corrupted by a number of artefacts resulting from the photophysical behavior of

the probes as well as the imaging and localization steps. Most problematic is the multiple appearance (multiple-blinking) problem whereby fluorophores undergo multiple on-off cycles before permanently bleaching, making a single flourophore appear as several localizations. This issue combined with the discretization effects that result from observing fluorescent signals on discrete camera frames[2]. The multiple-blinking problem produces data sets that are artificially clustered and overly populated (Figure 1a). As such, quantitative cluster analysis of SMLM data, including testing for spatial randomness of the underlying fluorophores, remains a challenge.

The most commonly employed method for correction of the multiple-blinking problem is to merge events that appear close in space and time[3-6], which we refer to as Dark Time Thresholding (DTT). Such methods require a means of determining the best spatial and temporal thresholds for merging. This determination typically relies on heuristic methods, since the blinking behavior of the fluorescent probes is often unknown. Apart from the challenges involved in determining optimal thresholds, these methods have variable performance, depending on the underlying protein organization and fluorophore blinking characteristics. Instead of attempting to produce a corrected version of the data which can then be used for any subsequent analysis, other approaches have looked to correct specific spatial statistics to account for multiple-blinking. For example, it is possible to use calibration data to estimate a multiple-blink corrected pair-correlation curve[7,8]. However this cannot then be used to find a cluster map.

Here, we present a new method for correction of multiple-blinking artefacts in PALM data, which estimates, directly from the sample data set, the parameters of a realistic model of fluorescent protein photophysics[9]. Cluster analysis of the spatial (x,y), temporal (t) and precision (σ) data set then allows computation of the marginal likelihood of any given blink-merge proposal, under a full generative model for the data. We select the most likely of several proposals generated using a customized hierarchical clustering algorithm. Finally, each blink cluster is consolidated into a single position, now free from multiple-blinking. As with other correction methodologies[5], an added benefit of this consolidation step is to improve localization precision. The overall effect is to convert the set of raw x,y,t,σ localisation data into a new set, x,y,σ, with enhanced resolution.

We evaluate the method on simulated PALM data, varying both the ground-truth organization and photophysical properties of the fluorescent proteins. In each case, we compare to the state-of-the-art method of Bohrer et al[10]., and DTT, outperforming alternatives in speed and accuracy. Our method allows for testing the completely spatially random (CSR) hypothesis at the correct significance level, whereas DTT fails to do so. We also validate the method on experimental data in which a notion of ground truth is available, namely nuclear pore complex (NPC) data in which the number of proteins per complex is known to be 32.

PALM is increasingly used in the biological sciences and owing to the properties of commonly used total internal reflection fluorescence (TIRF) illumination, the distributions of membrane proteins have been especially well studied. Despite this, because of artificial clustering resulting from multiple-blinking, the question of whether membrane proteins are randomly distributed or not has become increasingly contentious[11]. Using our validated method combined with subsequent testing of the corrected protein locations, we show that the adaptor protein Linker for Activation of T cells is clustered in the plasma membrane of T cell immune synapses[12,13].

Results

Description of the algorithm

We work with the space-time localizations and uncertainties that result from localization software (here ThunderSTORM[16]) that is run on the raw microscope data. We apply drift correction, but otherwise no pre-processing is used. The data points are then modeled as a collection of independent and identically distributed fluorophore blinking clusters, with times following a realistic 4-state model[17,18], discretized by the camera frames. The spatial locations for each cluster are independently drawn from a spherical Gaussian distribution of fixed center (the true molecule position) and variable but known standard deviation (the localization uncertainty). The centers are given a uniform prior over the region of interest (ROI).

We refer to our algorithm as model-based correction (MBC), and a schematic of its workflow is shown in Figure 1b. We first estimate the temporal rates governing the switching behavior of fluorescent proteins under the 4-state model[9], and the fraction of background noise points. This is done directly using the experimental data, requiring no additional calibration experiments. A recently developed mathematical technique extracts a component from the empirical mark and pair correlation functions which depends only on the spatio-temporal dynamics of the multiple-blinking process, and not the underlying protein distribution. The parameters of the 4-state model drive the theoretical shape of this component, and so they can be optimized to best fit the empirical version[9]. The rate estimates allow computation of the marginal likelihood of a sequence of timepoints posited to correspond to one multiple-blinking fluorescent protein, and further yields an estimate on the total number, N, of proteins and noise points in the ROI. Using a custom agglomerative hierarchical clustering (HC) algorithm[19], we split the data in the ROI into partitions with N categories. HC takes as input a dissimilarity matrix and a linkage criterion. The dissimilarity matrix determines the distances between pairs of points, and the linkage criterion determines the way to generalize this distance to pairs of clusters. To favor groups likely to correspond to multiple-blinking clusters, we first scale the temporal dimension by a time-dilation hyperparameter, S, and then compute the sum of Euclidean distances in space and in time. For linkage, we choose Ward's Minimum Variance Method[20], which is well-suited for Gaussian clusters, and consistently resulted in the most likely partitions across all tested linkage criteria. By varying S, we obtain a large sequence of blinking cluster proposals, and evaluate the marginal likelihood of each. Finally, using the best partition and the localization uncertainties, we optimally merge the clusters down to their estimated centers, using inverse-variance weighted averages, and update the uncertainty associated with that center (example shown in Supplementary Figure 1 and 3D rotatable version in Supplementary Data).

PALM data simulation setup

For a given set of protein positions, corresponding PALM data were generated as follows. We simulated fluorescent protein time traces according to the 4-state switching model (see Figure 1a), and the continuous signals were discretized to emulate a camera operating at 25 frames per second (40 ms integration time). This was done for 2 different sets of rates (given in Supplementary Table 1), with the light blinking resulting in 5.36 appearances per protein on average, and the heavy blinking resulting in 14.94 appearances. These were selected to approximate real blinking behavior of fluorescent proteins such as mEos2. For each of these appearances, the observed spatial coordinates were simulated by adding Gaussian localization noise to the ground-truth position of the associated fluorescent protein, with standard deviation following a Gamma distribution with mean

30 nm and standard deviation 13.4 nm, emulating the localization uncertainties that can be observed in real PALM data[12].

Recovery of the ground truth

We begin by comparing MBC to the method of Bohrer et al., which we take to be the state of the art at the task of recovering protein positions without calibration data. We will conduct this comparison with simulated data from regular (529 fluorescent proteins were regularly positioned on a ~3000 nm x 3000 nm grid), CSR (500 proteins were placed at random in a noiseless 3000 nm x 3000 nm ROI), and clustered distributions. For the latter, 500 ground-truth proteins were placed in a 3000 nm x 3000 nm ROI, with either 10 clusters of 10 molecules each, overlaid with 400 CSR molecules (light clustering) or 10 clusters of 40 molecules each, overlaid with 100 CSR molecules (heavy clustering). The metrics used for the comparison are Wasserstein distance, image error, and counting error. The Wasserstein distance can be thought of as the cost of transporting a standardized mass between two sets of points and is also known as the earth mover's distance. For a perfectly reconstructed dataset this distance is zero, whereas under a model in which each molecule was observed exactly once with localisation error drawn from the Normal-Gamma distribution above, this distance would be roughly 37 nm. In this way, whenever we report a Wasserstein distance below 37 nm, we can conclude that we have exploited, rather than suffered from, the multiple-blinking artefact. Relevant plots show this 37 nm benchmark. The image error measures the protein count error in 100 nm x 100 nm pixels, whereas the counting error refers to the total number of proteins per ROI, calculated as (corrected-truth)/truth.

In the main text we show results at the two extremes, regular (Figure 2) versus heavy clustering (Figure 3), with the two other conditions shown in Extended Data Figure 1. Our method is always superior in Wasserstein distance and image error, in counting error for heavy blinking, and never substantially worse. Our gains are most substantial under heavy clustering with heavy blinking, where Bohrer et al.'s method overcounts proteins by a factor of at least two (Figure 3). Our method is also much faster, taking one or two minutes per ROI on a standard desktop versus several hours.

As an alternative to Bohrer et al, we also consider an idealized version of DTT, in which points are considered to have come from the same fluorophore if they were separated by at most r in space and T in time, chosen to minimize either the counting error (iDTT_N) or Wasserstein distance (iDTT_W). The reason we consider this choice, which is unavailable in practice, is to cover all possible methods for choosing r and T, including those involving calibration data. Results are shown for regularly distributed data under light and heavy blinking in Extended Data Figure 2 respectively. Our method is always superior on Wasserstein distance and image error (even when compared to iDTT_W), and superior in counting error when compared to iDTT_W.

To test MBC in dense scenarios, we repeat all experiments with five times the density, chosen to mimic the maximum density observed in any of the experimental data analyzed in this paper (2500 proteins per 3000 nm x 3000 nm ROI). When it comes to denser scenarios, Bohrer et al's method can take ten days per ROI, and optimizing over space and time thresholds to perform iDTT is also computationally demanding. However, MBC is still functional, and its performance is compared to DTT across all conditions and metrics in Extended Data Figures 3-4. For DTT, we use parameters r equal to 4 times the mean localization uncertainty in space, and T determined for each ROI using the

method of Annibale et al[4]. For each of four scenarios; regular and CSR (Extended Data Figure 3) and light clustering and heavy clustering (Extended Data Figure 4), MBC substantially outperforms DTT in all three metrics (Wasserstein distance, image error and counting error).

Having established that MBC outperforms Bohrer et al.'s method and iDTT for recovery of ground truth in all conditions, at an extremely reduced computational cost, we proceed with comparing our method only to DTT due to its simplicity, speed, and wide uptake (including in commercial software).

Testing for complete spatial randomness

We first evaluate our algorithm for testing for complete spatial randomness of the underlying ground-truth proteins. In each run (n = 30 per condition), 500 proteins were placed at random in a noiseless 3000 nm x 3000 nm ROI. For each ROI, we compute the function $L(r)$-r (Figure 4b), where L is Besag's L function[21], testing its maximum (Figure 4c) under a CSR null hypothesis. The standard DTT correction method was unable to recover the ground-truth functions and resulted in rejection of the CSR null hypothesis in 24 and 30 out of the 30 regions, for light and heavy blinking respectively. On the other hand, MBC resulted in the CSR null hypothesis being rejected for 2 and 4 of the regions for light and heavy blinking respectively. These numbers are within the expected range at a 5% confidence level. Thus, we were able to reliably test the CSR hypothesis using MBC, but not using DTT. The estimated total number of fluorescent proteins in each ROI is shown in Figure 4d. Under CSR, DTT tends to overestimate the number of proteins in the ROI whereas MBC closely recovers the ground truth.

Cluster analysis

In this experiment, we demonstrate that a clustering algorithm can extract correct cluster descriptions from underlying clustered ground truth protein distributions when coupled with MBC, and we compare performance with DTT. We simulated data from protein distributions exhibiting light and heavy clustering (n = 30 per condition), as described in the recovery of ground truth section (Figure 5a). We used Bayesian cluster analysis[14] for detection of clusters in MBC and DTT corrected data sets. Only under MBC could we consistently recover the 10 clusters under varying degrees of blinking severity (Figure 5b). The failure of DTT to recover the correct number of clusters is even more evident in the case of heavy clustering (Figures 5c and d).

Determining optimal imaging conditions

As a final test using known ground-truth simulated data, we used Virtual-SMLM[22] to simulate raw camera frames. This allowed us to test the effect of varying both the camera frame rate and the intensity of the 405 nm activation laser on the performance of MBC. A ground truth of CSR fluorescent proteins were simulated (Extended Data Figure 5a), imaged using the virtual microscope and output analysis with ThunderSTORM. The camera integration time was set to either 10 ms or 40 ms and the 405 nm laser intensity either kept constant, or ramped to maintain a constant density of point spread functions (PSFs) per frame over the course of the acquisition. Raw localizations (Extended Data Figure 5a) were then corrected using MBC (Extended Data Figure 5b). The Wasserstein distance shows marginally superior performance of the reconstruction when using constant 405 nm laser power and when using longer, 40 ms frames. We attribute this to the lower density of PSFs per frame

in the constant-405 case leading to fewer overlapped PSFs during localization and to the increased localization precision offered by the longer frames (Extended Data Figure 5c, d). The performance of MBC itself is only weakly dependent on the imaging conditions, and in each condition we were able to recover the ground truth number of molecules to within around 10% error (Extended Data Figure 5e). We conclude therefore that when using MBC, PALM imaging conditions should be chosen to maximize conventional notions of data quality — low density of PSFs and high signal-to-noise ratio. Because of this, we also conclude that MBC is also backwards compatible with all historically acquired PALM data.

Model validation

MBC is based on a statistical model for multiple blinking, camera discretization, localization error, and a Bayesian prior distribution on the protein positions. In this section we address concerns about model misspecification. First, we test our method on simulated data in which the multiple-blinking process follows a different model. Second, we test our method on nuclear pore complex data in which there is a notion of ground truth, showing accurate recovery in a situation where the blinking dynamics have all the complexities of real fluorophores.

For the first experiment, the multiple blinking follows the 3-dark-state photokinetic model shown in Jensen et al[9] with parameters $r_F = 0.005$, $r_B = 2.5$, $r_{D1} = 4$, $r_{R1} = 0.25$, $r_{D2} = 4$, $r_{R2} = 1$, $r_{D3} = 4$ and $r_{R3} = 10$ (the numeric part of the subscript indicates the dark-state number). Overall, this produces slightly more total blinks than our previous heavy blinking case. Extended Data Figure 6 shows the results of MBC on regularly distributed data perturbed by this multiple blinking process. MBC still performs well when the blinking shows the extra complexity of multiple dark-states, still clearly outperforming DTT in this scenario. We believe the method should be robust, more generally, to misspecification of the blinking model as long as two assumptions can be made: first, the blinking behavior of different fluorophores is independent, and second, the process governing the blinking state switching is time-homogeneous.

Despite the performance on simulated data showing complex photoblinking behavior, real fluorophores in cells and experimental procedures such as imaging and localization have the potential to introduce many more complexities than we could simulate. We therefore investigate the performance of MBC on experimental reference data. We consider real biological cells from a recently developed cell line[23] expressing the nucleoporin Nup96, which forms nuclear pore complexes (NPC). These complexes have a well-characterized ring shape, composed of 32 Nup96 positioned into 8 identical corners. Knowing that each ring holds 32 proteins allows us, in principle, to compare corrected counts to a ground truth value.

We consider 16 independent recordings of cells expressing the fluorophore Nup96-mMaple, as available on the BioImage Archive. Following Thevathasan et al., 2019[23], we first filter out emitters of poor quality using the SMAP software[24]. In addition to filtering, SMAP also detects and segments out the NPCs present in each cell, and estimates the effective labeling efficacy (p_{ele}), the probability that a Nup96-mMaple is detected in the recording. For a given p_{ele}, which will vary for each recording, we can thus expect the number of detectable Nup96-mMaple in an NPC (N_{NPC}) to have a binomial distribution $N_{NPC} \sim \mathrm{Bin}(32, p_{ele})$.

After running MBC on each dataset, we cluster the corrected localizations into NPCs according to the closest NPC center. For each corrected NPC, we compute the number of localizations and divide by the p_{ele} relevant to that cell, to obtain a standardized count. Under an unbiased correction method this standardized count will have a mean value of 32 for all cells, allowing us to analyze and visualize all 16 cells simultaneously, see Extended Data Figure 7. By comparison to the theoretical distribution of ground-truth standardized counts (corresponding to perfect recovery) we see that MBC is closely emulating this optimal distribution. The percentage error on the mean number of proteins per NPC between MBC and the idealized case is only 1.1%, showing that MBC exhibits minimal bias. Our standard deviation is 6.16 and the ideal case is 4.6 (Bohrer et al. reported 11). Overall, experimental data acquired with a known ground truth show that MBC is accurate and unbiased in correcting multiple-blinking even through the full complexity of real fluorophore photophysics confounded by potential imaging and localization artefacts.

Analysis of experimental data

Nanoscale clustering is posited to play a role in regulating protein-protein interactions and therefore the efficiency of signaling propagation along pathways[25]. T cell microclusters of proximal signaling molecules have been widely documented by conventional microscopy[26,27]. Many have recently been studied by SMLM and shown to also cluster on the nanoscale[12,13,28,29]. This has proved controversial however, with counter-proposals that, in some circumstances, proteins may in fact be randomly distributed on the cell surface, with observed clustering attributed to multiple-blinking artefacts[11]. For PALM data, MBC should enable researchers to navigate this controversy.

We analysed the distribution of the adaptor protein LAT[30] in the plasma membrane of the Jurkat CD4+ Helper T cell line at an artificial immune synapse formed against an activating, antibody coated coverslip (see Online Methods). To assess the role of intracellular phosphorylation in maintaining this distribution, we also mutated intracellular tyrosine residues to phenylalanine (YF LAT). Both wild-type (WT) LAT and YF LAT were fused to the photoconvertible fluorescent protein mEos3.2 with cells imaged under TIRF illumination. Raw localizations were obtained using ThunderSTORM and then corrected using MBC, with the average localization uncertainty across experiments reduced by 40.52%. The resulting corrected localizations were then tested for spatial randomness using the L-function, and any regions found to be clustered subjected to Bayesian cluster analysis[14] (using default parameters).

Figure 6 shows WT and YF LAT-mEos3.2 from representative regions (from n = 12 - 25 ROIs from 3-6 cells) acquired from the central regions of the cell synapse and from the synapse periphery, both before (Figure 6a) and after (Figure 6b) correction using MBC. Clearly, the large, dense clusters evident in the uncorrected data in all conditions are reduced in the corrected regions. However, by analysing the L-function curves from the ROIs (Figure 6c) and extracting the maximum value of those curves (Figure 6d), we were able to perform significance testing on whether the LAT distributions were truly CSR. For the two WT LAT conditions, the null hypothesis that LAT is randomly distributed was rejected in most regions. Therefore, it is likely that WT LAT was clustered in most analyzed WT ROIs. This was not true for the YF mutant however, with the null hypothesis of randomly distributed LAT not rejected in the majority of peripheral regions (Figure 6d). This therefore may point to a role of intracellular tyrosine phosphorylation in maintaining LAT clustering.

For all regions where the CSR null hypothesis was rejected we then further interrogated the data using Bayesian cluster analysis. For WT LAT, the data showed no statistically significant difference in cluster membership between central and peripheral regions. However, the YF mutant showed a significant decrease in the number of molecules per cluster in peripheral regions, both when compared to YF central regions ($p = 0.026$) and WT peripheral regions ($p = 0.001$) (Figure 6e). Other outputs from the cluster analysis are shown in Extended Data Figure 8, with p-values summarized in Supplementary Table 2. The decrease in cluster membership and, in some ROIs, the loss of clustering altogether, in peripheral regions of the T cell synapse resulting from the YF mutation, is a strong indication that intracellular tyrosine phosphorylation is involved in maintaining LAT signaling clusters. Signaling phosphorylation events are known to originate in the synapse periphery and it is therefore consistent that the effect of the mutation is most pronounced there, compared to the central region where signaling is terminated[31].

Discussion

Super-resolution fluorescence microscopy by SMLM, such as PALM, results in a pointillist dataset representing an attempted realization of the underlying ground-truth fluorophore locations[1]. A common goal in the biological sciences is to test whether such underlying distributions are clustered or randomly distributed and, if clustered, to determine their clustering properties. Achieving this has proved difficult however, because the generated localizations are corrupted by artefacts, principally the repeated localization of the same fluorophore due to multiple-blinking[2]. This has led to controversy about whether proteins are truly clustered in cells, hindering our understanding of the causes and function of nanoscale protein clustering.

Here, we develop an algorithm, MBC, for correcting multiple-blinking that requires no user input, no additional calibration data, and is not limited to a specific analysis goal. We show that it can be used to reliably test for spatial randomness or recover other clustering properties from the ground truth.

A number of methods have been put forward to test for spatial randomness in SMLM data. These include methods based on varying the labelling density and observing the effects on specific cluster analysis outputs[32] or by labeling the same species with two different fluorophores allowing a cross-comparison to be made[33]. These, however, require multiple sample preparation rounds and are therefore more complex and time-consuming. Correction can also be made by measuring blinking behavior in a separate sample of well-isolated fluorophores[8], but this again adds complexity and experimental effort and requires the assumption that probe photophysics are maintained between the sample and the calibration. It is also possible to measure or simulate multiple-blinking using realistic photophysical models and use these to derive new CSR confidence intervals for the L-function curves[34]. It should be noted however that none of these methods produces a new set of corrected positions.

An approach with the same goal as ours is the method of Bohrer et al[10]. We perform a comprehensive comparison to this approach, showing MBC achieves superior accuracy at a fraction of the computational cost. Performance gains are often substantial in both measures, e.g., the method of Bohrer et al. can overcount the number of proteins by up to a factor of two, and can take more than a week per ROI (where our method takes 2 minutes). MBC demonstrates a sufficiently accurate recovery of ground truth to reliably test the CSR hypothesis. MBC therefore represents a new

capability - of obtaining a set of corrected ground-truth locations of sufficient quality that any subsequent statistical analysis can be conducted with assurance.

The limitations of MBC are as follows. The method is only applicable to photophysical models typical of PALM acquisitions, and therefore cannot be used to correct dSTORM or other SMLM modalities. Performance of the correction will decrease as the clustering of the ground-truth increases, however, it tested favorably with realistic and heavily clustered scenarios. The method also adds computational time to any analysis pipeline. For a 3000 nm x 3000 nm ROI containing 500 ground-truth proteins, we estimate the MBC step to take 2-4 minutes per ROI on a standard desktop computer. Of course, as it results in fewer points per ROI, subsequent analysis will typically be accelerated. While here the correction is limited to 2D data, it can in principle be adapted to 3D (x,y,z) coordinates. In this case, the algorithm would need to take account of the differing localization precision in z and potentially the lower detection efficiency of proteins deeper into the sample. Of course, there are many other sources of potential error in SMLM other than multiple-blinking. These include low detection efficiency[35,36], endogenous expression, overlapped PSFs, drift, etc which are not addressed by MBC. The resulting point patterns generated by MBC should therefore still be interpreted with these in mind.

In conclusion, MBC allows for accurate recovery of ground-truth fluorophore positions, with enhanced precision, from PALM data sets subjected to multiple-blinking artefacts. Corrected sets are of sufficient quality to allow accurate cluster analysis and the statistical testing for complete spatial randomness. We therefore believe that PALM combined with MBC will be an invaluable tool for addressing questions on the existence, determinants and functions of protein nanoscale clustering.

Acknowledgements

L.G.J was supported by the Centre for Stochastic Geometry and Advanced Bioimaging, funded by grant 8721 from the Villum Foundation. D.O. acknowledges funding from BBSRC grant BB/R007365/1. P.R.D. acknowledges funding from the BBSRC grant BB/R007837/1. We also acknowledge the use of the King's College London Nikon Imaging Centre (NIC). The authors would like to thank Ute Hahn (Aarhus University) for motivating this project in initial discussions. We acknowledge Jonas Ries and EMBL for supplying NPC data.

Author contributions

L.G.J. developed the software. L.G.J., T.Y.H., D.M.O. and P.R-D conceived the experiments. L.G.J. and T.Y.H. ran analysis. L.G.J. and T.Y.H. provided simulated data. D.J.W., J.G. and D.S. provided additional simulated and experimental data. L.G.J., T.Y.H., D.M.O. and P.R-D wrote the manuscript. L.G.J. and P.R-D. conceived the method.

Competing interests

The authors declare no competing interests

Figure legends

Figure 1: Illustration of the MBC workflow. a) During PALM image acquisition and subsequent localisation steps, the ground-truth protein positions are corrupted by multiple-blinking in combination with discretisation by the camera frames and scrambling by the localisation uncertainty,

resulting in a data set which is over-populated and over-clustered. b) Our algorithm (MBC) takes as input x,y,t,σ data and estimates the rate parameters of a 4-state photophysical model, from which it derives the total number of molecules in the ROI. This is then used as input to a hierarchical clustering step (experimental data shown with colours representing the clusters found), after which clusters are merged to their centres, creating a new dataset free from multiple-blinking and with enhanced localisation precision.

Figure 2: Comparison of MBC with the state-of-the-art, with molecules on a fixed grid. a) Representative ground truth, blinking and corrected data (from one of n = 50 realisations). b,c,d) Wasserstein distances (b), image error (c), counting error (d), between corrected data and ground truth. In b), the dashed horizontal line shows the 37nm benchmark and c,d) the dashed horizontal line shows the optimal value 0. The columns show different blinking conditions (LB: light blinking, HB: heavy blinking) and correction methods (MBC: our method, Boh: Bohrer et al.'s method). Our method has superior or comparable performance on all metrics. Box plots show median, 25th and 75th percentiles and min, max.

Figure 3: Comparison of MBC with the state-of-the-art, with heavily clustered molecules. a) Representative ground truth, blinking and corrected data (from one of n = 30 realisations). b,c,d) Wasserstein distances (b), image error (c), counting error (d), between corrected data and ground truth. In b), the dashed horizontal line shows the 37nm benchmark and c,d) the dashed horizontal line shows the optimal value 0. The columns show different blinking conditions (LB: light blinking, HB: heavy blinking) and correction methods (MBC: our method, Boh: Bohrer et al.'s method). Our method has superior performance on all metrics. Box plots show median, 25th and 75th percentiles and min, max.

Figure 4: Testing for spatial randomness. a) Representative simulated data of ground-truth CSR points with light or heavy blinking (LB or HB) either corrected by MBC or DTT as a comparison (n = 30 simulations). b) L(r)-r (mean in solid line) with pointwise 95% quantile bands (dashed line). c) max(L(r)-r) derived from these functions. Points in red correspond to ROIs that were rejected as CSR in a Monte-Carlo test ($p < 0.05$). Note that DTT often (and sometimes always) incorrectly rejects the CSR null hypothesis, whereas MBC does not. d) Number of molecules per ROI (log-scaled) showing superior correction of MBC compared to DTT in light and heavy blinking cases. Box plots show median, 25th and 75th percentiles and min, max. p-values calculated based on one-sided Monte-Carlo test using 10,000 CSR simulations.

Figure 5: Testing on clustered ground-truth data sets. a) Low levels of clustering with either light or heavy blinking, corrected by MBC or DTT (n = 30 simulations). b) Number of detected clusters (true number of clusters in dashed line) by Bayesian analysis. c) High levels of clustering with either light or heavy blinking, corrected by MBC or DTT. d) Number of detected clusters (true number of clusters in dashed line) by Bayesian analysis. MBC has superior performance in all cases except heavy clustering/light blinking, where results are comparable. Box plots show median, 25th and 75th percentiles and min, max.

Figure 6: Cluster analysis of LAT-mEos3.2 at the T cell immunological synapse. a) Representative, raw 3000 nm x 3000 nm ROIs from each of the 4 conditions (WT Centre (n = 16 regions), WT Periphery (n = 19 regions), YF Centre (n = 12 regions) and YF Periphery (n = 25 regions)). b) Representative MBC-corrected ROIs, on which analysis was conducted. c) L(r)-r (mean in solid line) with pointwise 95% quantile bands (dashed line). d) max(L(r)-r) derived from these functions. Points coloured in red correspond to ROIs where the CSR null hypothesis was rejected in a Monte Carlo test ($p < 0.05$). These ROI were then retained for subsequent Bayesian cluster analysis. e) Number of proteins per cluster detected by Bayesian analysis. Box plots show median, 25^{th} and 75^{th} percentiles and min, max. p-values calculated based on one-sided Monte-Carlo test using 10,000 CSR simulations.

Editor's summary

A model-based correction (MBC) algorithm offers fast and accurate correction of multiple-blinking artifacts in PALM data. MBC outperforms other algorithms in both speed and accuracy and improves quantitative downstream image analysis.

References

1. Betzig, E. et al. Imaging Intracellular Fluorescent Proteins at Nanometer Resolution. Science 313, 1642-1645 (2006).
2. Annibale, P., Scarselli, M., Kodiyan, A. & Radenovic, A. Photoactivatable Fluorescent Protein mEos2 Displays Repeated Photoactivation after a Long-Lived Dark State in the Red Photoconverted Form. The Journal of Physical Chemistry Letters 1, 1506-1510 (2010).
3. Annibale, P., Vanni, S., Scarselli, M., Rothlisberger, U. & Radenovic, A. Identification of clustering artefacts in photoactivated localisation microscopy. Nature Methods 8, 527-528 (2011).
4. Annibale, P., Vanni, S., Scarselli, M., Rothlisberger, U. & Radenovic, A. Quantitative photo activated localisation microscopy: unraveling the effects of photoblinking. PLoS One 6, e22678-e22678 (2011).
5. Lee, S.-H., Shin, J.Y., Lee, A. & Bustamante, C. Counting single photoactivatable fluorescent molecules by photoactivated localisation microscopy (PALM). Proc Natl Acad Sci U S A 109, 17436-17441 (2012).
6. Levet, F. et al. SR-Tesseler: a method to segment and quantify localisation-based super-resolution microscopy data. Nature Methods 12, 1065-1071 (2015).
7. Veatch, S.L. et al. Correlation Functions Quantify Super-Resolution Images and Estimate Apparent Clustering Due to Over-Counting. PLoS One 7, e31457 (2012).
8. Sengupta, P. et al. Probing protein heterogeneity in the plasma membrane using PALM and pair correlation analysis. Nature methods 8, 969-975 (2011).
9. Jensen, L.G., Williamson, D.J. & Hahn, U. Semiparametric point process modelling of blinking artefacts in PALM. BiorXiv (2021).

10. Bohrer, C.H. et al. A pairwise distance distribution correction (DDC) algorithm to eliminate blinking-caused artefacts in SMLM. Nature Methods 18, 669-677 (2021).
11. Rossboth, B. et al. TCRs are randomly distributed on the plasma membrane of resting antigen-experienced T cells. Nature Immunology 19, 821-827 (2018).
12. Williamson, D.J. et al. Pre-existing clusters of the adaptor Lat do not participate in early T cell signaling events. Nat Immunol 12, 655-662 (2011).
13. Lillemeier, B.F. et al. TCR and Lat are expressed on separate protein islands on T cell membranes and concatenate during activation. Nature Immunology 11, 90-96 (2010).
14. Rubin-Delanchy, P. et al. Bayesian cluster identification in single-molecule localisation microscopy data. Nature Methods 12, 1072-1076 (2015).
15. Griffié, J. et al. A Bayesian cluster analysis method for single-molecule localisation microscopy data. Nature Protocols 11, 2499-2514 (2016).
16. Ovesný, M., Křížek, P., Borkovec, J., Švindrych, Z. & Hagen, G.M. ThunderSTORM: a comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging. Bioinformatics 30, 2389-2390 (2014).
17. Fricke, F., Beaudouin, J., Eils, R. & Heilemann, M. One, two or three? Probing the stoichiometry of membrane proteins by single-molecule localisation microscopy. Scientific Reports 5, 14072 (2015).
18. Coltharp, C., Kessler, R.P. & Xiao, J. Accurate construction of photoactivated localisation microscopy (PALM) images for quantitative measurements. PLoS One 7, e51725 (2012).
19. Cormack R.M. A review on classification. J. Royal Stat. Soc. 134(3), 321-367 (1971).
20. Ward, J.H. Hierarchical Grouping to Optimize an Objective Function. Journal of the American Statistical Association 58, 236-244 (1963).
21. Diggle, P.J. On Parameter Estimation and Goodness-of-Fit Testing for Spatial Point Patterns. Biometrics 35, 87-101 (1979).
22. Griffié, J. et al. Virtual-SMLM, a virtual environment for real-time interactive SMLM acquisition. bioRxiv, 967893 (2020).
23. Thevathasan, J.V. Nuclear pores as versatile reference standards for quantitative super-resolution microscopy. Nature Methods 16, 1045-1053 (2019).
24. Ries, J. SMAP : A modular super-resolution analysis platform for SMLM data. Nature Methods 17, 870-872 (2020).
25. Roob, E., 3rd, Trendel, N., Rein Ten Wolde, P. & Mugler, A. Cooperative Clustering Digitizes Biochemical Signaling and Enhances its Fidelity. Biophys J 110, 1661-1669 (2016).
26. Grakoui, A. et al. The immunological synapse: a molecular machine controlling T cell activation. Science 285, 221-227 (1999).
27. Purbhoo, M.A. et al. Dynamics of Subsynaptic Vesicles and Surface Microclusters at the Immunological Synapse. Science Signaling 3, ra36-ra36 (2010).
28. Rossy, J., Owen, D.M., Williamson, D.J., Yang, Z. & Gaus, K. Conformational states of the kinase Lck regulate clustering in early T cell signaling. Nat Immunol 14, 82-89 (2013).
29. Razvag, Y., Neve-Oz, Y., Sajman, J., Reches, M. & Sherman, E. Nanoscale kinetic segregation of TCR and CD45 in engaged microvilli facilitates early T cell activation. Nature Communications 9, 732 (2018).

30. Balagopalan, L., Kortum, R.L., Coussens, N.P., Barr, V.A. & Samelson, L.E. The linker for activation of T cells (LAT) signaling hub: from signaling complexes to microclusters. J Biol Chem 290, 26422-26429 (2015).
31. Varma, R., Campi, G., Yokosuka, T., Saito, T. & Dustin, M.L. T cell receptor-proximal signals are sustained in peripheral microclusters and terminated in the central supramolecular activation cluster. Immunity 25, 117-127 (2006).
32. Baumgart, F. et al. Varying label density allows artefact-free analysis of membrane-protein nanoclusters. Nature Methods 13, 661-664 (2016).
33. Arnold, A.M. et al. Verifying molecular clusters by 2-color localisation microscopy and significance testing. Scientific Reports 10, 4230 (2020).
34. Platzer, R. et al. Unscrambling fluorophore blinking for comprehensive cluster detection via photoactivated localisation microscopy. Nature Communications 11, 4993 (2020).
35. Dursic, N. et al. Single-molecule evaluation of fluorescent protein photoactivation efficiency using an in vivo nanotemplate. Nature Methods 11, 156-162 (2014).
36. Shivanandan, A., Unnikrishnan, J. and Radenovic, A. Accounting for Limited Detection Efficiency and Localization Precision in Cluster Analysis in Single Molecule Localization Microscopy. Plos ONE 10(3): e0118767 (2015).

Methods

Sample preparation

For LAT images, Jurkat E6.1 cells (ECACC 88042803) expressing LAT-mEos3.2 (wild-type, WT LAT, or signalling deficient mutant, YF LAT) were introduced to anti-CD3 (at 2 μg/ml; eBioscience clone OKT3, 16-0037-81) and anti-CD28 (at 5 μg/ml; RnD Systems, clone CD28.2, 16-0289-85) coated glass-bottomed chamber slides (#1.5 glass, ibidi μSlides) at 50×10^3 cells/cm^2 in warm HBSS and incubated at 37°C for 5 minutes to allow for synapse formation. The chamber wells were gently washed with warm HBSS and then fixed in 3% paraformaldehyde in phosphate-buffered saline (PBS) for 20 minutes at 37°C. Fixed cells were washed five times in PBS and used immediately for PALM imaging.

Imaging

PALM image sequences were acquired on a Nikon N-STORM system in a TIRF configuration using a 100 × 1.49 NA CFI Apochromat TIRF objective for a pixel size of 160 nm and running NIS Elements software v4.6. Samples were continuously illuminated with 561 nm laser light at approximately 2 kW/cm^2 and 405 nm laser light (to induce photo-conversion) at approximately 2 W/cm^2. Images were recorded on an Andor IXON Ultra 897 EMCCD with an electron multiplier gain of 200 and pre-amplifier gain profile 3 to a centered 256 × 256 pixel region at 40 ms per frame for 5,000 to 15,000 frames.

Virtual microscope simulations

Raw camera frames were generated using Virtual-SMLM[22] operating in PALM mode (i.e., using a 4 state photophysical model). The frame rate was set to 25 or 100 frames per second. The activation laser (i.e. initial state transition) was either fixed or ramped up over the acquisition. In the first case, the number of fluorophores emitting per frame decreases over time. In the second case, it remains constant over the acquisition. Emission traces were generated independently for each fluorophore and imaging continued until all fluorophores had been imaged and bleached. All other state transition probabilities and photophysics properties were fixed to mimic mEos blinking characteristics. The PSFs were recorded on a virtual EMCCD camera, with an EM gain fixed at 300. Virtual-SMLM took as input ground truth maps of mEos2 positions. 5556 mEos proteins were placed randomly over a 10000 nm x 10000 nm 2D area. Generated camera frames were then analysed using ThunderSTORM and the data cropped into non-overlapping 3000 x 3000 nm regions.

Localisation

Localisation of fluorophore coordinates were reconstructed using ThunderSTORM[16] and corrected for sample drift using cross-correlation of images from 5 bins at a magnification of 5. No further post-processing was performed.

Mathematical details

1. ***Marginal likelihood of clusters***

We represent the observed process by a series of localisations $(X_i, Y_i)_{i=1}^n \in R$ with associated 'blink' times $T_1, \dots, T_n$, and localisation uncertainties $\sigma_1^2, \dots, \sigma_n^2$, where $R = [x_0, x_1] \times [y_0, y_1]$ is the region of interest.

For a given partition of the localisations into groups, we compute the marginal likelihood of the data as follows. Consider a group comprising the observations $1, \dots, m$, with $1 \leq m \leq n$, *posited to correspond to one, distinct molecule*. In particular, we defer until later the treatment of background noise.

The independence assumptions set out in the main article result in the following marginal likelihood factorisation:

$$p\{(X_i)_{i=1}^m, (Y_i)_{i=1}^m, (T_i)_{i=1}^m\} = p\{(X_i)_{i=1}^m\} p\{(Y_i)_{i=1}^m\} p\{(T_i)_{i=1}^m\}.$$

Denoting by $\mu = (\mu_X, \mu_Y)$ the true position of the molecule, the spatial components above have likelihood (given only for $(X_i)_{i=1}^m$)

$$\begin{aligned} p\{(X_i)_{i=1}^m | \mu_X\} &= \prod_{i=1}^m \sqrt{(2\pi\sigma_i^2)} \exp\left\{-\frac{1}{2}\left(\frac{X_i - \mu_X}{\sigma_i}\right)^2\right\} \\ &= (2\pi)^{-m/2} \left(\prod \sigma_i\right)^{-1} \exp\left\{-\frac{1}{2}\sum \eta_i (X_i - \mu_i)^2\right\}, \end{aligned}$$

where $\eta_i = 1/\sigma_i^2$. Defining the weighted mean

$$\tilde{X} = \frac{\sum \eta_i X_i}{\sum \eta_i}$$

We find

$$p\{(X_i)_{i=1}^m | \mu_X\} = (2\pi)^{-m/2} \left(\prod \sigma_i\right)^{-1} \exp\left\{-\frac{1}{2}\sum \eta_i \left(X_i - \tilde{X}\right)^2\right\} \times \exp\left\{-\frac{\sum \eta_i}{2}\left(\mu_i - \tilde{X}\right)^2\right\}.$$

Placing a uniform prior on μ_X, we find

$$\begin{aligned} p\{(X_i)_{i=1}^m | \mu_X\} &= (2\pi)^{-m/2} \left(\prod \sigma_i\right)^{-1} \exp\left\{-\frac{1}{2}\sum \eta_i \left(X_i - \tilde{X}\right)^2\right\} \\ &\times (x_1 - x_0)^{-1} \int_{x_0}^{x_1} \exp\left\{-\frac{\sum \eta_i}{2}\left(\mu_i - \tilde{X}\right)^2\right\} d\mu_X \\ &= (2\pi)^{-m/2} \left(\prod \sigma_i\right)^{-1} \exp\left\{-\frac{1}{2}\sum \eta_i \left(X_i - \tilde{X}\right)^2\right\} \\ &\times (x_1 - x_0)^{-1} \left(2\pi \sum \eta_i\right)^{1/2} \left\{\Phi\left(\frac{x_1 - \tilde{X}}{\sqrt{\sum \eta_i}}\right) - \Phi\left(\frac{x_0 - \tilde{X}}{\sqrt{\sum \eta_i}}\right)\right\}. \end{aligned}$$

The temporal component has likelihood

$$p((T_i)_{i=1}^m) = p(u)p(T_{min})\prod_k p(f_k)\prod_k p(d_k),$$

where each term is computed as follows. The blink times $T_1, \dots, T_m$ are not typically observed exactly, and instead one has access only to associated frame numbers $F_1, \dots, F_m$, taken to represent (small) windows of time containing them, We therefore consider a visit to the fluorescent state to be a block of $L \geq 1$ contiguous fluorescent frames (or consecutive frame numbers), and impute the length ofthis visit to be the time elapsed over $L - 1$ frames, to obtain auxiliary quantities

$$f_k : \text{time spent in fluorescent state } (k\text{th visit}).$$

Up to discrete approximation error, each f_k represents the minimum of two exponential random variables with respective rates r_D and r_B, with likelihood contribution

$$p(f_k) = (r_D + r_B)e^{-(r_D+r_B)f_k}.$$

Similarly, let d_k denote the time elapsed over the kth interval between non-contiguous frames, taken to represent

$$p(d_k) = r_R e^{-r_R d_k}\frac{r_D}{r_D + r_B}.$$

The initial switch of the fluorophore to the activated state happens at time T_{min}, computed simply as the minimum T_i value, and this contributes

$$p(T_{min}) = r_F e^{-r_F T_{min}},$$

to the likelihood.
Finally, let u denote the time since the last blink (a period during which it is unknown whether the process has entered a dark or bleached state). The final contribution is

$$p(u) = \frac{r_D}{r_B + r_D}e^{-r_R u} + \frac{r_B}{r_B + r_D}$$

To finalise calculations, one must account for background noise (in the case of $m = 1$). Such points are assumed to be uniform in spacetime. The complete marginal likelihood is

$$\pi((X_i)_{i=1}^m, (Y_i)_{i=1}^m, (T_i)_{i=1}^m) = \begin{cases} (1-\alpha)p((X_i)_{i=1}^m, (Y_i)_{i=1}^m, (T_i)_{i=1}^m) & m > 1 \\ \alpha V^{-1} + (1-\alpha)p((X_i)_{i=1}^m, (Y_i)_{i=1}^m, (T_i)_{i=1}^m) & m = 1 \end{cases}$$

where $V = T(x_1 - x_0)(y_1 - y_0)$, T is the length pf the period of observation, and α is the background probability.

2. *Identifying and summarising clusters*

For both MBC and DTT clustering, an expected number of clusters, *N*, is first estimated, and a version of agglomerative hierarchical clustering (AHC) is then used to partition the dataset into *N* clusters. In AHC, each point is initially considered to be a distinct cluster. Using a user-specified metric and a linkage criterion, a stepwise greedy merging of the closest clusters is repeated until a partition with a predetermined number of clusters is obtained, or until no more clusters can be merged with a distance less than some specified number. The metric determines the distances between pairs of points, and the linkage criterion generalises these to a distance between clusters. Once the final partition has been identified, we merge each cluster down to its estimated centre, and the uncertainty of the centre is computed. In the following, we use the notation

$$L = (X, Y)$$
$$d_l(L_1, L_2) = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2}$$
$$d_T(T_1, T_2) = |T_1 - T_2|$$

3. *MBC clustering*

For MBC, the number of desired clusters, *N*, is an output of the rate-estimation step, and is thus decoupled from the clustering problem. For the AHC step, we use the family of metrics

$$d_s\big((L_1, T_1), (L_2, T_2)\big) = \frac{d_l(L_1, L_2)}{(\sigma_1 + \sigma_2)} + S \cdot d_t(T_1, T_2)$$

For $S \geq 0$, which is simply the sum of the Euclidean distance between the locations and (scaled) times. For the linkage criterion we chose Ward's Minimum Variance Method, as implemented via the Lance-Williams formula[37], as it tends to find homogeneous clusters of spherical shape.
By varying *S* along a grid, we obtain a sequence of partitions, each slightly different but all chosen to have *N* clusters. By evaluating the marginal likelihood for the partitions obtained at each value of *S*, we then pick the most likely partition among those considered.

4. *DTT clustering*

DTT, or dark time thresholding, is a general idea in SMLM blinking correction literature, but implementation details are rarely discussed. The general principle is to merge locations that are close in space and time, with hard thresholds on the maximally allowed bridging distances in space and time. As a way to implement this idea in the AHC framework, we define the distance between two observations as

$$d_{\tau_s,\tau_t}\big((L_1, T_1), (L_2, T_2)\big) = d_l(L_1, L_2) + d_t(T_1, T_2) + \infty \cdot 1\big(d_l(L_1, L_2)\big) > \tau_s \text{ or } d_t(T_1, T_2) > \tau_t$$

where $\infty \cdot 0 = 0$. Although not strictly a metric, this distance measure allows us to implement the dark time thresholding idea. We use the single-linkage criterion for cluster merging, which considers the distance between two clusters to be the smallest pairwise distance between them. Combined with our metric, this means that the clustering algorithm is allowed to merge points and clusters, so long as they can be combined via paths that do not violate the hard thresholds. Finally, a clustering is achieved by continuing to merge clusters until only infinite distances between clusters remain (no

more legal merges can be made). For τ_s, we used 4 times the mean localisation uncertainty. The temporal threshold, τ_t, was determined as follows. First, the method of Annibale et al was used to determine N[4]. Next, τ_t was increased incrementally until the AHC algorithm produced a partition with N clusters, or as close to N as possible.

5. Cluster centres and uncertainty

Let $(X_i, Y_i)_{i=1}^m$ be the coordinates of an arbitrary cluster with centre μ. Once a particular clustering is given, it makes sense to treat the cluster centres as fixed parameters to be estimated. Thinking therefore of μ as fixed, the maximum likelihood estimator, $\hat{\mu}$, maximises the likelihood of the cluster coordinates

$$\log(p)\{(X_i, Y_i)_{i=1}^m\} = -\frac{\sum \eta_i}{2}\left(\left(\mu_X - \tilde{X}\right)^2 + \left(\mu_Y - \tilde{Y}\right)^2\right) + C$$

Where C does not depend on μ and it follows immediately that

$$\hat{\mu} = \left(\tilde{X}, \tilde{Y}\right) = \left(\frac{\sum \eta_i X_i}{\sum \eta_i}, \frac{\sum \eta_i Y_i}{\sum \eta_i}\right).$$

Using $\hat{\mu}$, we can estimate the position of the molecule associated with a given cluster. As the coordinates of $\hat{\mu}$ are i.i.d., the covariance matrix of $\hat{\mu}$ is given as

$$\mathbb{V}[\hat{\mu}] = \tilde{\sigma}^2 I_2$$

where $I_2$is the 2×2 identity matrix, and

$$\tilde{\sigma}^2 = \mathbb{V}[\tilde{X}] = \frac{\sum \eta_i^2 \sigma_i^2}{\left(\sum \eta_i^2\right)^2} = \frac{1}{\sum \frac{1}{\sigma_i^2}}$$

And the updates localisation uncertainty is then simply the associated standard deviation

$$\tilde{\sigma} = \frac{1}{\sqrt{\sum \frac{1}{\sigma_i^2}}}$$

Justification of optimisation strategy

It may seem wasteful to correct the localisations in two steps, rather than a more integrated approach. However if, in a single step, we optimise the model likelihood for blinking parameters and data partitions simultaneously, the maximum likelihood clustering overfits in a catastrophic way, giving each position its own cluster, and fitting a degenerate blinking model (infinite bleach rate). A more sophisticated attempt would be to take a fully Bayesian approach, with priors on the blinking dynamics, but this has two disadvantages: the first is added and potentially catastrophic sensitivity

to model misspecification, for example, multiple dark states; second, this approach introduces substantial computational issues, which can only be addressed by further layers of approximation (e.g. Monte Carlo, or variational inference) and analysis parameters (burnin, runtime, step-size, etc). The convergence of Markov Chain Monte Carlo is highly uncertain in any Bayesian clustering problem: to quote an influential source, "Although we may be somewhat presumptuous, we consider that almost the entirety of MCMC samplers implemented for mixture models has failed to converge!"[38].

We now demonstrate that, once the blinking parameters are estimated and held fixed, a) the likelihood does not overfit the data partition, and b) our simple search strategy gets close to the optimum. To do this, we implement a more flexible clustering approach, Bayesian hierarchical clustering (BHC), in which clusters are greedily merged according to the largest increase in likelihood. Despite an efficient implementation, e.g. exploiting the Lance-Williams update, this algorithm is too time costly to be suggested for general use, but it provides a benchmark for the performance that could be obtained with a more involved approach and more computational power. In Supplementary Figure 2, we show the likelihood obtained by BHC in a range of different conditions, noting that the algorithm starts with a partition separating all points and ends with all points in one partition (so obtains a likelihood for every possible number of clusters). We observe that a) the highest likelihood partition found by BHC does not routinely overfit (overestimate the number of clusters) and b) the likelihood of the MBC and BHC solutions are close. MBC and BHC are also comparable in terms of other metrics (e.g. each wins roughly half the time in Wasserstein distance).

Significance testing

The p-values reported in Table 1 are based on a two-sided permutation test of the absolute difference of means, using 10,000 simulations.

Data availability

A data simulator to recapitulate simulated data conditions and raw experimental data (point clouds) are available at https://github.com/Louis-Jensen/MBC-for-PALM. Raw experimental data (camera frames) available upon request.

Code availability

MBC code is available as Supplementary Material together with installation instructions and example simulated data sets. MBC code is also available at https://github.com/Louis-Jensen/MBC-for-PALM.

References

37. Ward, J.H. Hierarchical Grouping to Optimize an Objective Function. Journal of the American Statistical Association 58, 236-244 (1963).

38. Celeux, G., Hurn, M., and Robert, C.P. Computational and inferential difficulties with mixture posterior distributions. Journal of the American Statistical Association 95, 957-970 (2000).