

# INFORMATION THEORETIC SPACE OBJECT DATA ASSOCIATION METHODS USING AN ADAPTIVE GAUSSIAN SUM FILTER

Richard Linares\*, Vishwajeet Kumar†, Puneet Singla‡ and John L. Crassidis§

This paper shows an approach to improve the statistical validity of orbital estimates and uncertainties as well as a method of associating measurements with the correct space objects. The approach involves using an adaptive Gaussian mixture solution to the Fokker-Planck-Kolmogorov equation for its applicability to the space object tracking problem. The Fokker-Planck-Kolmogorov equation describes the time-evolution of the probability density function for nonlinear stochastic systems with Gaussian inputs, which often results in non-Gaussian outputs. The adaptive Gaussian sum filter provides a computationally efficient and accurate solution for this equation, which captures the non-Gaussian behavior associated with these nonlinear stochastic systems. This adaptive filter is designed to be scalable, relatively efficient for solutions of this type, and thus is able to handle the nonlinear effects which are common in the estimation of resident space object orbital states. The main purpose of this paper is to develop a technique for data association based on information theoretic approaches that are compatible with the adaptive Gaussian sum filter. The adaptive filter and corresponding measurement association methods are evaluated using simulated data in realistic scenarios to determine their performance and feasibility.

## INTRODUCTION

Commercial and military analyst routinely use both observation data and forward orbit propagation models to assist space surveillance. The U.S. Air Force collects the information for the purpose of space surveillance through a global network of radars and optical sensors to maintain a map of over 20,000 RSOs, both active and inactive. The information on the RSOs is stored in a catalog. However, because of the large number of resident space objects (RSOs) and the limited number of sensors available to track these objects, it is impossible to maintain persistent surveillance on all objects, and, therefore there is inherent uncertainty and latency in the catalog. Basically, the observational data is limited in terms of the kind and frequency of observations that can be taken and may only provide access to limited information about the RSO orbit. The recent IRIDIUM satellite collision clearly illustrates this fact.

On other hand, forward orbit propagation models are based upon equations of motion and involve modeling of space weather effects that produce disturbances on RSO motion. Examples of

---

\*Graduate Student, Department of Mechanical & Aerospace Engineering, University at Buffalo, State University of New York, Amherst, NY 14260-4400. Email: linares2@buffalo.edu.

†Graduate Student, Department of Mechanical & Aerospace Engineering, University at Buffalo, State University of New York, Amherst, NY 14260-4400. Email: kumarvis@buffalo.edu.

‡Assistant Professor, Department of Mechanical & Aerospace Engineering, Richard Linares State University of New York, Amherst, NY 14260-4400. Email: psingla@buffalo.edu.

§Professor, Department of Mechanical & Aerospace Engineering, University at Buffalo, State University of New York, Amherst, NY 14260-4400. Email: johnc@buffalo.edu.

such disturbances include solar radiation “wind,” atmospheric drag from air molecules, and higher-order gravity potentials due to a non-spherical Earth. The orbit of a RSO leads to different effects. For example, low-Earth orbiting (LEO) objects are more prone to atmospheric drag effects while higher-orbiting objects, such as geosynchronous-Earth orbiting (GEO) ones, are more prone to solar wind effects. Higher-order gravity effects produce a drift of nodes, such as the right ascension of the ascending node. Generally, a nominal model for orbit dynamics, including various space weather effects, is considered which involves errors due to truncation of the non-spherical gravity model and errors in model parameters such as incorrect values for RSO geometry, inertia properties, atmospheric drag or solar radiation pressure coefficient. These factors cause overall accuracy to degrade as the orbit model states evolve.

Nevertheless, commercial and military analysts need to make important decisions daily with this limited information with large uncertainties due to model and sensor observation errors. The rapid estimation of the orbit and identity of an RSO and the accurate assessment of confidence in that estimate can be critically important to the space situational awareness community and the warfighter. The fusion of observational data with orbit state models promises to provide greater understanding of physical phenomenon than either approach alone can achieve. However, one needs to associate the observation data with a particular RSO before fusing the orbit state model with observational data. Data association (DA) involves the matching of sensor measurements to specific tracks or targets. If measurement-to-target matching fails, then proper state estimation will not be possible. For large propagation and large initial errors the orbital uncertainty will likely be non-Gaussian, so non-Gaussian DA-based methods are required.

A critical step for successful DA involves the accurate characterization of orbit state probability density function (pdf). For stochastic continuous dynamic systems the exact evolution of the state pdf is given by the Fokker-Planck-Kolmogorov Equation (FPKE).<sup>1</sup> Park et al.<sup>2</sup> have discussed the use of the FPKE to analyze spacecraft trajectory statistics by incorporating higher-order Taylor series terms in the spacecraft dynamics. Analytical solutions for the FPKE exist only for a stationary pdf and are restricted to a limited class of dynamical systems.<sup>1,3</sup> Recently Terejanu et al. have developed an Adaptive Gaussian Sum Filter (AGSF) approach<sup>4,5</sup> for accurate uncertainty propagation through nonlinear dynamical systems while incorporating the solution to the FPKE. This approach has been successfully applied to propagate initial orbit uncertainty through a low-Earth orbit with nonconservative atmospheric drag<sup>6</sup> and has also been applied to the spacecraft attitude estimation problem.<sup>7,8</sup>

Particle filter-based methods are highly useful for DA involving non-Gaussian problems, which are typical for RSO tracking, but they have the significant disadvantage of being computational expensive. This is because particle filters are based on Monte Carlo sampling approaches. Even though methods have been proposed to reduce the computational load, such as replacing the standard importance sampling with a Markov Chain Monte Carlo (MCMC) method,<sup>9</sup> they generally are still not viable for actual RSO tracking applications. As illustrated in References 4, 5, 6, 8, the AGSF algorithm can produce the entire non-Gaussian pdf with much less computations than what is typically required in particle filters. To develop a successively refining pdf representation, it is important to define a metric for the data association error, so that improvement due to refinements can be assessed.

In this paper the metric for modeling error will be in terms of *information geometry*, and different information theoretic methods will be applied to the orbital data association problem. The combination of the AGSF with information theoretic approaches for data association is well suited for orbit

problems where the errors may be highly non-Gaussian. In References 10, 11, 12 an information theoretic divergence measure has been proposed to measure the confidence for fusion and tracking that has been lacking earlier in the literature. In Reference 6, a new approach is presented for data association for resident space object tracking which combines an adaptive Gaussian sum filter with the Kullback-Leibler (KL) divergence measure for effective data association. However, the efficient computation of the KL divergence poses significant computational challenges.

In this paper, we first investigate many information theoretic measures for DA and assess their relative performance and ease of computation of the measure. Once a computationally attractive divergence measure is found, a key question is how to compute the data association error. Two different approaches are proposed to quantify the data association error while making use of information theoretic measures. The first approach corresponds to computing the distance between the estimated likelihood distribution and the actual sensor distribution, derived from the known sensor statistics while the second approach involves the computation of mutual information content between the measurement data and the state vector.

The organization of this paper is as follows. First a review of the FPKE is provided. Then the AGSF is summarized followed by a review of information theoretic data association measures. Then two approaches for information theoretic data association are discussed. Finally, simulation results are presented to show the effectiveness of the proposed concepts.

## THE FOKKER-PLANCK-KOLMOGOROV EQUATION

In the standard orbit dynamic model, the orbit state variables (e.g. position and orientation of the RSO) are assumed to be deterministic quantities. Instead of solving for the point position and orientation of the RSO, we are interested in the probability distribution for their values due to uncertainty in input parameters, initial conditions and space weather effects. Thus, the position and orientation of the RSO are assumed to be a *random vector*,  $\mathbf{x}(t)$ , whose time evolution is given by the following stochastic differential equation:

$$\dot{\mathbf{x}} = \mathbf{f}(t, \mathbf{x}) + \mathbf{g}(t, \mathbf{x})\Gamma(t), \quad \mathbf{x}(t_0) = \bar{\mathbf{x}}_0 \quad (1)$$

The random function  $\Gamma(t)$  represents stochastic forcing terms that might be added to account for any modeling errors. These random functions may be modeled as a Gaussian process with specified correlation function  $\mathbf{Q}(t)$ . Finally, the nominal initial position and orientation is given by  $\bar{\mathbf{x}}_0$ , which may also be uncertain. The total uncertainty associated with the state vector  $\mathbf{x}(t)$  is characterized by the pdf  $p(\mathbf{x}(t))$ . A key idea is to *replace the time evolution of state vector  $\mathbf{x}(t)$  by the time evolution of the pdf  $p(\mathbf{x}(t))$* . By computing full probability density functions, we can better monitor the space-time evolution of the uncertainty, represent multi-modal distributions, incorporate complex prior models, and exploit Bayesian belief propagation.

The exact time evolution of the state pdf is given by the Fokker-Planck-Kolmogorov equation (FPKE):<sup>1</sup>

$$\frac{\partial}{\partial t}p(t, \mathbf{x}) = \mathcal{L}_{FP}p(t, \mathbf{x}) = \frac{\partial \mathbf{f}^T(t, \mathbf{x})p(t, \mathbf{x})}{\partial \mathbf{x}} + \frac{1}{2}Tr \left( \mathbf{g}(t, \mathbf{x}(t))\mathbf{Q}\mathbf{g}^T(t, \mathbf{x}(t)) \frac{\partial^2 p}{\partial \mathbf{x} \partial \mathbf{x}^T} \right) \quad (2)$$

The FPKE is a formidable equation to solve because of the following issues: 1) *positivity* of the pdf, 2) *normalization* constraint of the pdf:  $\int_{\mathbb{R}^n} p(t, \mathbf{x})d\mathbf{x} = 1$ , and 3) *no fixed solution domain*: how to impose boundary conditions in a finite region and restrict numerical computation to regions where  $p > \sim 10^{-9}$ .

Analytical solutions for the FPKE exist only for a stationary pdf and are restricted to a limited class of dynamical systems.<sup>1,3</sup> Numerical approximations to solve the FPKE,<sup>13,14,15,16,17</sup> generally using the variational formulation of the problem, suffer from the “*curse of dimensionality*.” Recently, an Adaptive Gaussian Sum Filter (AGSF) method<sup>4</sup> has been developed to accurately solve the FPKE. The key idea of the AGSF is to approximate the state pdf by a finite sum of Gaussian density functions whose mean and covariance are propagated from one time-step to the next using linear theory. The weights of the Gaussian kernels are updated at every time-step by minimizing the two-norm of FPKE error.<sup>4</sup> This methodology effectively decouples a large uncertainty propagation problem into many small problems. As a consequence, the solution algorithm can be parallelized on most high performance computing systems. Finally, a Bayesian framework can be used on the AGSF structure to assimilate (noisy) observational data with model forecasts.<sup>5</sup>

### Adaptive Gaussian Sum Filter

This subsection briefly summarizes the AGSF approach; details can be found in Reference 18,4,19. The Gaussian mixture model approximation (denoted by the caret  $\hat{\cdot}$ ) of the forecast pdf can be written as

$$\hat{p}(t, \mathbf{x}(t)) = \sum_{i=1}^N w_i(t) \mathcal{N}(\mathbf{x}(t) | \boldsymbol{\mu}_i(t), \mathbf{P}_i(t)) \quad (3)$$

In this equation,  $\boldsymbol{\mu}_i(t)$  and  $\mathbf{P}_i(t)$  represent the mean and covariance of the  $i^{th}$  component of the Gaussian pdf,  $\mathcal{N}(\mathbf{x}(t) | \boldsymbol{\mu}_i(t), \mathbf{P}_i(t))$ , respectively, and  $w_i$  denotes the amplitude of the  $i^{th}$  Gaussian in the mixture. The positivity and normalization constraint on  $\hat{p}(t, \mathbf{x})$  lead to the following conditions at every time-step:

$$\sum_{i=1}^N w_i(t) = 1 \text{ and } w_i(t) \geq 0, \forall i \quad (4)$$

In Reference 20, it is shown that because all the components of the mixture pdf in Eq. (3) are Gaussian, only estimates of their mean and covariance need to be maintained. These estimates can be propagated using linear system propagation methods such as the propagation part of the extended Kalman filter (EKF) or unscented Kalman filter (UKF):

$$\dot{\boldsymbol{\mu}}_i(t) = \mathbf{f}(t, \boldsymbol{\mu}_i(t)) \quad (5a)$$

$$\dot{\mathbf{P}}_i(t) = \mathbf{A}_i(t)\mathbf{P}_i(t) + \mathbf{P}_i(t)\mathbf{A}_i^T(t) + \mathbf{g}(t, \boldsymbol{\mu}_i(t))\mathbf{Q}(t)\mathbf{g}^T(t, \boldsymbol{\mu}_i(t)) \quad (5b)$$

where  $\mathbf{A}_i(t) = \left. \frac{\partial \mathbf{f}(t, \mathbf{x}_k)}{\partial \mathbf{x}} \right|_{\mathbf{x}=\boldsymbol{\mu}_i}$ . The propagation equations for the mean and covariance using an UKF are given as

$$\mathbf{X}_i = [\boldsymbol{\mu}_i \dots \boldsymbol{\mu}_i] + \sqrt{c} [\mathbf{0} \ \mathbf{A} \ -\mathbf{A}], \quad c = \alpha^2(n + \kappa) \quad (6a)$$

$$\dot{\boldsymbol{\mu}}_i = \mathbf{f}(\mathbf{X}_i)w_m \quad (6b)$$

$$\dot{\mathbf{P}}_i = \mathbf{X}_i \mathbf{W} \mathbf{f}^T(\mathbf{X}_i) + \mathbf{f}(\mathbf{X}_i) \mathbf{W} \mathbf{X}_i^T + \mathbf{g}(t, \boldsymbol{\mu}_i) \mathbf{Q} \mathbf{g}^T(t, \boldsymbol{\mu}_i) \quad (6c)$$

where  $\mathbf{X}_i$  is the  $n \times 2n + 1$  matrix of sigma-points and the weight vector,  $\mathbf{w}_m$ , and weight matrix,  $\mathbf{W}$ , are given by

$$\lambda = \alpha^2(n + \kappa) - n \quad (7a)$$

$$W_0^{(mean)} = \frac{\lambda}{n + \lambda} \quad (7b)$$

$$W_0^{(cov)} = \frac{\lambda}{(n + \lambda) + (1 - \alpha^2 + \beta)} \quad (7c)$$

$$W_j^{(mean)} = \frac{1}{(2(n + \lambda))}, j = 1, \dots, 2n \quad (7d)$$

$$W_j^{(cov)} = \frac{1}{(2(n + \lambda))}, j = 1, \dots, 2n \quad (7e)$$

$$\mathbf{w}_m = \left[ W_0^{(mean)} \dots W_{2n}^{(mean)} \right]^T \quad (7f)$$

$$\mathbf{W} = (\mathbf{I} - [w_m \dots w_m]) \times \text{diag} \left( W_0^{(cov)} \dots W_{2n}^{(cov)} \right) \times (\mathbf{I} - [w_m \dots w_m])^T \quad (7g)$$

The constants  $\alpha$ ,  $\beta$ , and  $\kappa$  in the above equations are constant parameters of the method. The spread of sigma points is determined by  $\alpha$  and is typically a small positive value, i.e.  $1 \times 10^{-4} \leq \alpha \leq 1$ . The parameter  $\beta$  is used to incorporate prior knowledge of the distribution, and is optimally chosen as 2 for a normal distribution. The parameter  $\kappa$  can be used to exploit knowledge of the distributions higher moments, and for higher order systems choosing  $\kappa = 3 - n$  minimizes the mean-squared-error up to the fourth order. The use of the UKF in the mixture model is especially advantageous since it does not require the computation of the Jacobian matrix  $\mathbf{A}_k$ .

Notice that the weights  $w_i$  of the Gaussian components are not known at time  $t$  and must be computed as part of the solution process. To determine the unknown weights, the two-norm of FPKE error is minimized subject to positivity and normality constraints of Eq. (4). This leads to the following convex optimization problem:<sup>21</sup>

$$\min_{w_i(t')} J = \frac{1}{2} \mathbf{w}(t')^T \mathbf{M}_c \mathbf{w}(t') + \mathbf{w}(t')^T \mathbf{N}_c \mathbf{w}(t), \text{ s.t. } \mathbf{1}_{N \times 1}^T \mathbf{w}(t') = 1, \mathbf{w}(t') \geq \mathbf{0}_{N \times 1} \quad (8)$$

Here  $\mathbf{w}(t')$  represents a vector of unknown weights at time  $t' = t + \Delta t$  while  $\mathbf{w}(t)$  represents the vector of known initial weights at time  $t$ ,  $\mathbf{1}_{N \times 1} \in \mathbb{R}^{N \times 1}$  is a vector of ones,  $\mathbf{0}_{N \times 1} \in \mathbb{R}^{N \times 1}$  is a vector of zeros, and the matrices  $\mathbf{M}_c \in \mathbb{R}^{N \times N}$  and  $\mathbf{N}_c \in \mathbb{R}^{N \times N}$  are given as:

$$M_{c_{ij}} = \frac{1}{\Delta t^2} |2\pi(\mathbf{P}_i + \mathbf{P}_j)|^{-1/2} \exp \left[ -\frac{1}{2} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T \times (\mathbf{P}_i + \mathbf{P}_j)^{-1} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j) \right] \quad (9a)$$

$$N_{c_{ij}} = \frac{1}{\Delta t} p_{g_i} \int_V \left( \frac{\partial p_{g_j}^T}{\partial \boldsymbol{\mu}_j} \dot{\boldsymbol{\mu}}_j + \text{Tr} \left[ \frac{\partial p_{g_j}}{\partial \mathbf{P}_j} \dot{\mathbf{P}}_j \right] - \frac{1}{\Delta t} p_{g_j} + \frac{\partial p_{g_j}^T}{\partial \mathbf{x}} \mathbf{f}(t, \mathbf{x}) + p_{g_j} \text{Tr} \left[ \frac{\partial \mathbf{f}(t, \mathbf{x})}{\partial \mathbf{x}} \right] - \frac{1}{2} \text{Tr} \left[ \mathbf{g}(t, \mathbf{x}) \mathbf{Q} \mathbf{g}^T(t, \mathbf{x}) \frac{\partial^2 p_{g_j}}{\partial \mathbf{x} \partial \mathbf{x}^T} \right] \right) d\mathbf{x}, p_{g_j} = \mathcal{N}(\mathbf{x}(t) | \boldsymbol{\mu}_j(t), \mathbf{P}_j(t)) \quad (9b)$$

In Reference 4, it is shown that the matrix  $\mathbf{M}_c$  is positive semi-definite and the cost function  $J$  is lower bounded. As a consequence of this, the aforementioned optimization problem can be posed as a convex optimization problem which is guaranteed to have a unique solution.<sup>22</sup>

Furthermore, Bayesian tools are employed to incorporate observational data. Given a prediction of the state variable  $\mathbf{x}_k$ , standard Bayesian algorithms assume a sensor model  $\mathbf{h}$  to obtain the measurement

$$\mathbf{y}_k = \mathbf{h}(t_k, \mathbf{x}_k) + \mathbf{v}_k, \quad \mathbf{x}_k \equiv \mathbf{x}(t_k) \quad (10)$$

Here  $\mathbf{v}_k$  is the measurement noise with prescribed likelihood function  $p(\mathbf{y}_k|\mathbf{x}_k)$ . Using the dynamic state evolution sketched above as a forecasting tool, the state pdf can be updated using Bayes' rule on the arrival of a measurement:

$$p(\mathbf{x}_k|\mathbf{Y}_k) = \frac{p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{Y}_{k-1})}{\int p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{Y}_{k-1})d\mathbf{x}_k} \quad (11)$$

Here,  $p(\mathbf{x}_k|\mathbf{Y}_{k-1})$  represents the prior pdf (the computed solution of the FPKE),  $p(\mathbf{y}_k|\mathbf{x}_k)$  is the likelihood that we observe  $\mathbf{y}_k$  given the state  $\mathbf{x}_k$ , and  $p(\mathbf{x}_k|\mathbf{Y}_k)$  represents the posterior pdf of  $\mathbf{x}_k$ . While both the state and the covariance matrix are updated using the EKF or UKF measurement update equations, the weights are updated using Bayes' rule:

$$w_{k+1|k+1}^i = \frac{w_{k+1|k}^i \beta_k^i}{\sum_{i=1}^N w_{k+1|k}^i \beta_k^i}, \quad \beta_k^i = p(\mathbf{y}_k|\boldsymbol{\mu}_{k+1|k}^i) \quad (12)$$

## INFORMATION THEORETIC DATA ASSOCIATION

In this section, an information-theoretic basis is developed to quantify the uncertainty metrics for data association. The central idea is to frame the data association metrics in terms of information theory. If one agrees that the pdf of a random variable is a representation of uncertainty, then it is meaningful to quantify the goodness of estimates in terms of the information theoretic metrics. The idea of using a pdf-based metric is compelling given the fact that both the likelihood and posterior pdf are generally non-Gaussian in nature.

Shannon's entropy is the most popular choice for measuring information value contained in a random variable,  $\mathbf{x}$ , from its pdf  $p(\mathbf{x})$ :

$$H(\mathbf{x}) = - \int p(\mathbf{x}) \log p(\mathbf{x}) d\mathbf{x} = E_{p(\mathbf{x})} \{ - \log p(\mathbf{x}) \} \quad (13)$$

Entropy is a nonnegative function and  $H(\mathbf{x}) = 0$  if  $p(\mathbf{x}) = 0$ . It reaches its maximum value for a uniform distribution, which is the most unstructured distribution over a given domain, and its minimum value for a Dirac delta distribution, which is the most structured distribution (the random variable  $\mathbf{x}$  is completely known in this case). Therefore, entropy can be interpreted as the degree of diversity for a realization of an outcome of a random variable. Fundamentally, entropy depends on the pdf of a random variable, not the values taken by the random variable and it can be evaluated within a probabilistic framework. Also, if  $p(\mathbf{x})$  is assumed to be Gaussian with covariance matrix  $\Sigma$ , then the entropy can be analytically computed as:

$$\int_{-\infty}^{\infty} p(\mathbf{x}) \ln p(\mathbf{x}) d\mathbf{x} = \ln \left( \sqrt{|2\pi e \Sigma|} \right) \quad (14)$$

It is well-known that Shannon's entropy is not a true metric because it is non-symmetric and it does not satisfy the triangle inequality. Kullback<sup>23</sup> and Kolmogorov<sup>24,25</sup> extended the notion of Shannon's entropy to measure the distance between two density functions by computing the relative entropy. The Kullback-Leibler (KL) divergence measure, or relative entropy, describes the information geometry for the space of density functions and allows one to measure the degree of deviation of one distribution from another. Considering two probability distributions  $p(\mathbf{x})$  and  $q(\mathbf{x})$ , the Kullback-Leibler divergence measure  $\mathcal{D}_{\text{KL}}(p, q)$  is defined as

$$\mathcal{D}_{\text{KL}}(p, q) = \int_{-\infty}^{\infty} p(\mathbf{x}) \ln \frac{p(\mathbf{x})}{q(\mathbf{x})} d\mathbf{x} = \int_{-\infty}^{\infty} p(\mathbf{x}) \ln p(\mathbf{x}) d\mathbf{x} - \int_{-\infty}^{\infty} p(\mathbf{x}) \ln q(\mathbf{x}) d\mathbf{x} \quad (15)$$

The first term in the aforementioned expression for the KL-divergence is simply the Shannon entropy associated with true pdf  $p$  and is the measure of uncertainty in  $p$  while the second term is the measure of uncertainty in  $q$  relative to the true pdf  $p$ . The divergence measure is non-symmetric, i.e.,  $\mathcal{D}_{\text{KL}}(p, q) \neq \mathcal{D}_{\text{KL}}(q, p)$ . If both  $p(\mathbf{x})$  and  $q(\mathbf{x})$  are assumed to be Gaussian with distributions  $\mathcal{N}(\mu_p, \Sigma_p)$  and  $\mathcal{N}(\mu_q, \Sigma_q)$ , respectively, then the KL-divergence has a closed-form expression given as:

$$\mathcal{D}_{\text{KL}}(p, q) = \frac{1}{2} \left[ \log\left(\frac{\|\Sigma_q\|}{\|\Sigma_p\|}\right) + \text{Trace}(\Sigma_q^{-1}\Sigma_p) - d + (\mu_p - \mu_q)^T \Sigma_q^{-1} (\mu_p - \mu_q) \right] \quad (16)$$

The KL divergence measure is non-negative and like the Shannon entropy, it does not satisfy the triangle inequality. Another important measure of distance between two pdf is provided by the Bhattacharyya divergence measure, denoted by  $D_{\text{B}}$ :<sup>26</sup>

$$D_{\text{B}}(p(\mathbf{x}), q(\mathbf{x})) = -2 \ln \left( \int \sqrt{p(\mathbf{x})q(\mathbf{x})} d\mathbf{x} \right) \quad (17)$$

Notice that unlike the KL divergence,  $D_{\text{B}}(p(\mathbf{x}), q(\mathbf{x}))$  is symmetric and satisfies the triangle inequality which makes it a proper metric.  $D_{\text{B}}(p(\mathbf{x}), q(\mathbf{x}))$  vanishes if and only if  $p(\mathbf{x}) = q(\mathbf{x})$  almost everywhere. If both  $p(\mathbf{x})$  and  $q(\mathbf{x})$  are assumed to be Gaussian with distributions  $\mathcal{N}(\mu_p, \Sigma_p)$  and  $\mathcal{N}(\mu_q, \Sigma_q)$  respectively, the Bhattacharyya-divergence has a closed-form expression given as

$$\mathcal{D}_{\text{B}}(p(\mathbf{x}), q(\mathbf{x})) = \left[ \frac{1}{2} \ln\left(\frac{\|\Sigma\|}{\|\Sigma_p \Sigma_q\|}\right) + \frac{1}{8} (\mu_p - \mu_q)^T \Sigma^{-1} (\mu_p - \mu_q) \right] \quad (18)$$

where  $\Sigma \equiv \frac{\Sigma_p + \Sigma_q}{2}$ . Furthermore, the Bhattacharyya coefficient,  $\rho = \exp\{-D_{\text{B}}\}$ , is a very popular statistical measure which provides upper and lower bounds on the probability of classification error.<sup>27</sup> The optimal Bayes classification error between the two classes is bounded by the following expression:

$$\epsilon \leq \frac{1}{2} \exp\{-D_{\text{B}}(p(\mathbf{x}), q(\mathbf{x}))\} \quad (19)$$

This is an important property for the Bhattacharyya coefficient since such a bound is not available for other information theoretic measures.

In 1955, Alfred Renyi introduced a generalized form of Shannon's entropy which leads to a generalized divergence measure known as the Renyi-divergence measure,<sup>28</sup> is given as

$$D_{\text{R}}(p(\mathbf{x}), q(\mathbf{x})) = \frac{1}{1-\alpha} \ln \left( \int [p(\mathbf{x})]^\alpha [q(\mathbf{x})]^{1-\alpha} d\mathbf{x} \right) \quad \alpha \neq 1 \quad 0 < \alpha < 1 \quad (20)$$

This divergence measure defines a spectral of divergence measures that are parameterized by  $\alpha$ . One can notice that the Bhattacharya distance corresponds to  $\alpha = \frac{1}{2}$  and it can be shown that the KL distance corresponds to  $\alpha = 1$ . An interesting measure is Renyi's quadratic entropy, where  $\alpha = 2$ , given by

$$D_{\alpha=2}(p(\mathbf{x}), q(\mathbf{x})) = -\frac{1}{2} \ln \left( \int \frac{p(\mathbf{x})^2}{q(\mathbf{x})} d\mathbf{x} \right) \quad (21)$$

A number of researchers have conducted systematic studies of the effect of Renyi's entropy order,  $\alpha$ , and have found that although Renyi's entropy provides similar performance to Shannon's entropy, the quadratic entropy (when  $\alpha = 2$ ) is preferred over Shannon's entropy because of computational benefits.<sup>29</sup> Renyi's quadratic entropy provides nearly identical values to Shannons entropy values with exponentially reduced computational complexity.<sup>29</sup> The computational cost of calculating Renyi's entropy  $H(X)$  and in turn the Renyi's divergence measure is now reduced to  $O(N^2)$ . The Renyi entropy measure between a gaussian mixture and one gaussian component can be written in closed form. To illustrate this lets consider a gaussian mixture pdf given by

$$p(\mathbf{x}) = \sum_{i=1}^N w_i \mathcal{N}(\mathbf{a}_i, \mathbf{C}_i) \quad (22)$$

where  $\mathbf{a}_i$  and  $\mathbf{C}_i$  are the  $i^{\text{th}}$  mean and covariance, respectively. Then assuming that  $q(\mathbf{x}) = \mathcal{N}(\mathbf{b}, \mathbf{M})$ , we can write the second term in Renyi's quadratic entropy expression as

$$q(\mathbf{x})^{-1} = |2\pi\mathbf{M}| \mathcal{N}(\mathbf{b}, -\mathbf{M}) \quad (23)$$

The distance between the Gaussian mixture pdf  $p(\mathbf{x})$  and the Gaussian pdf is given by

$$D_{\alpha=2}(p(\mathbf{x}), q(\mathbf{x})) = -\frac{1}{2} \ln \left( \int \frac{p(\mathbf{x})^2}{q(\mathbf{x})} d\mathbf{x} \right) \quad (24)$$

Multiplying this expression out gives

$$D_{\alpha=2}(p(\mathbf{x}), q(\mathbf{x})) = -\frac{1}{2} \ln \left( \int \sum_{i=1}^N \sum_{j=1}^N w_i w_j |2\pi\mathbf{M}| \mathcal{N}(\mathbf{x} - \mathbf{a}_i, \mathbf{C}_i) \mathcal{N}(\mathbf{x} - \mathbf{a}_j, \mathbf{C}_j) \mathcal{N}(\mathbf{x} - \mathbf{b}, -\mathbf{M}) d\mathbf{x} \right) \quad (25)$$

For two Gaussian pdfs the following property is seen:

$$\int_{-\infty}^{\infty} \mathcal{N}(\mathbf{y} - \mathbf{a}_i, \mathbf{C}_i) \mathcal{N}(\mathbf{y} - \mathbf{a}_j, \mathbf{C}_j) d\mathbf{y} = \mathcal{N}(\mathbf{a}_i - \mathbf{a}_j, \mathbf{C}_i + \mathbf{C}_j) \quad (26)$$

Using this property Renyi's quadratic entropy can be written in closed form as

$$D_{\alpha=2}(p(\mathbf{x}), q(\mathbf{x})) = -\frac{1}{2} \ln \left[ \sum_{i=1}^N \sum_{j=1}^N w_i w_j \left( \frac{|2\pi W|^{1/2} |2\pi \mathbf{M}|^{1/2}}{\prod_l |2\pi \mathbf{C}_l|^{1/2}} \Omega_{ij} \Omega_{iM} \Omega_{jM} \right) \right] \quad (27)$$

where

$$\Omega_{ij} = \exp\left\{-\frac{1}{2} (\mathbf{a}_i - \mathbf{a}_j)^T \mathbf{B}_{ij} (\mathbf{a}_i - \mathbf{a}_j)\right\} \quad (28a)$$

$$\Omega_{iM} = \exp\left\{-\frac{1}{2} (\mathbf{b} - \mathbf{a}_i)^T \mathbf{B}_{iM} (\mathbf{b} - \mathbf{a}_i)\right\} \quad (28b)$$

$$\mathbf{B}_{ij} = \mathbf{C}_i^{-1} \mathbf{W} \mathbf{C}_j^{-1} \quad (28c)$$

$$\mathbf{B}_{iM} = \mathbf{C}_i^{-1} \mathbf{W} \mathbf{M}^{-1} \quad (28d)$$

$$\mathbf{W} = \left(\sum_i^N \mathbf{C}_i^{-1}\right)^{-1} \quad (28e)$$

### Divergence Based Measures in Data Association

To understand the role of the divergence-based measures in data association, consider the problem of associating a measurement  $\mathbf{y}$  to one of the possible  $M$  targets. The variable  $\mathbf{x}^k$  denotes the state vector corresponding to the  $k^{\text{th}}$  target and  $p(\mathbf{x}^k)$  denotes the prior density function for the state vector  $\mathbf{x}^k$  corresponding to the  $k^{\text{th}}$  target before the measurement  $\mathbf{y}$  arrives. According to the AGSF structure,  $p(\mathbf{x}^k)$  is parameterized by a Gaussian mixture model as described in the previous section:

$$p(\mathbf{x}^k) = \sum_{i=1}^N w_i^k(t) \mathcal{N}(\mathbf{x}^k | \boldsymbol{\mu}_i^k(t), \mathbf{P}_i^k(t)) \quad (29)$$

Further, let us assume that  $\mathcal{L} = p(\mathbf{y}|\mathbf{x})$  represents the known likelihood function representing our confidence in the accuracy of the measurement data. Now, by propagating the prior state pdf  $p(\mathbf{x}^k)$  for each target through the measurement model of Eq. (10), we can obtain the estimated likelihood function,  $\hat{\mathcal{L}}^k$ , representing our confidence in the estimated measurement data assuming it belongs to the  $k^{\text{th}}$  target:

$$\hat{\mathcal{L}}^k = p(\hat{\mathbf{y}}|\mathbf{x}^k) = \sum_{i=1}^N w_i^k(t) \mathcal{N}(\mathbf{y} | \mathbf{h}(\boldsymbol{\mu}_i^k(t)), \mathbf{H}_i^k \mathbf{P}_i^k(t) \mathbf{H}_i^{kT} + \mathbf{R}) \quad (30)$$

Now, any information theoretic divergence metric such as the KL divergence, the Bhattacharyya metric or Renyi measure, can be used to discriminate between different targets. Basically, if the pdf  $\hat{\mathcal{L}}^k$  is very unlikely, then the information theoretic measure  $D(\mathcal{L}, \hat{\mathcal{L}}^k)$  should be large. The main steps of such an algorithm can be enumerated as follows:

1. Using a nonlinear filtering approach, such as the AGSF, obtain the prior target position distribution  $p(\mathbf{x}_k)$  at time  $t_k$ .
2. Use the sensor model to calculate the conditional distribution of the estimated observation  $\hat{\mathbf{y}}_k$  under the hypothesis that the target position is  $\mathbf{x}_k$ , i.e.  $p(\hat{\mathbf{y}}_k/\mathbf{x}_k)$ .
3. Compute the information theoretic divergence measure for each hypothesis by comparing the assumed distribution  $p(\hat{\mathbf{y}}_k/\mathbf{x}_k)$  with the actual sensor distribution  $p(\mathbf{y}_k/\mathbf{x}_k)$ , i.e., by computing  $\mathcal{D}(p(\mathbf{y}_k/\mathbf{x}_k), p(\hat{\mathbf{y}}_k/\mathbf{x}_k))$ .
4. Associate the measurement with the tracking hypothesis which leads to the minimum value of divergence measure,  $\mathcal{D}(p(\mathbf{y}_k/\mathbf{x}_k), p(\hat{\mathbf{y}}_k/\mathbf{x}_k))$ .

Ideally, any of the information theoretic divergence measures can be used to associate measurement data to a particular target state. However, the main challenge lies in computing these measures efficiently and accurately when the involved pdfs are parameterized by a mixture of Gaussian kernels. Notice that if the likelihood pdfs  $\mathcal{L}$  and  $\hat{\mathcal{L}}^k$  are parameterized by a mixture of Gaussians then the ratio or product of these pdfs can also be written as a mixture of Gaussians by making use of the fact that product of two Gaussian kernels also results in a Gaussian kernel. Hence, the multi-dimensional integral involved in the expression of the Renyi measure can be decomposed into a summation of many integrals involving only Gaussian kernels, particularly for  $\alpha = 0.5$  and  $\alpha = 2$ . This leads to an analytical expression for  $D(\mathcal{L}, \hat{\mathcal{L}}^k)$ , for  $\alpha = 0.5$  and  $\alpha = 2$ . However, the computation of the KL divergence poses a special challenge for the Gaussian mixture approximation due to the presence of the natural log function inside the integral expression. For other values of  $\alpha$ , one can make use of numerical integration schemes, such as Gauss-quadrature rules, to compute these integrals efficiently.

### Mutual Information in Data Association

In this section, an alternate method is discussed for associating the measurement data to a specific target state vector. The main idea is to compute the joint distribution of the target state vector and estimated measurement, i.e.,  $\mathcal{J}_1 = p(\hat{\mathbf{y}}_k, \mathbf{x}^k)$  by assuming that measurement  $\mathbf{y}$  belongs to the  $k^{\text{th}}$  target and comparing it to joint distribution  $\mathcal{J}_2 = p(\mathbf{y}_k)p(\mathbf{x}^k)$ , which is obtained by assuming that the measurement  $\mathbf{y}$  is completely independent of the  $k^{\text{th}}$  target state vector  $\mathbf{x}^k$ . This represents the reduction in uncertainty in the target state due to knowledge of the measurement.

The joint distribution  $\mathcal{J}_1$  can be computed easily by using the UKF structure for each Gaussian component. By making use of the sigma points for each Gaussian component, one can compute covariance matrices  $\mathbf{P}_{xx}$ ,  $\mathbf{P}_{xy}$  and  $\mathbf{P}_{yy}$  corresponding to each Gaussian component. By making use of the joint distribution for each component and the weights from the AGSF solution, the following Gaussian mixture model for the joint pdf can be written:

$$p(\mathbf{y}, \mathbf{x}^k) = \sum_{i=1}^N w_i^k(t) \mathcal{N}([\boldsymbol{\mu}_i^k \quad \mathbf{h}(t, \boldsymbol{\mu}_i^k)]^T, \bar{\mathbf{P}}_i^k(t)) \quad (31)$$

where

$$\bar{\mathbf{P}}_i^k(t) = \begin{bmatrix} \mathbf{P}_{\mathbf{xx}_i^k}(t) & \mathbf{P}_{\mathbf{xy}_i^k}(t) \\ \mathbf{P}_{\mathbf{xy}_i^k}^T(t) & \mathbf{P}_{\mathbf{yy}_i^k}(t) \end{bmatrix} \quad (32)$$

The actual sensor distribution is given by the measurement noise distribution:

$$p(\mathbf{y}) = \mathcal{N}(\tilde{\mathbf{y}}, \mathbf{R}(t)) \quad (33)$$

The state distribution is given by the AGSF solution:

$$p(\mathbf{x}^k) = \sum_{i=1}^N w_i^k(t) \mathcal{N}(\boldsymbol{\mu}_i^k, \mathbf{P}_i^k(t)) \quad (34)$$

Then the mutual information measure can be written in terms of these distributions and a divergence measure, given by

$$I(\mathbf{x}, \mathbf{y}) = \mathcal{D}(p(\mathbf{x}^k, \mathbf{y}), p(\mathbf{x}^k)p(\mathbf{y})) \quad (35)$$

It should be mentioned that the mutual information can be defined in terms of any divergence measure. The choice of the measure will not change the results, but as discussed in the previous section some measures are easier to calculate than others and some measures have probability bounds associated with them that provide more theoretical justification.

## NUMERICAL RESULTS

In this section, numerical results are presented validating the key ideas presented in this paper. To show the effectiveness of the proposed ideas, we consider the problem of tracking a high area-to-mass ratio (HAMR) object in a low-Earth orbit subject to nonconservative atmospheric drag. The planar equations of motion for a LEO object affected by nonconservative atmospheric drag forces are given by<sup>30</sup>

$$\begin{aligned}\ddot{x} + \frac{\mu x}{r^3} &= a_{D_x}(t, x, y, \dot{x}, \dot{y}), & a_D &= -\frac{1}{2} \frac{C_d A}{m} \rho v_{rel}^2 \frac{\mathbf{v}_{rel}}{|\mathbf{v}_{rel}|} \\ \ddot{y} + \frac{\mu y}{r^3} &= a_{D_y}(t, x, y, \dot{x}, \dot{y}), & \rho &= \rho_0 e^{-\frac{(r-R_\oplus)}{h}}\end{aligned}$$

where  $C_d$  is the coefficient of drag,  $A$  is the cross-sectional area,  $m$  is the mass of the object, and  $\rho$  is the atmospheric density at a given altitude. The atmospheric density model is assumed to be an exponential model with reference density  $\rho_0$ . It is also worth noting that the  $v_{rel}$  is not the velocity state vector, but rather the velocity relative to the Earth's atmosphere.

For simulation purposes, the value of the ballistic coefficient,  $B = \frac{C_d A}{m}$ , is chosen to be 1.4 which is consistent with a HAMR object.<sup>31</sup> Perfect knowledge of system dynamics is assumed, i.e., there is no process noise in the system. The initial state pdf is assumed to be Gaussian with the following mean and covariance:

$$\mu_0 = \begin{bmatrix} 6.6032 \times 10^6 \\ 0 \\ 0 \\ 7.7695 \times 10^3 \end{bmatrix} \quad \mathbf{P}_0 = \begin{bmatrix} 1.78 \times 10^6 & 0 & 0 & 0 \\ 0 & 2.50 \times 10^5 & 0 & 0 \\ 0 & 0 & 6.25 & 0 \\ 0 & 0 & 0 & 25 \end{bmatrix}$$

The mean of the initial pdf corresponds to a starting altitude of 225 km. The covariance matrix reflects a larger uncertainty in the radial position and the tangential velocity than in the in-track position and the radial velocity, respectively.

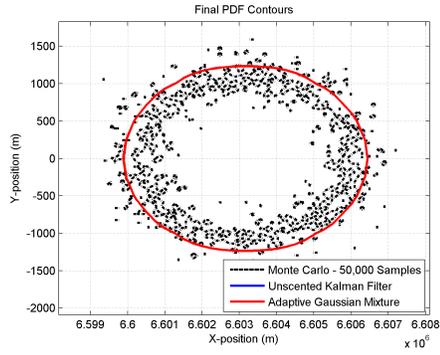
In this work it is assumed that range and angle observations are of an RSO and are denoted by  $\|\boldsymbol{\rho}\|$  and  $\Theta$ , respectively. The observation equations are given by

$$\|\boldsymbol{\rho}\| = (x^2 + y^2)^{1/2} \quad (36a)$$

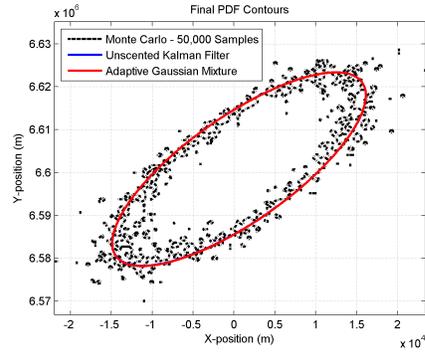
$$\Theta = \tan^{-1} \left( \frac{y}{x} \right) \quad (36b)$$

It is assumed the observation are corrupted with zero-mean white noise process with covariance matrix denoted by  $\mathbf{R} = \text{diag}([100 \text{ m}^2 \quad 0.0085 \text{ deg}^2])$ .

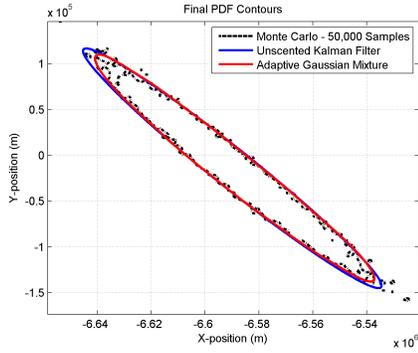
The initial state pdf is propagated through the orbit dynamics for a full orbit using the AGSF, as well as an UKF and the sequential Monte Carlo (SMC) method with 50,000 runs. In addition to the initial Gaussian pdf, 15 mixture components with zero weights are introduced along the principal axes of the the initial covariance matrix. Figure 1 shows the contour plots corresponding to 1% of



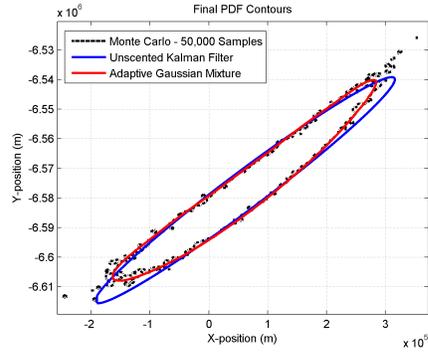
(a) Initial PDF Contours



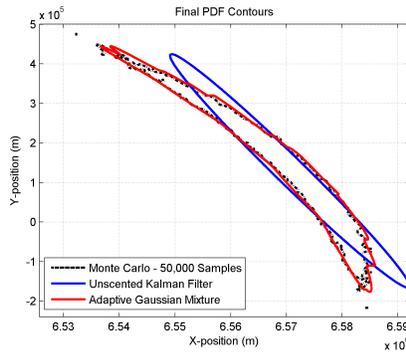
(b) 0.25 Orbits



(c) 0.5 Orbits



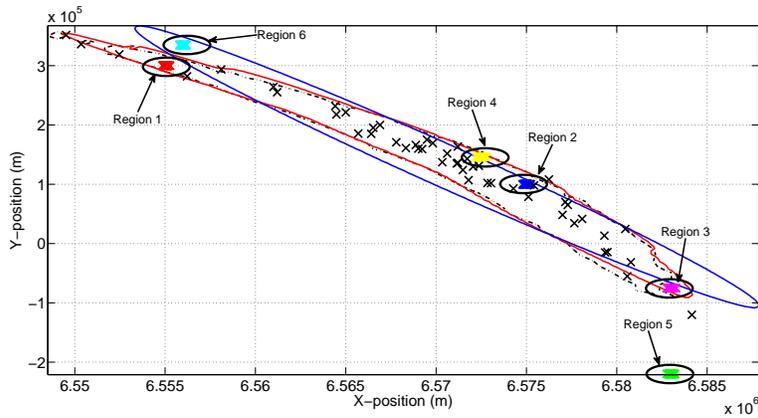
(d) 0.75 Orbits



(e) 1 Orbit

**Figure 1. State pdfs Propagation for a HAMR object in a LEO Orbit with Atmospheric Drag. The blue line represents the Gaussian approximation, red line represents the Gaussian mixture approximation while black dots represent the Monte Carlo particles.**

the state pdf's peak value during various times of the orbit. As expected the effects of nonlinearities and atmospheric drag skew the state pdf, which is accurately captured by the AGSF approximation and SMC runs. It is clear that the UKF and the AGSF pdfs are initially identical and remain similar for some time, however, the UKF no longer accurately represents the area of uncertainty given by the SMC samples contour at the end of one orbit. These plots clearly shows the effectiveness of the



**Figure 2. Density Contours After One Orbit and Simulated Measurements for Different Monte Carlo Runs**

AGSF method in capturing the non-Gaussian behavior of the state pdf.

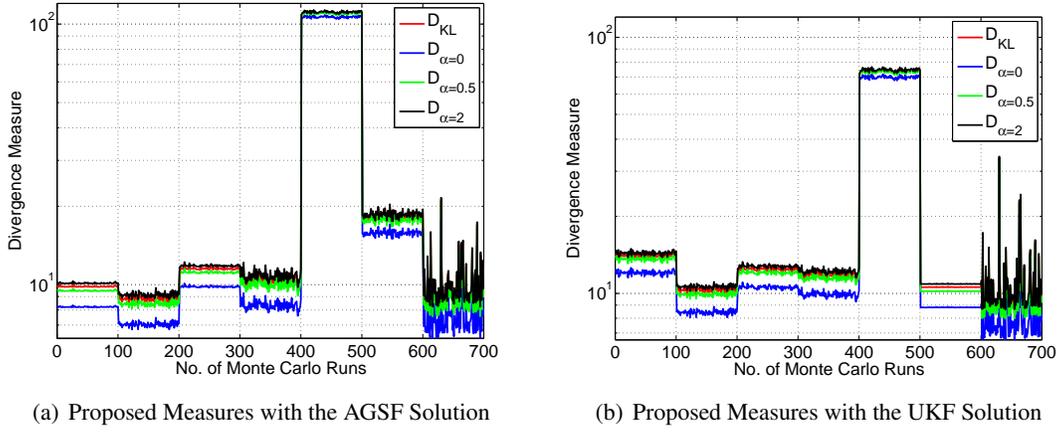
To show the effectiveness of the divergence measures for data association, sensor measurements are simulated at the end of an orbit. A total of 700 Monte Carlo runs are performed to generate different measurement data, as shown in Figure 2. The blue and red contour lines in Figure 2 correspond to contours for the state pdf propagated through the measurement model using the UKF and the AGSF approximations, respectively. To show the efficacy of capturing the non-Gaussian behavior, the simulated measurements are divided into four different regions (Regions 1–4) of varying degree of nonlinearity. Furthermore, measurements corresponding to false targets are also generated, denoted by Regions 5 and 6. The rest of the Monte Carlo runs correspond to the high probable region of the actual state pdf.

Renyi’s entropy divergence for  $\alpha = 0$  and  $\alpha = 2$ , KL-divergence (Renyi  $\alpha = 1$ ), and Bhattacharya distance (Renyi  $\alpha = 0.5$ ) are computed for both the AGSF and UKF approximated pdf by using 10,000 Gaussian quadrature points and assuming the true measurement pdf (likelihood function) to be Gaussian with standard deviation of 100 m. For notational sake, the divergence measure computed by using the AGSF approximation is represented by  $D_{AGSF}$  while the one using the UKF approximation is represented by  $D_{UKF}$ .

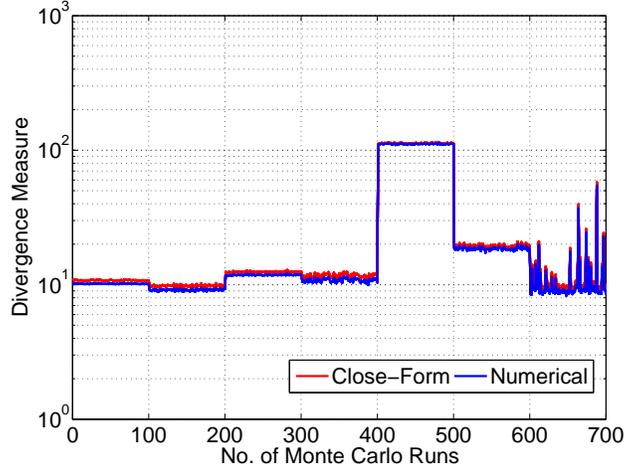
Figure 3 shows plots for both  $D_{AGSF}$  and  $D_{UKF}$  for different Monte Carlo runs of simulated measurements and for the different divergence measures. As expected the value of  $D_{AGSF}$  is consistently less than the value for  $D_{UKF}$  for measurements belonging to Regions 1–4 and 5. The divergences return similar results, as seen in Figure 3 where it is seen that all measures have similar values but  $\mathcal{D}_{KL}$ ,  $\mathcal{D}_{\alpha=0.5}$ , and  $\mathcal{D}_{\alpha=2}$  are very close. This is a very useful since  $\mathcal{D}_{KL}$  is much harder to evaluate than Renyi’s entropy divergence for  $\alpha = 2$ .

The results using the closed-form expression given in Eq. (27) for Renyi’s quadratic entropy divergence using the AGSF solution is given in Figure 4. The numerically calculated values plotted in Figure 3 are also shown in Figure 4. It is clearly that the numerical values are in good agreement with the closed-form results. This is beneficial because the closed-form expression offers a large reduction in computational cost for calculating the distance measure.

Furthermore, both the UKF- and AGSF-based measures correctly identify the false measurements corresponding to Region 5 by a sudden increase in the value of the divergence measure, although



**Figure 3. Data Association Results**



**Figure 4. Comparison Between Renyi's Entropy Divergence for  $\alpha = 2$  Determined Numerically and Determined Using the Closed-Form Expression**

the increase is much larger for the AGSF-based measures. However, the UKF-based measure fails to identify the false measurements corresponding to Region 6 due to the skewness of the actual pdf which is not captured by the UKF-based measure. This once again illustrates the benefit of capturing the actual non-Gaussian pdf. These plots clearly illustrate the effectiveness of the AGSF and divergence-based measures in correctly identifying targets under large uncertainty.

## CONCLUSION

In this paper an approach for data association of resident space object tracking was developed. The approach combines an adaptive Gaussian sum filter with information theoretic divergence measures to associate measurement data to a particular space object. This approach has several advantages over existing approaches, including: it is able to approximate the pdf associated with nonlinear systems well and it is computationally efficient so that it can be executed in realtime using modern-

day computers. Four different measures were considered: Renyi's entropy divergence for  $\alpha = 0$  and  $\alpha = 2$ , KL-divergence (Renyi  $\alpha = 1$ ), and Bhattacharya distance (Renyi  $\alpha = 0.5$ ). The numerical results presented in this paper show that all measures correspond to very similar results; however, the KL-divergence is much more computationally expensive to compute than the other measures.

## REFERENCES

- [1] H. Risken, *The Fokker-Planck Equation: Methods of Solution and Applications*. Springer, 1989, 1-12, 32-62.
- [2] R. S. Park and D. J. Scheeres, "Nonlinear Mapping of Gaussian Statistics: Theory and Application to Spacecraft Design," *AIAA Journal for Guidance, Control and Dynamics*, Vol. 29, Nov.-Dec. 2006.
- [3] A. T. Fuller, "Analysis of nonlinear stochastic systems by means of the Fokker-Planck equation," *International Journal of Control*, Vol. 9, 1969, p. 6.
- [4] G. Terejanu, P. Singla, T. Singh, and P. D. Scott, "Uncertainty Propagation for Nonlinear Dynamical Systems using Gaussian Mixture Models," *Journal of Guidance, Control, and Dynamics*, Vol. In Press, 2008.
- [5] G. Terejanu, P. Singla, T. Singh, and P. D. Scott, "A Novel Gaussian Sum Filter Method for Accurate Solution to Nonlinear Filtering Problem," *2008 International Conference on Information Fusion*.
- [6] D. Giza and P. Singla, "An Approach for Nonlinear Uncertainty Propagation: Application to Orbital Mechanics," *2009 AIAA Guidance Navigation and Control Conference*.
- [7] G. Terejanu, J. George, and P. Singla, "An Adaptive Gaussian Sum Filter For The Spacecraft Attitude Estimation Problem," *F. Landis Markley Astronautics Symposium*, Cambridge, MD, June, 2008.
- [8] G. Terejanu, J. George, and P. Singla, "An Adaptive Gaussian Sum Filter For The Spacecraft Attitude Estimation Problem," *Accepted for publication in the Journal of Astronautical Sciences*, May, 2009.
- [9] A. Khan, T. Balch, and F. Dellaert, "MCMC-Based Particle Filtering for Tracking a Variable Number of Interacting Targets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, Nov. 2005, pp. 1805–1819.
- [10] M. E. Liggins, II, M. A. Nebrich, and K. C. Chang, "Information theoretics for improved tracking and fusion performance," *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series* (I. Kadar, ed.), Vol. 5096 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Aug. 2003, pp. 379–386.
- [11] Y. Xu, Y. Tan, Z. Lian, and R. He, "An information theoretic approach based Kullback-Leibler discrimination for multiple target tracking," *Information and Automation, 2009. ICIA '09. International Conference on*, June 2009, pp. 1129–1134.
- [12] P. E. Berry and D. A. B. Fogg, "GAMBIT: Gauss-Markov and Bayesian Inference Technique for Information Uncertainty and Decision-making in Surveillance Simulations," *DSTO, Edinburgh, South Australia 2003*.
- [13] H. C. Lambert, F. E. Daum, and J. L. Weatherwax, "A Split-Step Solution of the Fokker-Planck Equation for the Conditional Density," *Signals, Systems and Computers, 2006. ACSSC '06. Fortieth Asilomar Conference on*, Oct.-Nov. 2006, pp. 2014–2018.
- [14] M. Kumar, P. Singla, S. Chakravorty, and J. L. Junkins, "A Multi-Resolution Approach for Steady State Uncertainty Determination in Nonlinear Dynamical Systems," *38th Southeastern Symposium on System Theory*, 2006.
- [15] M. Kumar, P. Singla, S. Chakravorty, and J. L. Junkins, "The Partition of Unity Finite Element Approach to the Stationary Fokker-Planck Equation," *2006 AIAA/AAS Astrodynamics Specialist Conference and Exhibit, Keystone, CO*, Aug. 21-24, 2006.
- [16] G. Muscolino, G. Ricciardi, and M. Vasta, "Stationary and non-stationary probability density function for non-linear oscillators," *International Journal of Non-Linear Mechanics*, Vol. 32, 1997, pp. 1051–1064.
- [17] M. D. Paola and A. Sofi, "Approximate solution of the Fokker-Planck-Kolmogorov equation," *Probabilistic Engineering Mechanics*, Vol. 17, 2002, pp. 369–384.
- [18] G. Terejanu, P. Singla, T. Singh, and P. D. Scott, "Uncertainty Propagation for Nonlinear Dynamical Systems using Gaussian Mixture Models," *Proceedings of the AIAA Guidance, Navigation and Control Conference and Exhibit*, Honolulu, HI, 2008.
- [19] P. Singla and T. Singh, "A Gaussian function Network for Uncertainty Propagation through Nonlinear Dynamical System," *18th AAS/AIAA Spaceflight Mechanics Meeting, Galveston, TX*, Jan. 27-31, 2008.
- [20] D. Alspach and H. Sorenson, "Nonlinear Bayesian estimation using Gaussian sum approximations," *Automatic Control, IEEE Transactions on*, Vol. 17, No. 4, 1972, pp. 439–448.

- [21] G. Terejanu, P. Singla, T. Singh, and P. D. Scott, "Decision Based Uncertainty Propagation Using Adaptive Gaussian Mixtures," *2009 International Conference on Information Fusion*.
- [22] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004, 67-78, 152-159.
- [23] S. Kullback and R. Leibler, "On Information and Sufficiency," *Ann. Math. Stat.*, Vol. 22, 1951, pp. 79-86.
- [24] A. N. Kolmogorov, "On the Shannon theory of information in the case of continuous signals," *IRE Trans. Inform. Theory*, Vol. 2, 1956, pp. 102-108.
- [25] A. N. Kolmogorov, *Information Theory and the Theory of Algorithms*, Vol. III of *Selected Work of A. N. Kolmogorov*. Moscow: Kluwer Academic Publisher, 1987.
- [26] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley Series in Telecommunications, New York: Wiley Interscience, 1991.
- [27] F. G. A. Djouadi, O. Snorrason, "The Quality of Training-Sample Estimates of the Bhattacharyya Coefficient," *EEE Trans. Pattern Analysis Machine Intell.*, p. 1990.
- [28] S. Intille, J. Davis, and A. Bobick, "Real-Time Closed-World Tracking," *IEEE Conf. on Comp. Vis. and Pat. Rec.*, pp. 697-703.
- [29] A. Alempijevic, S. Kodagoda, and G. Dissanayake, "Mutual information based data association," *2009 International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, Dec. 2009, pp. 97-102.
- [30] D. A. Vallado, *Fundamentals of Astrodynamics and Applications*. Microcosm, 2007.
- [31] T. F. T. Schildknecht, R. Musci, "Properties of the high area-to-mass ratio space debris population at high altitudes," *Advances in Space Research*, Vol. 41, 2008, pp. 1039-1045.