



Uncertainty modelling and computational aspects of data association

Jeremie Houssineau¹ · Jiajie Zeng² · Ajay Jasra³

Received: 8 September 2020 / Accepted: 1 August 2021 / Published online: 14 August 2021
© The Author(s) 2021

Abstract

A novel solution to the smoothing problem for multi-object dynamical systems is proposed and evaluated. The systems of interest contain an unknown and varying number of dynamical objects that are partially observed under noisy and corrupted observations. In order to account for the lack of information about the different aspects of this type of complex system, an alternative representation of uncertainty based on possibility theory is considered. It is shown how analogues of usual concepts such as Markov chains and hidden Markov models (HMMs) can be introduced in this context. In particular, the considered statistical model for multiple dynamical objects can be formulated as a hierarchical model consisting of conditionally independent HMMs. This structure is leveraged to propose an efficient method in the context of Markov chain Monte Carlo (MCMC) by relying on an approximate solution to the corresponding filtering problem, in a similar fashion to particle MCMC. This approach is shown to outperform existing algorithms in a range of scenarios.

Keywords Possibility theory · Markov chain Monte Carlo · Simulated annealing · Multi-target tracking

1 Introduction

We consider the problem of performing inference for multi-object dynamical systems under partial, corrupted and noisy observations. This class of problems, known as *multi-target tracking* in the engineering literature (Fortmann et al. 1980; Mahler 2003; Vo et al. 2014), arises in many applications, e.g. bio-imaging (Chenouard 2014), robotics (Mullane et al. 2011) and surveillance (Benfold and Reid 2011), which can all benefit from principled inference solutions in different ways: i) when the number of objects is too large to be treated by hand, ii) when the phenomena of interest take place on extended periods of time or, conversely, when an immediate response is needed, iii) when the data available about

each object are scarce and iv) when it is difficult to tell one object from another. One of the main difficulties with the considered type of system is that the number of objects is not known a priori and might vary in time due to a birth-death process. Also, objects are observed under multiple perturbations: i) each object might or might not be detected, ii) if an object is detected then its state is only partially observed and the observation is subject to noise and iii) observations not related to any object, referred to as false alarms, are also received. The main task when inferring the number of objects in a given system as well as their respective state is to solve the data association problem, that is, to estimate whether or not observations at different time steps originate from the same object. Each of the above-mentioned perturbations incurs a significant increase in the size of the set of all possible data associations, making it highly combinatorial. Due to this combinatorial nature, the task of estimating the current state of all objects based on all previous observations, referred to as multi-object filtering, is a difficult problem. It has been an active research topic for several decades and continues to be challenging in spite of the ever-increasing available computational resources (Fortmann et al. 1980; Vo et al. 2014). In this article, we aim to tackle the even harder problem of multi-object smoothing, that is, our objective is to keep evaluating the likelihood of data associations at previous times in light of newly received data. This is an important problem in prac-

This work was supported by Singapore Ministry of Education tier 1 Grant R-155-000-182-114. AJ was additionally supported by KAUST baseline funding.

✉ Jeremie Houssineau
jeremie.houssineau@warwick.ac.uk

¹ Department of Statistics, University of Warwick, Coventry CV4 7AL, UK

² Department of Statistics and Applied Probability, National University of Singapore, Singapore 119077, Singapore

³ Computer, Electrical and Mathematical Science and Engineering Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia

tice since the elicitation of objects' trajectories and origins is fundamental for the evaluation of the objects' identities and of the associated situational awareness. Indeed, knowing the current state of each object is not sufficient in many situations and maintaining an up-to-date estimate of their past trajectories is often crucial. For instance, in defence applications, if an object labelled as "ally" crosses path with another object labelled as "enemy", then being able to tell one from the other at a later time can be more critical than having an accurate estimate of their state at that time.

In the context of filtering, one of the most natural ways of improving the trajectory estimates over the last few time steps is referred to as fixed-lag smoothing, where a sliding window made of a given number of time steps is updated based on the latest observations. The advantage with fixed-lag smoothing is that the computational cost can easily be tuned by selecting an adequate lag. However, since our objective is to elicit particular events that might have taken place at arbitrary time steps, we consider instead a "batch" alternative where a user-defined time window of interest is fixed. It would be possible to extend our approach in order to make it suitable for online estimation, e.g. by running the proposed MCMC chain over a fixed lag for some or all iterations of a suitable online estimation algorithm in order to further explore the set of data associations; this extension is kept for future work. Further generalisations of our approach could consider smoothing on an infinitely growing time interval; however, this would require improving the current understanding of the statistical properties of multi-object systems (Houssineau et al. 2019b) in order to leverage the corresponding forgetting properties.

Defining a standard statistical model for representing multiple objects requires setting a number of probability distributions and parameters to characterise the different aspects of the problem, including highly uncertain phenomena such as false alarms. Such models also usually ignore the disparity between the different objects of interest in terms of behaviour and detection profile. In this article, we consider an alternative representation of uncertainty (Houssineau et al. 2019a; Houssineau 2018b), based on possibility theory (Dubois and Prade 2015), that allows for acknowledging the lack of information about the different aspects of multi-object dynamical systems with the objective of increasing the robustness to misspecification of the derived solutions. The considered representation of uncertainty has links with imprecise probabilities (Walley 1991) and Dempster–Shafer theory (Dempster 1967; Shafer 1976).

Our motivation for using possibility theory in this work stems from the provided ability to assess the *consistency* between a model and an observation via the marginal likelihood rather than the associated *fitness* of the model as with probability theory. The difference is that greater uncertainty in the model yields greater consistency but smaller fitness. This notion of consistency can be related informally to a

model-based spatio-temporal distance between observations, and is key when exploring the set of all possible data associations.

The use of MCMC to solve data association problems has been previously explored in Oh et al. (2009) as well as in Vu et al. (2014), Jiang et al. (2015) and Jiang and Singh (2018). The approach considered in these articles is based on local proposals in the set of data association, with Vu et al. (2014), Jiang et al. (2015) and Jiang and Singh (2018) additionally considering the estimation of the object's trajectories. The objective in this article is to show that a suitable trade-off can be found between proposing global moves and maintaining a reasonable probability of acceptance of each move. This result is achieved by leveraging the efficiency of an approximate multi-object filtering method. The use of MCMC in discrete spaces is discussed more generally in Zanella (2019). MCMC has also been used in conjunction with, or as a replacement of, sequential Monte Carlo in the context of filtering for single-object systems (Gilks and Berzuini 2001) and multi-object systems (Khan et al. 2005; Septier et al. 2009; Carmi et al. 2012; Maroulas and Stinis 2012; Bao and Maroulas 2017); however, this type of approach is less directly related to the method proposed in this article.

Overall, the contributions of the articles are as follows: i) a full multi-object model is defined in the context of possibility theory, building up on the components of single- and multi-object models of Ristic et al. (2020) and Houssineau (2018a); ii) a possibilistic analogue of the scalable solution to multi-object filtering of Houssineau and Clark (2018) is introduced; iii) the tools of possibility theory are used to define a suitable structure on the set of data associations; and iv) a new efficient MCMC-based solution for the multi-object smoothing problem is introduced and its performance is demonstrated.

The proposed statistical model for representing multi-object systems is described in Sect. 2. This is followed by the presentation of the proposed method for exploring the set of data association in Sect. 3, before considering an extension of this approach in Sect. 4. The performance of the proposed method is then assessed on simulated data in Sect. 5.

2 Model

We consider a fixed number K of time steps and assume without loss of generality that time steps take integer values between 1 and K . At each time step $k \in \{1, \dots, K\}$, a set of observations Z_k is received, containing both object-originated observations and false alarms. Each observation in the set Z_k is an element of an observation set Z , which is assumed to be a subset of \mathbb{R}^{d_Z} , where d_Z is the number of components of each observation, e.g. $d_Z = 3$ for a radar measuring range, azimuth and Doppler shift. In order to model

that an object might not be detected, we introduce the notation ϕ for the empty observation, that is, an object for which detection has failed is associated with the empty observation ϕ . We assume, as is standard, that an object cannot generate more than one observation at each time step. Therefore, denoting $\bar{Z}_k = Z_k \cup \{\phi\}$ the set of observations at time k augmented with the empty observation for any $k \in \{1, \dots, K\}$, any sequence of observations generated by an object through the K time steps of the scenario can be seen as an element of

$$\mathcal{O}_K = \bar{Z}_1 \times \dots \times \bar{Z}_K \setminus \{\phi\}^K$$

where the sequence of observation containing empty observations only is not considered. Elements of \mathcal{O}_K are also referred to as *observation paths* or simply as *paths*. Multi-object data association can then be seen as the problem of determining the probability for all the paths in a given subset of \mathcal{O}_K to be the true paths of objects in the system under consideration; i.e. we are interested in the joint detection and tracking of all objects. Another standard assumption about multi-object systems is that each observation cannot originate from more than one object; as a consequence, not all subsets of \mathcal{O}_K are considered feasible and we focus on the set \mathcal{A} of subsets of \mathcal{O}_K such that for any set $A \in \mathcal{A}$, any two different observations paths o and o' in A must verify that either $o_k = o'_k = \phi$ or $o_k \neq o'_k$ for all $k \in \{1, \dots, K\}$, where o_k denotes the k th element of the path o . Less formally, elements of \mathcal{A} only contain paths that are different where they are not both equal to the empty observation. In practice, it might happen that a single object generates multiple observations, e.g. if it spans several resolution cells, or that multiple objects generate a single observation, e.g. if they are within the same resolution cell, these generalisations significantly increase the complexity of the problem and are kept for future work. The set \mathcal{A} , in spite of being a strict subset of the power set of \mathcal{O}_K , has a large cardinality and evaluating the credibility of each of its elements by exhaustion can be difficult even when the number of observations at each time step is small. Assuming, for simplicity, that the number of observations at every time step is constant and equal to m , the number of elements in the power set of \mathcal{O}_K is equal to $2^{mK} - 1$, which is prohibitively large even for toy problems. It is generally difficult to devise algorithms that perform inference on a large discrete space such as \mathcal{A} , yet, MCMC methods can help to address part of this challenge since they only require being able to evaluate the credibility of a given association $A \in \mathcal{A}$ proposed via some user-defined transition kernel.

In practice, we also need to estimate the interval of existence of each object. For this purpose, we introduce a set \mathcal{T} which is similar to \mathcal{A} except that each path o will be paired with a time of appearance $m \in \{1, \dots, K\}$ and the last time of existence $n \in \{m, \dots, K\}$; the tuple (o, m, n) will be referred to as a *track* since all the observable characteristics of the cor-

responding hypothetical object can be estimated using this tuple. Formally, for any set $T \in \mathcal{T}$, any track $t = (o, m, n)$ in T must verify $o_k = \phi$ for any $k \notin \{m, \dots, n\}$ and, for any (o', m', n') in T different from (o, m, n) , it must hold that either $o_k = o'_k = \phi$ or $o_k \neq o'_k$ for all $k \in \{1, \dots, K\}$, as for data associations. We denote by κ the function extracting paths from tracks, that is, $\kappa(t) = o$ for any track t with path o .

2.1 Uncertain variable and possibility function

We consider the representation of uncertainty described in Houssineau et al. (2019a) which can be used as an alternative to subjective probabilities in a statistical model. The objective of this representation of uncertainty is to model information rather than randomness and therefore to address common issues with statistical modelling for complex systems and with the use of subjective probabilities. In the context of multi-object systems, some of these issues are:

- 1) the associated models are hierarchical which precludes the use of improper priors on the first level of this hierarchy;¹ however, there is often no prior information on the location of appearing objects which means that uninformative priors are needed;
- 2) as with many complex systems, there are a large number of parameters which are not necessarily known in practice and learning these parameters is both challenging computationally and potentially useless if they are likely to change drastically from one time step to the other; this is for instance the case with the probability of detection;

As will be shown in the next few sections, the proposed approach allows for addressing these issues while preserving most of the usual intuitive mechanisms in Bayesian inference.

We model a fixed but unknown quantity as a mapping \mathbf{x} from a sample space Ω to a set X , referred to as an *uncertain variable*. The difference with a random variable is that Ω is not defined as a probability space and the value x^* of the unknown quantity is not seen as the realisation of a “true” underlying probability distribution; instead, there is a reference element in Ω , denoted ω^* , such that $\mathbf{x}(\omega^*) = x^*$. This modification implies that uncertainty should be modelled differently and we consider the approach where the information about the true value of \mathbf{x} is represented by a non-negative function f_x on X verifying $\sup_{x \in X} f_x(x) = 1$, referred to as a *possibility function*. In this context, instead of defining the probability of an event $X \in A$ for some subset $A \subseteq X$ as

¹ An improper prior contains an arbitrary constant which is not compensated for in the hierarchical Bayes rule since other terms with proper priors also appear; this arbitrary constant therefore remains in the posterior, which is not acceptable.

$\int_A p_X(x)dx$ for some random variable $X \sim p_X$, we define the *credibility* of the event $x \in A$ as $\bar{P}(A) = \sup_{x \in A} f_x(x)$. The notion of credibility can be illustrated by the two extreme cases: if $\bar{P}(A) = 0$ then the event $x \in A^c = X \setminus A$ happened almost surely, whereas if $\bar{P}(A) = 1$ then there is no information against the event $x \in A$. (However, it might also hold that $\bar{P}(A^c) = 1$ as opposed to the probabilistic case.) Possibility functions are not characterised by their corresponding uncertain variables, and instead, we say that the possibility function *describes* the uncertain variable. If y is another uncertain variable in a set Y and if x and y are jointly described by the possibility function $f_{x,y}$ then y is described by the marginal possibility function

$$f_y(y) = \sup_{x \in X} f_{x,y}(x, y), \quad y \in Y.$$

Assuming that $f_y(y) > 0$ for some given $y \in Y$, the conditional possibility function describing x given that $y = y$ is defined for any $x \in X$ as

$$f_x(x|y) = \frac{f_{x,y}(x, y)}{f_y(y)} = \frac{f_y(y|x)f_x(x)}{\sup_{x' \in X} f_y(y|x')f_x(x')},$$

which is the analogue of Bayes' theorem for possibility functions (De Baets et al. 1999). In this context, we will refer to f_x and $f_x(\cdot|y)$ as the prior and posterior possibility functions, respectively, and $f_y(y|\cdot)$ will be called the likelihood function; similarly, $f_y(y)$ will be referred to as the marginal likelihood. If it holds that $f_{x,y}(x, y) = f_x(x)f_y(y)$ for all $(x, y) \in X \times Y$, then x and y are said to be independently described. This form of independence only implies that the information about x is not related to the information we hold about y .

A notion of expected value can be defined for possibility functions via the corresponding law of large numbers of Houssineau et al. (2019a) as

$$\mathbb{E}^*(x) = \arg \sup_{x \in X} f_x(x).$$

Another useful notion of expected value, which is the direct analogue of the standard expected value, can be defined for any real-valued function φ on X as

$$\bar{\mathbb{E}}(\varphi(x)) = \sup_{x \in X} \varphi(x) f_x(x).$$

The scalar $\bar{\mathbb{E}}(\varphi(x))$ can be interpreted as the maximum expected value of $\varphi(x)$. If the expected value $\bar{\mathbb{E}}(x)$ is interpreted as the first moment, then $\mathbb{E}^*(x)$ can be seen as the argument of the supremum in the zeroth moment $\bar{\mathbb{E}}(x^0) = \sup_{x \in X} f_x(x) = 1$.

Many concepts and results holding for probability distributions can be used for possibility functions. For instance,

if $Y = X = \mathbb{R}$ and if the likelihood function is a normal possibility function, i.e.

$$f_y(y|x) = \exp\left(-\frac{1}{2\sigma^2}(y-ax)^2\right) \doteq \bar{N}(y; ax, \sigma^2)$$

for some $a \in \mathbb{R}$ and some $\sigma > 0$, then one can show that the posterior is also a normal possibility function if the prior f_x is normal. In other words, the concept of conjugate priors makes sense. This result can be extended to the multivariate case, and it has been shown in Houssineau and Bishop (2018) that the posterior expected value and variance of the Kalman filter can be recovered with possibility functions.

If the objective is to find the subjective probability $p(B)$ of some event $x \in B$ for some measurable subset B of X , then the credibility $\sup_{x \in B} f_x(x)$ can be seen as an upper bound for this subjective probability, and we find that

$$1 - \sup_{x \in B^c} f_x(x) \leq p(B) \leq \sup_{x \in B} f_x(x). \quad (1)$$

Since x is not random, we emphasise that $p(B)$ is a subjective probability; that is, $p(B)$ only captures a degree of belief for the event $x \in B$ rather than a probability in the strict sense (e.g. as a limiting frequency for that event). Equation (1) implies that the possibility function **1**, which is equal to 1 everywhere on X , is the least informative; this uninformative possibility function is well defined even when X is unbounded. Another consequence of (1) is that an alternative possibility function f'_x for x can be seen as being less informative than f_x if $f'_x(x) \geq f_x(x)$ for any $x \in X$; indeed, the bounds in (1) will be wider with f'_x . This is related to the notion of *credal set* that is common in possibility theory (Dubois and Prade 2015). It is also possible to interface uncertain variables and random variables in order to introduce more sophisticated representations of uncertainty involving both lack of information and randomness (Houssineau 2018b). However, we will argue that all the elements of the introduced statistical model can be seen as deterministic so that only possibility functions will be used. Our approach has connections with the treatment of single-object dynamical systems based on random set theory, see, e.g. (Mahler 2007) where generalised likelihood functions are suggested. Our contribution is the modelling of the entire multi-object system with possibility functions.

2.2 Multi-object model

We first introduce the assumptions and notations for modelling the way objects appear, behave and disappear in Sect. 2.2.1 before moving on to the considered sensor modelling in Sect. 2.2.2. Most of the assumptions are standard in the field of multi-object estimation.

2.2.1 Object and population dynamics

We consider the case where there is no information about some or all of the components of the state of appearing objects. Typically, there might be no prior information about the position of objects, whereas assumptions can be made about the velocity components. Denoting $m \in \{1, \dots, K\}$ the time step at which a given object has appeared, the state at this time step is represented by an uncertain variable \mathbf{x}_m in a space $\mathbf{X} \subseteq \mathbb{R}^{d_x}$ described by a possibility function f_0 . With probabilistic modelling, improper priors might be required in order to model the absence of information about appearing objects; however, the hierarchical nature of multi-object estimation implies that improper priors cannot be used without adding heuristics at the level of data association (Houssineau and Laneuville 2010; Ristic et al. 2012).

We consider that there is a non-negligible heterogeneity between the dynamics of the different objects and that the characteristics of the objects' motion is not necessarily well known. As a consequence, and recalling that n is the last time step of existence of the object, we model the trajectory of an object as a sequence of uncertain variables $\{\mathbf{x}_k\}_{k=m+1}^n$ on \mathbf{X} such that, for any $k \in \{m+1, \dots, n\}$, \mathbf{x}_k is described by a possibility function $f_{\mathbf{x}_k}(\cdot | \mathbf{x}_m, \dots, \mathbf{x}_{k-1})$ satisfying

$$f_{\mathbf{x}_k}(\mathbf{x}_k | \mathbf{x}_m, \dots, \mathbf{x}_{k-1}) = g_k(\mathbf{x}_k | \mathbf{x}_{k-1}), \quad \mathbf{x}_k \in \mathbf{X},$$

for some possibility function $g_k(\cdot | \mathbf{x}_{k-1})$ on \mathbf{X} . This is an analogue of the Markov property for uncertain variables.

We take into account the fact that objects might completely disappear from the scene before the last time step, in which case we say that the object has “not survived”. This could be seen as a convenient way of dealing with objects that are no longer detectable by the sensor(s). Since object's dynamics is modelled via uncertain variables, we also model object survival as deterministic. The respective credibilities for an object with state $x \in \mathbf{X}$ to survive or not survive to the next time step are denoted $\alpha_s(x)$ and $\alpha_{ns}(x)$. These credibilities must verify $\max\{\alpha_s(x), \alpha_{ns}(x)\} = 1$ for any $x \in \mathbf{X}$. We consider the case where $\alpha_s = \mathbf{1}$, i.e. $\alpha_s(x) = 1$ for any $x \in \mathbf{X}$, since we want to model that objects are unlikely to disappear right after appearing, for which we need to set $\alpha_{ns}(x) \ll 1$ for any $x \in \mathbf{X}$. The subjective probability of survival for an object with state $x \in \mathbf{X}$ is therefore restricted to the interval $[1 - \alpha_{ns}(x), 1]$.

Given the introduced model and notations, the joint credibility of a trajectory $x_{m:n} = (x_m, \dots, x_n) \in \mathbf{X}^{n-m+1}$ and of the corresponding last time of existence $n \in \{1, \dots, K\}$ for an object that is known to appear at time step m can be characterised by the possibility function

$$g(x_{m:n}, n | m) = f_0(x_m) \alpha_{ns}(x_n)^{\mathbb{1}(n < K)} \prod_{k=m+1}^n g_k(x_k | x_{k-1}),$$

where $\mathbb{1}(e)$ equals 1 if e is true and 0 otherwise.

There are several possible models for the number of appearing objects per time step. The simplest is to assume that the credibility for an object to appear at time $k \in \{1, \dots, K\}$ is $\alpha_{k,+}$ and that this aspect can be independently described for all objects. The credibility $f_{k,+}(M)$ for M objects to appear at time step k is then $f_{k,+}(M) = \alpha_{k,+}^M$. Additional information might, however, be available about appearing objects, such as a maximum number $M_{k,+}$ at time step k , in which case credibility for M objects to appear would be $\mathbb{1}(M \leq M_{k,+}) \alpha_{k,+}^M$.

2.2.2 Observation

Most sensors acquire information about the objects of interest by measuring some signal over an array of resolution cells. This is the case for cameras, where these resolution cells are pixels, but also for most radars and sonars (Skolnik 1990). Considering for instance the case of a radar measuring range and azimuth, the internal processing of the radar image yields a set of resolution cells where the strength of the signal suggests the presence of an object in the corresponding directions and at the specified distances. In addition, objects are often extended and the signal can originate from different edges and/or surfaces depending on their (unknown) orientations. As a consequence, we model the observation process via uncertain variables and consider the following form for the likelihood function:

$$\ell_k(z | x) = \exp \left(-\frac{1}{2} (z - h_k(x))^T R_k^{-1} (z - h_k(x)) \right) \doteq \bar{N}(z; h_k(x), R_k),$$

for any $z \in \mathbf{Z}$, where R_k is a $d_z \times d_z$ symmetric positive-definite matrix related to the size and shape of the resolution cells (assumed constant in \mathbf{Z}). The difference between this normal possibility function and the corresponding normal probability distribution would not matter in a standard single-object tracking scenario since normalising constants would simplify in Bayes' theorem; however, in multi-object tracking, these constants are important since they appear in the assessment of data associations. The credibility for an object with state $x \in \mathbf{X}$ to be detected is denoted $\alpha_d(x)$ and, similarly, the credibility of a detection failure is denoted $\alpha_{nd}(x)$.

Since it must hold that $\max\{\alpha_d(x), \alpha_{nd}(x)\} = 1$ for any $x \in X$, we will assume that $\alpha_d = 1$ and $\alpha_{nd}(x) \ll 1$ so that it is unlikely for an object to remain undetected. Given a trajectory $x_{m:n}$ of an object appearing at time step m and disappearing after time step n , we assume that the uncertain variables $\{z_k\}_{k=m}^n$ modelling the observation in $Z \cup \{\phi\}$ are conditionally independent given $x_{m:n} = x_{m:n}$, it follows that the likelihood function for a path $o \in \mathcal{O}_K$ is

$$\ell(o | x_{m:n}, m, n) = \prod_{k=m}^n \alpha_{nd}(x_k)^{\mathbb{1}(o_k=\phi)} \ell_k(o_k | x_k)^{\mathbb{1}(o_k \neq \phi)}.$$

Overall, we have the analogue of a HMM for each object where the sequence of uncertain variables $\{x_k\}_{k=m}^n$ has the Markov property and where observations are conditionally independent given $\{x_k\}_{k=m}^n$.

The credibility for an observation $z \in Z$ at time $k \in \{1, \dots, K\}$ to be a false alarm is denoted $\alpha_{k,fa}(z)$, which will be assumed to be strictly lesser than 1; indeed, if there is no information on false alarms, then the data association with highest credibility will be the one defining all observations as false alarms. The credibility for a given finite subset Z of observations in Z to be false alarms is then

$$f_{k,fa}(Z) = \prod_{z \in Z} \alpha_{k,fa}(z).$$

As a possibility function on sets, $f_{k,fa}$ must verify that $\sup_{Z \subseteq Z} f_{k,fa}(Z) = 1$.

2.3 Target possibility function

We now introduce the posterior possibility function on the set \mathcal{T} describing the unknown set of tracks, based on the model detailed in Sect. 2.2. For this purpose, we consider a track $t = (o, m, n)$ and start by defining the credibility $\pi(o, n | m)$ of the pair (o, n) given the time of appearance $m \in \{1, \dots, K\}$ as

$$\pi(o, n | m) = \sup_{x_{m:n} \in X^{n-m+1}} \ell(o | x_{m:n}, m, n) g(x_{m:n}, n | m).$$

Other aspects such as false alarms and initial observations must be considered jointly. We denote by f_{fa} the function defined on \mathcal{A} as

$$f_{fa}(A) = \prod_{k=1}^K f_{k,fa}(Z_{k,fa}(A)), \quad A \in \mathcal{A},$$

where $Z_{k,fa}(A) = \{z \in Z_k : \forall o \in A, z \neq o_k\}$ is the set of false alarms induced by A at time step k . We also introduce

f_+ as the function on \mathcal{T} defined as

$$f_+(T) = \prod_{k=1}^K f_{k,+}(M_k(T))$$

where $M_k(T) = \sharp\{(o, m, n) \in T : m = k\}$ is the number of objects appearing at time step $k \in \{1, \dots, K\}$. The functions f_{fa} and f_+ , defined, respectively, on \mathcal{A} and \mathcal{T} , are not possibility functions; instead, they are simply the joint credibility for observations that are not in a given element of \mathcal{A} to be false alarms and for tracks that are in a given element of \mathcal{T} to have appeared at the indicated time steps. The target possibility function, i.e. the posterior possibility function Π on \mathcal{T} describing the unknown set of tracks, is then expressed for any $T \in \mathcal{T}$ as

$$\Pi(T) \propto f_{fa}(\kappa(T)) f_+(T) \prod_{(o,m,n) \in T} \pi(o, n | m), \quad (2)$$

and is such that $\max_{T \in \mathcal{T}} \Pi(T) = 1$. The marginal likelihood for the set of paths $A = \kappa(T)$ is then defined as

$$\hat{\Pi}(A) = \max\{\Pi(T) : T \in \mathcal{T}, \kappa(T) = A\}. \quad (3)$$

3 MCMC for data association

3.1 Computational aspects of possibility theory

Approximation methods for possibility functions must be devised in order to solve the corresponding inference problems in general. Grid-based methods have the same shortcomings as in the probabilistic case since it is often difficult to anticipate where the posterior possibility function will take non-negligible values. We propose to use MCMC with a suitably designed proposal distribution in order to explore credible data associations and attempt to find the global maximum and/or regions of high credibility. In this case, there is no requirement of targeting a given probability distribution and there is no concern regarding the independence between samples. One of the consequences is that low-discrepancy sequences can be used instead of pseudorandom numbers. The generated chain, say $\{X_n\}_{n \geq 1}$, will simply be used to approximate the expected value

$$\bar{\mathbb{E}}(\varphi(\mathbf{x})) \approx \max_{n \geq 1} \varphi(X_n) f_{\mathbf{x}}(X_n),$$

for any real-valued function φ on X . As opposed to the standard Monte Carlo approximation, the possibility function $f_{\mathbf{x}}$ appears explicitly in the expression of $\bar{\mathbb{E}}(\varphi(\mathbf{x}))$ since the density of samples in a given area conveys no information about $f_{\mathbf{x}}$; instead, the chain $\{X_n\}_{n \geq 1}$ simply provides

support points for the approximation of f_x as a function. Although the empirical distribution of the samples generated by the MCMC chain is not relevant in this context, the use of MCMC remains beneficial since the underlying mechanisms yield an efficient exploration of the set of data associations which would otherwise be challenging. This is an example of synergy between probability theory and possibility theory where the former helps computing quantities appearing in the latter by leveraging random exploration. Once the most likely data association A^* has been found, the credibility of any other data association $A \in \mathcal{A}$ can be computed exactly; this would require summing over all data associations in the probabilistic case, which is not usually feasible. If only a local maximum $A' \neq A^*$ is found, the credibility of other data associations will be over-estimated, which is conservative according to the interpretation associated with (1) and therefore acceptable in practice.

If only the expected value $\mathbb{E}^*(x)$ of x is of interest, then the possibility function f_x^γ for some $\gamma > 1$ can be used instead. The considered power can also be increased during the execution of the MCMC, leading to a simulated annealing. Conversely, if one is interested in identifying the subset of X containing at least $100(1 - \alpha)\%$ of the subjective probability mass defined in (1), then areas where f_x has value α must also be explored, hence justifying the use of a power γ strictly lesser than 1.

When using the possibility function f_x^γ in a MCMC algorithm, it is the probability distribution on X defined as the renormalised version of f_x^γ that is targeted (assuming f_x^γ is integrable). This is not, however, the only possible approach. Indeed, (1) suggests that a possibility function can be seen as inducing an upper bound for probability distributions. It follows that selecting the sampling distribution from the set of upper-bounded probability distributions is also meaningful. A particular choice that is appropriate in many settings is to follow the maximum-entropy principle (Jaynes 1957) and consider the maximum-entropy distribution that is upper bounded by f_x as in (1), as proposed in Houssineau and Ristic (2017). When X is discrete, it is possible to further increase the entropy by replacing the set-wise upper bound of (1) by a point-wise upper bound of the form $p(x) \leq f_x(x)$, $x \in X$, with p a probability mass function on X . This approach will be particularly useful in the context of multi-object inference since it will lead to an increase of the diversity of explored data associations when compared to sampling from the distribution proportional to f_x .

Overall, our approach provides a trade-off between the standard probabilistic modelling where all aspects of the problem must be characterised and model-free methods, see, e.g. (Sgouralis et al. 2017), which rely on minimalistic assumptions about the underlying dynamics. This trade-off can be beneficial in situations where one wants to leverage the

available information without describing phenomena such as false alarms which are often challenging to characterise.

3.2 Problem formulation

The objective in the remainder of this section is to design a proposal distribution for identifying the mode of the possibility function $\hat{\Pi}$ defined in (3) via the Metropolis–Hastings algorithm. We assume for the moment that this proposal distribution is given and express it as a Markov kernel Φ from \mathcal{A} to itself. A natural starting point for exploring the set \mathcal{A} is to consider the case where all observations are false alarms, that is, we start from the element $A = \emptyset \in \mathcal{A}$. We first assume that $\hat{\Pi}$ can be evaluated everywhere so that, given a previous sample A , a new sample A' can be obtained from the probability distribution $\Phi(\cdot | A)$ and accepted with probability

$$\hat{\alpha}_t(A, A') = \min \left(1, \frac{\hat{\Pi}(A')^{\rho_t} \Phi(A | A')}{\hat{\Pi}(A)^{\rho_t} \Phi(A' | A)} \right), \quad (4)$$

where t is the current iteration and ρ_t is the inverse temperature defined by $\rho_0 = 1$ and $\rho_t = \rho_{t-1}/(1 - c)$ for some constant c .

The main difficulty with the Metropolis–Hastings algorithm in the context of interest is to design a proposal distribution Φ with adequate properties. In particular, there are two issues with this approach which we will aim to solve in the remainder of this section:

- i) The possibility function $\hat{\Pi}$ on \mathcal{A} is highly multimodal in general so that moves that are local both in space and time are unlikely to yield a sufficient exploration of the space.
- ii) Implementing moves on entire paths in the set \mathcal{O}_K would be more global in nature; however, this requires the non-trivial introduction of additional structure on this set.

These two issues will be addressed in Sects. 3.3 and 3.4, respectively. Section 3.5 will then detail the construction of the proposal distribution Φ . Extensions of the MCMC algorithm introduced for $\hat{\Pi}$ to the possibility function Π on \mathcal{T} will be covered in Sect. 4.

3.3 Approximate multi-object filtering

As is usual with MCMC algorithms, the objective is to design a proposal distribution that explores the set \mathcal{A} as quickly as possible while maintaining a reasonable probability of acceptance. To this effect, we propose to use a multi-object filter to ensure that any proposed association is meaningful from the viewpoint of the model. The motivation for leveraging the capabilities of an approximate filtering algorithm to

solve the corresponding smoothing problem is very similar to the one behind particle MCMC (Andrieu et al. 2010). The corresponding moves that we will construct will be global in the sense that they might affect all time steps but local in the sense that only a restricted number of paths will be (re)assigned. The considered filtering algorithm should have a low complexity in order to limit the computational cost of the overall MCMC algorithm. A possible candidate could therefore be the probability hypothesis density (PHD) filter (Mahler 2003) or its analogue in the context of possibility theory (Houssineau 2018a). However, the PHD filter does not solve the data association problem explicitly and, as a consequence, cannot be formally used to propose paths. Instead, we consider an analogue of the hypothesised filter for stochastic populations (Houssineau and Clark 2018), or HISP filter, which is of the same complexity as the PHD filter and which allows for distinguishing objects.

At time step $k \in \{1, \dots, K\}$, the HISP filter extends a set of paths $O_{k-1} \subseteq \mathcal{O}_{k-1} = \bar{Z}_1 \times \dots \times \bar{Z}_{k-1} \setminus \{\phi\}^{k-1}$ with the new observations in \bar{Z}_k and computes the marginal probability of each association between paths in O_{k-1} and observations in \bar{Z}_k . The standard version of the algorithm would consider all data associations with non-negligible marginal probability and proceed to the next time step. Instead, we consider a modified version of the algorithm where a single observation in \bar{Z}_{k-1} is selected at random for each path in O_{k-1} ; this allows for keeping constant the number of considered paths through all time steps. We also use the modelling based on possibility functions introduced in the previous sections instead of the probabilistic modelling considered in Houssineau and Clark (2018). The different steps of this modified HISP filter are given in the following sections.

The context is as follows: denoting A the current state of the MCMC chain, we aim to reassign a subset A_r of A by using the HISP filter. This implies that the other paths in $A \setminus A_r$ are not to be modified. However, since a given observation can only appear in a single path, it follows that the observations contained in the paths in $A \setminus A_r$ cannot be used when reassigning paths in A_r . As a consequence, only a subset of all observations can be used within the HISP filter and this subset is denoted Z_k^- at time step $k \in \{1, \dots, K\}$. We will assume in this section that the credibility α_{nd} of detection failure and the credibility α_{ns} of non-survival are constant over the state space for the sake of simplicity; as opposed to the probabilistic case, this can be achieved in general by selecting the (constant) credibility of detection failure to be $\sup_{x \in X} \alpha_{nd}(x)$ and similarly for the credibility of non-survival. This operation can be seen as a voluntary loss of information with the purpose of gaining a property of interest: we forgo specific spatial information regarding detection and survival in order to make α_{nd} and α_{ns} constant and allow a Gaussian implementation to be used.

3.3.1 Initialisation

We assume that, when reassigning the set A_r of paths, the MCMC algorithm specifies when and at which observations the new paths should start. The corresponding set is denoted $Z_c = \{(z_c^i, k_c^i)\}_{i=1}^{N_c}$ with, for any $i \in \{1, \dots, N_c\}$, z_c^i the observation where one of the new paths should start and with k_c^i the corresponding time step. These observations might or might not be at the same time step but the pairs (z_c^i, k_c^i) , $i \in \{1, \dots, N_c\}$, are assumed to be different from each other. A path will be initialised every time one of these observations is encountered in the sets of observation Z_0^-, \dots, Z_K^- .

3.3.2 Prediction

We denote by O_{k-1} the set of paths at time $k-1$, that is, the subset of \mathcal{O}_{k-1} composed of paths that have been selected so far as potential sequences of object-originated observations. To each path $o \in O_{k-1}$ corresponds a possibility function $f_{k-1}(\cdot | o)$ on the state space X . Recalling that g_k is the Markov transition from X to itself describing the objects' dynamics, we obtain the predicted possibility function

$$f_{k|k-1}(x | o) = \sup_{x' \in X} g_k(x | x') f_{k-1}(x' | o), \quad x \in X.$$

Such a prediction only considers the event where the object survives to the k th time step although it is possible for objects to disappear. We postpone considerations of this aspect of the prediction to a further stage in the algorithm.

3.3.3 Update

At time step k , the set of observations Z_k^- is available to update the existing paths. For any path o in the set O_{k-1} of previously selected paths and for any new observation $z \in Z_k^- \cup \{\phi\}$, the posterior possibility function associated with the extended path $o : z$, with “:” denoting concatenation, is defined as

$$f_k(x | o : z) = \begin{cases} \frac{\ell_k(z | x) f_{k|k-1}(x | o)}{\sup_{x' \in X} \ell_k(z | x') f_{k|k-1}(x' | o)} & \text{if } z \in Z_k^- \\ f_{k|k-1}(x | o) & \text{if } z = \phi. \end{cases}$$

We can then select which observation in $Z_k^- \cup \{\phi\}$ will be used to propagate the path based on the credibility of the corresponding association. However, before expressing the latter, we first have to introduce the prior credibility of presence, which depends on the consecutive number of time steps for which the path under consideration has not been detected. Indeed, there is some remaining ambiguity whenever the empty observation ϕ is selected for a path since it is unclear in this case whether the detection has failed for the correspond-

ing object or the object has not survived the last prediction step. We purposefully maintain this ambiguity and postpone the decision in order to better estimate which of these two events occur. Indeed, the credibility of non-survival is most often much lower than the credibility of detection failure, e.g. $\alpha_{nd} = 0.1$ and $\alpha_{ns} = 0.001$, so that terminating a track after a single detection failure is unlikely. Yet, if detection failures keep occurring for l time steps, then the credibility of the corresponding events, i.e. α_{nd}^l for the case where the object remains and α_{ns} for the case where the object has disappeared, will rapidly favour a disappearance as opposed to a sequence of detection failures. For any path $o \in O_{k-1}$, we denote by l_o the number of consecutive time steps for which ϕ has been selected, e.g. if o is of the form $(o_1, \dots, o_{k-3}, \phi, \phi)$ with $o_{k-3} \neq \phi$ then $l_o = 2$. We then compute the credibility that the corresponding object has survived/not survived since the last detection as

$$\hat{\alpha}_s(o) = \frac{\alpha_{nd}^{l_o}}{\alpha_{ns} \vee \alpha_{nd}^{l_o}}, \quad \hat{\alpha}_{ns}(o) = \frac{\alpha_{ns}}{\alpha_{ns} \vee \alpha_{nd}^{l_o}},$$

with $a \vee b \doteq \max\{a, b\}$ for any $a, b \in \mathbb{R}$. The binary operator \vee is assumed to have lower precedence than multiplication, i.e. $a \vee bc = a \vee (bc)$ for any $a, b, c \in \mathbb{R}$.

We can now express the marginal credibility of association on $Z_k^- \cup \{\phi\}$ for the path $o \in O_{k-1}$ as

$$\gamma_k(z|o) \propto \Gamma_k(Z_k^- \setminus \{z\} | O_{k-1} \setminus o) L_k(z|o)$$

for any observation $z \in Z_k^- \cup \{\phi\}$, with

$$L_k(z|o) = \begin{cases} \hat{\alpha}_s(o) \sup_{x \in X} \ell_k(z|x) f_{k|k-1}(x|o) & \text{if } z \in Z_k^- \\ \hat{\alpha}_{ns}(o) \vee \hat{\alpha}_s(o) (\alpha_{ns} \vee \alpha_{nd}) & \text{otherwise} \end{cases}$$

the marginal likelihood for the observation z and with $\Gamma_k(Z|O)$ the credibility for paths in the set $O \subseteq O_{k-1}$ to be associated with observations in the set $Z \subseteq Z_k^-$, which can be expressed as

$$\Gamma_k(Z|O) = \max_{\sigma: O \rightarrow Z \cup \{\phi\}} f_{fa}(Z \setminus \sigma(O)) \prod_{o \in O} L_k(\sigma(o)|o),$$

where the maximum is over all mappings σ from O to $Z \cup \{\phi\}$ that are injective on Z and where $\sigma(O)$ is the image of O by σ , i.e. $\sigma(O) = \{\sigma(o) : o \in O\}$. Although the number of simultaneously reassigned paths will be limited in the context of interest, the number of observations in Z can be extremely large so that the computation of $\Gamma_k(Z|O)$ can be challenging. Yet, it is possible to rewrite this term by assuming that any two paths in O are unlikely to obtain large marginal likelihoods from a single observation in Z , that is, for any $o, o' \in O$ such that $o \neq o'$ and any $z \in Z$, there

exists $z' \in Z$ such that

$$L_k(z|o) L_k(z|o') \leq L_k(z|o) L_k(z'|o'). \quad (5)$$

In the probabilistic version of this assumption (Houssineau and Clark 2018), the left-hand side needs to be equal to 0, which is more constraining. It follows that $\Gamma_k(Z|O)$ can be expressed as

$$\Gamma_k(Z|O) = f_{fa}(Z) \prod_{o \in O} \left[L_k(\phi|o) \vee \max_{z \in Z} \frac{L_k(z|o)}{\alpha_{fa}(z)} \right].$$

This result can be proved easily by developing the product in the approximated expression and removing the terms where a single observation is associated with several tracks. Using this expression, all the terms $\Gamma_k(Z_k^- \setminus \{z\} | O_{k-1} \setminus o)$, for any $z \in Z_k^- \cup \{\phi\}$ and any $o \in O_{k-1}$, can be calculated with a computational complexity of order $|O_{k-1}| |Z_k^-|$. The approach is similar to the one detailed in Houssineau and Clark (2018) for the probabilistic case.

We then randomly select an observation in $Z_k \cup \{\phi\}$ for each of the paths in O_{k-1} by sampling from the maximum entropy distribution induced by the marginal credibility of association $\gamma_k(\cdot|o)$. There are two ways of enforcing the modelling assumption that paths cannot contain the same observation:

- i) Use a rejection sampling strategy at the level of the HISP algorithm to ensure that only acceptable data associations are proposed and
- ii) Completely reject the proposed data association at the level of the MCMC algorithm if it contains overlapping paths.

The main drawback with the first option is that calculating the probability of proposing a given acceptable data association is combinatorial in nature and becomes a computational bottleneck when the number of observations is large. We therefore consider the second option.

Finally, we initialise a new path for any (z_c^i, k_c^i) , $i \in \{1, \dots, N_c\}$, such that $k_c^i = k$. This path is of the form $o = (\phi, \dots, \phi, z_c^i)$.

At the last time step, a decision is taken for all observations paths, even the one ending with empty observations, and a set A_c is defined as the set of all created paths. The conditional probability for generating the set of paths A_c given the initial observations Z_c and the available observations Z_1^-, \dots, Z_K^- is denoted $P_c(A_c | Z_c, Z_{1:K}^-)$.

3.4 Structure on the set of paths

In order to help exploring the set of data associations \mathcal{A} , it is useful to equip the underlying set of paths \mathcal{O}_K with

additional structure; in particular, the objective is to identify which pairs of points in the observation space Z are likely to be consecutive observations of the same object. The only natural structure on \mathcal{O}_K is the one inherited from the fact that the set Z is a subset of an Euclidean space. This is not, however, sufficient since simply measuring the distance between two observations z_k and $z'_{k'}$ at two different time steps $k \neq k'$ as $\|z_k - z'_{k'}\|$, with $\|\cdot\|$ the Euclidean norm, does not take into account the structure of the problem. Moreover, the notion of distance is very model dependent and what is considered as “close” or “far” would need to be adjusted for each scenario. For instance, depending on how large the speed of objects is likely to be, observations at two consecutive time steps that are 100 meters apart might be very likely or very unlikely to originate from the same object; similarly, when the credibility of detection failure increases, the distance between observations should also decrease as objects are more likely to be undetected for several time steps and therefore move further before being detected again. One partial solution would be to consider the Mahalanobis distance which can be seen as a rescaling of the Euclidean distance based on a covariance matrix; we aim to go further and use all the aspects of the objects’ dynamical and observation model as a reference and relate observations via the credibility for these observations to be generated by the same object. These observations can be seen as *consistent* if that credibility is close to 1 and *inconsistent* if it is close to 0. In order to simplify the calculations, we assume the existence of an upper bounding function g for the Markov transition g_k such that $g_k(x|x') \leq g(x|x')$, for any $x, x' \in X$ and for any $k \in \{1, \dots, K\}$, with g of the form

$$g(x|x') = \bar{N}(x; Fx', Q), \quad x, x' \in X,$$

for some $d_X \times d_X$ matrices F and Q . We also introduce a lower bound a_{nd} for the credibility of non-detection, i.e. a_{nd} is such that $\alpha_{\text{nd}}(x) \geq a_{\text{nd}}$ for any $x \in X$.

We consider two time steps $k, k' \in \{1, \dots, K\}$ such that $k < k'$ as well as two observations z and z' at time steps k and k' , respectively, and introduce $f_{k'|k}(z'|z)$ as the possibility for an object initialised from z at time step k to be next observed at time step k' at z' , that is,

$$f_{k'|k}(z'|z) = a_{\text{nd}}^{l-1} \sup_{x, x' \in X} \ell_k(z'|x') g^l(x'|x) f_k(x|z)$$

where $l = k' - k$, where g^l is the l fold convolution of the transition g , that is, for any $x_k, x_{k'} \in X$,

$$g^l(x_{k'}|x_k) = \sup_{x_{k+1}, \dots, x_{k'-1} \in X} g(x_{k'}|x_{k'-1}) \dots g(x_{k+1}|x_k)$$

and where $f_k(\cdot|z)$ is the posterior possibility function defined as

$$f_k(x|z) = \frac{\ell_k(z|x) f_0(x)}{\sup_{x' \in X} \ell_k(z|x') f_0(x')}, \quad x \in X.$$

The possibility function $g^l(\cdot|x_k)$ is an upper bound for the convolution of the Markov transitions $g_{k+1}, \dots, g_{k'}$. Assuming that $f_k(\cdot|z) = \bar{N}(m_z, \Sigma_0)$ and denoting by Σ_l the covariance matrix after l predictions, e.g. $\Sigma_1 = F \Sigma_0 F^\top + Q$, then the possibility function $f_{k'|k}(\cdot|z)$ can be written

$$f_{k'|k}(z'|z) = a_{\text{nd}}^{l-1} \bar{N}(z'; H_{k',z,l} F^l m_z, H_{k',z,l} \Sigma_l H_{k',z,l}^\top + R_{k'})$$

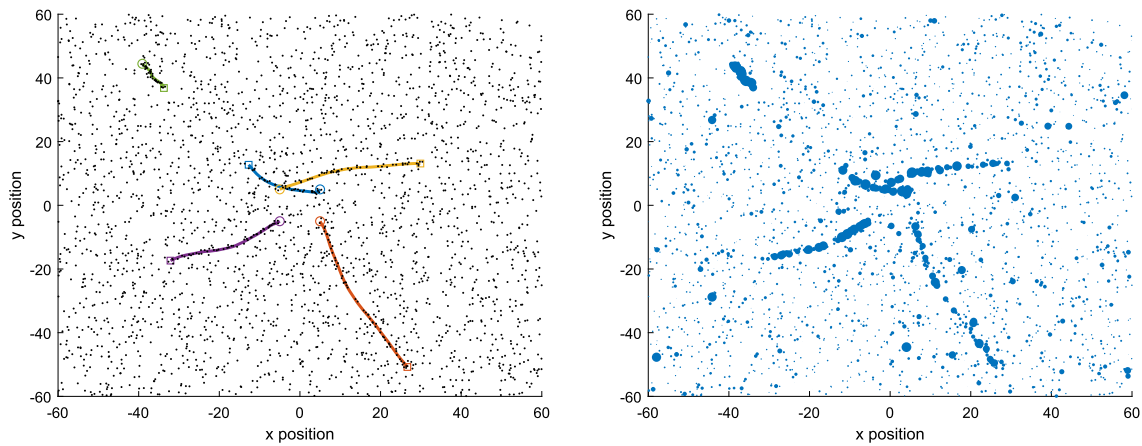
where $H_{k',z,l}$ is the Jacobian of $h_{k'}$ at the point $F^l m_z$. The function $f_{k'|k}$ can be easily extended to $Z \cup \{\phi\}$ by defining $f_{k'|k}(\cdot|\phi) = f_{k'|k}(\phi|\cdot) = 0$.

Example 1 To illustrate the use of the notion of consistency provided by $f_{k'|k}$, a simple scenario consisting of five objects is considered as in Fig. 1a. For each observation $z \in Z_k$ at some time step $k \in \{1, \dots, K\}$, we compute a marginal credibility for z as

$$\hat{f}_k(z) = \max_{k': k' > k} \left(\max_{z' \in Z_{k'}} f_{k'|k}(z'|z) \right). \quad (6)$$

The scalar $\hat{f}_k(z)$ can be interpreted as the credibility for z to be followed by another observation in $Z_{k'}$ for some $k' > k$. When creating a new track, we can then define the probability for selecting z as the first observation of the new track as a function of $\hat{f}_k(z)$. A scatter plot displaying these credibilities for all observations is shown in Fig. 1b. As mentioned in the caption of Fig. 1, the superposition of observations at all times means that the spatio-temporal aspect of the data is not represented; in this sense, Fig. 1b better captures the complexity of the problem by showing how “close” are observations originating from a given target and how “far” are false alarms, which are two important factors in the assessment of the complexity of a given scenario. The probabilistic analogue of the scalar $\hat{f}_k(z)$ would penalise detection failures more strongly since the fitness of the model would decrease quickly with the lag $k' - k$, especially in the presence of a large dynamical noise; this means that our approach leads to a higher probability of proposing the initialisation a new path for objects with challenging detection profiles, e.g. when there are gaps between most observations.

The advantage of relating observations in this way is that it can be easily extended to paths. Indeed, we can define the consistency between an observation $z \in Z_k$ at some given time k with a path o in \mathcal{O}_K as



(a) True trajectories as coloured lines, initial/final positions as circles/squares and observations as black dots.

(b) Scatter plot of the observations with a size proportional to the corresponding credibilities of selection.

Fig. 1 Scenario with 5 objects as represented in **a** with the corresponding scatter plot of the probability for each observation to be selected in **b**. Observations at all time steps are superimposed in these figures, which do not represent the available spatio-temporal that is crucial in

solving the corresponding data association problem. Although the trajectories of two objects cross, the objects do not reach the crossing point at the same time

$$\hat{f}_k(z, o) = \max \left\{ \max_{k': k' < k} f_{k|k'}(z | o_{k'}), \max_{k': k' > k} f_{k'|k}(o_{k'} | z) \right\}, \quad (7)$$

where the initial observation is either z or one of the observations in o . Similarly, the consistency between two paths o and o' in \mathcal{O}_K is defined as

$$\hat{f}(o, o') = \max \left\{ \max_{k, k': k' < k} f_{k|k'}(o_k | o'_{k'}), \max_{k, k': k' > k} f_{k'|k}(o'_{k'} | o_k) \right\}.$$

We can now propose to modify a given data association by changing nearby paths, and therefore focus the computational power on moves that are likely to be accepted. Although the approach considered here is not standard, it has two main advantages: it relates observations together and applies to nonlinear cases as long as a Gaussian upper-bounding function can be found.

In practice, it might be necessary to reduce the time required for computing $f_{k'|k}(z' | z)$ between any pair (z, z') of observations, especially if the scenario runs over many times steps or if the number of observations at every time step is large. In that case, one can define a threshold τ' such that if $a_{nd}^l < \tau'$ then any observations that are l time steps apart will be arbitrarily assigned a credibility of 0.

3.5 Design of the proposal distribution

When designing a proposal distribution Φ for our MCMC algorithm, several requirements need to be considered: it should be possible to

- i) Reassign several paths simultaneously in order to address crossings and track fragmentation,

- ii) Reassign both the initial observation of a path and subsequent observations and
- iii) Create a new path.

Requirement i) can be easily fulfilled by using the approach presented in Sect. 3.3; however, instead of simply choosing the paths at random, it is more efficient to focus on nearby paths. In order to simultaneously reassign the initial observations of a given set of paths A_r as needed in Requirement ii), we consider the notion of consistency defined in (7). Once a new initial observation has been selected, the approach of Sect. 3.3 can be used to reassign the rest of the chosen path. Finally, the marginal consistency defined in (6) can be used for Requirement iii) in order to identify observations that are likely to be initial observations.

The general objective is to find a proposal distribution Φ that is as simple as possible and such that the associated MCMC kernel is irreducible and reversible. Starting from a given set of paths A of size $s = |A|$, we suggest to proceed as follows:

- 1) Sample a number N_r of paths to reassign from a probability mass function (p.m.f.) $p_r(\cdot | s)$ such that $N_r \leq s$ almost surely (a.s.), e.g. a truncated Poisson distribution. Then, sample the number N_c of paths to be created from the p.m.f. $p_c(\cdot | N_r)$ on the set of non-negative integers \mathbb{N} defined as

$$p_c(n | N_r) \begin{cases} \delta_1(n) & \text{if } N_r = 0 \\ \tilde{p}_c(n - N_r) & \text{otherwise,} \end{cases}$$

with \tilde{p}_c a p.m.f. on $\{-1, 0, 1\}$ to be defined. With this model, there will be one created path a.s. when none are reassigned (there is limited interest in creating several paths at once in this case) and the number of paths will be increased by one, kept constant or decreased by one in case of reassignment. Reducing the number of paths by one will address the issue of track fragmentation, keeping the number of paths constant is appropriate when considering objects with crossing trajectories, and leaving the possibility of increasing the number of paths is required to ensure reversibility. Indeed, when evaluating the probability of the reverse proposal, created paths will become reassigned paths and vice versa.

- 2) If $N_r = 0$, then define $A_r = \emptyset$ and proceed to the next step; otherwise, select the set $A_r = \{o_i\}_{i=1}^{N_r}$ of paths to be reassigned as follows: the first path o_1 is picked uniformly at random from the set of paths A and then the $N_r - 1$ remaining paths, if any, are selected based on their distance to o_1 :

$$o_i \sim \mathcal{P}_{o_{1:i-1}}(\hat{f}(o_1, \cdot))$$

for any $1 < i \leq N_r$, where $\mathcal{P}_{o_{1:i-1}}(\cdot)$ is a function transforming possibility functions into probability distributions, e.g. the maximum-entropy distribution upper-bounded point-wise by $\hat{f}(o_1, \cdot)$, which we assume to verify

$$\sum_{o \in A} \mathcal{P}_{o_{1:i-1}}(\hat{f}(o_1, \cdot))(o) = 1$$

$$\tilde{p}_c(Z_c | N_c, Z_{1:K}^-) = \begin{cases} \sum_{\sigma \in S_{N_c}} \prod_{i=1}^{N_c} \mathcal{P}_{(\hat{z}_c^{\sigma(1)}, \dots, \hat{z}_c^{\sigma(i-1)})}(\hat{f}_c^{N_r})(\hat{z}_c^{\sigma(i)}) & \text{if } N_c > 0 \\ \delta_{\emptyset}(Z_c) & \text{otherwise.} \end{cases}$$

and $\mathcal{P}_{o_{1:i-1}}(\hat{f}(o_1, \cdot))(o_j) = 0$ for all $j \in \{1, \dots, i-1\}$. Therefore, the set of paths A_r is sampled without replacement from the set A . When evaluating the probability $P_r(A_r | N_r)$ for sampling the subset A_r of A , all possible ways of obtaining such a subset must be taken into account, that is,

$$P_r(A_r | N_r, A) = \begin{cases} s^{-1} \sum_{\sigma \in S_{N_r}} \prod_{i=2}^{N_r} \mathcal{P}_{o_{\sigma(1):\sigma(i-1)}}(\hat{f}(o_1, \cdot))(o_{\sigma(i)}) & \text{if } N_r > 0 \\ \delta_{\emptyset}(A_r) & \text{otherwise,} \end{cases}$$

where S_n is the set of permutations of $\{1, \dots, n\}$. Although the computational complexity for this term is combinatorial, N_r is usually small so the actual computational time is limited.

- 3) If $N_c = 0$ then define $Z_c = \emptyset$ and proceed to the next step; otherwise, select the N_c initial observations $Z_c = \{(z_c^i, k_c^i)\}_{i=1}^{N_c}$ from the set $\bigcup_{k=1}^K \{(z, k) : z \in Z_k^-\}$ of available observations, with Z_k^- defined for any $k \in \{1, \dots, K\}$ as

$$Z_k^- = \{z \in Z_k : \forall (o, k_-) \in A \setminus A_r, z \neq o_k\}.$$

The selection of the initial observations is performed without replacement as

$$\hat{z}_c^i \sim \mathcal{P}_{(\hat{z}_c^1, \dots, \hat{z}_c^{i-1})}(\hat{f}_c^{N_r})$$

where, for any $j \in \{1, \dots, N_c\}$, \hat{z}_c^j stands for the pair (z_c^j, k_c^j) , and where the possibility function $\hat{f}_c^{N_r}$ is equal to the marginal consistency defined in (6) if $N_r = 0$ and as the consistency defined in (7) with the future observation in the paths in A_r otherwise. Indeed, when reassigning $N_r > 0$ paths, it is more efficient to propose new paths in the same area rather than initialising paths in random locations, especially during the burn-in period of the MCMC when observations in different places are likely to originate from objects. The probability of proposing the subset Z_c of observations takes a similar form as for path reassignment and can be expressed as

The comment regarding computational complexity made about $P_r(\cdot | N_r)$ applies equally here.

- 4) Apply the approximate multi-object filter of Sect. 3.3 to the set of initial observations Z_c and with the sets of

available observations Z_1^-, \dots, Z_K^- and denote A_c the generated set of paths. If $A_c \cap A_r \neq \emptyset$, then we reject the proposal; otherwise, the proposed set of paths is $A' = (A \setminus A_r) \cup A_c$. The reason for rejecting the proposal when

$A_c \cap A_r \neq \emptyset$ is to ensure that A_c and A_r can be recovered from A and A' as $A_c = A' \setminus A$ and $A_r = A \setminus A'$.

If the proposal has not been already rejected during its construction, the probability $\Phi(A' | A)$ to go from the previous set of paths A to the new set of paths A' is computed as

$$\Phi(A' | A) = P_c(A_c | Z_c, Z_{1:K}^-) \tilde{P}_c(Z_c | N_c, Z_{1:K}^-) \times P_r(A_r | N_r, A) p_c(N_c | N_r) p_r(N_r | s).$$

The probability $\hat{\alpha}(A, A')$ of accepting the proposed set of paths A' can then be computed using (4).

4 MCMC on the set of tracks

We now want to design a MCMC algorithm that targets the possibility function Π as introduced in (2). In this case, the Metropolis–Hastings acceptance ratio is

$$\alpha_t(T, T') = \min \left(1, \frac{\Pi(T')^{\rho_t} \Psi(T | A) \Phi(A | A')}{\Pi(T)^{\rho_t} \Psi(T' | A') \Phi(A' | A)} \right),$$

with A and A' the set of paths in T and T' , respectively. We therefore have to propose a time of appearance and a last time of existence for each path in A . These time steps will be sampled independently from their previous values in T .

4.1 Proposing the interval of existence

The objective in this section is to propose a time of appearance m and a last time of existence n for a given path $o \in \mathcal{O}_K$, using the different quantities introduced in Sect. 3.3. We consider a path $o \in \mathcal{O}_K$ of the form $o' : \phi$. One can sample the lag corresponding to the last time of appearance according to the probability mass function $p_-(\cdot | o)$ on $\{0, \dots, l_o\}$ defined as the maximum-entropy distribution bounded by $l \mapsto \alpha_{ns}^{\mathbb{1}(l < l_o)} \alpha_{nd}^l$. The last time of existence is set to $n = K - l_o + L_-$. For the time of appearance m associated with a path $o \in \mathcal{O}_K$, we can simply sample a lag L_+ from the maximum-entropy distribution $p_+(\cdot | o)$ bounded by $l \mapsto \alpha_{nd}^l$ and set $n = k_+(o) - L_+$ with $k_+(o)$ the time of the first observation in o . The probability distribution $\Psi(\cdot | A)$ is then associated with the proposal of a time of appearance and a last time of existence for each path in a given set $A \in \mathcal{A}$, i.e.

$$\Psi(T | A) = \delta_A(\kappa(T)) \prod_{(o, m, n) \in T} [p_+(m | o) p_-(n | o)].$$

4.2 Evaluating the marginal likelihood

So far, the proposed approach does not assume a specific model for the dynamics and for the observation process. Indeed, although the likelihood $\ell_k(\cdot | x)$ is assumed to take the form of a Gaussian possibility function, the function h relating states to observations is general. We will, however, distinguish two different cases for the evaluation of the marginal likelihood: the linear-Gaussian case in which Kalman filtering can be used and the nonlinear case where sequential Monte Carlo techniques are a natural alternative.

4.2.1 Linear-Gaussian case

If the Markov transition g_k is of the form $g_k(\cdot | x') = \bar{N}(F_k x', Q_k)$ for some $d_X \times d_X$ matrices F_k and Q_k and for any $k \in \{1, \dots, K\}$ and if the observation function h_k is of the form $h_k(x) = H_k x$, then the posterior distribution of the state at any time step can be computed analytically via the Kalman filter. In particular, for a given path $o \in \mathcal{O}_{k-1}$, we denote by m_k^o and Σ_k^o the mean and variance of the state at time $k \in \{1, \dots, K\}$ given the observations in the path o . The only difference with the standard Kalman filtering equation is the marginal likelihood which, due to the form of the likelihood, is expressed at time step k as

$$\begin{aligned} \hat{\ell}_k(z | o) &= \sup_{x \in X} \ell_k(z | x) f_{k|k-1}(x | o) \\ &= \bar{N}(z; H_k m_k^o, H_k \Sigma_k^o H_k^\top + R_k) \end{aligned}$$

for any $z \in Z_k$.

4.2.2 Nonlinear case

If either the objects' dynamics or the observation function is not linear, then there is no analytical form for the filtering distributions at different time steps in general. Sequential Monte Carlo (SMC) methods are an alternative to the Kalman filter in this case. An analogue of the bootstrap particle filter Gordon et al. (1993) can be used as in Houssineau and Ristic (2017), see also Ristic et al. (2018) and Ristic et al. (2020). In particular, for a given path $o \in \mathcal{O}_{k-1}$, we denote by $\{(w_{k-1,i}^o, x_{k-1,i}^o)\}_{i=1}^N$ the indexed family of weighted particles approximating the predicted possibility function $f_{k|k-1}(\cdot | o)$, i.e.

$$\bar{\mathbb{E}}(\varphi(x_k) | o) \approx \max_{1 \leq i \leq N} w_{k-1,i}^o \varphi(x_{k-1,i}^o)$$

for any real-valued function φ on X , with the uncertain variable x_k being described by $f_{k|k-1}(\cdot | o)$. Then

$$\bar{\mathbb{E}}(\varphi(x_k) | o : z) \approx \frac{\max_i w_{k,i}^{o:z} \varphi(x_{k-1,i}^o)}{\max_i w_{k,i}^{o:z}}$$

for any $z \in Z_k$, where $w_{k,i}^{o:z} = w_{k-1,i}^o \ell_k(z | x_{k-1,i}^o)$ for any $i \in \{1, \dots, N\}$. In this situation, the marginal likelihood at time step k can be approximated by

$$\hat{\ell}_k(z | o) \approx \max_{1 \leq i \leq N} w_{k,i}^{o:z}.$$

5 Simulations

In all the cases to be considered, $K = 50$ and $\mathbf{X} = \mathbb{R}^4$. States at time step k are of the form $x_k = (\mathbf{x}_k, \mathbf{y}_k, \dot{\mathbf{x}}_k, \dot{\mathbf{y}}_k)^\top$, where \mathbf{x}_k and \mathbf{y}_k are the coordinates of the position in the 2-dimensional Euclidean space and where $\dot{\mathbf{x}}_k$ and $\dot{\mathbf{y}}_k$ are the coordinates of the velocity. The duration of one time step is denoted Δ and the motion model is assumed to be of the form

$$q_k(x_k | x_{k-1}) = \mathcal{N}(x_k; Fx_{k-1}, Q)$$

with

$$F = \begin{bmatrix} 1 & 0 & \Delta & 0 \\ 0 & 1 & 0 & \Delta \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

and

$$Q = \sigma_a^2 \begin{bmatrix} \Delta^4/4 & 0 & \Delta^3/2 & 0 \\ 0 & \Delta^4/4 & 0 & \Delta^3/2 \\ \Delta^3/2 & 0 & \Delta^2 & 0 \\ 0 & \Delta^3/2 & 0 & \Delta^2 \end{bmatrix},$$

where σ_a is the standard deviation of the zero-mean random acceleration, which is considered as a noise term. This model is referred to as the nearly constant velocity model. We will consider in particular the case where $\Delta = 1$ and $\sigma_a = 0.05$.

For the sake of simplicity, the observation model is assumed to be linear; the position $(\mathbf{x}_k, \mathbf{y}_k)^\top$ of an object is observed directly, which leads to $h(x_k) = Hx_k$ with

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

The variance R is of the form $\sigma^2 \mathbf{I}_2$ with $\sigma > 0$ and \mathbf{I}_2 the identity matrix of dimension 2. This model is useful when tracking directly in the coordinate systems defined by a sensor such as the image plane of a camera. Other situations where this model arises are when multiple sensors provide complex observations which can be combined into a single observation before being used in a tracking algorithm such as with GPS or with multiple-input multiple-outputs sensor systems (Bekkerman and Tabrikian 2006; Haimovich et al. 2007; Pailhas et al. 2016). We will consider in particular the case where $\sigma = 0.3$ and $\mathbf{Y} = [-60, 60] \times [-60, 60]$.

5.1 Parametrisation of the proposed algorithm

If the probability of detection is p_d then the possibility of detection failure is set to $\alpha_{nd} = 1 - p_d$ and the possibility of detection α_d is set to 1. The same approach is used with the probability of survival. The possibility function $\alpha_{k,fa}$ is assumed to be constant and equal to 10^{-2} for all scenarios; this is in spite of the fact that the number of false alarms will vary significantly across the considered settings. The reason for this is that $\alpha_{k,fa}^n$ is seen as an upper bound for the probability of having n false alarms. A similar approach is used for appearing objects with $f_{k,+}(n) = \alpha_+^n$ with $\alpha_+ = 10^{-4}$. The other model parameters such as σ and σ_a are assumed to be known.

The proposed approach is compared to the MCMC for Data Association (MCMC-DA) method introduced in Oh et al. (2009). In order to make the two methods comparable, the possibility function Π is used to evaluate the log-likelihood of the proposed sets of tracks. However, as opposed to the proposed approach, MCMC-DA is provided with the true parameters of the model in the design of the corresponding proposal distribution.

The main differences between MCMC-DA and our approach lies in the construction of the proposal distribution: MCMC-DA has a number of simple moves, whereas our approach focuses on one sophisticated move. The main consequence is the complexity of a single step of the corresponding MCMC algorithms: each step in MCMC-DA has a constant complexity, whereas each step in our approach has a complexity of order K ; yet, this is compensated by the more efficient exploration of the set of data associations that our approach yields, as demonstrated below in a range of scenarios.

5.2 Choice of parameter

We assume that the current sample from Π is $T \in \mathcal{T}$ and denote by $A = \kappa(T)$ the corresponding set of paths. We then comment on the choice of parameters for the different steps in the proposal mechanism.

The number N_r of tracks to reassign is chosen from a Poisson distribution with parameter $\lambda_r = 1$, truncated to the interval $\{0, \dots, |A|\}$. The parameter λ_r can be adjusted depending on the considered scenario: if objects are expected to be very close to each other and to frequently have crossing trajectories, then λ_r could be increased to raise the average number of tracks that are reassigned at once. Large reassignments are, however, less likely to be accepted so that a trade-off between exploration and mixing must be found, as is usual with MCMC.

The distribution $\tilde{p}_c(\cdot | N_r)$ on $\{-1, 0, 1\}$ drives the increase or decrease of the number of tracks in the proposal step. Since

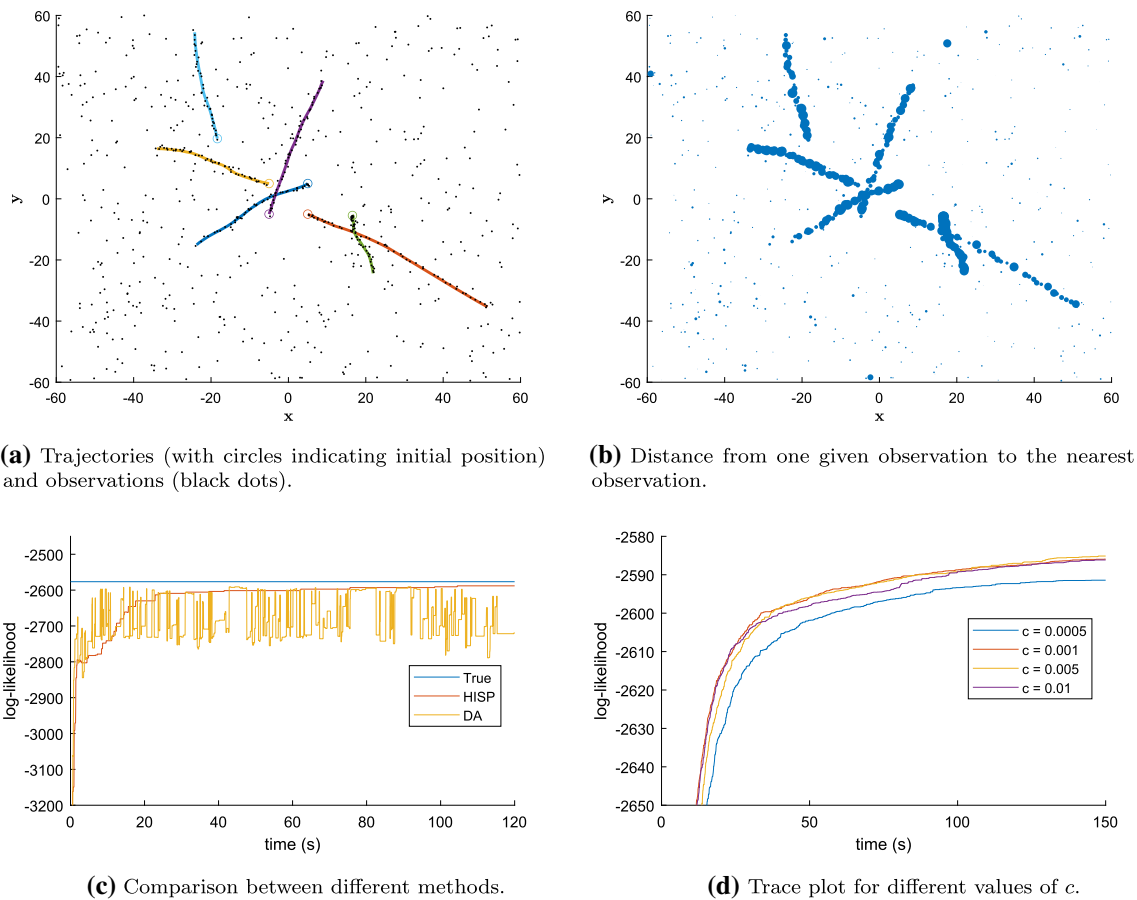


Fig. 2 Simple scenario with performance comparison for different parameter choices

one of the main issues with the MCMC approach for data association is track fragmentation, i.e. the representation of a single object by a series of shorter tracks, it is generally helpful to focus on reducing the number of tracks. We therefore consider the following parametrisation:

$$\tilde{p}_c(\delta | N_r) = \begin{cases} \frac{1}{2} & \text{if } \delta = -1 \\ \frac{1}{4} & \text{if } \delta = 0 \\ \frac{1}{4} & \text{if } \delta = 1. \end{cases}$$

5.3 MCMC on the data association set

The choice of parameter and the performance of the proposed approach are assessed on different scenarios.

5.3.1 Simple scenario

We first consider a simple scenario, as shown in Fig. 2a, with ten false alarms and 0.1 appearing objects per time step on average and with a probability of detection of $p_d = 0.9$. The simplicity of the scenario is illustrated in Fig. 2a where it appears that most of the false alarms are far from any other

observation and, conversely, object-originated observations are close to each other.

The performance of the two considered approaches is first assessed on a single run in Fig. 2c where the evolution of the log-likelihood is displayed as a function of the computational time. “HISP” refers to the proposed approach whereas “DA” refers to the MCMC-DA. The difference in behaviour between the proposed approach and MCMC-DA is due to the use of the simulated annealing in the former. Both methods provide satisfactory results in this case and the MCMC-DA’s chain mixes well. Figure 2d, which displays the performance averaged over 50 repeats, shows that setting the parameter c in the inverse temperature ρ_t to 0.001 provides the best performance throughout the duration of the runs.

5.3.2 Scenario with high false-alarm rate

We consider a first type of challenging scenario, depicted in Fig. 3a, with the following challenging characteristics: there are 100 false alarms and 0.5 appearing objects per time step on average and the probability of detection p_d is equal to 0.8. In this case, it is the large number of false alarms that make the estimation difficult due to the fact that they are

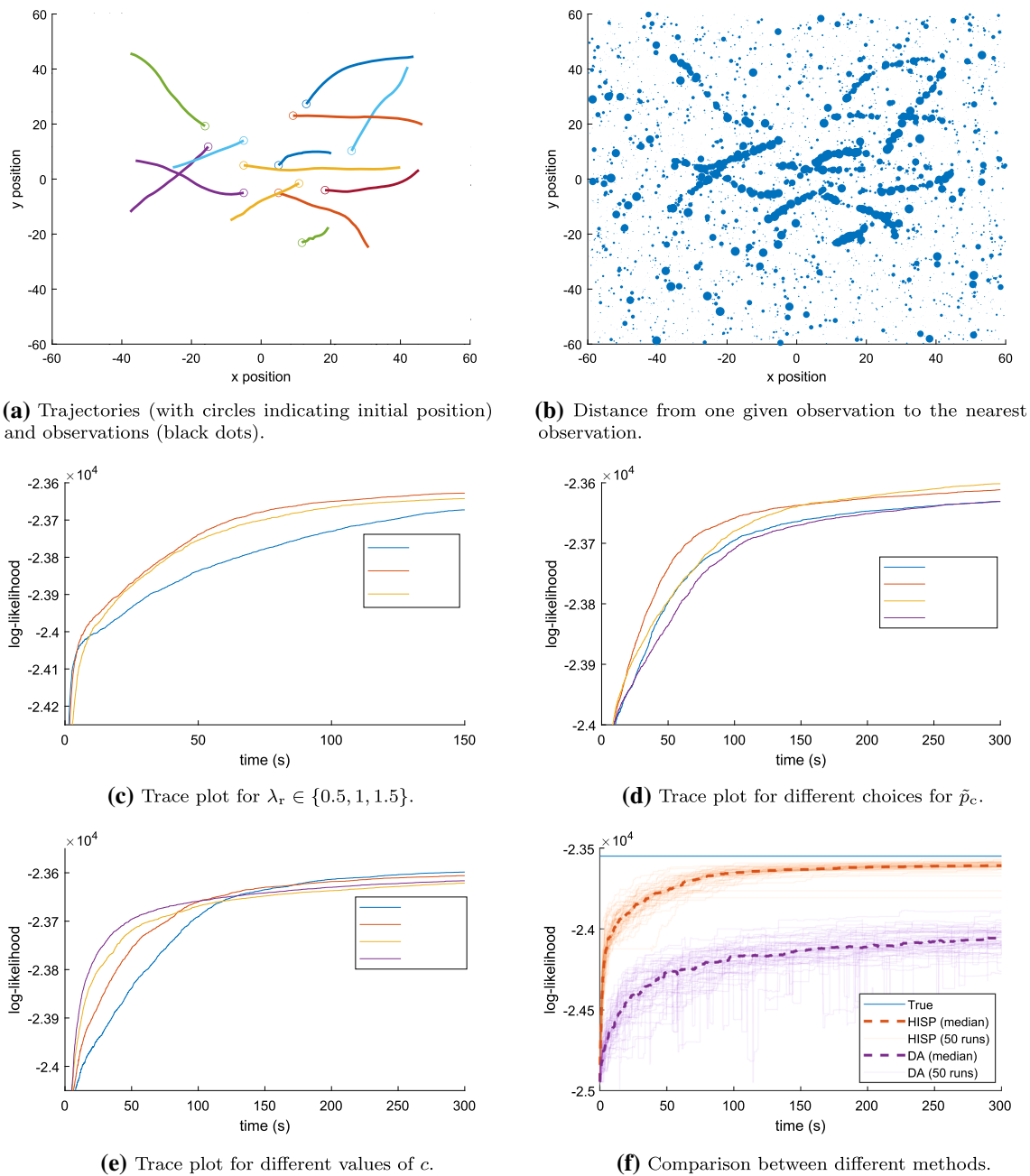


Fig. 3 Scenario with high false-alarm rate

likely to form coherent observation sequences over 2 to 3 time steps. This aspect is illustrated in Fig. 3b where many false alarms can be seen to be near other observations. Figure 3c considers different choices for the Poisson parameter λ_r with the log-likelihood being once again averaged over 50 repeats. The choice $\lambda_r = 1$ allows for rapidly creating tracks while proposing the simultaneous reassignment of 2 tracks often enough to prevent track fragmentation, whereas setting λ_r to 0.5 or 1.5 does not perform as well. Finally, a few options are compared in Fig. 3d for the distribution \tilde{p}_c ,

with the log-likelihood being averaged over 50 repeats. The assessed options are

$$\tilde{p}_c((-1, 0, +1) | N_r) = \begin{cases} (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}) & \text{as "uniform"} \\ (\frac{1}{2}, \frac{1}{4}, \frac{1}{4}) & \text{as "focus on } -1" \\ (\frac{1}{4}, \frac{1}{2}, \frac{1}{4}) & \text{as "focus on } 0" \\ (\frac{1}{4}, \frac{1}{4}, \frac{1}{2}) & \text{as "focus on } +1", \end{cases}$$

where $\tilde{p}_c((\delta_1, \delta_2, \delta_3) | N_r) = (p_1, p_2, p_3)$ is a shorthand notation for $\tilde{p}_c(\delta_i | N_r) = p_i$ for $i \in \{1, 2, 3\}$. The results

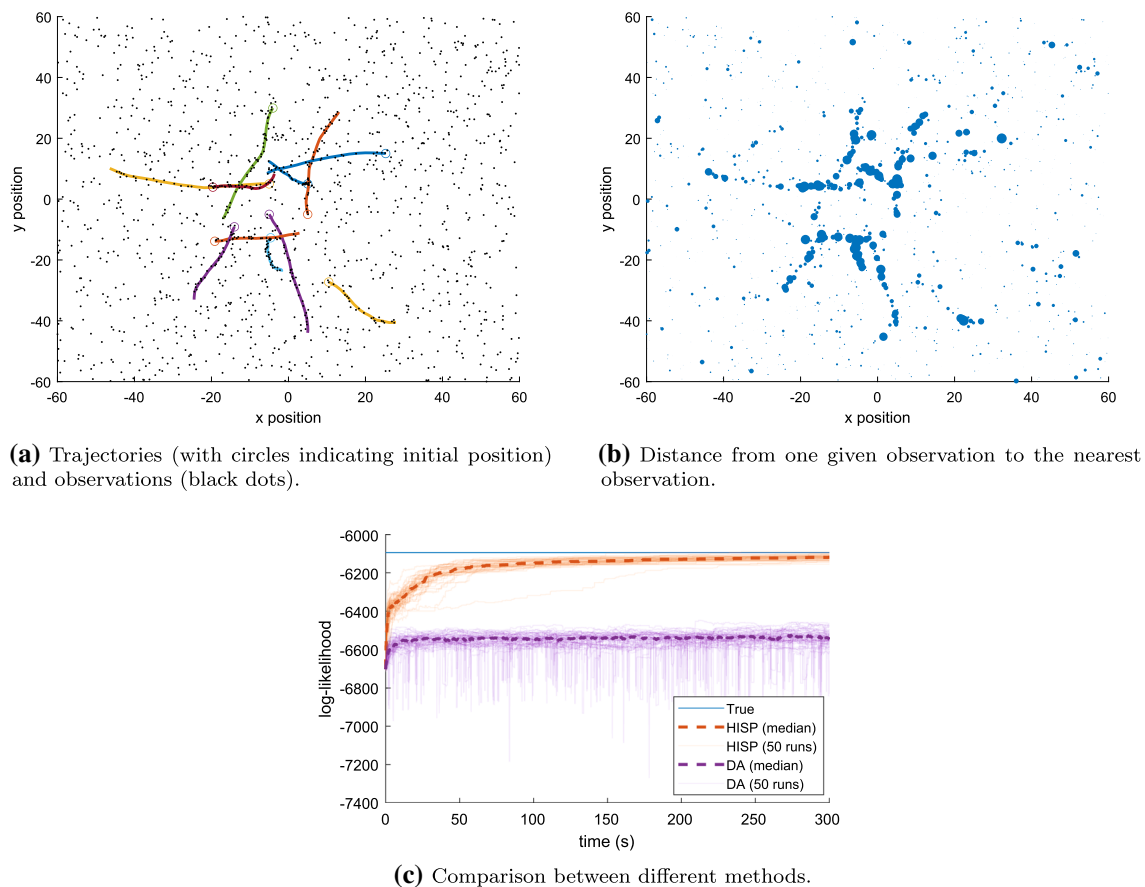


Fig. 4 Scenario with low probability of detection

in Fig. 3d show that focusing on $\delta = -1$ yields a slightly better performance, followed by focusing on $\delta = 0$. Once again, this can be attributed to the reduction in track fragmentation. The influence of the parameter c is considered once more in Fig. 3e where it appears that $c = 0.0005$ gives the best long-run performance. However, $c = 0.001$ still provides good performance throughout the run time and is considered for the other simulations. Figure 3f compares the performance of the proposed approach with MCMC-DA and shows that the latter does not mix as well as in the first scenario and fails to identify most of the tracks. The fact that the proposed approach does not reach the true log-likelihood can be attributed to local maxima in the posterior possibility function Π as well as to identifiability issues. The trace plots are shown for 50 repeats, and the median of these repeats is also plotted in order to better illustrate the behaviour of both approaches. In this scenario, it appears that it can be beneficial to run the simulated annealing several times in order to ensure that a good local maxima is found, as is usual with this type of algorithm.

5.3.3 Scenario with low probability of detection

To further assess the performance of the considered approach, we consider another challenging scenario, as shown in Fig. 4a, with the following characteristics: there are 25 false alarms and 0.5 appearing objects per time step on average and the probability of detection p_d is equal to 0.5. The difficulty of this scenario is illustrated in Fig. 4b where it appears that the inter-observation distance is not sufficient to clearly identify the objects; in particular, the observations belonging to the object at the bottom right barely appear in Fig. 4b, emphasising the fact that a probability of detection of 0.5 is not sufficient to guarantee the spatio-temporal consistency between observations. Figure 4c shows that the proposed approach can capture most of the structure of the scenario, whereas the MCMC-DA did not identify the majority of tracks in the allocated time.

5.3.4 Performance of multi-object filtering

We further assess the performance of the proposed approach by comparing the proposed multi-object filtering method against existing algorithms of the same complexity. The

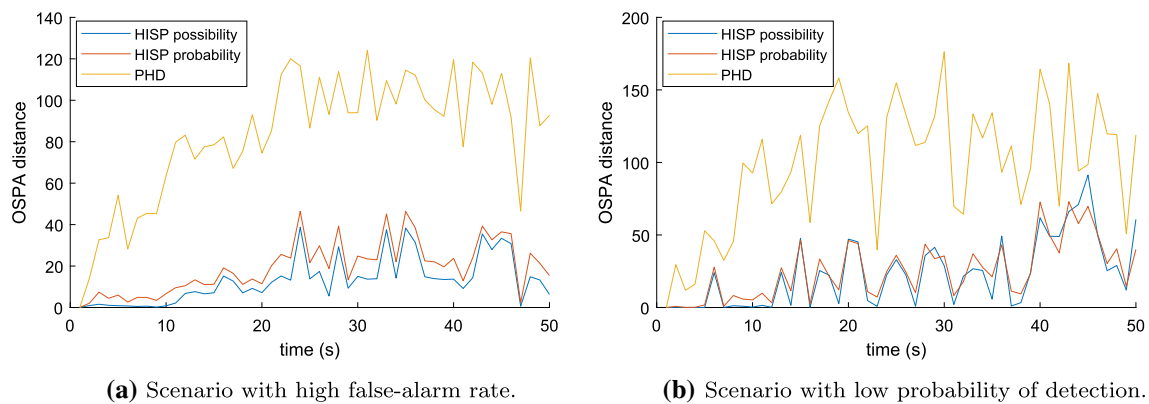


Fig. 5 Comparison between multi-object filtering methods with the first observation of each path being given. The measure of performance is the OSPA distance (Schuhmacher et al. 2008) with parameters $c = 25$ and $p = 2$ averaged over 1000 repeats

method detailed in Sect. 3.3 is an analogue of the HISP filter (Houssineau and Clark 2018) in the context of possibility theory, and both the original and the proposed approach are compared against each other as well as against the PHD filter (Mahler 2003). For the latter, paths are extracted from the underlying Gaussian mixture as is common, see, e.g. Pace and Del Moral (2013), and data associations are sampled from a distribution based on the components' weight in the Gaussian mixture.

Performance assessment is based on the OSPA distance (Schuhmacher et al. 2008), which provides a suitable metric on sets and is used here to evaluate the distance between the set of true object states and the corresponding estimate at each time step. For each scenario, the different methods are initialised with the first observation of a given object and run until the last time step. The results are then averaged over all objects and over 1000 repeats. Figure 5 shows that the proposed analogue of the HISP filter outperforms the original version in general, and especially when the false-alarm rate is high. The PHD filter does not yield reliable estimates in either of the considered scenarios. The performance of the proposed approach, when compared to the original HISP filter, could be the consequence of the assumption in (5) which relaxes the one in the original version.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Andrieu, C., Doucet, A., Holenstein, R.: Particle Markov chain Monte Carlo methods. *J. Royal Stat. Soc. Series B (Stat. Method.)* **72**(3), 269–342 (2010)
- Bao, F., Maroulas, V.: Adaptive meshfree backward SDE filter. *SIAM J. Sci. Comput.* **39**(6), A2664–A2683 (2017)
- Bekkerman, I., Tabrikian, J.: Target detection and localization using mimo radars and sonars. *IEEE Trans. Sig. Proc.* **54**(10), 3873–3883 (2006)
- Benfold B, Reid I (2011) Stable multi-target tracking in real-time surveillance video. In: *CVPR 2011, IEEE*, pp 3457–3464
- Carmi, A., Septier, F., Godsill, S.J.: The gaussian mixture mcmc particle algorithm for dynamic cluster tracking. *Automatica* **48**(10), 2454–2467 (2012)
- Chenouard N, et al. (2014) Objective comparison of particle tracking methods. *Nature methods* **11**(3)
- De Baets, B., Tsiporkova, E., Mesiar, R.: Conditioning in possibility theory with strict order norms. *Fuzzy Sets Syst.* **106**(2), 221–229 (1999)
- Dempster, A.P.: Upper and lower probability inferences based on a sample from a finite univariate population. *Biometrika* **54**(3–4), 515–528 (1967)
- Dubois D, Prade H (2015) Possibility theory and its applications: Where do we stand? In: *Springer Handbook of Computational Intelligence*, Springer, pp 31–60
- Fortmann TE, Bar-Shalom Y, Scheffe M (1980) Multi-target tracking using joint probabilistic data association. In: *1980 19th IEEE Conference on Decision and Control including the Symposium on Adaptive Processes*, IEEE, pp 807–812
- Gilks, W.R., Berzuini, C.: Following a moving target - Monte Carlo inference for dynamic Bayesian models. *J. Royal Stat. Soc: Series B (Stat. Methodol.)* **63**(1), 127–146 (2001)
- Gordon NJ, Salmond DJ, Smith AF (1993) Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In: *IEEE proceedings F (radar and signal processing)*, IET, vol 140, pp 107–113
- Haimovich, A.M., Blum, R.S., Cimini, L.J.: MIMO radar with widely separated antennas. *IEEE Sig. Proc. Mag.* **25**(1), 116–129 (2007)
- Houssineau J (2018a) A linear algorithm for multi-target tracking in the context of possibility theory. *arXiv preprint arXiv:1801.00571*
- Houssineau J (2018b) Parameter estimation with a class of outer probability measures. *arXiv preprint arXiv:1801.00569*
- Houssineau, J., Bishop, A.N.: Smoothing and filtering with a class of outer measures. *SIAM/ASA J. Uncert. Quant.* **6**(2), 845–866 (2018)

- Houssineau, J., Clark, D.E.: Multitarget filtering with linearized complexity. *IEEE Trans. Sig. Proc.* **66**(18), 4957–4970 (2018)
- Houssineau J, Laneuville D (2010) PHD filter with diffuse spatial prior on the birth process with applications to GM-PHD filter. In: 13th Conference on Information Fusion
- Houssineau J, Ristic B (2017) Sequential Monte Carlo algorithms for a class of outer measures. *arXiv preprint arXiv:1708.06489*
- Houssineau J, Chada NK, Delande E (2019a) Elements of asymptotic theory with outer probability measures. *arXiv preprint arXiv:1908.04331*
- Houssineau, J., Singh, S.S., Jasra, A.: Identification of multiobject dynamical systems: Consistency and Fisher information. *SIAM J. Control Optim.* **57**(4), 2603–2627 (2019b)
- Jaynes, E.T.: Information theory and statistical mechanics. *Phys. Rev.* **106**(4), 620 (1957)
- Jiang, L., Singh, S.S.: Tracking multiple moving objects in images using Markov Chain Monte Carlo. *Stat. Comput.* **28**(3), 495–510 (2018)
- Jiang, L., Singh, S.S., Yildirim, S.: Bayesian tracking and parameter learning for non-linear multiple target tracking models. *IEEE Trans. Sig. Proc.* **63**(21), 5733–5745 (2015)
- Khan, Z., Balch, T., Dellaert, F.: MCMC-based particle filtering for tracking a variable number of interacting targets. *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(11), 1805–1819 (2005)
- Mahler, R.P.S.: Multitarget Bayes filtering via first-order multitarget moments. *IEEE Trans. Aerospace Elect. Syst.* **39**(4), 1152–1178 (2003)
- Mahler, R.P.S.: Statistical multisource-multitarget information fusion. Artech House, Inc, UK (2007)
- Maroulas, V., Stinis, P.: Improved particle filters for multi-target tracking. *J. Comput. Phys.* **231**(2), 602–611 (2012)
- Mullane J, Vo BN, Adams MD, Vo BT (2011) A random-finite-set approach to Bayesian SLAM. *IEEE T on Robotics* 27(2)
- Oh, S., Russell, S., Sastry, S.: Markov chain Monte Carlo data association for multi-target tracking. *IEEE Trans. Automat. Cont.* **54**(3), 481–497 (2009)
- Pace, M., Del Moral, P.: Mean-field phd filters based on generalized feynman-kac flow. *IEEE J. Select. Top. Sig. Proc.* **7**(3), 484–495 (2013)
- Pailhas, Y., Houssineau, J., Petillot, Y.R., Clark, D.E.: Tracking with MIMO sonar systems: applications to harbour surveillance. *IET Radar Sonar Nav.* **11**(4), 629–639 (2016)
- Ristic, B., Clark, D., Vo, B.N., Vo, B.T.: Adaptive target birth intensity for PHD and CPHD filters. *IEEE Trans. Aerospace Electr. Syst.* **48**(2), 1656–1668 (2012)
- Ristic, B., Houssineau, J., Arulampalam, S.: Robust target motion analysis using the possibility particle filter. *IET Radar Sonar Nav.* **13**(1), 18–22 (2018)
- Ristic, B., Houssineau, J., Arulampalam, S.: Target tracking in the framework of possibility theory: The possibilistic Bernoulli filter. *Inf. Fus.* **62**, 81–88 (2020)
- Schuhmacher, D., Vo, B.T., Vo, B.N.: A consistent metric for performance evaluation of multi-object filters. *IEEE Trans. Sig. Process.* **56**(8), 3447–3457 (2008)
- Septier F, Pang SK, Carmi A, Godsill S (2009) On MCMC-based particle methods for Bayesian filtering: Application to multitarget tracking. In: 2009 3rd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), pp 360–363
- Sgouralis, I., Nebenfuhr, A., Maroulas, V.: A Bayesian topological framework for the identification and reconstruction of subcellular motion. *SIAM J. Imag. Sci.* **10**(2), 871–899 (2017)
- Shafer, G.: A mathematical theory of evidence, vol. 42. Princeton University Press, USA (1976)
- Skolnik, M.I.: Radar handbook, 2nd edn. McGrawHill, New York (1990)
- Vo, B.N., Vo, B.T., Phung, D.: Labeled random finite sets and the Bayes multi-target tracking filter. *IEEE Trans. Sig. Proc.* **62**(24), 6554–6567 (2014)
- Vu, T., Vo, B.N., Evans, R.: A particle marginal Metropolis-Hastings multi-target tracker. *IEEE Trans. Sig. Proc.* **62**(15), 3953–3964 (2014)
- Walley, P.: Statistical reasoning with imprecise probabilities. Chapman and Hall, UK (1991)
- Zanella G (2019) Informed proposals for local MCMC in discrete spaces. *J. Am. Stat. Assoc.*

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.