Information Theoretic Criteria for Observation-to-Observation Association

Islam I. Hussein, Christopher W. T. Roscoe and Matthew P. Wilkins
Applied Defense Solutions, Columbia, Maryland
Paul W. Schumacher, Jr.
Air Force Research Laboratory, Kihei, Hawaii

ABSTRACT

There are three types of data association problems. The first is the observation-to-track association (OTTA) problem, where given an observation with some known measurement statistics and a set of existing candidate (uncertain) resident space object (RSO) tracks the analyst seeks to associate each observation with a unique track (or none). The second association problem is where we have multiple tracks at different time instances and wish to determine whether any of the tracks belong to the same RSO. This is the track-to-track association (TTTA) problem. The final association problem is where we are given a set of observations at different time instances and wish to determine which of these observations were generated by the same RSO. This is the observation-to-observation association (OTOA) problem. The focus of our paper is the OTOA problem. In this paper, we tackle the OTOA problem by using an appropriate initial orbit determination (IOD) method as well as criteria from information theory. The two main criteria we use in this paper are mutual information and information divergence. We demonstrate how these two criteria can be used within an unscented transform framework as well as with a particle-based approach. The information theoretic solutions described in this paper can be adjusted to address the other (OTTA and TTTA) association problems, which will be the focus of future research. We will demonstrate the main result in simulation.

1. INTRODUCTION

Consider a set of indistinguishable objects moving continuously under the influence of a common set of deterministic dynamics and stochastic environmental influences. One or more of these objects appear randomly in the field of view (FOV) of one or more sensors (i.e. they are detectable above the background sensor noise). While these objects persist in the sensor FOV and remain detectable, the sensor provides a set of noisy measurements of the object states and their time stamp, which typically includes a subset of false detections and clutter. The essence of the multi-object tracking problem is to find tracks from these noisy sensor measurements and to rule out clutter from resident space objects (RSOs). The literature is replete with techniques on state estimation if the sequence of measurements associated with each object is known. However, the association between measurement observations and objects is not always known, leading to the well known problem of un-correlated tracks (UCTs) when attempting to update the space catalog of observed RSOs. The crux of modern space surveillance from an algorithmic point of view is to solve the data association problem and determine which measurements were generated by which objects.

In general, there are three types of data association problems. The first is the observation-to-track association (OTTA) problem described above, where the analyst seeks to associate each observation with a unique track (or none) given an observation with some known measurement statistics and a set of existing candidate (uncertain) resident space object (RSO) tracks. The second association problem is where we have multiple tracks at different time instances from one or more sensors and wish to determine whether any of the tracks belong to the same RSO. This is the track-to-track association (TTTA) problem. The final association problem is where we are given a set of observations at different time instances and wish to determine which of these observations were generated by the same RSO. This is the observation-to-observation association (OTOA) problem, which this paper will focus on.

Most methods for observation-to-track data association fall into one of two broad categories: multiple hypothesis tracking (MHT) and joint probabilistic data association (JPDA). MHT has been studied extensively in the literature and is a methodology that uses some or all reasonable measurements to update a track and delay the decision on which observation was correctly associated to a track until all measurements have been processed. Alternatively, JPDA calculates a weighted sum of reasonable observations in order to create and update a track [1]. These weights are the probability that the observation originated from a specific object, also known as the probability of detection or appearance probability. While conceptually attractive, both MHT and JPDA suffer from the curse of dimensionality where the computational burden increases exponentially as the number of potential objects increases. Modern im-

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
SEP 2014		**00-00-2014 to 00-00-2014**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Information Theoretic Criteria for Observation-to-Observation Association	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Air Force Research Laboratory,Kihei,HI,96753**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
In the Advanced Maui Optical and Space Surveillance Technologies (AMOS) Conference, 9-12 Sep 2014, Maui, HI.

14. ABSTRACT
There are three types of data association problems. The first is the observation-to-track association (OTTA) problem where given an observation with some known measurement statistics and a set of existing candidate (uncertain) resident space object (RSO) tracks the analyst seeks to associate each observation with a unique track (or none). The second association problem is where we have multiple tracks at different time instances and wish to determine whether any of the tracks belong to the same RSO. This is the track-to-track association (TTTA) problem. The final association problem is where we are given a set of observations at different time instances and wish to determine which of these observations were generated by the same RSO. This is the observation-to-observation association (OTOA) problem. The focus of our paper is the OTOA problem. In this paper, we tackle the OTOA problem by using an appropriate initial orbit determination (IOD) method as well as criteria from information theory. The two main criteria we use in this paper are mutual information and information divergence. We demonstrate how these two criteria can be used within an unscented transform framework as well as with a particle-based approach. The information theoretic solutions described in this paper can be adjusted to address the other (OTTA and TTTA) association problems, which will be the focus of future research. We will demonstrate the main result in simulation.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**10**	

plementations of near optimal JPDA algorithms and multi-frame adaptations of MHT have recently been introduced which run in polynomial time, greatly improving the computational effectiveness of JPDA based approaches [2–5]. For an extensive literature review and discussion of the nuances of different MHT and JPDA techniques, please refer to Ref. [6]. For an extensive discussion of JPDA as applied to classical filtering (JPDAF), please refer to Ref. [7].

Many data association techniques rely upon an assumption of Gaussianity of measurement errors; however, this is not typical of the space surveillance problem. For this reason, particle filter-based methods are attractive for data association but they have the significant disadvantage of being computationally expensive due to the fact that they are based upon Monte Carlo sampling approaches. In the broader multiobject multitracking community, Markov Chain Monte Carlo (MCMC) methods have been proposed to reduce the computational load [6] but, until recently, the space surveillance community has largely shied away from these types of approaches because the full implications of proper statistical sampling are yet to be fully understood [8].

The track-to-track problem can be thought of as the problem where one track is propagated forward to the time of the subsequent track and then determining an efficient gating strategy for associating the two tracks [9]. The classic space object cataloging approach has used fixed rectangular gates in Cartesian position and velocity space leading to the generation of large numbers of UCTs as the space object population has grown significantly in recent decades. This fixed gating approach was improved upon recently by Alfriend and then Hill with the introduction of covariance based track association (CBTA) [10, 11]. CBTA has been shown to be more effective than classic gating techniques but is fundamentally constrained by the validity and realism of the 2^{nd} moment of the measurement errors. In many cases, sensors belonging to the Space Surveillance Network (SSN) do not provide individual observation or track errors requiring the analyst to construct an appropriate representation and hope that it is adequate to perform CBTA.

This paper considers the final problem of observation-to-observation association. This problem can be thought of as determining the statistical dependence of observations of which there are numerous metrics to choose from. A recent approach combines an adaptive Gaussian sum filter with the Kullback-Leibler (KL) divergence measure for effective data association [12]. However, the KL divergence does not satisfy all the properties of a true distance metric making analysis of results all the more challenging in addition to the fact that computing the KL divergence is computationally demanding. In general, we want the chosen statistical dependence metric between two observations to identify a non-linear higher-than-second order dependence between measurements in order to claim a pair-wise association between observations. Mutual information (MI) has been identified in the literature as the most promising metric of statistical dependence for OTOA but there are alternative non-parametric techniques such as Kendall tau, Cross Correlograms, and Independent Component Analysis that could also be considered [13]. Ideally, the measure of statistical dependence should be valid without any assumptions of an underlying probability density function and should be extended to high dimensionality of input measurements. In this paper, we further explore both information divergence (ID) and mutual independence indexes for the OTOA problem.

In this paper we demonstrate the promise of these indexes for data association and their efficacy in solving a simple OTOA problem with two closely spaced RSOs. In Section 2., we first describe the overall procedure and how it relates to initial orbit determination (IOD). In Section 3., we describe the various information theoretic indexes one may consider. We demonstrate efficacy of the methods in a numerical example in Section 4. We summarize the main results of the paper in Section 5.

2. INITIAL ORBIT DETERMINATION AND THE OTOA PROBLEM

The core idea in the proposed OTOA approach is to use an appropriate initial orbit determination (IOD) method to generate an uncertain track from a set of measurements (that we wish to test for association) and their known statistics (see Fig. (1)). One can then, for example, compare (in some information theoretic sense) the amount of information shared between the generated orbit statistics and the measured output statistics. The more consistent the estimated track is with the measurements, the more likely the observations were generated by the same physical phenomenon. One may also compare the degree of consistency between the generated orbit's output and the measured output statistics. The former method of comparison is based on the notion of mutual information between the IOD-based orbit statistics and the measured observation statistics (see Fig. (2)). The second method is based on some notion of information divergence or "distance" between the measured observation statistics and the IOD-based reconstructed

observation statistics (see Fig. (3)). Clearly, the performance of the proposed data association schemes depends on the performance of the underlying IOD method. We will describe in detail the two approaches in the next section.

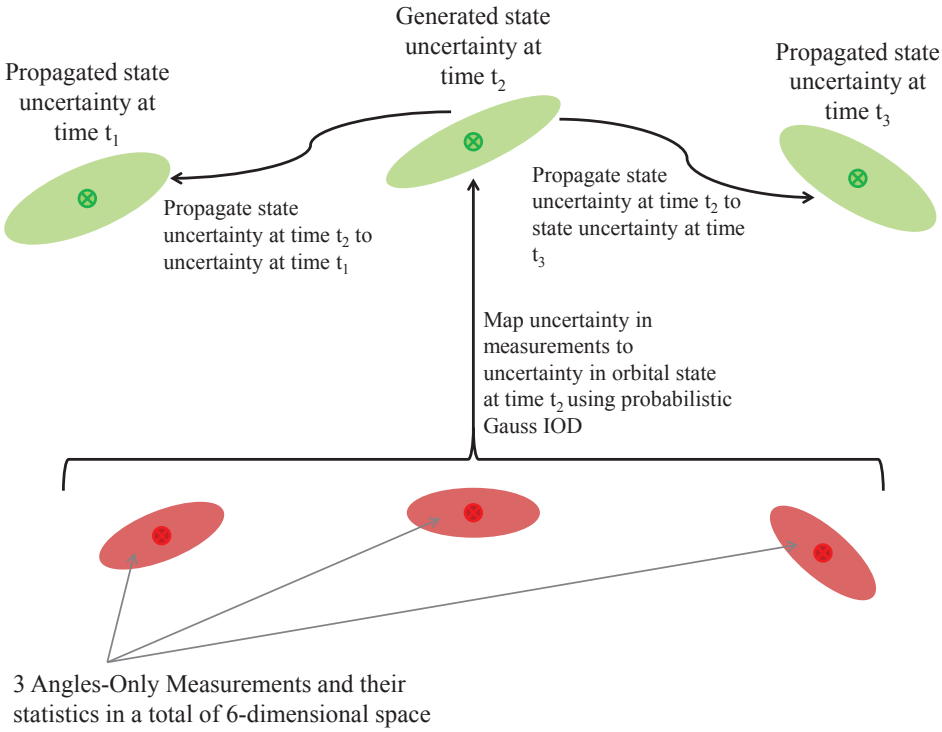

Fig. 1. The general probabilistic IOD approach developed by Hussein et al. [14].

For illustration purposes, we will assume that the observations are angles-only. We will use the classical Gauss method approach for the IOD step (see Ref. [15]). In general, assume we are given a set of n observations $\{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n\}$ taken at observation times $t_1, \ldots, t_n$ (times are assumed distinct, without any loss of generality). Of these we will choose 3 observations (as required by the Gauss method) to test whether they were generated by the same RSO. The procedure is applied to all combinations of observations. How to computationally address the combinatorics problem is beyond the scope of this paper and we will only address the problem of determining whether a set of three observations were generated by the same RSO. But in the general n observation problem, those observations that were deemed to be associated will now form tracks and the remainder of this problem becomes a TTTA problem that results in "stringing"' these tracks together to form a set of non-redundant new tracks. We do not address the TTTA problem here.

Given three distinct measurements $\boldsymbol{z} = (\boldsymbol{z}_1, \boldsymbol{z}_2, \boldsymbol{z}_3)$ taken at t_1, t_2, and t_3, the Gauss IOD method produces a candidate orbit described by the six-dimensional state $\boldsymbol{x}_2 = \boldsymbol{g}(\boldsymbol{z})$ at time t_2, where $\boldsymbol{g}(\cdot)$ is the function that maps a set of three angles-only measurements to orbital space coordinates. The state may be specified in orbital elements, Cartesian coordinates, etc. Furthermore, let $\boldsymbol{f}_{ij}$ be the function that propagates the state $\boldsymbol{x}_i$ defined at time t_i to the state $\boldsymbol{x}_j$ defined at time t_j. Then $\boldsymbol{x}_1 = \boldsymbol{f}_{21}(\boldsymbol{x}_2)$ is the (backward) propagated state at time t_1 and $\boldsymbol{x}_3 = \boldsymbol{f}_{23}(\boldsymbol{x}_2)$ is the propagated state at time t_3 given the state $\boldsymbol{x}_2$ at time t_2. Finally, let $\boldsymbol{h}_i$ be the function that maps the state at time t_i to an observation $\boldsymbol{z}_i = \boldsymbol{h}_i(\boldsymbol{x}_i)$.

In order to analyze the uncertainty in the orbital space resulting from a given uncertainty in the measurement space, we will use both the unscented transform (UT) and the Monte Carlo (MC) method (see Hussein et al. [14]). For the UT analysis, the measurement process is assumed Gaussian and the generated orbit uncertainty will, therefore, also be Guassian. Likewise, the measurement statistics in the MC approach will be assumed Gaussian. 5000 particles are generated from the 6-dimensional measurement uncertainty probability density function (pdf) $p_M(\boldsymbol{z})$ and mapped into orbital space coordinates at the time of the second measurement t_2. The resulting Gaussian distribution from the UT

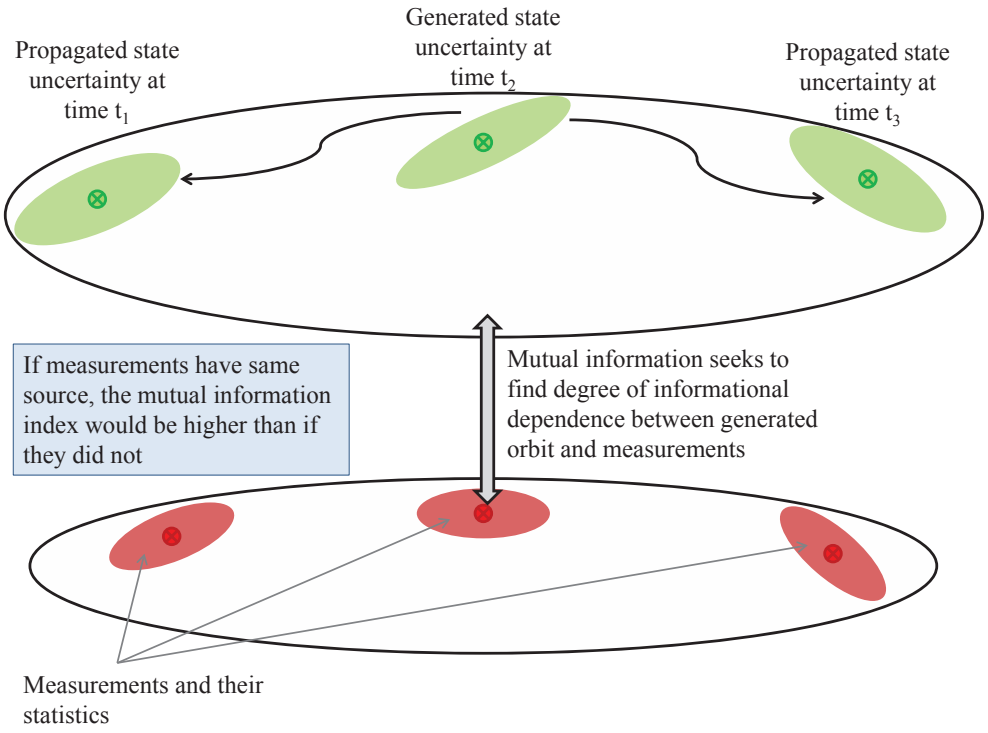

Fig. 2. Mutual information can be used as an index of how much dependence exist between a set of measurements and the orbit they generate if they were associated. The "more" unassociated the observations are the smaller the mutual information will be, and vice versa.

analysis and the particle cloud from the MC analysis in the orbital space are approximations of the actual uncertainty pdf $p_O^2(\boldsymbol{x}_2)$ of the IOD solution. In the UT method, there are a total of $13 = 2 \times 6 + 1$ sigma points $\{\mathcal{Z}^{(j)}\}$ since the measurement space has a dimension of 6.

As noted by Hussein et al. [14], there is a subtle difference between this description of the problem compared to other approaches which have appeared previously in the literature. The nonlinear mapping $\boldsymbol{g}(\cdot)$ is a map from the entire three-measurement space to the orbital space, and is *not* a map of an individual measurement $\boldsymbol{z}_i$, $i = 1, 2, 3$, to the orbital space. Therefore, samples of the measurement uncertainty should be drawn from the distribution in the six-dimensional measurement space defined with the global measurement variable $\boldsymbol{z}$ and not from the individual distributions defined on the individual measurement variables $\boldsymbol{z}_i$, $i = 1, 2, 3$. For the UT method, in particular, this will result in the correct number (13) of sigma points being generated to describe the uncertainty distribution in six-dimensional orbital space.

For example, an MC sample and UT sigma points are shown in Fig. (4) for a Gaussian uncertainty with each dimension independently distributed with $3\sigma = 10$ arcsec. Each colored set of dots represents a two dimensional projection of the MC sample onto each individual measurement plane (i.e., there are $5000 \times 3 = 15\,000$ dots shown, but the MC sample only contains 5000 particles). The remaining dimensions of the sigma points overlap each other since the uncertainty is the same in each direction. The histograms show the marginal distributions of the MC azimuth and elevation uncertainty samples for one of the measurement times.

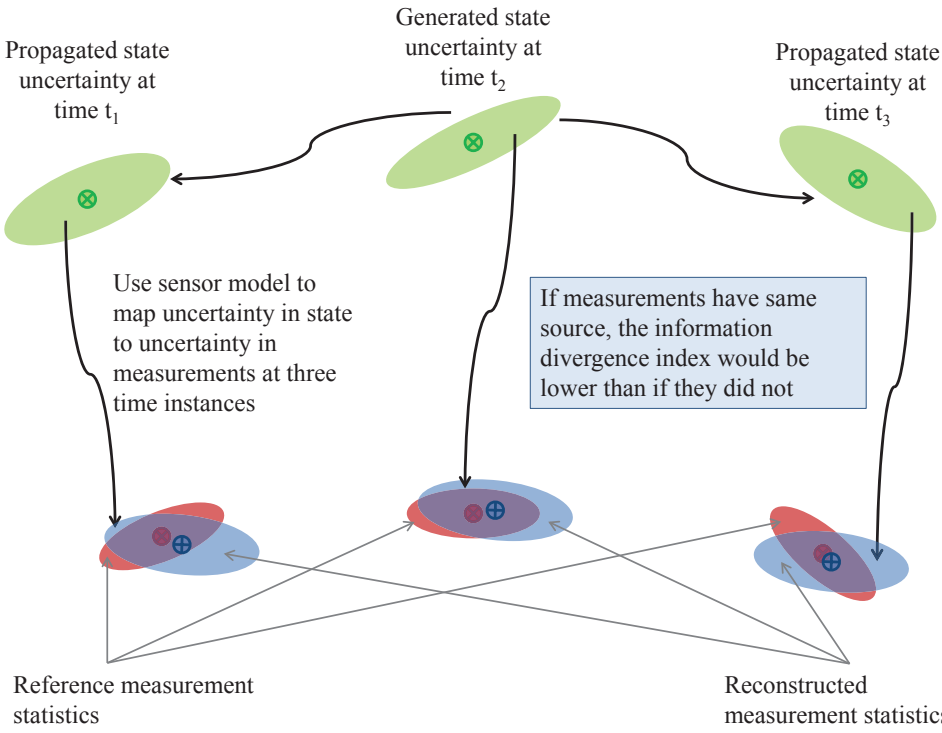

Fig. 3. Information divergence can be used to measure how different the measurement statistics are from that of the reconstructed measurement statistics under the hypothesis that the observations are associated.

3. INFORMATION THEORETIC CRITERIA FOR OTOA

3.1 Information Divergence

Assuming that the measurements are statistically independent, let $p_M(\boldsymbol{z}_1, \boldsymbol{z}_2, \boldsymbol{z}_3) = p_M^1(\boldsymbol{z}_1)p_M^2(\boldsymbol{z}_2)p_M^3(\boldsymbol{z}_3)$ be the joint pdf in the three measurements. With abuse of notation, let $p_O^2(\boldsymbol{x}_2) = \boldsymbol{g}(p_M(\boldsymbol{z}_1, \boldsymbol{z}_2, \boldsymbol{z}_3))$ be the orbital IOD-based reconstructed pdf at t_2, and $p_O^1(\boldsymbol{x}_1) = \boldsymbol{f}_{21}(p_O^2(\boldsymbol{x}_2))$ and $p_O^3(\boldsymbol{x}_3) = \boldsymbol{f}_{23}(p_O^2(\boldsymbol{x}_2))$ be the propagated pdfs at times t_1 and t_3, respectively. Finally, let $\tilde{p}_M^i(\boldsymbol{z}_i) = \boldsymbol{h}(p_O^i(\boldsymbol{x}_i))$ be the uncertainty in the measurement as mapped from the uncertainty in the state at time t_i. If

1. all three observations were generated by the same RSO,
2. the IOD method perfectly produces the exact orbit that generated the measurements, and
3. the orbital propagation and measurement models perfectly match the true orbital motion and measurement processes

then $\tilde{p}_M^i(\boldsymbol{z}_i)$ should be equal to $p_M^i(\boldsymbol{z}_i)$. In this paper we will assume that the third condition holds. However, the second condition is guaranteed not to be true since the Gauss IOD method is not exact. Therefore, any mismatch between $\tilde{p}_M^i$ and p_M^i would be the result of the IOD method's errors or if the hypothesis that the three observations were generated by the same RSO does not hold. *The key insight is that among the different associations between a set of measurements $\{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n\}$ the ones that result in the least amount of discrepancy between $\tilde{p}_M^i$ and p_M^i, above and beyond the discrepancy introduced by the IOD method that is shared among all associations*, are the more likely ones to be true associations.* The key question is: How can one measure the discrepancy between two pdfs?

*This assumes that the performance of the IOD method is independent of the measurements and their statistics. To the best of our knowledge, such a dependence has not been studied in the literature.

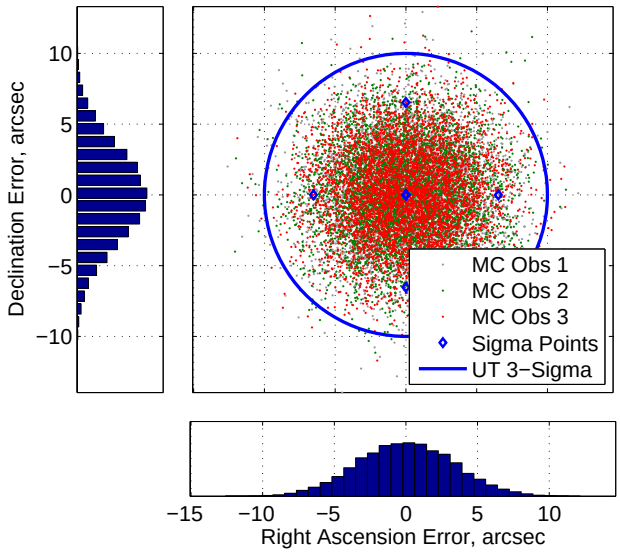

Fig. 4. Example of sampling a six-dimensional Gaussian measurement uncertainty.

One such method is the use of information divergence [16]. While there are many definitions of information divergence [17, 18], we will use the Kullback-Leibler (KL) divergence [16]. The KL divergence between two pdfs $p(\boldsymbol{x})$ and $q(\boldsymbol{x})$ is given by:

$$D_{KL}(p(\boldsymbol{x})||q(\boldsymbol{x})) = \int p(\boldsymbol{x}) \log \frac{p(\boldsymbol{x})}{q(\boldsymbol{x})} \mathrm{d}\boldsymbol{x} \tag{1}$$

If both p and q are Gaussian, then one can compute $D_{KL}(p||q)$ in closed form and is given by (see for example, Ref. [12])

$$D_{KL}(p(\boldsymbol{x})||q(\boldsymbol{x})) = \frac{1}{2}\left[\log\left(\frac{\|\boldsymbol{\Sigma}_q\|}{\|\boldsymbol{\Sigma}_p\|}\right) + \mathrm{Tr}\left(\boldsymbol{\Sigma}_q^{-1}\boldsymbol{\Sigma}_p\right) - d + \left(\boldsymbol{\mu}_p - \boldsymbol{\mu}_q\right)\cdot\boldsymbol{\Sigma}_q^{-1}\cdot\left(\boldsymbol{\mu}_p - \boldsymbol{\mu}_q\right)\right], \tag{2}$$

where $\boldsymbol{\mu}_p$ and $\boldsymbol{\Sigma}_p$ (resp., $\boldsymbol{\mu}_q$ and $\boldsymbol{\Sigma}_q$) are the mean and covariance of the pdf p (resp., q), and where d is the dimension of the underlying space. In the case where p and q are Gaussian mixtures, there is no known closed-form expression for the KL divergence. However, other definitions of divergence can be used to obtain a closed-form expression. One such divergence is the Cauchy-Schwarz divergence [19, 20].

3.1.1 Unscented Transform: Following the UT procedure described in Ref. [14], one starts by assuming Gaussian statistics for the observations: $p_M^i(\boldsymbol{z}_i) = p_g(\boldsymbol{z}_i; \boldsymbol{\mu}_{\boldsymbol{z}_i}, \boldsymbol{\Sigma}_{\boldsymbol{z}_i})$, where $p_g(\boldsymbol{r}; \boldsymbol{\mu}_{\boldsymbol{r}}, \boldsymbol{\Sigma}_{\boldsymbol{r}})$ is a Gaussian distribution in $\boldsymbol{r}$ with mean $\boldsymbol{\mu}_{\boldsymbol{r}}$ and covariance $\boldsymbol{\Sigma}_{\boldsymbol{r}}$. Using the Gauss IOD method, one can obtain a Gaussian approximation $\tilde{p}_O^2(\boldsymbol{x}_2) = p_g(\boldsymbol{x}_2; \boldsymbol{\mu}_{\boldsymbol{x}_2}, \boldsymbol{\Sigma}_{\boldsymbol{x}_2})$ of the pdf $p_O^2(\boldsymbol{x}_2)$. As one would normally do in an Unscented Kalman Filter (UKF), this pdf can be propagated backward (respectively, forward) in time to obtain the Gaussian pdf approximation $\tilde{p}_O^1(\boldsymbol{x}_1) = p_g(\boldsymbol{x}_1; \boldsymbol{\mu}_{\boldsymbol{x}_1}, \boldsymbol{\Sigma}_{\boldsymbol{x}_1})$ (respectively, $\tilde{p}_O^3(\boldsymbol{x}_3) = p_g(\boldsymbol{x}_3; \boldsymbol{\mu}_{\boldsymbol{x}_3}, \boldsymbol{\Sigma}_{\boldsymbol{x}_3})$). From these and using the assumed-known measurement model, one can generate the pdf for the measurements at these three times based on the IOD-based orbit under the assumption that the observations are associated: $\tilde{p}_M^i\left(\boldsymbol{z}_i; \tilde{\boldsymbol{\mu}}_{\boldsymbol{z}_i}^i, \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{z}_i}^i\right) = \boldsymbol{h}(\tilde{p}_O^i(\boldsymbol{x}_i))$, $i = 1, 2, 3$. We may now compute the information divergence $D_{KL}^i(\tilde{p}_M^i(\boldsymbol{z}_i)||p_M^i(\boldsymbol{z}))$ between $\tilde{p}_M^i$ and p_M^i for each of the three times t_i, $i = 1, 2, 3$. The overall criterion for assessing association would then be

$$D_{KL} = \sum_{i=1}^{3} D_{KL}^i(\tilde{p}_M^i(\boldsymbol{z}_i)||p_M^i(\boldsymbol{z})). \tag{3}$$

The association that maximizes this index among all possible combinations of three angles-only measurements is one way to assess association. We call this solution to the OTOA problem the *UT-ID solution* to be referenced in the

numerical example section later in the paper. This index is similar in spirit to the divergence-based (OTTA) association discussed in Ref. [12].

3.1.2 Particle Based Methods: The underlying assumption of Gaussianity in the UT approach, while analytically appealing, is not necessarily a faithful representation of the uncertainty in the reconstructed state from the IOD method. A more accurate representation would be to represent uncertainty using a Monte Carlo (MC) particle cloud. However, computing divergence based on a particle cloud poses analytical challenges. To see this, consider a particle cloud $\left\{\mathcal{Z}^{(i)}\right\}$ with weights $\left\{w_{\boldsymbol{z}}^{(i)}\right\}$ drawn from $p_M(\boldsymbol{z}_1,\boldsymbol{z}_2,\boldsymbol{z}_3) = p_g^1(\boldsymbol{z}_1;\boldsymbol{\mu}_{\boldsymbol{z}_1},\boldsymbol{\Sigma}_{\boldsymbol{z}_1})p_g^2(\boldsymbol{z}_2;\boldsymbol{\mu}_{\boldsymbol{z}_2},\boldsymbol{\Sigma}_{\boldsymbol{z}_2})p_g^3(\boldsymbol{z}_3;\boldsymbol{\mu}_{\boldsymbol{z}_3},\boldsymbol{\Sigma}_{\boldsymbol{z}_3})$. This cloud can be mapped into the state space using the Gauss IOD method and that, in return, mapped to the measurement space using the observation function $\boldsymbol{h}(\cdot)$. Denote the resulting cloud by $\{\tilde{\mathcal{Z}}^{(i)}\}$ with weights $\left\{\tilde{w}_{\boldsymbol{z}}^{(i)}\right\}$. If the IOD method was exact and the motion and measurement models were known perfectly, the particle set $\{\tilde{\mathcal{Z}}^{(i)}\}$ should be identical to the set $\{\mathcal{Z}^{(i)}\}$ and the KL divergence can be shown to be:

$$
\begin{aligned}
D_{KL}(\tilde{p}_M^i(\boldsymbol{z}_i)||p_M^i(\boldsymbol{z})_i) &= \int \tilde{p}_M^i(\boldsymbol{z}_i)\log\left(\frac{\tilde{p}_M^i(\boldsymbol{z}_i)}{p_M^i(\boldsymbol{z}_i)}\right)\mathrm{d}\boldsymbol{z}_i \\
&= E_{\tilde{p}_M^i(\boldsymbol{z}_i)}\left[\log\left(\frac{\tilde{p}_M^i(\boldsymbol{z}_i)}{p_M^i(\boldsymbol{z}_i)}\right)\right] \\
&\simeq \sum_i \tilde{w}^i \log\left(\frac{\tilde{p}_M^i(\tilde{\mathcal{Z}}^{(i)})}{p_M^i(\mathcal{Z}^{(i)})}\right) \\
&= \sum_i \tilde{w}^i \log\left(\frac{\sum_j \tilde{w}^j \delta_{\tilde{\mathcal{Z}}^{(j)}}(\tilde{\mathcal{Z}}^{(i)})}{\sum_k w^k \delta_{\mathcal{Z}^{(k)}}(\tilde{\mathcal{Z}}^{(i)})}\right) \\
&= \sum_i \tilde{w}^i \log\left(\frac{\tilde{w}^i}{w^i}\right) \\
&= \sum_i \tilde{w}^i \log\left(\frac{w^i}{w^i}\right) = 0,
\end{aligned}
$$

which is what we would expect under perfect knowledge of the system model and with a perfect IOD method. However, because of errors in the IOD solution, the two sets $\{\tilde{\mathcal{Z}}^{(i)}\}$ and $\{\tilde{\mathcal{Z}}^{(i)}\}$ will not match and the delta functions in the denominator in the previous equation will evaluate to zero for all i and we will get $D_{KL} \propto \log(1/0) = \log\infty = \infty$. Even if one uses other definitions of divergence, the value of the divergence would not correspond to the actual divergence between $p_M^i(\boldsymbol{z}_i)$ and $\tilde{p}_M^i(\boldsymbol{z})$. *Note, however, that if p_M is analytically known, can be sampled and can be evaluated, then information divergence can be computed when $\tilde{p}_M^i$ is represented by a cloud of particles.* We will demonstrate this in a future publication. We point out, however, that in the next subsection we will be able to compute the mutual information between two particle clouds and not run into the issue mentioned above for information divergence.

An alternative approach to computing the divergence directly using the particle cloud is to approximate the particle cloud using some analytic pdf for which we can compute the divergence. One such approach would be to use the expectation maximization algorithm to convert the particle cloud into a Gaussian Mixture Model [21]. We will not describe the details of the method here. Instead we refer the reader to an EM algorithm described in Ref. [22], which is a robust and versatile EM algorithm, called the FJ-EM algorithm. The method allows for the user to specify a maximum number of GMM components and it selects the number of components that best represent the particle cloud. The user is also not required to choose a good initial guess of the components. In other words, the method does not require a careful initialization of the algorithm (other EM algorithms do require a very careful initialization.)

As an illustration, consider the 2000 particle cloud generated from the four component planar GMM with weights:

$$w_i = 0.25,\ i = 1,2,3,4,$$

means:

$$\begin{aligned}\boldsymbol{\mu}_1 &= [10.0\ 0.0]^{\mathrm{T}} \\ \boldsymbol{\mu}_2 &= [-10.0\ 0.0]^{\mathrm{T}} \\ \boldsymbol{\mu}_3 &= [0.0\ 10.0]^{\mathrm{T}} \\ \boldsymbol{\mu}_4 &= [0.0\ -10.0]^{\mathrm{T}}\end{aligned}$$

and covariances:

$$\boldsymbol{\Sigma}_i = \begin{bmatrix} 2.0 & 0.0 \\ 0.0 & 2.0 \end{bmatrix},\ i = 1, 2, 3, 4.$$

The generated particle cloud was then fed into the EM algorithm and the following 5-component GMM initial guess was used for initializing the algorithm:

$$\begin{aligned}w_1^0 &= 0.1 \\ w_2^0 &= 0.4 \\ w_3^0 &= 0.2 \\ w_4^0 &= 0.2 \\ w_5^0 &= 0.1\end{aligned}$$

$$\begin{aligned}\boldsymbol{\mu}_1^0 &= [1.0\ 0.0]^{\mathrm{T}} \\ \boldsymbol{\mu}_2^0 &= [-1.0\ 0.0]^{\mathrm{T}} \\ \boldsymbol{\mu}_3^0 &= [0.0\ 1.0]^{\mathrm{T}} \\ \boldsymbol{\mu}_4^0 &= [-1.0\ -1.0]^{\mathrm{T}} \\ \boldsymbol{\mu}_5^0 &= [-5.0\ -5.0]^{\mathrm{T}}\end{aligned}$$

$$\boldsymbol{\Sigma}_i^0 = \begin{bmatrix} 1.0 & 0.0 \\ 0.0 & 1.0 \end{bmatrix},\ i = 1, 2, 3, 4, 5.$$

The algorithm resulted in the following set of GMM weights:

$$\begin{aligned}w_1^f &= 0.25 \\ w_2^f &= 0.24849 \\ w_3^f &= 0.26156 \\ w_4^f &= 0.23995 \\ w_5^f &= 0.0\end{aligned}$$

where we notice that one component has been eliminated. The final set of means were found to be

$$\begin{aligned}\boldsymbol{\mu}_1^f &= [9.96428\ 0.02116]^{\mathrm{T}} \\ \boldsymbol{\mu}_2^f &= [-9.87450\ -0.02471]^{\mathrm{T}} \\ \boldsymbol{\mu}_3^f &= [0.03449\ 10.06634]^{\mathrm{T}} \\ \boldsymbol{\mu}_4^f &= [0.04656\ -9.96678]^{\mathrm{T}}\end{aligned}$$

and the final set of covariances were found to be

$$\boldsymbol{\Sigma}_1^f = \begin{bmatrix} 1.92951 & 0.02156 \\ 0.02156 & 1.92172 \end{bmatrix}$$
$$\boldsymbol{\Sigma}_2^f = \begin{bmatrix} 2.10782 & -0.04989 \\ -0.04989 & 1.92587 \end{bmatrix}$$
$$\boldsymbol{\Sigma}_3^f = \begin{bmatrix} 1.89439 & 0.05116 \\ 0.05116 & 2.35834 \end{bmatrix}$$
$$\boldsymbol{\Sigma}_4^f = \begin{bmatrix} 1.99178 & 0.01623 \\ 0.01623 & 1.85653 \end{bmatrix}$$

The resulting GMM is graphically shown in Fig. 5 against the particles. Note that a component has a weight of zero (i.e., eliminated) resulting in an effective number of components of 4. While the true values of the weights, means and covariances were not recovered exactly, the error in these parameters is quite small.

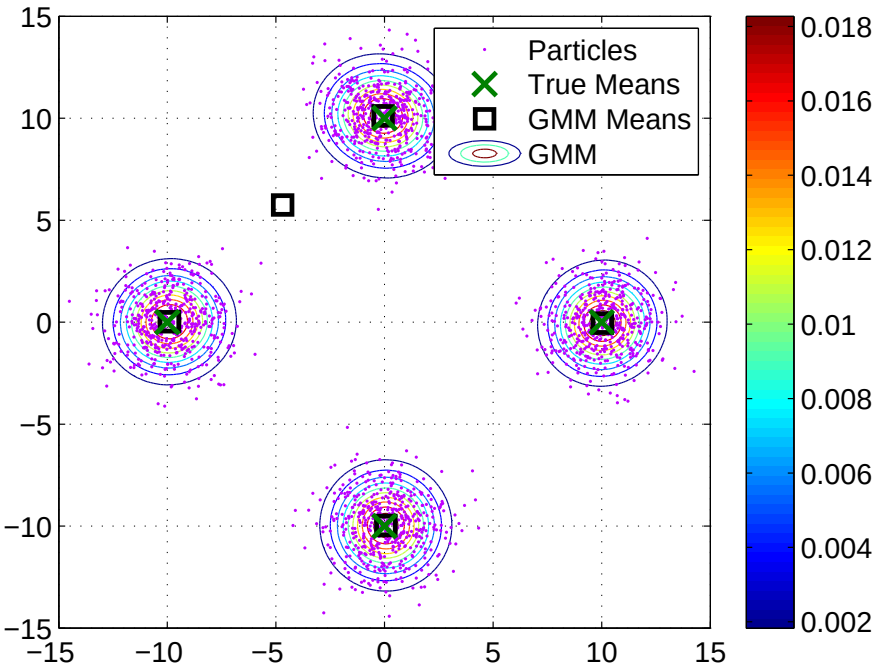

Fig. 5. An example showing the original means (equally weighted) that generated the shown particle cloud. The cloud was then fed into the FJ-EM algorithm that generated the GMM approximation shown. The FJ-EM algorithm was able to very accurately reconstruct the true GMM that generated the particle cloud.

To summarize starting with the two particle clouds $\{\mathcal{Z}^{(i)}\}$ and $\{\tilde{\mathcal{Z}}^{(i)}\}$, one can convert each into a GMM to approximate $p_M(\boldsymbol{z}_i)$ and $\tilde{p}_M(\boldsymbol{z}_i)$, $i = 1, 2, 3$. There is, however, no known closed-form expression for KL divergence between two GMMs. The Cauchy-Schwarz (CS) divergence, however, can be computed for GMMs [19, 20]. The general expression for CS divergence, K_{CS}, between two pdfs $p(\boldsymbol{x})$ and $q(\boldsymbol{x})$ is given by

$$D_{CS}(p(\boldsymbol{x})||q(\boldsymbol{x}) = \frac{1}{2}\log\left[\frac{\left(\int p^2(\boldsymbol{x})\mathrm{d}\boldsymbol{x}\right)\left(\int q^2(\boldsymbol{x})\mathrm{d}\boldsymbol{x}\right)}{\left(\int p(\boldsymbol{x})q(\boldsymbol{x})\mathrm{d}\boldsymbol{x}\right)^2}\right],$$

Our ability to compute the CS divergence for GMMs in closed-form lies in the fact that if $p(\boldsymbol{x}) = \sum_{i=1}^{N_p} w^i p_g(\boldsymbol{x}; \boldsymbol{\mu}_p^i, \boldsymbol{\Sigma}_p^i)$

and $q(\boldsymbol{x}) = \sum_{i=1}^{N_q} w^i p_g(\boldsymbol{x}; \boldsymbol{\mu}_q^i, \boldsymbol{\Sigma}_q^i)$ one can show that

$$\int p(\boldsymbol{x})q(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = \sum_{i=1}^{N_p}\sum_{j=1}^{N_q} w^i w^j p_g(\boldsymbol{\mu}_p^i; \boldsymbol{\mu}_q^j, \boldsymbol{\Sigma}^i + \boldsymbol{\Sigma}^j).$$

This expression can be used to compute $\int p^2(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$ and $\int q^2(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$, and hence $D_{CS}(p(\boldsymbol{x})||q(\boldsymbol{x}))$ when both p and q are GMMs.

Thus, the two particle clouds $\{\mathcal{Z}^{(i)}\}$ and $\{\tilde{\mathcal{Z}}^{(i)}\}$ are first converted to GMMs using the FJ-EM algorithm, which are then subsequently used to compute the approximate CS divergence between $p_M^i(\boldsymbol{z}_i)$ and $\tilde{p}_M^i(\boldsymbol{z}_i)$. The overall divergence-based index for the MC case is the sum of the three divergences:

$$D_{CS}^{\mathrm{tot}} = \sum_{i=1}^{3} D_{CS}(p_M^i(\boldsymbol{z}_i)||\tilde{p}_M^i(\boldsymbol{z}_i)). \tag{4}$$

We call this solution to the OTOA problem the *MC-ID solution* to be referenced in the numerical example section later in the paper.

3.2 Mutual Information

The mutual information, $I(\boldsymbol{x}, \boldsymbol{z})$, between two random variables $\boldsymbol{x}$ and $\boldsymbol{z}$ is a measure of the degree of dependence between these two variables. Formally, it is given by

$$I(\boldsymbol{x}, \boldsymbol{z}) = \int\int p(\boldsymbol{x}, \boldsymbol{z}) \log\left(\frac{p(\boldsymbol{x}, \boldsymbol{z})}{p(\boldsymbol{x})p(\boldsymbol{z})}\right) \mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{z}, \tag{5}$$

where $p(\boldsymbol{x}, \boldsymbol{z})$ is the joint distribution of $\boldsymbol{x}$ and $\boldsymbol{z}$ and where $p(\boldsymbol{x})$, respectively $p(\boldsymbol{z})$, is the marginalization of $p(\boldsymbol{x}, \boldsymbol{z})$ with respect to $\boldsymbol{z}$, respectively $\boldsymbol{x}$. Two important properties are to be noted here. Firstly, mutual information is a symmetric function of $\boldsymbol{x}$ and $\boldsymbol{z}$. Secondly, if $\boldsymbol{x}$ and $\boldsymbol{z}$ are independent then $p(\boldsymbol{x}, \boldsymbol{z}) = p(\boldsymbol{x})p(\boldsymbol{z})$ and the mutual information is zero.

The mutual information index we propose to use is as follows. First, consider the mutual information $I_i(\boldsymbol{x}_i, \boldsymbol{z}_i)$ between the state $\boldsymbol{x}_i$ and the measured output $\boldsymbol{z}_i$ at time t_i. The overall mutual information index would then be $I_{\mathrm{tot}} = I_1 + I_2 + I_3$. Other indexes based on mutual information can also be considered. For example, one can consider the mutual information between the *joint* orbital state variable $\boldsymbol{x}_{\mathrm{joint}} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3)$ and the joint measurement variable $\boldsymbol{z}_{\mathrm{joint}} = (\boldsymbol{z}_1, \boldsymbol{z}_2, \boldsymbol{z}_3)$. Such indexes will be the subject of future research.

It can be shown that the mutual information $I_i(\boldsymbol{x}_i, \boldsymbol{z}_i)$ can be expressed in terms of the KL divergence:

$$I_i(\boldsymbol{x}_i, \boldsymbol{z}_i) = D_{KL}(p(\boldsymbol{x}_i, \boldsymbol{z}_i)||p(\boldsymbol{x}_i)p(\boldsymbol{z}_i)). \tag{6}$$

Therefore, if the joint pdf $p(\boldsymbol{x}_i, \boldsymbol{z}_i)$ is Gaussian, one can compute I_i in closed-form. However, for the GMM case, no closed-form solution exists because the mutual information is related to the KL divergence and not a divergence for which a GMM has a closed-form solution.

3.2.1 Unscented Transform: Following the UT procedure described in Ref. [14], one can obtain a Gaussian approximation of the pdf of the state $\boldsymbol{x}_2$. One can then use the UKF to obtain the joint distribution $p(\boldsymbol{x}_i, \boldsymbol{z}_i)$ after propagating and updating with the corresponding measurement $\boldsymbol{z}_i$, $i = 1, 2, 3$. It can be shown that this distribution is Gaussian with the following mean:

$$\boldsymbol{\mu}_i^{\mathrm{joint}} = \begin{bmatrix} \boldsymbol{\mu}_{\boldsymbol{x}_i} \\ \boldsymbol{\mu}_{\boldsymbol{z}_i} \end{bmatrix}$$

and covariance:

$$\boldsymbol{\Sigma}_i^{\mathrm{joint}} = \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{x}_i} & \boldsymbol{\Sigma}_{\boldsymbol{x}_i,\boldsymbol{z}_i} \\ \boldsymbol{\Sigma}_{\boldsymbol{x}_i,\boldsymbol{z}_i}^{\mathrm{T}} & \boldsymbol{\Sigma}_{\boldsymbol{z}_i} \end{bmatrix},$$

where $\boldsymbol{\mu}_{\boldsymbol{x}_i}$ is the UKF *updated* state mean at time t_i, $\boldsymbol{\mu}_{\boldsymbol{z}_i}$ is the measurement mean at time t_i, $\boldsymbol{\Sigma}_{\boldsymbol{x}_i}$ is the covariance in the state at time t_i, $\boldsymbol{\Sigma}_{\boldsymbol{x}_i}$ is the measurement covariance at time t_i, and $\boldsymbol{\Sigma}_{\boldsymbol{x}_i,\boldsymbol{z}_i}$ is the cross-covariance between the state and the measurement at time t_i.

Since the joint pdf is Gaussian, the marginalization with respect to $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$ are both Gaussian and have means $\boldsymbol{\mu}_{\boldsymbol{x}_i}$ and $\boldsymbol{\mu}_{\boldsymbol{z}_i}$, respectively, and covariances $\boldsymbol{\Sigma}_{\boldsymbol{x}_i}$ and $\boldsymbol{\Sigma}_{\boldsymbol{z}_i}$, respectively. Now that we have Gaussian approximations of $p(\boldsymbol{x}_i, \boldsymbol{z}_i)$ and $p(\boldsymbol{x}_i)p(\boldsymbol{z}_i)$, one can use Eq. (6) and Eq. (2) to compute the mutual information $I^i(\boldsymbol{x}_i, \boldsymbol{z}_i)$. We call this solution to the OTOA problem the *UT-MI solution* to be referenced in the numerical example section later in the paper. This index is similar in spirit to the mutual information-based (OTTA) association discussed in Ref. [12].

3.2.2 Particle Based Methods: As mentioned previously, converting particle clouds to GMMs to compute MI is not feasible since MI is defined in terms the KL divergence Eq. (6) for which we do not have a closed-form GMM-based expression. Rather surprisingly, however, we can actually use the particle cloud to compute a faithful evaluation of the mutual information index. This solution is motivated by the MC approach in Ref. [23]. While this solution is faithful to the particle nature of the densities, it is computationally expensive. At the time of publication we were in the process of parallelizing the particle-based solution on a graphics processing unit (GPU). However, results are not readily available at the present time. Instead we will describe the mathematics behind the solution and provide a non-SSA example to verify that the particle-based expression for mutual information is correct.

Again, we begin with a particle cloud $\left\{\mathcal{Z}^{(j)}\right\}$ with weights $\left\{w_{\boldsymbol{z}}^{(j)}\right\}$ for the joint distribution $p_M(\boldsymbol{z}_1, \boldsymbol{z}_2, \boldsymbol{z}_3) = p_M^1(\boldsymbol{z}_1)p_M^2(\boldsymbol{z}_2)p_M^3(\boldsymbol{z}_3)$. This cloud can then be fed through the Gauss IOD method to produce a particle cloud $\left\{\boldsymbol{X}_2^{(j)}\right\}$ and weights $\left\{w_{\boldsymbol{x}_2}^{(j)}\right\}$ representing the uncertainty in the state at time t_2. This cloud can then be propagated backward and forward in time to obtain the state particle clouds $\left\{\boldsymbol{X}_1^{(j)}\right\}$ and $\left\{\boldsymbol{X}_3^{(j)}\right\}$ (with weights $\left\{w_{\boldsymbol{x}_1}^{(j)}\right\}$ and $\left\{w_{\boldsymbol{x}_3}^{(j)}\right\}$, respectively) at times t_1 and t_3, respectively.

We begin by following the same procedure as in Ref. [23] by noting that

$$I_i(\boldsymbol{x}_i, \boldsymbol{z}_i) = H(\boldsymbol{x}_i) - H(\boldsymbol{x}_i|\boldsymbol{z}_i)$$

where $H(\boldsymbol{x}_i)$ is the entropy of $\boldsymbol{x}_i$ and $H(\boldsymbol{x}_i|\boldsymbol{z}_i)$ is the conditional entropy of $\boldsymbol{x}_i$ given $\boldsymbol{z}_i$. Notice that given a particle-based distribution $p_M^i(\boldsymbol{z}_i) = \sum_{j=1}^{N_{\boldsymbol{z}}} w_{\boldsymbol{z}}^j \delta_{\boldsymbol{Z}_i^{(j)}}(\boldsymbol{z}_i)$, we can obtain an expression for $p_O^i(\boldsymbol{x}_i)$ by marginalization and using the basic MC integration approximation we obtain:

$$\begin{aligned} p_O^i(\boldsymbol{x}_i) &= \int p(\boldsymbol{x}_i|\boldsymbol{z}_i)p(\boldsymbol{z}_i)\mathrm{d}\boldsymbol{z}_i \\ &= E_{p_M^i(\boldsymbol{z}_i)}\left[p(\boldsymbol{x}_i|\boldsymbol{z}_i)\right] \\ &\simeq \sum_{j=1}^{N_{\boldsymbol{z}}} w_{\boldsymbol{z}}^j p(\boldsymbol{x}_i|\boldsymbol{Z}_i^{(j)}). \end{aligned}$$

Therefore, the entropy $H(\boldsymbol{x}_i)$ is given by:

$$H(\boldsymbol{x}_i) = -\int p_O^i(\boldsymbol{x}_i)\log p_O^i(\boldsymbol{x}_i)\mathrm{d}\boldsymbol{x}_i \tag{7}$$

$$= -\int \left(\sum_{j=1}^{N_{\boldsymbol{z}}} w_{\boldsymbol{z}}^j p(\boldsymbol{x}_i|\boldsymbol{Z}_i^{(j)})\right)\log\left(\sum_{l=1}^{N_{\boldsymbol{z}}} w_{\boldsymbol{z}}^l p(\boldsymbol{x}_i|\boldsymbol{Z}_i^{(l)})\right)\mathrm{d}\boldsymbol{x}_i \tag{8}$$

$$= -\sum_{j=1}^{N_{\boldsymbol{z}}} w_{\boldsymbol{z}}^j \int p(\boldsymbol{x}_i|\boldsymbol{Z}_i^{(j)})\log\left(\sum_{l=1}^{N_{\boldsymbol{z}}} w_{\boldsymbol{z}}^l p(\boldsymbol{x}_i|\boldsymbol{Z}_i^{(l)})\right)\mathrm{d}\boldsymbol{x}_i \tag{9}$$

$$= -\sum_{j=1}^{N_{\boldsymbol{z}}} w_{\boldsymbol{z}}^j E_{p(\boldsymbol{x}_i|\boldsymbol{Z}_i^{(j)})}\left[\log\left(\sum_{l=1}^{N_{\boldsymbol{z}}} w_{\boldsymbol{z}}^l p(\boldsymbol{x}_i|\boldsymbol{Z}_i^{(l)})\right)\right] \tag{10}$$

$$= -\sum_{j=1}^{N_{\boldsymbol{z}}} w_{\boldsymbol{z}}^j \sum_{k=1}^{N_{\boldsymbol{x}_i}} w_{\boldsymbol{x}_i}^k \log\left(\sum_{l=1}^{N_{\boldsymbol{z}}} w_{\boldsymbol{z}}^l p(\boldsymbol{X}_i^{k|j}|\boldsymbol{Z}_i^{(l)})\right), \tag{11}$$

where $\{\boldsymbol{X}_i^{k|j}\}$, $k = 1, \ldots, N_{\boldsymbol{x}_i}$, are $N_{\boldsymbol{x}_i}$ (assumed to be the same for all j) particles drawn from $p(\boldsymbol{x}_i|\boldsymbol{Z}^{(j)})$. Accounting for possible errors in the IOD solution and dynamic propagation and modeling them as additive noise, we have $\boldsymbol{x}_i = f_{2i}(g(\boldsymbol{z}_1, \boldsymbol{z}_2, \boldsymbol{z}_3)) + \boldsymbol{v}$, where $\boldsymbol{v}$ is an additive noise term with some density $p_{\boldsymbol{v}}(\boldsymbol{v})$. In this case, the particles $\boldsymbol{X}_i^{k|j}$ are drawn such that $\boldsymbol{X}_i^{k|j} = f_{2i}(g(\boldsymbol{Z}^{(j)})) + \mathcal{V}^{(k)}$ where $\mathcal{V}^{(k)} \sim p_{\boldsymbol{v}}(\boldsymbol{v})$ and where $\boldsymbol{Z}^{(j)} = (\boldsymbol{Z}_1^{(j)}, \boldsymbol{Z}_2^{(j)}, \boldsymbol{Z}_3^{(j)})$. However, an open area of research is the characterization of the error introduced in the IOD method. If one ignores these errors and assume that the state and the measurements are deterministically related, then we can simply propagate the particles $\boldsymbol{Z}^{(j)}$ through the various transformations without sampling. However, this will result in $N_{\boldsymbol{x}_i}$ identical particles $\{\boldsymbol{X}_i^{k|j}\}$ which can then be simply replaced by a single particle: $\{\boldsymbol{X}_i^{1|j}\} = f_{2i}(g(\boldsymbol{Z}^{(j)}))$.

Following a similar procedure as above one can show that

$$H(\boldsymbol{x}_i|\boldsymbol{z}_i) = -\sum_{j=1}^{N_{\boldsymbol{z}}} w_{\boldsymbol{z}}^j \sum_{k=1}^{N_{\boldsymbol{x}_i}} w_{\boldsymbol{x}_i}^k \log\left(p(\boldsymbol{X}_i^{k|j}|\boldsymbol{Z}^{(j)})\right) \tag{12}$$

One can now use Eq. (7) and Eq. (12) to compute the mutual information $I_i(\boldsymbol{X}_i, \boldsymbol{Z}_i)$. The overall index, as above, would then be $I_{\mathrm{tot}} = I_1 + I_2 + I_3$.

As a way to verify that the above equations, we consider a simple non-SSA problem for which we have a closed-form solution to compare against. Let some variable $\boldsymbol{x}$ be related to another variable $\boldsymbol{z}$ via the linear relationship: $\boldsymbol{x} = \boldsymbol{H}\boldsymbol{z} + \boldsymbol{v}$, where $\boldsymbol{z} \sim p_g(\boldsymbol{\mu}_{\boldsymbol{z}}, \boldsymbol{\Sigma}_{\boldsymbol{z}})$ and $\boldsymbol{v} \sim p_g(\boldsymbol{\mu}_{\boldsymbol{v}}, \boldsymbol{\Sigma}_{\boldsymbol{v}})$. One can use Eq. (2) and Eq. (6). It was found that the exact mutual information is given by 5.5254529. Fig. (6) shows the percentage error of the MC-based computation from the true MI value. While performing this analysis, it was found that the accuracy of the method depended heavily on the number of $\boldsymbol{z}$ particles and was very insensitive to the number of $\boldsymbol{x}$ sub-samples. In the figure we show the percentage error for a fixed number of $\boldsymbol{x}$ samples $N_{\boldsymbol{x}} = 100$. As can be seen the error is converging to zero with a larger number of samples.

4. SIMULATION RESULTS

In this simulation we compare the performance of the three OTOA solutions: UT-ID, MC-ID, and UT-MI. The MC-MI solution described at the end of last section will be tested in a future publication. For testing, we consider two objects, with identification numbers 0 and 1, in close proximity of each other. They both have the same orbital elements shown in Table 1 except for the value of true anomaly at the initial time. A set of two observations from the two RSos are collected at three different times. If we arbitrarily index the two measurements with 0 and 1, then the question is which sequence of tags are the correct ones? There are eight possible combinations of tags: 000 (i.e., observations with tags 0 at the three time instances are associated to one of the two objects and so on), 001, 010, 100, 011, 101, 110 and 111. The observations were tagged such that the two correct ones are 000 (all coming from RSO number 0) and

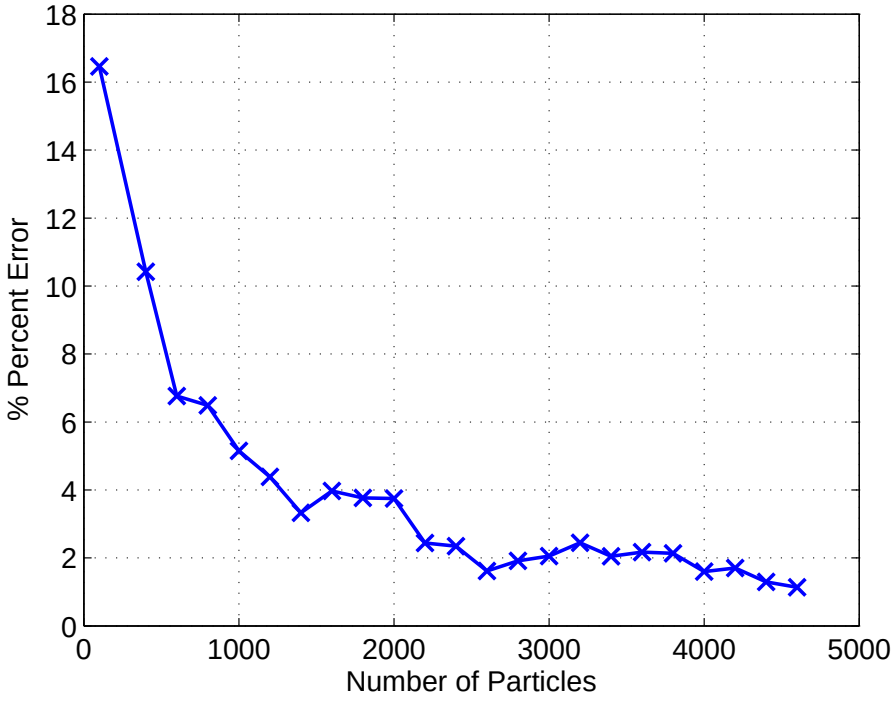

Fig. 6. Percentage error from true value vs. number of z samples in the MC-based MI computation for the linear Gaussian problem.

111 (all coming from RSO number 1). The criterion to evaluate the three OTOA methods will be how close the two objects are (in true anomaly) before the method being tested fails to return the two correct associations as the two most likely ones. Note that when a method "fails", while the correct associations may not be the one most highly ranked, they would rank very close to the top and as the separation distance between the two objects decreases further, the correct association are further from being top ranked and are more or less arbitrarily ranked as all solutions become indistinguishable. The sensor is assumed to be the Socorro, NM, ground sensor. Topocentric azimuth and elevation observations were collected at the rate of one observation every 20 minutes. The measurement noise is assumed to be Gaussian with an angular standard deviation of 2 arcsec for both azimuth and elevation.

Table 1. Orbital Elements of the True Orbit

Orbital Element	**Value**
Semi-major Axis (km)	26 600.
Eccentricity	0.2
Inclination (deg)	55.0
Perigee (deg)	-120.0
Right Ascension of the Ascending Node (deg)	-13.24

For each of the methods, the separation in angular anomaly between the two RSOs is steadily decreased until the method fails to report the correct associations as the most likely ones. The association picked by each method is the one that produces the maximum value of MI index or minimum ID index. As can be seen in Table 2, the UT-MI index is the one that performs the best among the ones considered in this paper. It has a resolution of 0.30997 degrees separation in true anomaly. It is important to note that while this number is relatively large ($\sim$ 1000 arcsec) compared to typical sensor resolutions, one has to remember that the OTOA problem being solved is far more challenging that the classical OTTA problem. With the methods proposed in this paper, and others to be considered in future work, one can sift through a collection of observations, not correlated to any object or uncorrelated track, and sort out which

ones belong to each other under minimal amount of knowledge of the observation statistics.

Table 2. True Anomaly Difference at which Association Method Fails

Method	True Anomaly Difference (deg)
UT-ID	0.64458
MC-ID	0.68182
UT-MI	0.30997

5. CONCLUSION

In this paper we considered four different information theoretic approaches to the observation-to-observation association problem. Three of these methods (UT-ID, MC-ID and UT-MI) have been implemented and the fourth (MC-ID) has been demonstrated to work on a non-SSA problem and is currently being developed for implementation in a GPU environment. It was shown that the mutual information method is the one that performed the best among the methods tested. Variations of these techniques exist and may prove to be even more accurate than the ones tested in this paper. It is generally conjectured that MC-based methods, on the long run, will prove to be more efficient as they more faithfully represent the underlying distributions. We further conjecture that the use of mutual information will be more accurate than divergence-based approaches as they more accurately capture degree of dependence between variables. This paper barely touches the tip of the iceberg, and its most basic goal is to induce the space community to further consider the use of information theoretic measures for solving the various data association problems (OTOA, OTTA, and TTTA).

REFERENCES

[1] J. Stauch, M. Jah, J. Baldwin, T. Kelecy, and K. A. Hill, "Mutual Application of Joint Probabilistic Data Association, Filtering, and Smoothing Techniques for Robust Multiple Space Object Tracking," *AIAA/AAS Astrodynamics Specialist Conference*, American Institute of Aeronautics and Astronautics, August 2014.

[2] J. Roecker, "A Class of Near Optimal JPDA Algorithms," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. AES-30, April 1994, pp. 504–510.

[3] S. Coraluppi and C. Carthel, "Multi-Stage Multiple-Hypothesis Tracking," *Journal of Advances in Information Fustion*, Vol. 6, June 2011.

[4] K. J. DeMars, M. K.Jah, and P. W. Schumacher, "Initial Orbit Determination Using Short-Arc Angle and Angle Rate Data," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 48, July 2012, pp. 2628–2637.

[5] N. Singh, J. T. Horwood, J. M. Aristoff, A. B. Poore, C. Sheaff, and M. K.Jah, "Multiple Hypothesis Tracking (MHT) for Space Surveillance: Results and Simulation Studies," *Advanced Maui Optical and Space Surveillance Technologies Conference*, 2013.

[6] S. Oh, S. Russell, and S. Sastry, "Markov Chain Monte Carlo Data Association for Mutli-Target Tracking," *IEEE Transactions on Automatic Control*, Vol. 54, No. 3, 2009, pp. 481–497.

[7] Y. Bar-Shalom, F. Daum, and J. Huang, "The Probabilistic Data Association Filter: Estimation in the Presence of Measurement Origin Uncertainty," *IEEE Control Systems Magazine*, December 2009, pp. 82–100.

[8] J. S. McCabe and K. J. DeMars, "Particle Filter Methods for Space Object Tracking," *AIAA/AAS Astrodynamics Specialist Conference*, San Diego, CA, August 2014.

[9] J. Collins and J. Uhlmann, "Efficient Gating in Data Association with Multivariate Distributed States," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 28, July 1992, pp. 909–916.

[10] K. Hill, K. T. Alfriend, and C. Sabol, "Covariance-based Uncorrelated Track Association," *AIAA/AAS Astrodynamics Specialist Conference* (AIAA, ed.), Vol. 7211, Honolulu, HI, August 2008.

[11] K. Hill, C. Sabol, and K. Alfriend, "Comparison of Covariance Based Track Association Approaches Using Simulated Radar Data," *The Journal of the Astronautical Sciences*, Vol. 59, No. 1-2, 2012, pp. 281–300, 10.1007/s40295-013-0018-1.

[12] R. Linares, V. Kumar, P. Singla, and J. L. Crassidis, "Information theoretic space object data association methods using an adaptive gaussian sum filter," *Advances in the Astronautical Sci-ences*, Vol. 140, 2011, pp. 665–680.

[13] A. Alempijevic, S. Kodagoda, and G. Dissanayake, "Mutual information based data association," *Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 2009 5th International Conference on*, IEEE, 2009, pp. 97–102.

[14] I. I. Hussein, C. W. T. Roscoe, M. P. Wilkins, and P. W. Schumacher, Jr., "Probabilistic Admissibility in Angles-Only Initial Orbit Determination," *Proceedings of the 24th International Symposium on Space Flight Dynamics*, Laurel, MD, May 5–9 2014.

[15] P. R. Escobal, *Methods of Orbit Determination*. John Wiley & Sons, Inc., 1965. (Reprint with corrections by Krieger Publishing Company, 1976).

[16] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley-Interscience, 1991.

[17] A. Rényi, "On measures of entropy and information," *Proceedings of the fourth Berkeley Symposium on Mathematics, Statistics and Probability 1960*, 1961, pp. 547–561.

[18] A. O. Hero, D. Castañón, D. Cochran, and K. Kastella, *Foundations and Applications of Sensor Management*. Springer, 2007.

[19] K. J. DeMars and M. K. Jah, "Evaluation of the information content of observations with application to sensor management for orbit determination," *AAS/AIAA Astrodynamics Specialist Conference*, 2011.

[20] K. Kampa, E. Hasanbelliu, and J. C. Principe, "Closed-form Cauchy-Schwarz PDF divergence for mixture of Gaussians," *Proc. of the 2011 International Joint Conference on Neural Networks (IJCNN 2011)*, July–August 2011, pp. 2578–2585.

[21] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society, Series B*, Vol. 39, 1977, pp. 1–38. 1.

[22] M. A. T. Figueiredo and A. K. Jain, "Unsupervised Learning of Finite Mixture Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, 2002, pp. 381–396. 3.

[23] G. M. Hoffmann and C. J. Tomlin, "Mobile Sensor Network Control Using Mutual Information Methods and Particle Filters," *IEEE Transactions on Automatic Control*, Vol. 55, January 2010, pp. 32–47.