A global sampler of single particle tracking solutions for single molecule microscopy

Michael Hirsch[1], Richard Wareham[2], Ji W. Yoon[3], Daniel J. Rolfe[1*], Laura C. Zanetti-Domingues[1] Michael P. Hobson[4], Peter J. Parker[5,6], Marisa L. Martin-Fernandez[1], Sumeetpal Singh[2],

1 Central Laser Facility, Science and Technologies Facilities Council, UK Research and Innovation, Didcot, Oxfordshire, UK
2 Department of Engineering, University of Cambridge, Cambridge, UK
3 Center for Information Security Technology, Korea University, Seoul, South Korea
4 Department of Physics, University of Cambridge, Cambridge, UK
5 School of Cancer and Pharmaceutical Sciences, King's College London, London, UK
6 Protein Phosphorylation Laboratory, The Francis Crick Institute, London, UK

* daniel.rolfe@stfc.ac.uk

Abstract

The dependence on model-fitting to evaluate particle trajectories makes it difficult for single particle tracking (SPT) to resolve the heterogeneous molecular motions typical of cells. We present here a global spatiotemporal sampler for SPT solutions using a Metropolis-Hastings algorithm. The sampler does not find just the most likely solution but also assesses its likelihood and presents alternative solutions. This enables the estimation of the tracking error. Furthermore the algorithm samples the parameters that govern the tracking process and therefore does not require any tweaking by the user. We demonstrate the algorithm on synthetic and single molecule data sets. Metrics for the comparison of SPT are generalised to be applied to a SPT sampler. We illustrate using the example of the diffusion coefficient how the distribution of the tracking solutions can be propagated into a distribution of derived quantities. We also discuss the major challenges that are posed by the realisation of a SPT sampler.

1 Introduction

Single molecule imaging is increasingly facilitating high-resolution investigations of molecular motion at the plasma membrane of cells (e.g. [1–4]). Hidden in these data there is crucial information on local environments, transport mechanisms, and the dynamic interactions that regulate protein networks and cell homeostasis. The analysis of particle motion is currently based on fitting trajectories with competing mathematical models, most commonly based on particle mean square displacements (MSD), whose deviations from the linearity characteristic of pure diffusion are interpreted in terms of standard types of particle motion like confined or directed (e.g. [5–9]), but also on hidden Markov calculations [10]. However, the heterogeneity often showed by the trajectories is not easily resolved by model fitting. For the latter to be effective the particles must either maintain the same type motion for multiple consecutive frames (typically >50 frames [8]) and/or display sufficiently long tracks, [5, 10]. This limits SPT to stationary-like conditions or to labelling with quantum dots or fluorescent beads that do not photobleach.

To evaluate particle motion in general one must measure the instantaneous values of motion parameters as they fluctuate along the particle trajectory and ultimately requires single frame sensitivity. The most accurate way to achieve this is from the globally optimal spatiotemporal solution to SPT. In this idealised approach each possible choice of particle reconnections and associated motion parameter values is considered and their consequences compared along the entire length of the tracks, therefore automatically exploiting all the information in the data to output the most likely empirical estimate of reconnections and parameters, and places confidence limits on them. Achieving the globally optimal solution has been the goal of SPT for decades but it has proven to be computationally prohibitive because of the colossal size of the configuration space of particle reconnection possibilities at high particle density, low signal-to-noise ratio (SNR) and fast particle movement typical of single molecule images in cells [11]. A wide range of methods has been developed to address this problem [12]. Naïvely one may deduce that it roughly scales as the factorial of the number of particles (thousands), motion parameters (dozens), and frames (hundreds). To make the problem tractable previous algorithms reduced the size of the configuration space both by imposing a priori narrow bounds on the parameters, from modelling or previous knowledge, and by approaching the globally optimal solution by taking many locally optimal solutions (e.g. [13–16]). This typically produces 'tracklets' separated by gaps, after which longer tracks may be recovered, for example, via minimal path techniques (e.g. [17, 18]), or maximum likelihood methods (e.g. [3, 4, 19, 20]). Although these algorithms addressed many of the challenges from high particle density and low signal-to-noise, it is difficult to ascertain how sensitive the results are to their choice of parameters [11], and the loss of temporal globality hinders access to the very statistical information one requires to evaluate dynamic motion.

Here we present the Biggles tracker, an automatic Bayesian Inference-based, Gibbs-sampler, GLobal EStimator of particle tracks and parameters that converges towards the globally optimal spatiotemporal solution in a computationally time practical for real-world tracking. Biggles allows to estimate the uncertainty in the tracking solution and finds probable alternative solutions. It therefore opens the possibility to propagate the tracking error to the estimation of derived biophysical quantities such as diffusion coefficients.

2   Material and Methods

Biggles uses some data, Y, which are the spatial and temporal (x, y, t) coordinates of the single particle spots detected in the images (referred to as *observations*), a hypothesis for the assignment of these observations to tracks, the track partition ω, and some global parametrised model, θ, for the properties of the system. (Full details of the algorithm are in Supporting information S1 Appendix). We write N for the number of observations, T for the number of time points, which in many applications is the number of images taken, and Ω for the set of all valid track partitions. We can fold into the algorithm data for any imaging detector and allow for a set of spurious measurements in each frame. We use a flexible yet simple model for particle motion: a random-walk [21]. Observations that are deemed to be spurious are collectively referred to as clutter. The clutter is treated as part of the track partition in the sense that in any partition each observation is assigned either to exactly one track or to the clutter. A track is defined in terms of the observations assigned to it, which must number at least two, and in terms of its first and last time points, which need not be associated with observations. At any time point a track may have at most one observation. A pair of consecutive observations within a track, which need not occur at consecutive time points, is referred to as a *link*. Therefore each track has at least one link.

Biggles finds the probability of any given set of tracks and motion parameters given the data. In a Bayesian framework [22], this allows one to explore via sampling the joint track and parameters empirical (or posterior) probability function:

$$P(\omega, \theta | Y), \tag{1}$$

with the aim of identifying the most probable set of tracks and parameters, together with the uncertainties on them. We note that, given that tracks are defined by observation and first and last time points, two different sets of tracks may correspond to the same partitioning of the observations. Although Biggles formally samples tracks, we will usually ignore this distinction in the subsequent discussion and denote them simply by the (track) partition ω. To avoid having to sample from the posterior simultaneously Biggles uses a Gibbs sampler [23], which allows one to draw samples ω_i and θ_i alternately from two conditional distributions:

$$\omega_i \quad \sim \quad P(\omega | \theta_{i-1}, Y), \tag{2}$$
$$\theta_i \quad \sim \quad P(\theta | \omega_i, Y). \tag{3}$$

An overview is shown in Fig 1A. We note that the partition sampler (steps ①-③ in Fig 1A) draws samples from a *probability mass function* (PMF), while the parameter sampler (steps ④ & ⑤ in Fig 1A) draws samples from a *probability density function*. The parameter space θ has 7 dimensions, while Ω has no intrinsic dimensionality, but is of finite size. While the parameters θ can be sampled directly because their posterior

Fig 1. Algorithm overview. A) Biggles runs two sampling chains with different initial partitions ω_0. Each chain is a Gibbs sampler alternately sampling partitions, steps ①-③, and parameters, steps ④-⑤ . B) Schematic of sampling and evaluating the proposal distribution $Q(\omega^* | \omega_{i-1})$, step ① in panel A. The sampling is realised by descending a sampling tree; the root of the tree is the current partition ω_{i-1} and the leaves are the partitions ω^* that can be reached, i.e. where $Q(\omega^* | \omega_{i-1}) > 0$. Each branch has a certain probability given the parent branch, so that the probability of a leaf is the product of the probabilities of the branches traversed during the descent. A descent down the reduce move branch is sketched. Other move types are executed in a similar manner, but with different branching operations. C) Cartoon of the move type pairs. Each move type has a positive probability to undo any modification of its partner, i.e. $Q(\omega^* | \omega_{i-1}) > 0$ if and only if $Q(\omega_{i-1} | \omega^*) > 0$. D) Cartoon of the observation likelihood calculation in step ② in panel A, using the update move example of panel C. The Kalman filter estimates the particle states (black dots) of the track model; the red line illustrates a possible course. The likelihood of the observations assigned to the track (yellow) is calculated using the filter's observation model. The change in the track assignment by the update move leads to different state estimates and hence to different observation likelihoods. For full details see Supplementary Notes.

distribution is known and separable, sampling from the track partition is non-trivial, as only a small number of analytical PMFs have known direct sampling algorithms. We therefore use the Metropolis-Hastings algorithm [24], which can draw samples from almost any PMF, and which, on convergence, yields candidate-sets of tracks whose distribution matches the track partition posterior $P(\omega | \theta, Y)$.

The *acceptance probability* of the Metropolis-Hastings sampler (Fig.1A step ③) is given by

$$\min\left(\frac{P(\omega^* | \theta_{i-1}, Y)}{P(\omega_{i-1} | \theta_{i-1}, Y)} \frac{Q(\omega_{i-1} | \omega^*)}{Q(\omega^* | \omega_{i-1})}, 1 \right), \tag{4}$$

where the proposed new track partition ω^* is sampled from the *proposal mass function* $Q(\omega^*|\omega)$, Fig 1A step ①. The sampling uses a step-by-step approach which can be described as descending a tree, Fig 1B. The root of the tree is the last partition sample ω_{i-1}. The nine move types are the main branches of which one is randomly chosen. Cartoons of the move types are shown in Fig 1C. The further steps depend on the sampled move type. For example, in the execution of a "reduce" move, the track to be shortened is chosen, then the end of the track is sampled (front or back) and finally the time point within the track where the cut happens is sampled, Fig 1B. Each step has a probability that depends on the previous step. The probability to sample the proposal partition is the product of the probability of each step. In the example $Q(\omega^*|\omega_{i-1}) = 1/9 \times 1/(\text{number of tracks}) \times 1/2 \times 1/(\text{number of time points in the selected track where the cut is allowed})$. The sampling structure ensures that

$$\sum_{\omega \in \Omega} Q(\omega|\omega') = 1 \quad \text{for any } \omega' \in \Omega, \tag{5}$$

which is a necessary condition for Q being a PMF. As already mentioned above, for a valid track we require that it contains observations at a minimum of two time points (moves that create tracks with fewer than two observations are not allowed), and we note that two tracks having the same observations and therefore the same links may not be equal, since a track may have unobserved states before the first or after the last observation. To improve the performance of the algorithm, we also limit the maximum speed of the particle (to a value that is larger than any value that can be physiologically expected).

It is possible that the execution of a move does not lead to a valid track partition. For instance, at the end of the birth move, the created track may contain no observations purely by chance. In such a case ω^* is set to ω_{i-1}. Since in such a case accepting and rejecting leads to the same result, $\omega_i = \omega_{i-1}$, we regard the proposal as *identity* as opposed to *accepted* or *rejected*. It means that $Q(\omega|\omega) > 0$ for most ω. We tried to minimise the identity proposals in the move design, e.g. we do not attempt a death move if the partition contains no tracks or we do not attempt to split a track that has only three observations and so on. The occurrence of the identity proposal is not a theoretical novelty. $Q(\omega|\omega) > 0$ also occurs if Q is a Gaussian distribution. However since the Gaussian distribution is continuous rather than discrete, it is very unlikely that a proposal sampled from it is equal to the last sample and special considerations of such a case are unnecessary. The distinction only takes effect in the calculation of the acceptance rate, where identity proposals are treated as rejected proposals even though their acceptance probability is equal to one.

To evaluate the target distribution, $P(\omega|\theta_{i-1}, Y)$, of the Metropolis-Hastings sampler, Fig. 1A step ②, we expand it into three different components using Bayes' theorem,

$$P(\omega|\theta, Y) \propto P(Y|\omega, \theta)P(\omega|\theta)P(\theta), \tag{6}$$

where $P(\theta)$ are the parameter priors and $P(\omega|\theta)$ is the probability of the track partition, which takes the assignment of observations to tracks into account but not their physical properties. $P(\omega|\theta)$ is assumed to depend separably on four parameters in θ, according to closed-form distributions. The likelihood $P(Y|\omega, \theta)$ is factorised into the likelihood of the clutter observations $P(Y^0|k_0, \theta)$ and the product of the likelihoods of the observations assigned to tracks, $\prod_{i=1}^{K} P(Y^i|k_i, \theta)$, where K is the number of tracks and Y^i are the observations assigned to track k_i. The likelihood $P(Y^i|k_i, \theta)$ is evaluated using a state space approach [25]. The (unobserved) particle state encompasses position and velocity, $X = (x, \dot{x}, y, \dot{y})^T$. The particle motion is a modelled as random walk in the positions plus a velocity term, where the velocity follows its own random walk. This

Fig 2. Principle of the observation likelihood calculation. The partition sampler has assigned the observations (yellow) to a track. From those observations, the Kalman filter estimates the most likely states (black) and the state estimation errors (not shown) given the model of particle motion. Under this fitted model, each observation assigned to the tracks has a specific likelihood, which is evaluated from a normal distribution that is centred in the projection of the state into the observation space and whose standard deviation is a combination of the observation error and the state estimation error. When the partition sampler reassigns an observation at time t_2 of the track by executing an update move, the fresh application of the Kalman filter results in new estimates for all states. That leads in turn to different observation likelihoods even for those observations whose assignment has not changed.

allows for both directed motion and undirected motion, but also more complicated types ₁₃₆ of motion. We use the Kalman filter [26] to estimate the states of the particles, both at ₁₃₇ time points where the track was observed and at time points without observations, e.g. ₁₃₈ due to fluorophore blinking. To base our estimates on all observations assigned to the ₁₃₉ track, we apply the Rauch-Tung-Striebel backwards smoothing filter [27]. The ₁₄₀ observation likelihood is calculated by three principle steps: ₁₄₁

1. Assigning the observations to tracks (done by the partition sampler), ₁₄₂

2. Estimating the track's states from the observations of the track using the Kalman ₁₄₃ filter (i.e. inferring the model), ₁₄₄

3. Calculating the likelihood of the observations given the model (i.e. states and the ₁₄₅ error estimates). ₁₄₆

An example is depicted in Fig. 2. The figure highlights the change in the observation ₁₄₇ likelihood caused by an update move. The sample from the proposal distribution ₁₄₈ contains the track given by four observations in Fig. 2A. A fifth observation at time ₁₄₉ point t_2 is considered clutter. The Kalman filter estimates the particle states from the ₁₅₀ observations of the track, which are marked by black dots in the state space. The states ₁₅₁ and their error estimates in turn imply how likely it is to find the observations at their ₁₅₂ actual positions. The likelihood of the individual observation Y_j^i is given by a normal ₁₅₃ distribution $\mathcal{N}(Y_j^i; B\hat{X}_j^i, \hat{S}_j^i)$, where $B\hat{X}_j^i$ is the projection of the state X_j^i that is ₁₅₄ associated with Y_j^i into the observation space and $\hat{S}_j^i$ is the innovation error, which is ₁₅₅ composed of the observation error R and the state estimation error. The full equations ₁₅₆ are given in the supplied appendix 5 sections 3.2 and 3.3. In Fig. 2B, the partition ₁₅₇ sampler has swapped the two observations at time t_2; the observation that previously ₁₅₈ was clutter is now part of the track and the other observation is now considered clutter. ₁₅₉ The Kalman filter provides new state estimates of all observations of the track, which ₁₆₀ leads in turn to new observation likelihood estimates. For example the likelihood of the ₁₆₁ observation at t_1 is reduced after the update move, while the likelihood of the ₁₆₂ observation at t_3 is increased. ₁₆₃

The 2×2 covariance matrix of the observation noise R that contributes to the ₁₆₄ innovation covariance is sampled as part of the parameter sampling stage of the Gibbs ₁₆₅ sampler, Fig. 1A step ⑤. The details about the calculation of $P(\omega|\theta_{i-1}, Y)$ are ₁₆₆ described in the supplementary notes. ₁₆₇

In addition to the three parameters that determine R, the further four parameters in ₁₆₈ θ control the track partitioning. The birth rate λ_b is the average number of tracks that ₁₆₉ begin at time point t in a normalised area A. The clutter rate λ_c is the average number ₁₇₀ of spurious observations that are found at time t in a normalised area A. We use the ₁₇₁ symbol "$\mathcal{E}$" for the unit of the rates and set $1\mathcal{E}$ equal to 1 event per frame and ₁₇₂

100pixels×100pixels. The observation probability p_o is the probability that a particle is observed at time t and the survival probability p_s is the probability that a track that is present at time $t-1$ is also present at time t. We note this implies the assumption that the track length is exponentially distributed with mean $1/(1-p_s)$. This is not a limitation for imaging fluorophores, since it reflects the bleaching behaviour of fluorescent molecules. Quantum dots, which are used in bioimaging as well, do have for practical purposes an infinite life time. The posterior distribution $P(\theta|\omega_i, Y)$ can be sampled directly and is separable, Fig. 1A step ④-⑤. In particular, the rates λ_b, λ_c and the probabilities, p_o, p_s are assumed to follow Gamma and Beta distributions, respectively, where the parameters for the Gamma and Beta distributions are derived from the current partition sample by counting the number of tracks, their lengths and so on, Fig. 1A step ④. The observation noise of the Kalman filter, R, is sampled from an inverse Wishart distribution, where the parameters are derived from the track observations and the track model. In summary we have

$$\theta = (\lambda_b, \lambda_c, p_o, p_s, R). \qquad (7)$$

The initial samples θ_0 are initialised to some values. In principle we need only specify loose "plausible" bounds of θ_0. Our current implementation initialises θ_0 to uncontroversial typical values. The rates λ_b and λ_c have improper uninformative priors on them being positive, p_s and p_o have uniform priors over $(0,1)$ and R has an inverse Wishart prior $\mathcal{W}(\Phi, s)$, see also Supplementary Notes. Following the parameter sampling, the cycle is complete and a new track partition is sampled from the proposal distribution.

A Biggles chain is ergodic, which is shown in the supplementary information. The initial part of the sampling before the sampling chain has reached the limiting distribution (the target distribution) is called burn-in phase. In order to assess the convergence to the limiting distribution of the Metropolis-Hastings sampler we run two chains. The first chain starts with the partition without any links, i.e. where all observations are assigned to the clutter. This is the *minimum partition*, ω^0. For the second chain, we use a randomised greedy algorithm to assign as many observations as possible to tracks to create the initial partition. No further links can be added to such a partition, which is therefore referred to as a *maximum partition*. Usually, Ω has many maximum partitions. The sole purpose of this approach is to get two starting partitions that are far away from each other. To assess the convergence of the two chains we implemented two tests. The first test is based on the similarity of partitions. We consider a track partition as a graph, where the observations are the nodes and the links are the edges. We use the graph edit distance (GED) [28] as similarity measure, where link insertion and link deletion are the graph edit operations. In other words, we define distance between two track partitions as the number of links in which the two partitions differ. For convergence we demand that the average cross-chain GED does not exceed the sum of the averages of the two inner chain GEDs. Second, the Gelman-Rubin [29, 30] statistics is implemented for the parameters $\lambda_b, \lambda_c, p_o$ and p_s. Fig.3 shows example data for the convergence.

Fig 3. Convergence. The panels show: the posterior density of the two chains; the Gelman-Rubin statistics for the four tracking control parameters and the GED criteria, (average-cross-chain-GED)/(sum-of-average-inner-chain-GEDs).

By design the Metropolis-Hastings partition sampler will yield correlated samples. That reduces the effective sample size [30] and increases the total number of samples required. On the other hand the number of samples that can be recorded is constrained by the computational resources. Specifically, partition samples can be dozens or

hundreds kB large, depending on the number of observations, N. Therefore we record 1 in every n samples, thereby *thinning* the chain, where n linearly depends on the acceptance rate at the end of the burn-in phase. In the present implementation every 1 in 8 samples would be recorded if the acceptance rate were 25%.

The burn-in phase is longer than the target distribution sampling phase. Experience shows that the ratio of the number of target distribution samples to the number of burn-in samples has a median of about 3.5% and a mean of 5.5%. The maximal observed value is 25%.

At the point of convergence, the Gibbs sampler starts to sample candidate-sets of tracks and parameters whose distribution matches that of the joint posterior distribution. We may therefore interpret results such as the sample mode and sample variance as maximum empirical estimates and experimental errors respectively. The ability of Biggles to directly return a representative sample of tracking results and parameters, and thus to place confidence limits on these, is powerful and, as far as we are aware, it is novel in this field.

To demonstrate Biggles we use synthetic and single molecule microscopy (SMM) data sets. For our simulations, we first generate tracks with a given birth rate λ_b^* and survival probability p_s^*, and the states of the particles at each time point of the tracks is determined by the state dynamics. The tracks of states are referred to as ground truth. Next, track observations are created with probability p_o^* from the states using a Gaussian observation model, $\mathcal{N}(\mathbf{0}, R^*)$, and the clutter observations are created, governed by λ_c^*. Finally, the data is cropped to the field of view. This final result is the *realisation of the ground truth* (GTR). That means we have three stages of ground truth; the parameters, the tracks states simulated from these parameters and the observations created from the states. We enable a range of behaviours in the simulation; random walk, directed motion, a combination of both and track splitting.

If not mentioned otherwise, the unit of x and y is pixel, where one pixel is 160nm×160nm for the SMM data sets that we present. The unit of time is the frame index. The time lag between two frames is 0.05 seconds for the SMM data sets.

3 Results

We created synthetic data sets to test the correctness of the Biggles sampling. We simulated two series of 10 data sets each that differ in the birth rate, with $\lambda_b^* = 0.1\mathcal{E}$ and $\lambda_b^* = 1.6\mathcal{E}$ respectively. To get a similar observation count for both series we reduced the field of view in the series with the higher birth rate. For each birth rate, Fig 4 shows the posterior distribution of λ_b, λ_c, p_o and p_s for two of these data sets in comparison with parameter samples given the GTR, $P(\theta|GTR, Y)$. The ground truth value for the parameter is indicated by a vertical line. First we see that the distributions derived from the GTR are are well distributed around the ground truth parameter values, albeit with some bias in the survival probability. Moreover, we observe very good agreement between the Biggles posterior distributions and those derived from the GTR (see also the Q-Q plot in the support material S1 Fig). There are some offsets in the birth rate and the observation probability, and also in the survival probability in the higher crowding case. In other words Biggles has a slight bias to fewer, longer tracks with less observations (i.e. long tracks with many dark states), while the total number of observations in tracks remains equal to that for the GTR. For example, if two short tracks with a temporal gap are merged, then survival rate goes up, the observation rate and the birth rate go down, while the number of observations in tracks remains unaffected.

We calculate the frequency with which any two observations have been linked and

Fig 4. Recovery of the simulation parameters. The histograms of the parameter samples for the GTR (black) and the parameter samples created by Biggles (red). A pair of data sets is shown with low track density (left) and a pair with high track density. The ground truth parameters for each pair are indicated by black vertical lines. See Fig 6 for example data sets and supplementary information for Q-Q plots for series of 10 such data sets.

use it as probability estimate, $\hat{p}(l)$, for the occurrence of a link, l,   267

$$\hat{p}(l) = (\text{number of records containing } l)/(\text{total number of records}). \qquad (8)$$

To assess the tracking results we adopted performance measure from [31]. However a   268
direct usage of these measure is not possible since there are not designed for tracking   269
PMF. We adopted the Jaccard similarity coefficient (JSC), in the following way. For a   270
given ratio, p_{min}, we consider a link, l, as predicted if $p_{min} \leq \hat{p}(l)$. We express the   271
results in terms of the confusion matrix as true and false positives and negatives; TP,   272
FP, TN and FN. For example, let $p_{min} = 0.6$ and let l occur in the GTR. If $\hat{p}(l) = 0.7$ 273
then l counts towards the number of TP. If $\hat{p}(l) = 0.5$ then l counts towards the number   274
FN. However, if $p_{min} = 0.5$ and $\hat{p}(l) = 0.5$ then l counts as TP. For any p_{min} we can   275
calculate JSC = TP/(TP + FP + FN), and also recall = TP/(TP + FN) and precision   276
= TP/(TP + FP). All measures have a range between 0 and 1, where 1 is best and 0 is   277
worst. If $p_{min} = 1$ then only links that occur in all records are considered positives. The   278
number of FP is lowest and FN is highest. When p_{min} is reduced, FP will increase and   279
FN will drop. If $p_{min} = 0$ then any link that at least occurred in one sample will be   280
considered as positive, which gives a large number of FP, while ideally the number of   281
FN should go to zero. In other words, if a link is in the GTR then we expect it will at   282
least occur in one sample. The two left panels of Fig 5 show the JSC response for the   283
two series of data sets with $\lambda_b = 0.1\mathcal{E}$ and $\lambda_b = 1.6\mathcal{E}$ respectively. We observe that the   284
JSC generally is highest for $0.4 < p_{min} < 0.6$, while it significantly drops for p_{min} near   285
0 and 1. For the less crowded data set we find higher JSC (between 0.97 and 0.99 at   286
$p_{min} = 0.5$) than for more crowded data sets (between 0.92 and 0.98 at $p_{min} = 0.5$).   287
The two right panels of Fig 5 show the precision vs. recall plots. The values at   288
$p_{min} = 0.5$ are marked with a dot. We observe that with lower p_{min} the recall increases,   289
i.e. less GTR links are missed, while with higher p_{min} the precision increases, i.e. less   290
false predictions are made. For the data sets with lower track density we observe recall   291
and precision higher than 0.99 at $p_{min} = 0.5$ and for data sets with higher track density   292
we observe recall and precision values of at least 0.96 at $p_{min} = 0.5$. For some low track   293
density data sets we observe almost perfect recall at $p_{min} = 0$ and almost perfect   294
precision at $p_{min} = 1$. However also for the higher density data sets we get recall and   295
precision of at least 0.99 at the extreme ends of p_{min}. We did not calculate the receiver   296
operating characteristic (ROC) curve, since we did not calculate the number of TN.   297
However, the plot recall versus precision is a similar visualisation; rather than assessing   298
how many negatives have been falsely classified as positive as in ROC, we assess how   299
many of the classified positives are true positives.

Fig 5. Performance of Biggles tracking for data sets with low and high crowding. From left to right: JSC for less crowded data sets, JSC for more crowded data sets, recall against precision (lower crowding) and recall against precision (higher crowding). The dots in the two right panels mark the values at $p_{min} = 0.5$.

                                                                                              300
We use the synthetic data of Fig 4 to illustrate the dependency of the posterior   301
distribution on the track density, see Fig 6. The links shown on the left of Fig 6 are   302

from one of ten data sets with low birth rate (3429 observations) and, on the right is one of the data sets with high birth rate (3558 observations). The histograms in the middle panels shows the percentage of links against their estimated probability of occurrence. The low birth rate data sets have a higher proportion of uncontroversial links, with 7 data sets having 95% or more compared to the data sets with high birth rate with 9 data sets having less than 90% of uncontroversial links. We see more links with low probability. The increase in links with medium probability indicates that we observe an increased uncertainty in our estimate of the tracking result.

Fig 6. Change of the distribution depending on the track density. The left histogram shows the link probability of 10 data sets with lower birth rate. One of the data sets is shown in the left 3D plot. The histogram on the right shows the link probability of 10 data sets with higher birth rate. In comparison with the low-birth-rate data sets, the high-birth-rate data sets have fewer links with very high probabilities, while the number of links with medium and low probability is increased. This higher uncertainty of some links is due to the higher crowding.

Single molecules show a variety of modes of motion. While we did not fully explore the behaviour of our algorithm under such conditions, we do provide some illustrative tracking examples in the supplementary figures; molecules that change the mode of motion from random walk to direction motion or the other way around (S2 Fig.); a mixture of molecules some of which move in a random walk and some of which have a directed motion (S3 Fig.); a mixture of molecules that move in a random walks with two different diffusion constants (S4 Fig.) and random walks of molecules with different local densities (S5 Fig.). All these synthetic data sets have been analysed without special adjustments of the algorithm.

We demonstrate the sampling of a derived quantity using the example of the diffusion rate [32], see Fig 7. We created 100 data sets as before with a birth rate of $1.6\mathcal{E}$. We calculated a single diffusion coefficient, D, for the GTR as well as for every recorded sample. The diffusion coefficient was calculated from the mean squared displacement, $\langle r^2(\tau) \rangle$, for time lag τ. The estimate is calculated from the positions of the observations assigned to the tracks. The mean was taken over all tracks, see Supplementary Notes. Each of the two panels show the data of five realisations of a ground truth partition (GTR), i.e. 5 sets of observations generated from the same ground truth states. The diffusion coefficient of each GTR is shown by a vertical line. They are different for each realisation due to the random nature of the generation. Each set of observations was used as input for Biggles and D was estimated for each recorded sample, 4000 samples per run. The D of the samples are shown as histograms in the same colour as the related D of the GTR. We calculated to each data set confidence intervals (CI) and counted how often it contains the D of the GTR. The CI to confidence level $X\%$ is calculated as the smallest interval that contains at least $X\%$ of the samples. We found an agreement of 52% for a 70% CI and 75% for a 95% CI. This is a very encouraging result. However the procedure somewhat underestimates the errors. This will be subject to future investigations.

We compared Biggles with uTrack [20]. For the comparison we simulated particles at a range of different conditions Fig.8. The track density was simulated by assuming different birth rates with spatially uniformly distributed first observations. The average nearest neighbour (NN) distance of the observed particles was determined from the GTR. It varies roughly from $4px$ to $12px$. Simulated particles moved in a random walk ($D = 0.02px^2$/frame and $D = 0.32px^2$/frame). After the generation of the GT particle trace the observation model was applied. A particle was observed at any time with a given probability, $p_o \in \{0.9, 0.7, 0.5\}$. The particle observation was sampled from a normal distribution with the GT particle location as mean. We assumed two different

Fig 7. The distribution of diffusion coefficient, D, calculated from the Biggles samples. Each panel shows the result from a single GT simulation, i.e simulated particle states governed by birth rate and the survival probability. From each GT ten GTR have been created, governed by the observation rate, the observation error and the clutter rate. The ground truth $D = 0.08$, D calculated from the GTR is shown as vertical, dashed lines whose colours correspond to the colour of the histograms. The two numbers indicate how often the GTR D lies within 1σ and 2σ of the sample average of the respective Biggles output. Since there are 10 GTR per panel, possible values are multiples of 10.

localisation errors, $0.1px$ and $0.4px$. The life time of the particle tracks was exponentially distributed. We added uniformly distributed spurious detection which resulted in a density of 0.4 ± 0.1 observations per $100px\times100px$ and frame. The resulting data sets had on average 2003 ± 575 observations.

Fig 8. Comparison between Biggles and uTrack. The uTrack results are shown in blue, one point per tracking result, the points connected by a line. The Biggles results are shown as vertical black lines, each line represents a single tracking result connecting the value of best with the value of the worst sample. The GED to the ground truth (number of wrong links + number of missed links) is normalised by the number of observations of the data set.

The results where compared with the GTR. The number of links in which the trackers differed from the GTR, the GED, was normalised by the number of observations in the input data. Note that the value can be larger than 1, with 2 being an upper boundary. A value of 0 indicates total agreement. For uTrack a single value per data set is shown in Fig.8 (blue), for Biggles each tracking result is represented by a vertical black line representing the range of all samples (2000 per data set). As expected, both trackers perform well under good conditions and the performance slides if the condition get worse. For a localisation error of $0.1px$ and high observation probability both trackers perform very well. Biggles remains stable for a low localisation error, even if the observation probability drops. The uTrack tracker on the other hand drops in performance if the observation rate goes down. For a localisation error of $0.4px$ the tracking results are consistently worse. However, for the highest observation probability the performance of both trackers is still good. As before, with lower observation probability the performance of uTrack drops faster than the performance of Biggles. In general Biggles has a better performance than uTrack.

To illustrate Biggles we have chosen a typical SMM data set from our lab, which has been imaged for a co-localisation experiment. The data sets was acquired by total internal reflection fluorescence (TIRF) microscopy using organic dyes (enhanced green fluorescent protein) [33]. We imaged Epidermal Growth Factor (EGF) Dyomics 549-P1 on CHO cells stably expressing wild type EGF receptors at a level of about 50000 receptors per cell, transiently transfected with PLCd-PH-eGFP. The data set has a $16\mu m\times16\mu m$ (100pixels$\times$100pixels) field of view, in $30s$ 600 image frames have been acquired and 9761 observations have been detected using an in-house algorithm [34]. We recorded 4000 tracking samples. The result is shown in Fig 9. In total 8773 links occurred, of which 7923 appear in every sample and can be considered certain. The histogram of the 850 links with $\hat{p} < 1$ is shown in Fig 10. The total number of samples drawn in both chains is $34\,273\,600$.

The tracking result is plausible. All obvious tracks are found and there are no obviously wrong links with high probability. Biggles considers most of the tracks as uncontroversial. About 10% of the links do not appear in all samples. Most of those

Fig 9. Two views on the tracking result for a SMM data set with 9761 observations. 4000 samples have been recorded. Links that appear in all samples are shown in blue. The colour of the other links indicates their frequency of occurrence as shown by the colour bar. Observations that are assigned to the clutter in all samples are shown as grey dots.

links have either a small probability or a high probability. We found 273 links (3.1%) with $0.2 \leq \hat{p} \leq 0.8$, see Fig. 10. Those links may indicate locally high crowding, track splitting or merging, large gaps between two track pieces and more. The final acceptance rate of the Metropolis-Hastings sampler is low, between 0.4% and 1.4% for the different move types (Table 1). Identity moves could be avoided for 5 of the move types. Relatively many identity moves occur for the birth type (5.3%) and the update-type (3.5%). However, for these moves we also observe below-average proposal rejections so that the acceptance rate is not affected.

Fig 10. Histogram of the 850 links with estimated probability less than 1 for the data set in Fig 9. The number of links with $\hat{p} = 1$ is 7923.

Table 1. Statistics of the move types for one sampler chain. Data set SMM–9761.

move type	rejected	identity	accepted	total
Birth	94.3%	5.3%	0.4%	1 928 159
Death	99.6%	0.0%	0.4%	1 926 784
Extend	98.8%	0.0%	1.2%	1 927 725
Reduce	98.0%	0.8%	1.2%	1 924 570
Split	99.5%	0.0%	0.5%	1 926 823
Merge	99.1%	0.4%	0.5%	1 926 366
Update	95.1%	3.5%	1.4%	1 928 992
Transfer	98.8%	0.0%	1.2%	1 927 295
Cross-over	99.6%	0.0%	0.4%	1 926 086

4   Discussion

Biggles expands the concept of the single particle tracker by sampling the posterior distribution of possible tracking solutions and their governing parameters. Therefore, with Biggles we enable the calculation of errors and other descriptive statistics for tracking solutions. The set of all partitions does not come with an canonical distance measure, which is needed for some statistics. There are several possibilities to introduce a distance measure such as: treating the tracking solutions as graphs and employing the GED, or treating tracking solutions as vectors of a high dimensional space where each pair of linkable observations contributes one dimension. We used the GED as part of the convergence assessment.

The knowledge of the most likely solution remains of limited use as long as we do not know *how* certain the solution is and how likely alternative solutions are. With Biggles we have now the means to do what we would do in any other measurement process: evaluate the error on our solution.

Biggles does not need to input parameters that control the tracking process. On the contrary these parameters are a part of the solution. We show a example of parameter distributions in Fig 4. There are some slight biases in the recorded parameter samples.

The design of the survival probability, p_s, implies a exponential distribution for the track length. However, tracks of length 0 or 1 are not allowed and all tracks are limited to be no longer than the number of imaged frames. The samples stem in fact from a truncated exponential distribution. Since each observation is either explained as clutter or as observed track the biases in the clutter rate, λ_c and observation probability, p_o, are opposing. However overall there is a very good agreement of the sampled parameters with those of the GTR, see Fig 4.

A direct validation of the Biggles samples is not easy since the ground truth *distribution* is not known. We treated the GTR links as if they would be certain under the given model, which is not true, since a particle can move in an unlikely manner. However, it should remain the exception that the ground truth is unlikely under the model, since the model shall explain the motion of the particles. In fact, we found very good agreement of the sampled links with the GTR. As expected in the case of higher track density we found more uncertainty in the links. This is not a shortcoming of the algorithm, but its point. In high crowding situations the track assignment is less clear and a higher temporal and spatial resolution would be required to achieve more confidence in a specific tracking solution. With Biggles we quantify our confidence in a specific solution and produce representative samples of alternatives. For the data sets in Fig 5 we found that for the vast majority of the cases links in the GTR and frequent links in recorded samples coincide. If we consider recall and precision as functions of the minimal estimated link probability, $Rec(p_{min})$ and $Pre(p_{min})$ respectively, we do expect to see $Rec(p_{min}) \to 1$ for $p_{min} \to 0$ and $Pre(p_{min}) \to 1$ for $p_{min} \to 1$. That means, we expect to find any real link in at least in one sample and we expect not to find the same false link in all samples. For the low density in Fig 5 this seems to be the case. For higher density data sets we still find $Rec(0) > 0.99$ and $Pre(1) > 0.99$.

Even though the definition of a track does not include merging or splitting of tracks, splits and merges are represented in Biggles sampling, see Fig 11 and supplementary figure S6. The data set we used to demonstrate this is composed of straight lines with some added white noise. At time points 5 and 15, there are four points where three lines intersect. At time point 10 there are three intersections of two lines and at time points 0 and 20 there are three points where two lines meet. Track splitting (or merging) events appear as regions with larger uncertainty in the links. In the course of the sampling, the observations at the intersection are assigned to different tracks. Fig 11 demonstrates also the tracking of directed tracks including change of direction. We made no special adjustments to track this data set.

Fig 11. Tracking complicated data sets. The left panel shows the link probabilities. Certain links are in blue, highly probable links are in red, 50-50 links are coloured orange and unlikely links are in green. The middle panel shows the histogram of the links that are not certain, i.e. $\hat{p}(l) < 1$. The right panel shows the maximum posterior solution.

Another approach to verify the results of Biggles uses derived quantities. We demonstrated this with the diffusion coefficient. Our results show that on average Biggles accurately reproduces the diffusion constant of the GTR. However for individual data sets the results may deviate from the GTR, significantly in terms of the sample standard deviation.

We have demonstrated in Fig 9 that Biggles can analyse real-world microscopy data sets. For large data sets the sampling process slows down from thousands of samples per second for small data sets to a few samples per second. The final sample rate for the SMM-9761 data set was about 500 samples per seconds using a fast commercial desktop computer. There are moderate improvements possible using faster computers, however software engineering of the code promises the most significant improvement. This

concerns for example the handling of data of significant size, which effects the speed the algorithm and the number of samples that can be held in the chains. The sampling of the moves depends on internal book keeping to get a high efficiency in identifying viable proposals. Solutions for implementation problems are independent from the algorithm development and are not discussed here.

There is another possibility to improve performance for large data sets. Our SMM data sets can contain about $500\,000$ observations. We are developing a chunked version of the algorithm that divides the data into overlapping spatial chunks, tracks the observations in each chunk separately and reconciles the results. This approach will also allow the usage of computing clusters or cloud computing resources.

The sampling process still lacks efficiency. On one hand the track partition samples are correlated which reduces the effective sampling size. The correlation between the track samples lies in the nature of the Metropolis-Hastings proposals. The vast majority of links of the proposal will be identical with the links of the last sample. If $Q_m(\omega, \omega') > 0$, then the GED between ω and ω' is at most 6 links, if m is any of the move types merge/split (1 link difference), transfer (6 links), cross-over (4 links), or update (max. 4 links). If the move type m is any of birth/death or extend/reduce then both partitions differ in one track only, with a GED less then T links. All moves at most modify 2 tracks at a time.

On the other hand, the acceptance rate of the sample recording of data set SMM–9761 is very low as we showed in Table 1. All moves are likely to propose changes to the partition which have a low target density. In many cases, we have $Q(\omega|\omega') \gg 0$ even if $P(\omega|\theta, Y) \ll P(\omega'|\theta, Y)$. If for example the update move is applied to a long track, then the update will be attempted at any time point with equal probability. However, often it is the case that there are only very few time points for which the move would produce an acceptable proposal, making the acceptance rate very low. In this regard the proposal mass function is wide in comparison with the target distribution. In the current implementation, Biggles therefore suffers from both a slow exploration speed and a low acceptance rate. Modifications to improve the proposal mass function both increase the complexity of the calculation of the proposal mass ratio and increase the complexity of the proposal creation itself due to the employment of smarter algorithms, extensive internal book keeping and so on.

The size of a single partition sample is linear in N. It is therefore not possible to keep a large number of samples in memory. This affects the number of chains that can be kept and their size, but also the number of samples that can be recorded. Especially for multimodal distributions it is important that the chains are long enough to cover the relevant part of Ω. If the chains are too short the assessment of convergence may go wrong. The assessment can go wrong in two ways; the chains have reached the stationary distribution, but are in different parts of it because they are too short to cover the whole support. It is also possible that the chains have not yet reached the stationary distribution, but accidentally it appears as such. The choice of the starting partition is therefore of some importance. Our approach runs two chains with the minimal partition ω^0 and a maximal partition ω^{max} as starting points.

The huge size of Ω seems to imply than any practicable sample size is far too small to explore Ω. However, experience shows that the vast majority of links are uncontroversial with $p(l) > 0.99$. The interesting subset $\Omega^* \subset \Omega$ is much smaller, and often focuses on small spatio-temporal regions. If such regions are disconnected, they could be sampled separately, which further reduces the size of Ω^*. Still, the size of Ω^* can be substantial. It is very difficult to say how many samples are required to cover it and it depends on the data set.

5   Conclusion

We present in this paper a prototype of a novel approach to single particle tracking (SPT) that samples from the combined probability mass function/probability density function of the track partitions and its governing parameters. This enables not just the estimation of the most likely tracking solution but also provides us with a measure of uncertainty of this solution and likely alternative solutions. Thus Biggles normalises SPT with standard measurements that provide measured value and error estimate. Our approach also has the potential be used to estimate the error on derived quantities as we demonstrated on the diffusion rate. The algorithm can handle different condition without special adjustment, such as random walk, directed motion, change of direction and track branching. We demonstrated that Biggles can analyse data sets with about 10 000 observations. The implementation of Biggles is complex. Smarter algorithms, optimised convergence control and sample recording and professional software engineering will improve the performance of the algorithm. We also indicated other potential improvements. Biggles opens a new direction in SPT.

Supporting information

S1 Fig.   Recovering of the simulation parameters. The Q-Q plots of the parameter samples for the GTR and the parameter samples created by Biggles. Shown are a series of ten data sets with low track density(top) and a series of data sets withhigh track density (bottom).

S2 Fig.   Biggles tracking example. A mixture between directed motion and random walks. Tracks have a chance to change the mode of motion. The grey dots mark the clutter observations.

S3 Fig.   Biggles tracking example. A mixture between directed motion and random walks. Tracks have a chance to change the mode of motion. The grey dots mark the clutter observations.

S4 Fig.   Biggles tracking example. A 50-50 mixture between random walks with two different diffusion coefficients, $d_1 = 0.45 pix/fr$ and $d_2 = 0.9 pix/fr$. Each track has one mode of mode of motion.

S5 Fig.   Biggles tracking example. Random walks with regions of different densities. Each track has one mode of mode of motion.

S6 Fig.   Alternative views of the complicated data set. Shown are the link probabilities.

S7 Fig.   Dependency on the process noise. A step length of $1 pixel/frame$ is equivalent to $D \approx 0.26 \mu m^2/s$ at a pixel size of $160 nm$ and a frame rate of $20 Hz$

S8 Fig.   Example of observation and estimated states of a track.

S1 Appendix.   Algorithm notes. Detailed information about the algorithm, discussions of properties of the space of valid track partitions Ω and proof that the partition sampler is ergodic.

Acknowledgments ₅₄₂

References

1. Mortensen KI, Churchman LS, Spudich JA, Flyvbjerg H. Optimized localization analysis for single-molecule tracking and super-resolution microscopy. Nat Methods. 2010;7:377–359. doi:10.1038/nmeth.1447.

2. Manley S, Gillette JM, Patterson GH, Shroff H, Hess HF, Betzig E, et al. High-density mapping of single-molecule trajectories with photoactivated localization microscopy. Nat Methods. 2008;5:155–157. doi:10.1038/nmeth.1176.

3. Low-Nam ST, Lidke KA, Cutler PJ, Roovers RC, van Bergen en Henegouwen PMP, Wilson BS, et al. ErbB1 dimerization is promoted by domain co-confinement and stabilized by ligand binding. Nat Struct Mol Biol. 2011;18:1244–1288. doi:10.1038/nsmb.2135.

4. Cutler PJ, Malik MD, Liu S, Byars JM, Lidke DS, Lidke KA. Multi-Color Quantum Dot Tracking Using a High-Speed Hyperspectral Line-Scanning Microscope. PLoS One. 2013;8. doi:10.1371/journal.pone.0064320.

5. Lidke DS, Lidke KA, Rieger B, Jovin TM, Arndt-Jovin DJ. Reaching out for signals: filopodia sense EGF and respond by directed retrograde transport of activated receptors. J Cell Biol. 2005;170:619–626. doi:10.1083/jcb.200503140.

6. Saxton MJ. Single-particle tracking: The distribution of diffusion coefficients. Biophys J. 1997;72:1744–1753. doi:10.1016/S0006-3495(97)78820-9.

7. Kusumi A, Sako Y, Yamamoto M. Confined lateral diffusion of membrane receptors as studied by single particle tracking (nanovid microscopy). Effects of calcium-induced differentiation in cultured epithelial cells. Biophys J. 1993;65:2021–2040. doi:10.1016/S0006-3495(93)81253-0.

8. Bouzigues C, Dahan M. Transient directed motions of GABA(A) receptors in growth cones detected by a speed correlation index. Biophys J. 2007;92:654–660. doi:10.1529/biophysj.106.094524.

9. Dietrich C, Yang B, Fujiwara T, Kusumi A, Jacobson K. Relationship of lipid rafts to transient confinement zones detected by single particle tracking. Biophys J. 2002;82:274–284. doi:10.1016/S0006-3495(02)75393-9.

10. Das R, Cairo CW, Coombs DA. Hidden Markov Model for Single Particle Tracks Quantifies Dynamic Interactions between LFA-1 and the Actin Cytoskeleton. PLOS Comput Biol. 2009;5. doi:10.1371/journal.pcbi.1000556.

11. Saxton MJ. Single-particle tracking: connecting the dots. Nat Methods. 2008;5:671–672. doi:10.1038/nmeth0808-671.

12. Smal I, Meijering E. Quantative comparison of multiframe data assoication techniques for particle tracking in time-lapse fluorescence microscopy. Med Image Anal. 2015;24:163–189. doi:10.1016/j.media.2015.06.006.

13. Chetverikov D, Verestoy J. Feature point tracking for incomplete trajectories. Computing. 1999;62:321–338. doi:10.1007/s006070050027.

14. Shafique K, Shah M. A noniterative greedy algorithm for multiframe point correspondence. IEEE Trans Pattern Anal Mach Intell. 2005;27:51–65. doi:10.1109/tpami.2005.1.

15. Sbalzarini IF, Koumoutsakos P. Feature point tracking and trajectory analysis for video imaging in cell biology. J Struct Biol. 2005;151:182–195. doi:10.1016/j.jsb.2005.06.002.

16. Feng L, Xu Y, Yang Y, Zheng X. Multiple dense particle tracking in fluorescence microscopy images based on multidimensional assignment. J Struct Biol. 2001;173:219–228. doi:10.1016/j.jsb.2010.11.001.

17. Bonneau S, Dahan M, Cohen LD. Single quantum dot tracking based on perceptual grouping using minimal paths in a spatiotemporal volume. IEEE Trans Image Process. 2005;14:1384–1395. doi:10.1109/tip.2005.852794.

18. Sage D, Neumann FR, Hediger F, Gasser SM, Unser M. Automatic tracking of individual fluorescence particles: Application to the study of chromosome dynamics. IEEE Trans Image Process. 2005;14:1372–1383. doi:10.1109/tip.2005.852787.

19. Serge A, Bertaux N, Rigneault H, Marguet D. Dynamic multiple-target tracing to probe spatiotemporal cartography of cell membranes. Nat Methods. 2008;5:687–694. doi:10.1038/nmeth.1233.

20. Jaqaman K, Loerke D, Mettlen M, Kuwata H, Grinstein S, Schmid SL, et al. Robust single-particle tracking in live-cell time-lapse sequences. Nat Methods. 2008;5:695–702. doi:10.1038/nmeth.1237.

21. Hughes BD. Random walks and random environments. vol. 1: random walks. Clarendon Press; 1995.

22. Yoon JW, Bruckbauer A, Fitzgerald WJ, Klenerman D. Bayesian Inference for Improved Single Molecule Fluorescence Tracking. Biophys J. 2008;12:4932–4947. doi:10.1529/biophysj.107.116285.

23. Geman S, Geman D. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. IEEE Trans Pattern Anal Mach Intell. 1984;6:721–741. doi:10.1016/B978-0-08-051581-6.50057-X.

24. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. J Chem Phys. 1953;21:1087–1092. doi:10.1063/1.1699114.

25. Durbin J, Koopman SJ. Time series analysis by state space methods. Oxford University Press; 2001.

26. Shumway RH, Stoffer DS. Time Series Analysis and Its Applications. 4th ed. Springer International Publishing; 2017.

27. Rauch HE, Tung F, Striebel CT. Maximum likelihood estimates of linear dynamic systems. AIAA J. 1965;3(8):1445–1450. doi:10.2514/3.3166.

28. Gao X, Xiao B, Tao D, Li X. A survey of graph edit distance. Pattern Anal Appl. 2010;13:113–129. doi:10.1007/s10044-008-0141-y.

29. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian data analysis. 3rd ed. CRC press; 2013.

30. Givens GH, Hoeting JA. Computational Statistics, 2nd edition. Wiley; 2013.

31. Chenouard N, Smal I, de Chaumont F, Maška M, Sbalzarini IF, Gong Y, et al. Objective comparison of particle tracking methods. Nat Methods. 2014;11:281–289. doi:10.1038/nmeth.2808.

32. Bräuchle C, Lamb DC, Michaelis J, editors. Single Particle Tracking and Single Molecule Energy Transfer. Wiley-VCH; 2009.

33. Clarke DT, Botchway SW, Coles BC, Needham SR, Roberts SK, Rolfe DJ, et al. Optics clustered to output unique solutions: A multi-laser facility for combined single molecule and ensemble microscopy. Rev Sci Instrum. 2011;82. doi:10.1063/1.3635536.

34. Rolfe DJ, McLachlan CI, Hirsch M, Needham SR, Tynan CJ, Webb SED, et al. Automated multidimensional single molecule fluorescence microscopy feature detection and tracking. European Biophysics Journal. 2011;40(10):1167–1186. doi:10.1007/s00249-011-0747-7.

A global sampler of single particle tracking solutions for single molecule microscopy

S1 Appendix

Contents

1 Observations, tracks and clutter

Single particle tracker track *particles* such as molecules. The appearance of a particle in an image (or other measurement) is a feature of that image. Biggles takes as input a set of features specified as spatial x, y and temporal t coordinates. These correspond to the output of a feature detector and are collectively called *observations*, $Y = \{Y_t = [x, y]^T\}$, of a set of particles. We assume that the time lag between images is constant, so that $t \in \{1, \ldots, T\}$, where T is the number of image frames taken. This is not a limitation of the algorithm, but it makes implementation easier. Missing frames mean that there are no observations at the time points in question. The set of observations that have all been generated by the same particle are called a *track*. Observations may be deemed to have arisen spuriously from the feature detector and not from a particle of interest. Spurious observations are collectively termed the *clutter*, $k_0 \equiv Y^0$. The subset of observations assigned to track k_i, $i = 1 \ldots K$, is Y^i, We have $Y^i \cap Y^j = \emptyset$ if $i \neq j$ and $\cup_{i=0}^{K} Y^i = Y$. Because of these properties, the set of all tracks and the clutter is termed a (track) *partition*, $\omega = \{k_0, \ldots, k_K\}$, of the input, where ω contains K tracks and the clutter k_0.

Condition 1 *A track fulfils the following conditions:*

  1. *a track has at most one observation at each time point,*

  2. *the number of observations of the track is at least 2,*

  3. *the spatial distance of any two observations in a track has a limit that linearly depends on the time lag between the observations.*

A track can be written as

$$k_i = (t_0^i, t_1^i, Y^i),$$

where t_0^i and t_1^i are the first and the last time point of the track respective. This also is the computational representation of a track. It is allowed that first and last time points do not have observations. The length of the track is the number of its time points, $t_1^i - t_0^i + 1$. With condition 1.2 follows that the length of any track is in $[2, T]$. The third condition means that a maximum velocity of the particle is assumed.

2 Gibbs sampler

Our aim is to explore the distribution $P(\omega, \theta | Y)$, where ω is a partition, θ is a set of model parameters and Y are the observations. Internally, the parameters θ are represented by a tuple containing floating point values and one matrix: $(\lambda_b, \lambda_c, p_s, p_o, R)$ where

  - λ_b: the mean number of new tracks appearing per frame and area.

  - λ_c: the mean number of clutter observations per frame and area.

  - p_s: the probability that a track will survive from frame t to $t + 1$, i.e. $p(\text{"track present at } t+1\text{"} \mid \text{"track present at } t\text{"})$.

  - p_o: the probability that a track will generate an observation at a time point.

  - R: a 2×2 matrix giving the assumed covariance of observation error of the Kalman filter (section 3.2).

We use a Gibbs sampler to avoid having to sample from $P(\omega, \theta | Y)$ directly. We draw a train of samples ω_i, θ_i by alternately drawing them from the two conditional distributions:

$$\omega_i \sim P(\omega | \theta_{i-1}, Y) \quad \text{and then} \quad \theta_i \sim P(\theta | \omega_i, Y),$$

where ω_0 and θ_0 are initialised to some values. In principle we are free to choose ω_0 and θ_0. Practically we avoid extreme values for θ_0 and fix it to uncontroversial moderate values. We are running two chains and initialises ω_0 in the first chain to a state where all features are assumed to be spurious detections. The second chain initialises ω_0 to a maximum partition, which is created by a greedy algorithm. The method how the maximum partition is created is not critical. In principal it works as follows:

1. Group and sort the available observations by time and start with the earliest group,

2. pick a random observation from the current time group and use it as first observation of a new track,

3. try to randomly add a new observation to the track from the time group just after the current last observation of the track, respecting condition 1.2 and allowing a gap of up to 5 time points,

4. repeat step 3 until its options are exhausted,

5. if the resulting track has more than one observations, save it and remove its observations from their time groups,

6. go to step 2, if the current time group is empty, pick the next time group, if all time groups are used, finish,

7. assign all remaining observations to the clutter.

Thus the Gibbs sampler alternates between two steps:

1. Sampling from the track partition using the Metropolis-Hastings algorithm (section 3)

2. Sampling the parameters (section 4)

3 Metropolis-Hastings Sampling from the track partition

The heart of the current implementation is a Metropolis-Hastings sampler. The target partition of the sampler is given by $P(\omega | \theta, Y)$ and a new partition, ω^*, is sampled from the proposal partition $Q(\omega^* | \omega)$. The proposal density Q is given by a set of moves that transform ω to ω^* and their probability to be executed.

When proposing ω^* from ω we choose uniformly from one of seven different types of move. In the following description, a "no observation" is an item that represents a time point of a track without observation. All moves enforce the formal constraints for tracks given in definition 1.

- Birth: Two time points are randomly chosen as first and last time point. Beginning with the first, for each time point a list of all candidate observations is assembled each with a weight of 1. A "no observation" is added to the list with the weight of $1/10$. An element of the list is randomly sampled and added to the track. This moves increases the number of tracks by 1 and reduces the number of clutter observations by at least 2.

- Death: a track is chosen uniformly and at random from the partition and is demoted to the clutter. This move reduces the number of tracks by 1 and increases the number of clutter observations by at least 2.

- Extend: a track is chosen uniformly and at random from the partition. It is chosen uniformly and at random in which temporal direction the track is to be extended and the track is extended as in the birth move. This move increases the length of the track by at least 1, it may or may not add observations to the track and hence may or may not reduce the number of clutter observations.

- Reduce: a track is chosen uniformly and at random from the partition. It is chosen uniformly and at random in which temporal direction the track is to be truncated and a time point within it is sampled uniformly and at random. The portion of the track extending from the sampled time point in the sampled direction is demoted to the clutter. This move will reduce the length of the track by at least 1, it may or may not reduce the number of observations of the track and add those observations to the clutter.

- Merge: a pair of nearby but non-temporally-overlapping tracks are sampled uniformly and at random from all such pairs and are merged into one by simple concatenation. This move reduces the number of tracks by 1, the clutter remains unaffected. The length of the new track is larger than or equal to the sum of the lengths of the tracks merged. The number of observations of the new track is equal to the sum of the number of observations of the merged tracks.

- Split: a track is chosen uniformly and at random from the partition. An time point is sampled uniformly and at random from within the track and the track is split at this point into two new tracks. The earlier track is randomly shortened by 0 or more final time points without observations. This moves increases the number of tracks by 1, the clutter remains unaffected. The sum of the lengths of the new tracks is less than or equal to the length of the split track. The sum of the observations of the new tracks is equal to the number of observations of the split track.

- Update: a track is chosen uniformly and at random from the partition. A time point is sampled uniformly and at random from the time span of the track. The track may or may not have an observation at the sampled time point. A new observation is sampled uniformly and at random from the set of all spatially nearby observations plus a "no observation". The original observation is demoted to clutter and the sampled observation (if any) is included into the track. This moves leaves the number of tracks and the length of the updated track unaffected. The number of observations of the updated track changes by -1 ("drop"), 0 ("swap") or 1 ("pick-up"), the size of the clutter changes by 1 (drop), 0 (swap), -1 (pick-up). Even in the swap case, the observations of the updated track and the clutter are changed. Any empty swap is prevented.

- Transfer: a pair of tracks is chosen from the set tracks that permit that move at one or more time points. One track is the donor the other track is the acceptor. A time point is uniformly sampled from the set of time points at which a valid transfer is possible. The observation of the donor at this time point is removed and added to the acceptor. This move leaves clutter, number of tracks and track lengths unaffected. The number of observations of the donor is reduced by 1 and the number of observations of the acceptor is increased by 1.

- Cross-over: a pair of tracks is chosen from the set tracks that permit that move at one or more time points. A time point is uniformly sampled from the set of time points at which a valid cross-over is possible. Both tracks are split into two branches at the time point so that each branch has at least one observation. Two new tracks are formed each using one branch of each old track. If the old tracks are denoted by (a_1, a_2) and (b_1, b_2) than the new tracks are denoted by (a_1, b_2) and (b_1, a_2). This move leaves the number of tracks and the clutter unaffected. The sum of the length of the new tracks is equal to the sum of the lengths of the old tracks.

There are 6 pairs of moves that can undo each other: birth – death, extend – reduce, merge – split, update – update, transfer – transfer and cross-over – cross-over. From a given partition ω any new proposal, ω^*, can be achieved by one move type only, e.g. a merge move can never have the same result as a extend move or birth move. The moves were so chosen such that the appropriate proposal density, $Q(\omega^*|\omega)$, is straightforward to compute. For example, given a birth move, $Q(\omega^*|\omega)$ is the likelihood of the birth that leads from ω to ω^*, while $Q(\omega|\omega^*)$ is the probability of the death that leads from ω^* to ω. In some cases a move does not create a valid partition. Example 1: A birth move has proposed a track that only contains 1 observation. Example 2: A reduce move is attempted but all tracks have length 2. Such cases are treated as $\omega^* = \omega$. Hence $Q(\omega|\omega) > 0$. However, efforts have been made to keep $Q(\omega|\omega)$ small.

We wish to find the mode of the posterior distribution $P(\omega|\theta, Y)$. We can re-arrange this posterior using the law of conditional probability which states $P(\omega|\theta, Y) = P(\omega, \theta|Y)/P(\theta|Y)$ combined with Bayes' theorem:

$$P(\omega, \theta|Y) = \frac{P(Y|\omega, \theta)P(\omega, \theta)}{P(Y)} = \frac{P(Y|\omega, \theta)P(\omega|\theta)P(\theta)}{P(Y)}$$

and hence

$$P(\omega|\theta, Y) = \frac{P(Y|\omega, \theta)P(\omega|\theta)P(\theta)}{P(\theta|Y)P(Y)} = \frac{P(Y|\omega, \theta)P(\omega|\theta)P(\theta)}{P(\theta, Y)}.$$

Since we use the Metropolis-Hastings algorithm which requires only a value proportional to this density. Removing terms independent of ω we obtain:

$$P(\omega|\theta, Y) \propto P(Y|\omega, \theta)P(\omega|\theta)P(\theta).$$

The $P(\theta)$ term is simply the prior on the model parameters. The $P(\omega|\theta)$ term is the likelihood of the tracks (the grouping of observations), independent of the physical properties (position, velocity and their errors) of the observations within them. This is a function of the model parameters and birth and death times for the tracks only. Since the data is partitioned between tracks and the clutter with no overlap, we may factorise the data likelihood term as follows:

$$P(Y|\omega, \theta) = P(Y^0|k_0, \theta) \prod_{i=1}^{K} P(Y^i|k_i, \theta)$$

where Y^0 is used to represent the clutter observations and $Y^1, ..., Y^K$ are the K sets of observations corresponding to the K tracks in the partition. $P(Y^0|\theta)$ is the likelihood of having seen the clutter observations, if we have given the parameters. $P(Y^i|k_i, \theta)$ is the likelihood of track k_i having generated the data we saw. Together we have

$$P(\omega|\theta, Y) \propto P(Y^0|k_0, \theta) \prod_{i=1}^{K} P(Y^i|k_i, \theta)P(\omega|\theta)P(\theta). \tag{1}$$

Taking logarithms, we obtain the final log density function used within the Metropolis-Hastings sampler:

$$\ell(\omega|\theta, Y) = \kappa + \ell(Y^0|k_0, \theta) + \sum_{i=1}^{K} \ell(Y^i|k_i, \theta) + \ell(\omega|\theta) + \ell(\theta) \tag{2}$$

where $\ell(\cdot)$ is used to denote the logarithm of probability density function (PDF) and κ is some arbitrary normalising offset. Thus the PDF has the following components:

- the partition conditional probability $P(\omega|\theta)$, section 3.1

- the track observations likelihood $P(Y^i|k_i, \theta)$, section 3.3

- the clutter observations likelihood $P(Y^0|k_0, \theta)$, section 3.4

- the parameter priors $P(\theta)$, section 3.5

The calculation of the priors $P(\theta)$ is not necessary for the acceptance probability since they does not depend on the partition and the two occurrences cancel each other.

3.1 Partition conditional probability

In this section we discuss evaluation of $P(\omega|\theta)$, that is the probability of the track partition given the model parameters. The physical properties of the observations, notably their position is not taken into account, only their grouping into tracks and clutter. We have no reason to assume a specific mutual dependency of the model parameters, we therefore may factorise the PMF as follows:

$$P(\omega|\theta) \propto P(\omega|p_s)P(\omega|p_o)P(\omega|\lambda_b)P(\omega|\lambda_c).$$

Note that $P(\omega)$ does not depend on R, since R has no bearing on the observation *assignment*.

We recall that Ω is the set of all possible partitions. Partitions only differ with each other in the assignment of observations to tracks and clutter. Each observation has a label, exchanging observations between tracks will result in a different partition. There is only one way (if any) to combine a given number of observations to a track. Reordering the observations in the clutter does not change the partition. Reordering of tracks does not change the partition. Each track is uniquely identified by the observations it contains and the first and last time point. The partition probability given the survival probability is the product of the probability of each survival and death. It has the general form of a binomial distribution, but does not feature the binomial coefficient. For a binomial distribution the sequences $(0, 1, 1)$ and $(1, 0, 1)$ mean both "2 out of 3", while for a partition it matters which track survived. Similarly, the partition probability given the observation probability does care about which observations have been made. It has therefore the same form. The birth rate and clutter rate dependencies are modelled by Poisson distributions.

Each of these terms may defined in terms of the following values:

- N_t^t : the number of tracks present at t;

- N_t^s : the number of tracks that survive from $t-1$ to t ;

- N_t^b : the number of tracks that newly appeared at t;

- N_t^o : the number of observations assigned to a track at t;

- N_t^c : the number of observations deemed spurious at t.

The numbers are counted from ω. The number of track that survive at t is given by: $N_t^s = N_t^t - N_t^b$. Each of the individual parameter PDF terms can be written based on the definition of the corresponding parameters:

- $P(\omega|p_s) = \prod_{t=2}^{T} p_s^{N_t^s} (1 - p_s)^{N_{t-1}^t - N_t^s}$;

- $P(\omega|p_o) = \prod_{t=1}^{T} p_o^{N_t^o} (1 - p_o)^{N_t^t - N_t^o}$;

- $P(\omega|\lambda_b) = \prod_{t=1}^{T} \mathcal{P}(N_t^b; \lambda_b)$;

- $P(\omega|\lambda_c) = \prod_{t=1}^{T} \mathcal{P}(N_t^c; \lambda_c)$,

where $\mathcal{P}(x; \lambda)$ is the probability mass function of the Poisson distribution with mean λ evaluated at x and $N_t^y = N_t^o + N_t^c$. The corresponding log terms become:

$$\ell(\omega|p_s) = \sum_{t=2}^{T} N_t^s \log(p_s) + (N_{t-1}^t - N_t^s) \log(1 - p_s) \tag{3}$$

$$\ell(\omega|p_o) = \sum_{t=1}^{T} N_t^o \log(p_o) + (N_t^t - N_t^o) \log(1 - p_o) \tag{4}$$

$$\ell(\omega|\lambda_b) = \sum_{t=1}^{T} N_t^b \log(\lambda_b) - \lambda_b - \log(\Gamma(1 + N_t^b)) \tag{5}$$

$$\ell(\omega|\lambda_c) = \sum_{t=1}^{T} N_t^c \log(\lambda_c) - \lambda_c - \log(\Gamma(1 + N_t^c)) \tag{6}$$

3.2 Kalman filter as time evolution model

We use a Kalman filter with a Rauch-Tung-Striebel backwards smoothing filter to estimate the underlying states of the particles which generated the observed features so that the track posterior may be evaluated.

The state estimation consists of a prediction step that is based on the physical model followed by an update step that compares the prediction to the observation. In our system, the state is the position and the instantaneous velocity of a particle:

$$X = [x, \dot{x}, y, \dot{y}]^T,$$

where

$$\dot{x} = \frac{dx}{dt}, \quad \dot{y} = \frac{dy}{dt}$$

States evolve using first order dynamics:

$$X_t = AX_{t-1} + \boldsymbol{w}_t, \quad A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

where $\boldsymbol{w}_t \sim \mathcal{N}(0, Q)$. The observation of the particle is modelled by

$$Y_t = BX_t + \boldsymbol{v}_t, \quad B = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

where $Y \in \mathbb{R}^2$ is the vector of the observed position, B is the observation model and $\boldsymbol{v}_t \sim \mathcal{N}(0, R)$ is the observation noise. The covariance matrix R is internally treated as a part of of the model parameters that are sampled by the Gibbs sampler (see section 4).

In the prediction step, the current estimate of the hidden state at time index t, $\hat{X}_t$, is evolved via the state evolution matrix A. In the update step, this estimate is refined by using any observations made. In this implementation the lack of an observation causes this update step to be skipped and the refined estimate is assumed to be equal to the prediction.

The prediction step is represented by the following recurrence relations:

$$\hat{X}_{t|t-1} = A\hat{X}_{t-1|t-1}, \quad P_{t|t-1} = AP_{t-1|t-1}A^T + Q$$

where $\hat{X}_{t-1|t-1}$ is the state estimate after the update step at time point $t-1$ and $\hat{X}_{t|t-1}$ is the state estimate at time t after the prediction step, but before the update step. $P_{t|t-1}$ and $P_{t|t}$ are, respectively, our prediction of the state estimation error and our refined prediction of the state estimation error.

The updated state at time t, $\hat{X}_{t|t}$, is represented by the following recurrence relations:

$$\hat{X}_{t|t} = \hat{X}_{t|t-1} + K_t\tilde{Z}_t, \quad P_{t|t} = (I - K_tB)P_{t|t-1}$$

where

$$
\begin{aligned}
\tilde{Z}_t &= Y_t - B\hat{X}_{t|t-1}, & \text{(Innovation)} \\
K_t &= P_{t|t-1}B^T S_t^{-1}, & \text{(Kalman gain)} \\
S_t &= BP_{t|t-1}B^T + R. & \text{(Innovation covariance)}
\end{aligned}
$$

Here Y_t is the observed position of feature at time t and $\tilde{Z}_t$ is the residual with respect to the predicted feature position (innovation). We initialise the filter by choosing some arbitrary initial state estimate, $\hat{X}_{0|0}$, and setting the initial state covariance matrix, $P_{0|0}$, to some sufficiently large multiple of I so as to specify almost no certainty on the initial estimate. We use the recurrence relations to compute estimates of states and estimation error covariances up until the last time index for the track.

The covariance matrix of the process noise Q is set to the following by default:

$$
Q = \begin{bmatrix}
0.3^2 & 0 & 0 & 0 \\
0 & 0.03^2 & 0 & 0 \\
0 & 0 & 0.3^2 & 0 \\
0 & 0 & 0 & 0.03^2
\end{bmatrix}.
$$

It is planned to sample Q in a similar fashion as the sampling of R. Supplementary figure S7 shows the agreement for the ground truth for random walks with different step lengths and the chosen positional component of the process noise.

In addition to the Kalman filter, a Rauch-Tung-Striebel backwards smoothing filter is applied. Once the forward prediction-update step has been completed for a Kalman filter, we can use a Rauch–Tung–Striebel smoother to refine our earlier state estimates. This is a backwards step which starts from the final estimated state (i.e. the one which has been influenced by all observations) and works backwards creating optimal estimates of the hidden state, $\hat{X}_{t|T}$, and estimation error covariance, $P_{t|T}$. Note that these estimates have been computed given all observations.

The estimates are computed via the following recurrence relations:

$$\hat{X}_{t|T} = \hat{X}_{t|t} + L_t(\hat{X}_{t+1|T} - \hat{X}_{t+1|t}), \quad P_{t|T} = P_{t|t} + L_t(P_{t+1|T} - P_{t+1|t})L_t^T$$

where $L_t = P_{t|t}A^T P_{t+1|t}^{-1}$.

Biggles uses these estimates for the evaluation of the log-likelihoods on track observations.

3.3 Track observations likelihood

The Kalman smoother outputs, for each candidate track, a set of maximally likely states. We assume that the evolution of states, beyond first order velocity, is due to the Gaussian noise process and so it is straightforward to write down the probability density for an individual observation, Y, corresponding to a predicted state, $\hat{X}$ by noting that $\hat{S}$, the innovation covariance or total error in the predicted observation, $\hat{Y} = B\hat{X}$, is given by $\hat{S} = B\hat{P}B^T + R$, see section 3.2. We therefore calculate the likelihood of Y assuming a Gaussian model:

$$P(Y|\hat{X}, \hat{P}, R) = \mathcal{N}(Y; B\hat{X}, \hat{S}).$$

The overall log probability density is therefore simply the sum of all the log probability densities for each observation associated with a track,

$$\ell(Y^i|k_i, \theta) = \sum_{j=1}^{N_o^i} \log P(Y_j^i|\hat{X}_{t(j)}, \hat{P}_{t(j)}, R), \tag{7}$$

where N_o^i is the number of observation of track k_i (i.e. the size of Y^i) and $t(j)$ is the time point of observation Y_j^i.

3.4 Clutter observations likelihood

The value of $\ell(Y^0|k_0, \theta)$ may be computed by assuming that each clutter observation arises independently and is uniformly likely over the field of view. Letting V be the total area (in square pixels) of the field of view, it follows that:

$$P(Y^0|k_0, \theta) = \prod_{t=1}^{T} \left(\frac{1}{V}\right)^{N_t^c}. \tag{8}$$

3.5 Parameter prior

The $\ell(\theta)$ term is computed as the sum of the (independent) parameter priors where:

- λ_b and λ_c have improper uninformative priors on them being positive.

- p_s and p_o have uniform priors over $(0, 1)$.

- R has an inverse Wishart prior $\mathcal{W}(\Phi, s)$ with parameters $\Phi = 2I$ and $s = 5$.

4 Sampling the parameters

Unlike track configuration, the parameters θ may be sampled directly since their posterior distribution is known and separable.

Sampling these parameters forms one half of the Gibbs sampler which implements the Biggles tracking algorithm; we sample a set of model parameters, θ, from the posterior distribution $P(\theta|\omega, Y)$ where ω is the partition and Y is the observed data.

The only parameter which depends directly on the observations is R, the remainder depend only on the track birth and death times and overall count of clutter observations. We may thus sample them from known conjugate parameter distributions for the distributions Specifically:

- $P(p_s|\omega) = f_\beta(p_s; 1 + N^s, 1 + N^d)$;

- $P(p_o|\omega) = f_\beta(p_o; 1 + N^o, 1 + N^s + N^b - N^o);$

- $P(T\lambda_b|\omega) = f_\gamma(T\lambda_b; 1 + N^b, 1);$

- $P(T\lambda_c|\omega) = f_\gamma(T\lambda_c; 1 + N^c, 1),$

where T is the total frame count, $f_\beta(\cdot)$ is the PDF of the Beta distribution, $f_\gamma(\cdot)$ is the PDF of the Gamma distribution, $N^s = \sum_{t=1}^{T} N_t^s$ and similarly for N^d, N^o, N^b and N^c. Closed form solutions exist to sample from these distributions.

Sampling R is more difficult as it depends on the actual observations. To sample R, we first sample the underlying states of the dynamic model using the state estimates and covariance matrices generated by the Kalman smoother. This provides us with a state sample, X, drawn from $P(X|R,\omega,Y)$ with $\omega = \{k_0, \ldots, k_K\}$ where k_0 is the clutter. The usual distribution for sampling a covariance matrix is an inverse Wishart distribution, $P(R|X,\omega,Y) = \mathcal{W}^{-1}(R; \Phi', s')$ whose parameters may be calculated as:

$$\Phi' = \Phi + \sum_{i=1}^{K} \sum_{t \in T_i} (y_{i,t} - Bx_{i,t})(y_{i,t} - Bx_{i,t})^T, \quad s' = s + \sum_{i=1}^{K} l_k$$

where K is the number of tracks in the current partition, T_i is the set of time indices with an associated observation in track k_i, $y_{i,t}$ is the observation with time index t in track k_i, $x_{i,t}$ is the estimate of state with time index t in track k_i, B is the state observation matrix from the Kalman smoother and l_i are the number of observations in track k_i. The parameters of the prior are $s = 5$, $\Phi = 2I$.

5 Notes on partition distances, ergodicity and move design

Initially we note that whether the acceptance probability of the Metropolis-Hastings sampler is larger than zero for any two partitions ω' and ω and therefore if the sampling chain can move from ω to ω' only depends on the proposal density $Q(\omega'|\omega)$ if . The target distribution for any partition is always larger than zero for most parameter values, i.e. $P(\omega|Y) = \int P(\omega, \theta|Y)d\theta > 0$ for all ω and data Y. This can be confirmed from equation (1).

5.1 Distances between partitions

In the following we address two similar questions; what is the maximum graph edit distance (GED) between any two partitions and what is the number of Biggles moves required to connect two partitions that are furthest away in terms of the Biggles moves. If we write $d(\omega, \omega')$ for the GED between ω, ω' and $b(\omega, \omega')$ for the minimum number of Biggles moves required to go from ω to ω', we are looking for

$$\max_{\omega, \omega' \in \Omega} d(\omega, \omega')$$

and

$$\max_{\omega, \omega' \in \Omega} b(\omega, \omega')$$

When applying the GED, we identify all tracks that have the same observations, i.e. we disregard eventual differences in the birth and death time as long as the observations are the same. The GED measures the distance between two partitions by counting the number of links in which the partitions differ. To transform one partition ω into

another ω' can always be achieved by first deleting all links of ω, which results in the minimal partition ω^0, and then adding all links of ω'. The maximum possible GED in the set of all partitions, Ω, therefore may occur between two maximal partitions, ω and ω', that have no links in common. In the examples presented here, the maximal possible number of links in a partition is near but less than the number of observations N. That means the GED can be close to but will always be smaller than $2N$. Or more formally:

$$\max_{\omega,\omega'\in\Omega} d(\omega,\omega') \leq \max_{\omega\in\Omega} d(\omega^0,\omega) + \max_{\omega'\in\Omega} d(\omega^0,\omega') \leq N + N \leq 2N.$$

What is the maximum of the Biggles distance, $b(\omega,\omega')$? The shortest chain from the minimum partition ω^0 to any other partition has as many moves as the number of tracks in the target partition, since no move can change the number of tracks by more than one and the birth move can create any valid track. The maximum number of tracks occurs in such partitions where all tracks have at most 3 observations, since a track with 4 observation could be split into two, removing one link. Roughly speaking, the maximum possible number of tracks is near but always less than $N/2$ (depending on the data set), i.e. $b(\omega^0,\omega) < N/2$ for any ω. Since we can always move between any two partitions via the minimum partition, not more than N moves are needed to go from any partition to any other partition,

$$\max_{\omega,\omega'\in\Omega} b(\omega,\omega') < N$$

.

5.2 Distance based on Biggles moves

A proposal mass function Q could be constructed on the basis of graph edit operations. In our case we would get two moves that add or remove a single link, respectively. On the other hand any Q may imply a distance measure on Ω, if some weak assumptions are fulfilled. The Biggles proposal moves correspond to a distance measure on Ω, i.e. $b(\omega,\omega')$, the minimum number of moves required to go from ω to ω', can be used as a distance function. We will see that the condition

$$Q(\omega|\omega') > 0 \iff Q(\omega'|\omega) > 0$$

is key for that. Biggles moves are defined via Q. If $Q(\omega'|\omega) > 0$ then it is possible that the Markov chain continues with the ω' given that the last sample was ω. Or, if the minimum number of Biggles moves required to go from ω to ω' is one, then it must be possible to directly draw ω' from Q given that the last sample is ω. We therefore have

$$b(\omega,\omega') = 1 \implies Q(\omega'|\omega) > 0 \quad.$$

However, in the current implementation of Biggles it is possible that the move construction does not lead to a new partition and the original partition is returned, hence

$$b(\omega,\omega') = 0 \implies Q(\omega'|\omega) > 0 \quad.$$

If the minimum number of Biggles moves that is required to go from one partition to another is larger than one, then the proposal density is zero otherwise the partitions in question could be connected with a single move:

$$b(\omega,\omega') > 1 \implies Q(\omega'|\omega) = 0 \quad.$$

Therefore

$$Q(\omega'|\omega) > 0 \iff b(\omega,\omega') \leq 1 \quad. \tag{9}$$

To be a distance function $b(\omega,\omega')$ needs to fulfil four conditions:

1. $b(\omega, \omega') \geq 0$ for all $\omega, \omega' \in \Omega$. This follows immediate from the definition of b (the minimum number of Biggles move to go from any partition ω to any other partition ω' is always zero or larger of course).

2. $b(\omega, \omega') = 0 \iff \omega = \omega'$. This follows immediately from the definition of b too (if the minimum number of Biggles moves to go from ω to ω' is zero, then ω is ω' and if ω is ω', then there are no moves required to go from ω to ω').

3. $b(\omega, \omega') = b(\omega', \omega)$ (*symmetry*). Since Q is designed such that $Q(\omega|\omega') > 0 \iff Q(\omega'|\omega) > 0$, it follows with equation (9) that $b(\omega, \omega') \leq 1 \iff b(\omega', \omega) \leq 1$ and with condition 2 that $b(\omega, \omega') = 0 \iff b(\omega', \omega) = 0$. Therefore $b(\omega, \omega') = 1 \iff b(\omega', \omega) = 1$. If $(\omega_i)_{i=0}^{k}$ is a partition sequence from ω_0 to ω_k, with $b(\omega_i, \omega_{i+1}) = 1$ for $i = 0 \ldots k-1$, then the reverse sequence from ω_k to ω_0 fulfils $b(\omega_{i+1}, \omega_i) = 1$ for $i = 0 \ldots k-1$. If one sequence is a shortest of such sequences (there can be more than one shortest sequence), then its reverse sequence is a shortest sequence as well and we have $b(\omega_0, \omega_k) = b(\omega_k, \omega_0) = k$. Therefore, $b(\omega, \omega')$ is symmetric.

4. $b(\omega, \omega'') \leq b(\omega, \omega') + b(\omega', \omega'')$ (*subadditivity*). Let $(\omega_i)_{i=0}^{k}$ with $\omega_0 = \omega$ and $\omega_k = \omega'$ be a shortest move sequence from ω to ω' and let $(\omega_i)_{i=k}^{k+l}$, with $\omega_{k+l} = \omega''$ be a shortest move sequence from ω' to ω''. Then $(\omega_i)_{i=0}^{k+l}$ is a (not necessarily shortest) move sequence from ω to ω'' of length $b(\omega, \omega') + b(\omega', \omega'') = k + l$. Therefore $b(\omega, \omega'')$, the length of a shortest move sequence from ω to ω'', is at most $b(\omega, \omega') + b(\omega', \omega'')$.

However, a distance measure based on the Biggles move is more complicated to implement and less general, therefore we didn't apply it. Biggles moves include all possibilities to add or remove a single link. Adding a single link can by accomplished by the birth move, if both observations are clutter, by the extend move, if one observation is clutter and the other at the end of the track and the merge move, if both observations are part of a track. In other words, if ω and ω' only differ in 1 link, then it always holds $Q(\omega|\omega') > 0$. It always holds that

$$b(\omega, \omega') \leq d(\omega, \omega')$$

for all $\omega, \omega' \in \Omega$.

5.3   Ergodicity

We prove that the partition sampler is ergodic by following the argument by K. Murphy [1]. In order to show that the limiting distribution exists, we need to show that biggles chains are *irreducible* and *aperiodic*. A chain is irreducible if we can get from any partition to any other partitions, which we have shown above. A chain is aperiodic, if at least one partition in this chain is aperiodic, which in turn means there are two move sequences with length l_1 and l_2 that start and end in this partition so that the greatest common divisor of l_1 and l_2 is 1. Lets take a track with three observations. The following two sequences are always possible: death-birth with a length of 2 and reduce-death-birth with a length of 3. Therefore Biggles chains are aperiodic and the limiting distribution exists. Since Ω is finite and Biggles chains are irreducible, all ω are *recurrent* and *non-null*. Therefore Biggles chains are ergodic.

5.4   Number of modes of the target distribution and move design

Our first implementation of the algorithm did not include the transfer move and the cross-over move. To do a cross-over from ω to ω' without a special move requires two

split moves and two merge moves, i.e. it requires to remove two links and to add two different links, creating intermediate partitions ω_1, ω_2 and ω_3. While it may be that $P(\omega|\theta, Y) \approx P(\omega'|\theta, Y)$, we may find $P(\omega_1|\theta, Y) \ll P(\omega|\theta, Y)$ and therefore the proposal ω_1 is likely to be rejected and it may take a large number of samples to reach ω'. The absence of the cross-over move creates a probability mass trench between ω and ω'. In other words, without the cross-over move the target distribution may have two modes that merge into one mode if the cross-over move is added. Hence, the modification of the proposal moves changes the number of local extrema of the target distribution. The value of target density of any partition remains unaffected, of course.

The design of the moves requires great care. For example during the execution of the birth move, rather than sampling the death time we could continuously sample observation until this fails, using a Poisson distribution to determine the time gaps between observations. However, the probability of sampling such a track would contain not just the probabilities of sampling the observations, but also the probability of failing to sample any observation at the end of the track. This fail probability is practically difficult to determine.

The design of the proposal mass function is critical for the performance of Biggles. While a short maximum distance between any two partitions is favourable, more important is the reduction of the number of modes and the reduction of $Q(\omega|\omega')$ for less acceptable proposals ω given ω'.

6   Technical Notes

6.1   Diffusion coefficient

We calculated the mean squared displacement $\langle r^2(\tau) \rangle$ for a partition ω as

$$\langle r^2(\tau) \rangle = \frac{1}{K} \sum_{i=1}^{K} \frac{1}{N_o^i - \tau} \sum_{j=1}^{N_o^i - \tau} d(Y_{j+\tau}^i, Y_j^i)^2 \quad .$$

As before, K is the number of tracks, N_o^i is the number of observations of track k_i and $d(Y_{j+\tau}^i, Y_j^i)$ is the Euclidean distance between observations $Y_{j+\tau}^i$ and Y_j^i. For readability, this equation assumes that there are no observation gaps in the tracks. The actual calculation skips expressions where the track has no pair of observations with time lag τ and the scaling factor $N_o^i - \tau$ is reduced accordingly. The diffusion coefficient D was estimated from

$$\langle r^2(\tau) \rangle = 4D\tau \,, \quad \tau \in \{1, 2\}.$$

The estimation uses the slope yielded by linear regression.

References

1. Kevin P. Murphy. 2012. Machine Learning. A probabilistic Perspective. MIT Press.
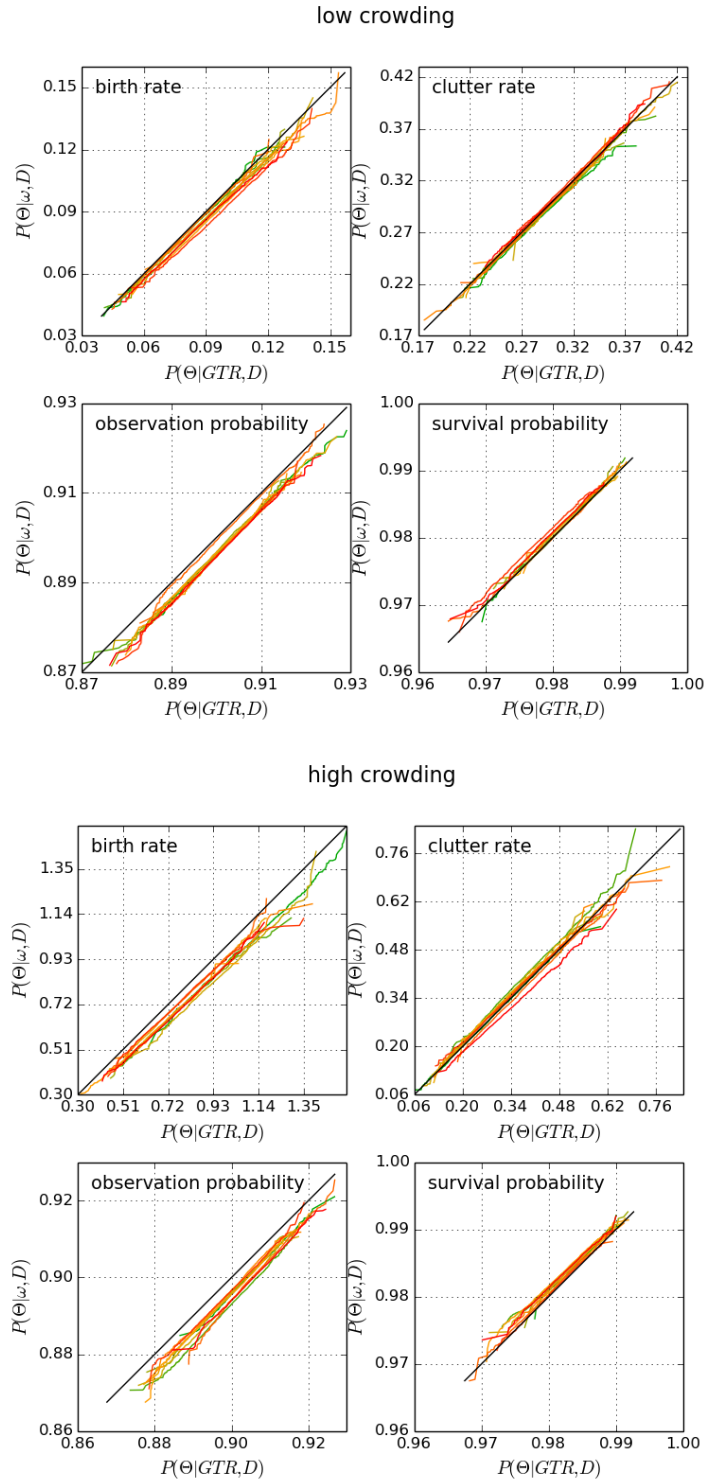
Contents

1 S1 Figure

Figure 1: Recovering of the simulation parameters. The Q-Q plots of the parameter samples for the GTR and the parameter samples created by Biggles. Shown are a series of ten data sets with low track density (top) and a series of data sets with high track density (bottom).
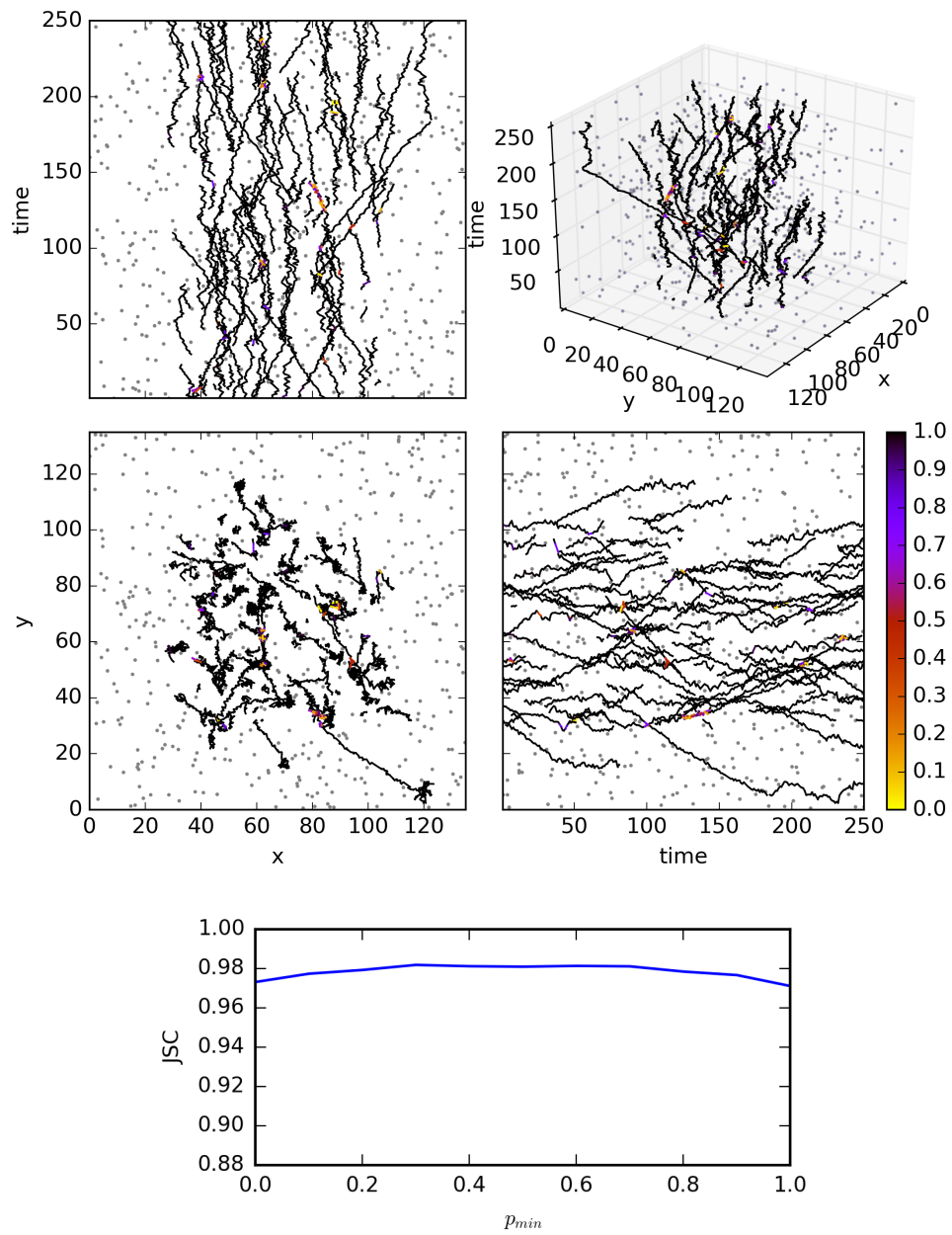
2 S2 Figure

Figure 2: Biggles tracking example. A mixture between directed motion and random walks. Tracks have a chance to change the mode of motion. The grey dots mark the clutter observations.
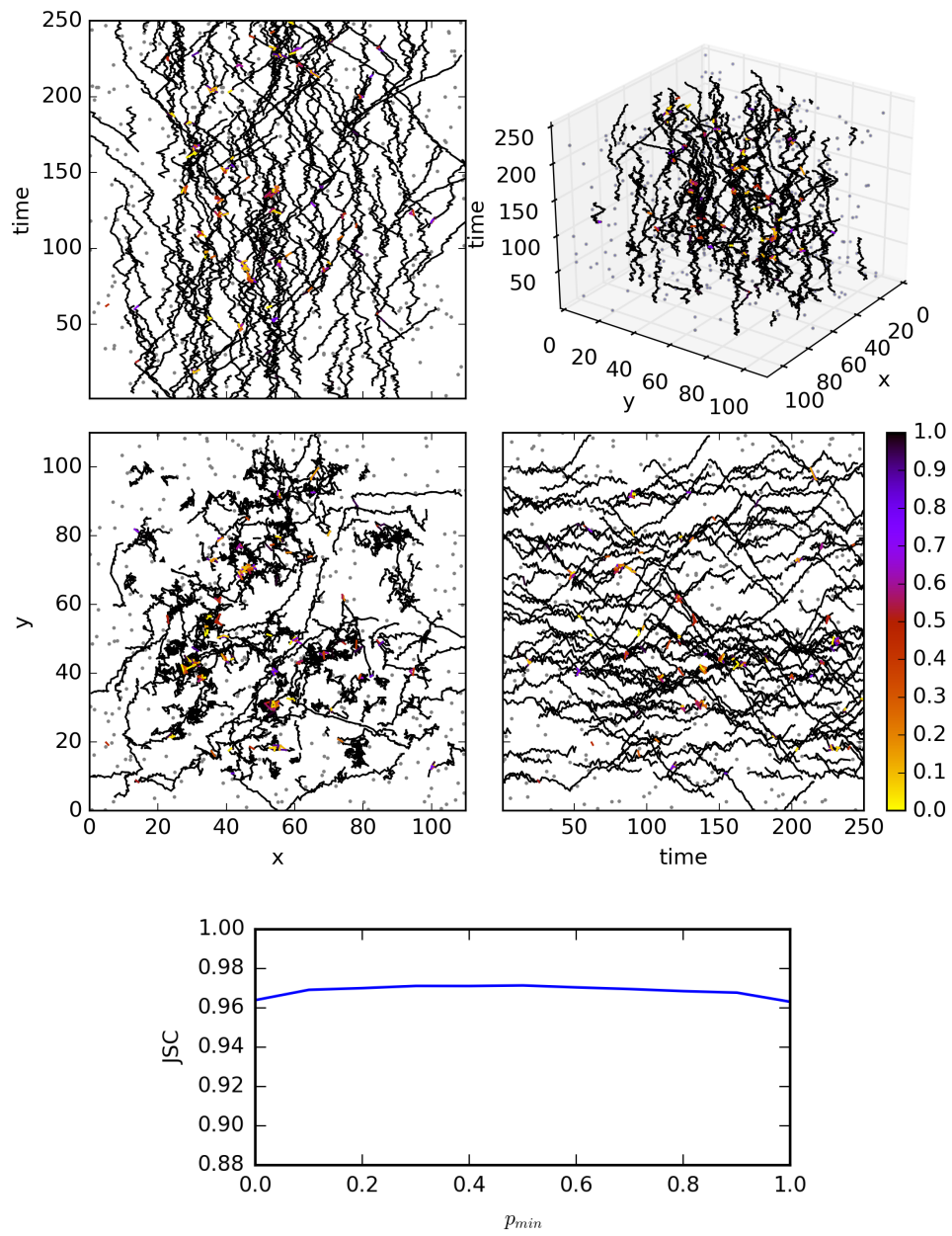
3 S3 Figure

Figure 3: Biggles tracking example. A 50-50 mixture between directed motion and random walks. Each track has one mode of mode of motion.

4 S4 Figure
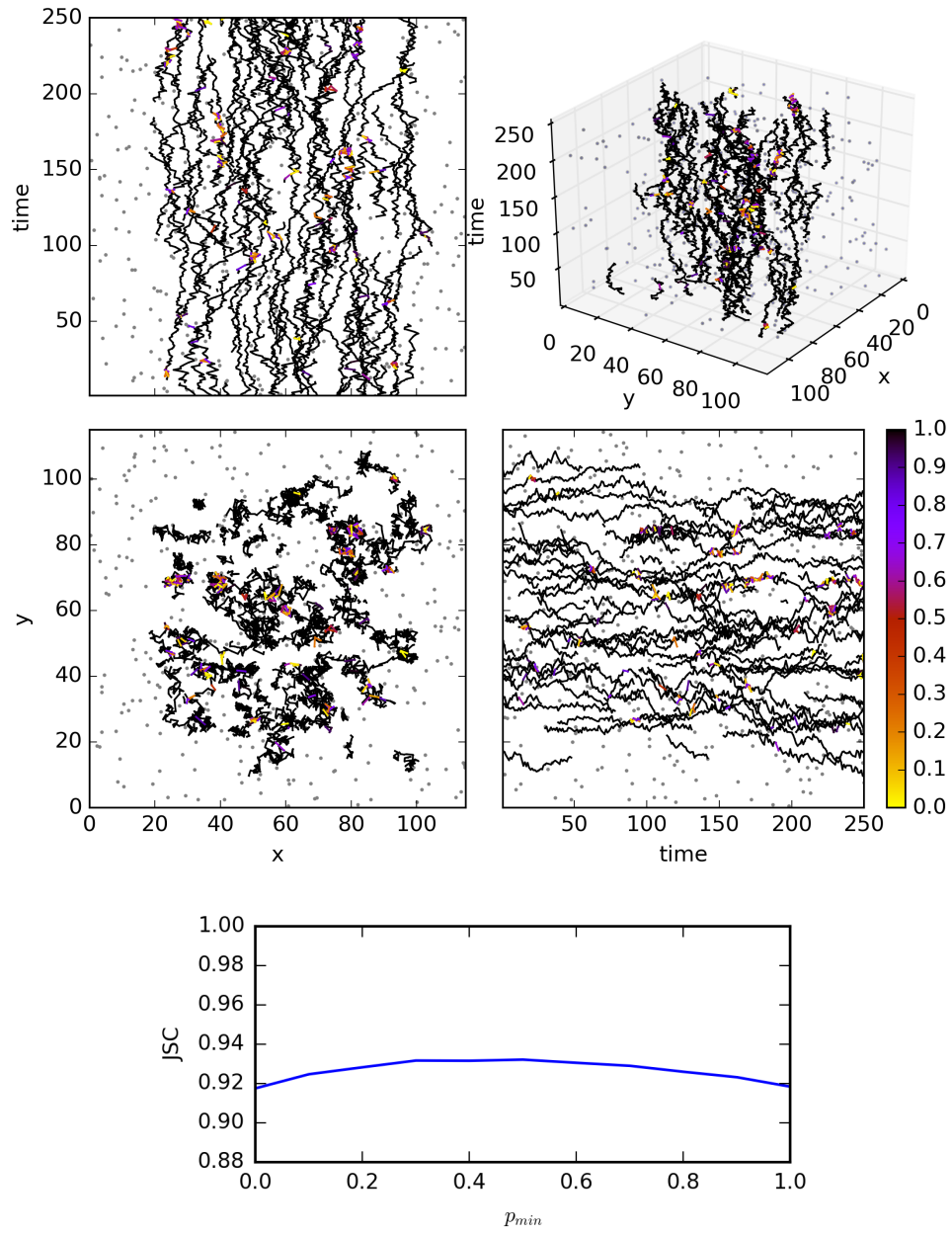
Figure 4: Biggles tracking example. A 50-50 mixture between random walks with two different diffusion coefficients, $d_1 = 0.45pix/fr$ and $d_2 = 0.9pix/fr$. Each track has one mode of mode of motion.

5   S5 Figure
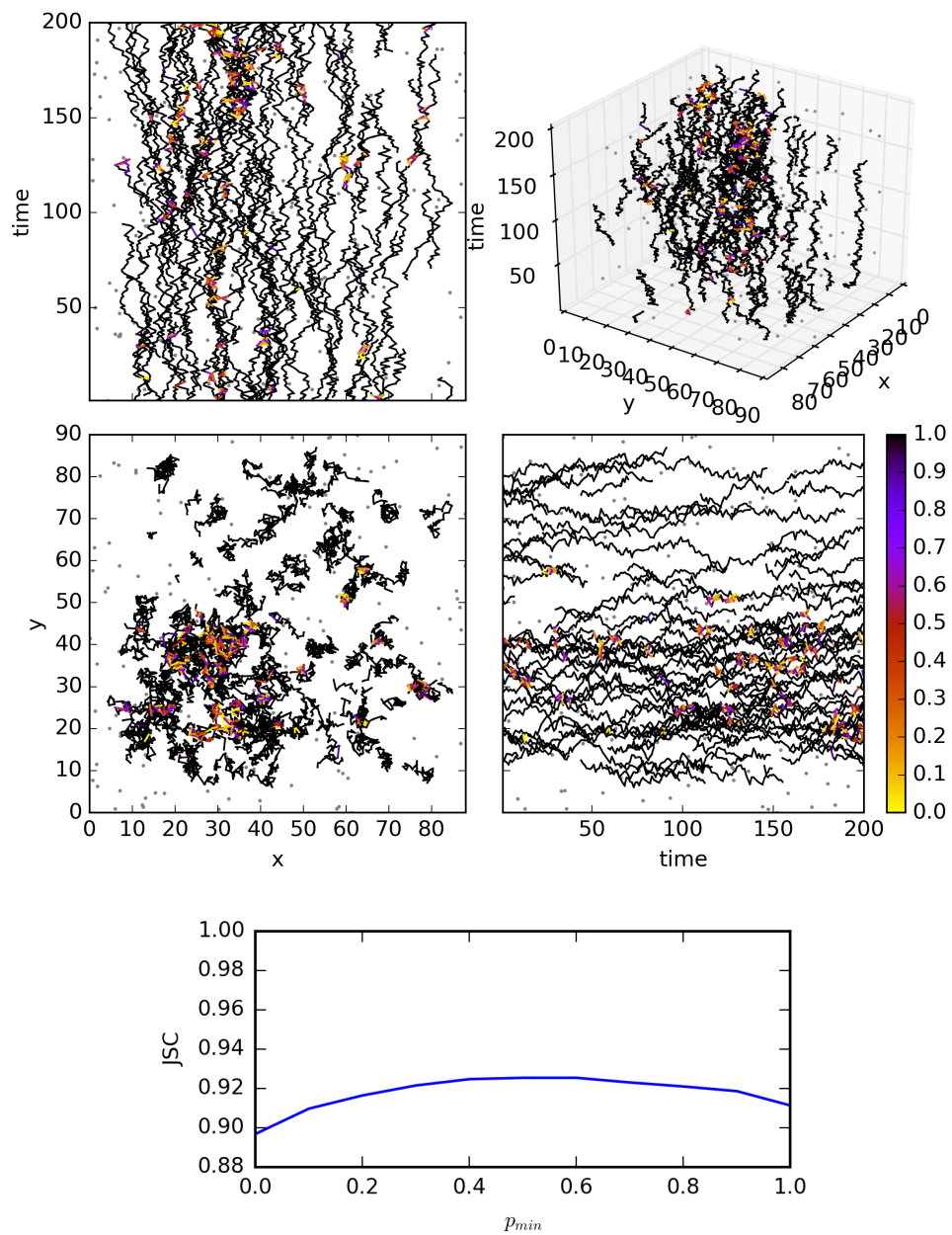
Figure 5: Biggles tracking example. Random walks with regions of different densities. Each track has one mode of mode of motion.
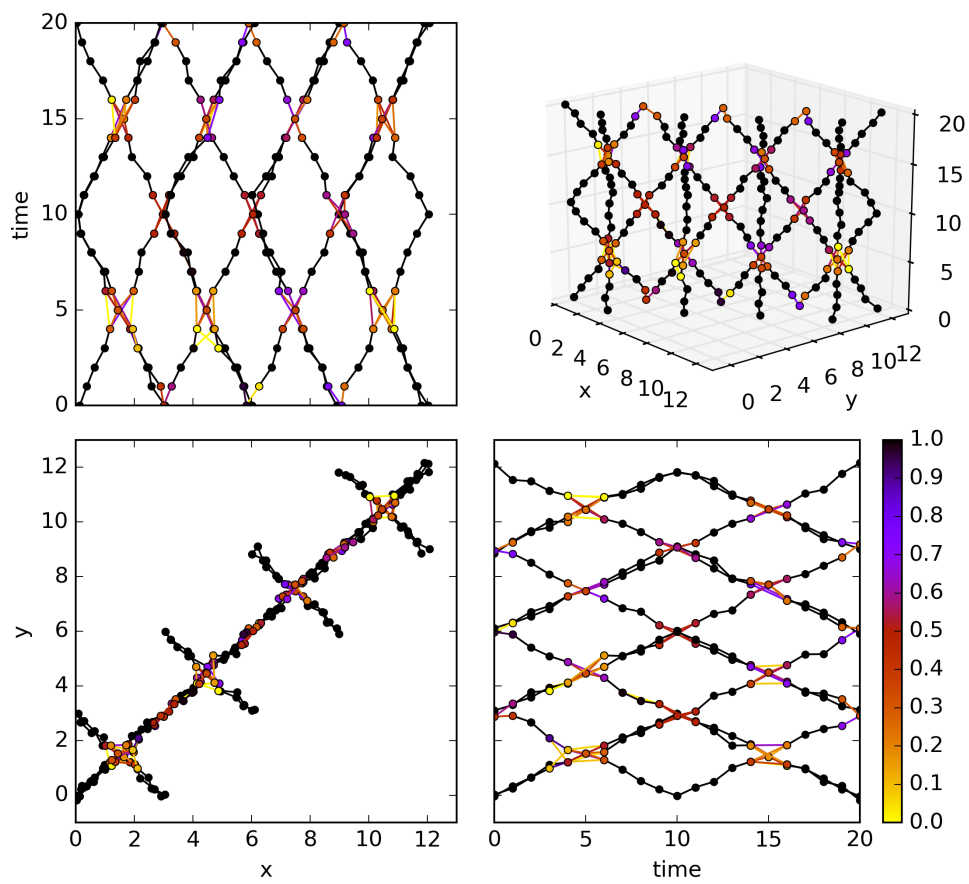
6 S6 Figure

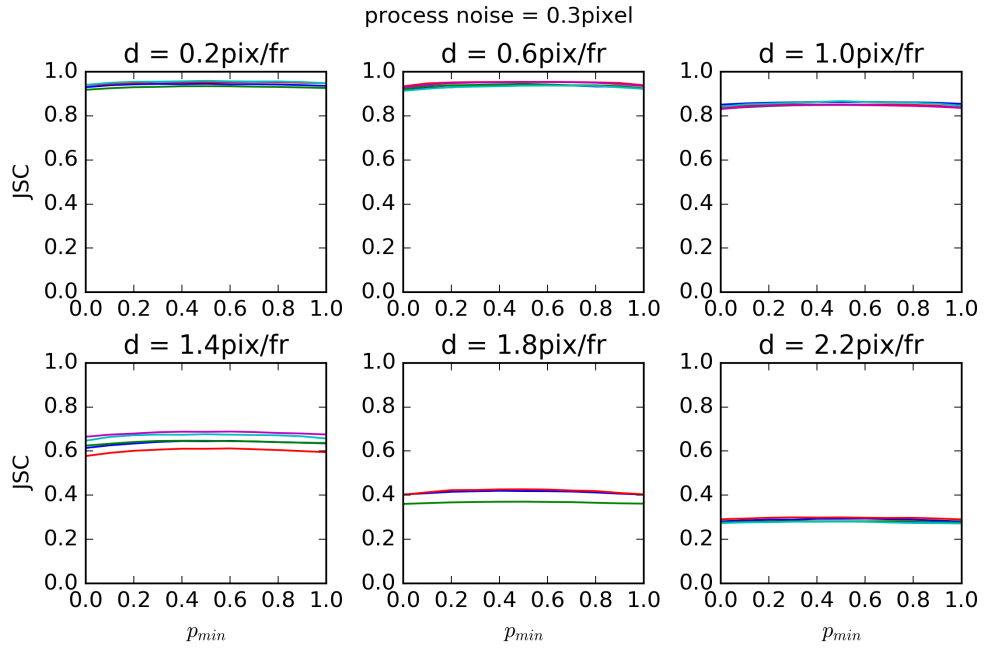Figure 6: Alternative views of the complicated data set. Shown are the link probabilities.

7 S7 Figure

Figure 7: Dependency on the process noise. A step length of $1 pixel/frame$ is equivalent to $D \approx 0.26 \mu m^2/s$ at a pixel size of $160nm$ and a frame rate of $20Hz$
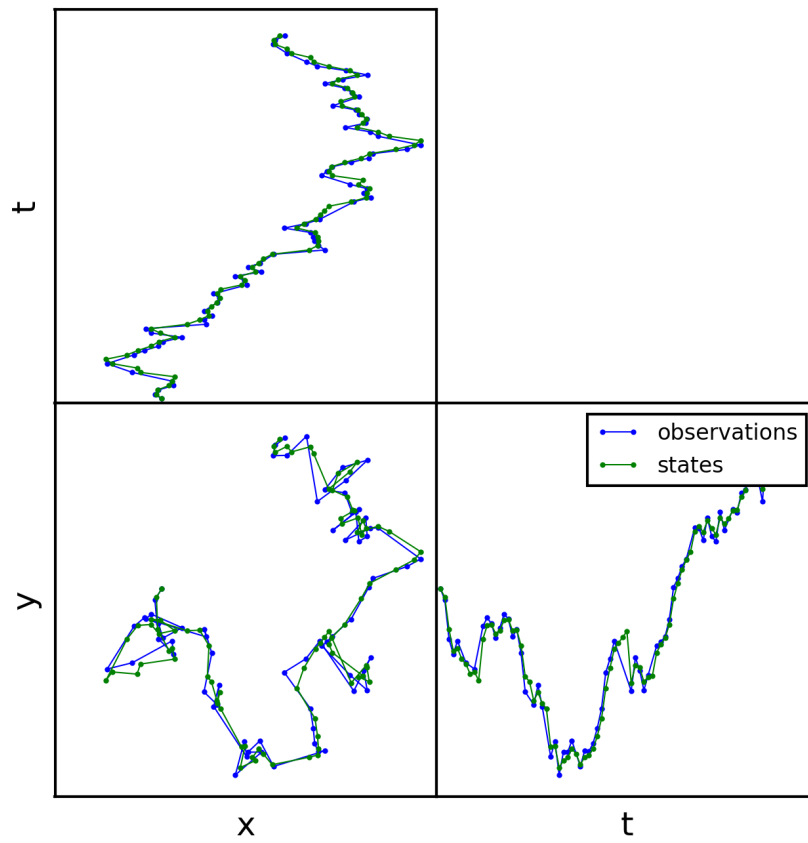
8 S8 Figure

Figure 8: Example of observation and estimated states of a track.