Establishing Consensus Between Implicitly Updated Decentralized Probability Distribution Functions

Juan F. Gutierrez
KBR Inc.
Waqar Zaidi
L3Harris

ABSTRACT

This paper explores the concept of consensus between implicitly updated decentralized probability distribution functions (PDFs). A prior PDF tied to an angles-only optical observation of a space object is formed using a probabilistic method of admissible regions. This PDF is then implicitly updated at a time of three decentralized radar measurements using a closed-form solution derived from conservation equations. The implicit update produces three independent bimodal PDFs where one mode is eliminated based on the sensor field of view. We then apply a Gaussian Mixture Model (GMM) to the second non-Gaussian mode of each PDF and study information divergence between the components of PDF's GMM representation. Specifically, the Burbea-Rao and Bhattacharyya centroids are utilized to understand difference in distributions as well as perform a fusion step between GMM components to produce a new mean and covariance. In verifying the statistical distances from selection of GMM components establishes a consensus approach to obtain distributions closer to truth state.

1. INTRODUCTION

Space Traffic Management (STM) is the planning and coordination of activities that enhance sustainability of operations in the space environment while Space Situational Awareness (SSA) is the knowledge and characterization of space objects and their operational environment to support safe, stable, and sustainable space activities [5]. Important features of a worldwide STM construct must include data provenance, immutability, and attribution that support established norms of spacecraft behavior. Accordingly, consensus between decentralized sources of SSA data is an integral part of corroborating or contesting such norms of behavior.

In this paper we explore consensus between implicitly updated decentralized probability distribution functions (PDFs). We specifically use a Probabilisitic Admissible Region (PAR) method and an angles-only optical measurement at an initial time to build a prior PDF as in [13]. For brevity, the PAR equations are not repeated in this paper. We then implicitly update the prior PDF at the time of a second radar measurement using a Newton-Raphson iteration on a closed-form solution formed derived from the conservation of energy and angular momentum equations. That is, the prior PDF at the time of the first measurement is not updated through the traditional Chapman-Kolmogorov equation but instead through a root-finding scheme that produces conserved orbit solutions that link the original PDF with a second radar measurement.

In adopting this process, we need to form initial guesses for possible optical measurements at the time of the second radar measurement update to complete an attributable vector that can be mapped to position and velocity coordinates. Applying this root-finding scheme ends up producing optimal solutions that conserve energy and angular momentum linked to the prior PDF. We adopt this approach with three different decentralized radar sensors with various measurement noise characteristics in order to set up the consensus problem.

Since PARs are built with a single angles-only measurement, the distributions they produce are highly non-Gaussian in nature. Thus, we leverage MATLAB's Gaussian Mixture Model (GMM) and clustering algorithms to fit multiple Gaussians to the implicitly updated PDFs. We then shift our focus on studying the similarity between the components of these posterior non-gaussian PDFs and explore an unorthodox, non-Bayesian step to fuse them together. Specifically, we first leverage information divergence metrics such as Burbea-Rao and Bhattacharrya distance measures to

explicitly determine if there is consensus between the decentralized GMM components of each posterior PDF and then we use centroid calculations to produce a fused mean and covariance. From these centroid calculations the statistical distances are used again to show initial correlation between closeness to the truth and minimal group distances to the centroids. Thus we have demonstrated optimization methods to both implicitly update a prior PDF using conservations equations and three decentralized sensors and fused the results using statistical centroid methods.

2. ANGULAR MOMENTUM AND ENERGY SOLUTION SPACE

Given an attributable vector, $[\rho_1,\ \dot{\rho}_1,\ \alpha_1,\ \dot{\alpha}_1,\ \delta_1,\ \dot{\delta}_1]$, associated to a PAR particle at the time of an initial optical measurement, $[\alpha_1,\ \delta_1]$, this section looks at the utilization of the Newton-Raphson method in order to implicitly update the vector using the energy and angular momentum conservation. Thus, we are able to find solutions containing a given subsequent radar measurement, $[\rho_2,\ \dot{\rho}_2]$, to find the right ascension α_2 and declination δ_2 along with the respective rates $\dot{\alpha}_2$ and $\dot{\delta}_2$ to build an implicitly updated attributable vector $[\rho_2,\ \dot{\rho}_2,\ \alpha_2,\ \dot{\alpha}_2,\ \delta_2,\ \dot{\delta}_2]$. Eqs. (1) and (2) provide the equations for the angular momentum vector as functions of range and range-rate, then in terms of the right ascension and declination rates.

$$\mathbf{h}(\rho,\dot{\rho}) = \mathbf{h}_1\dot{\rho} + \mathbf{h}_2\rho^2 + \mathbf{h}_3\rho + \mathbf{h}_4 \tag{1}$$

$$\mathbf{h}(\dot{\alpha},\dot{\delta}) = \gamma_1\dot{\alpha} + \gamma_2\dot{\delta} + \gamma_3 \tag{2}$$

Now Eqs. (3) through (6) provide the functions utilized in Eq. (1). The o vector is the inertial position of the sensor and $\dot{o}$ is the interial velocity.

$$\mathbf{h}_1 = \mathbf{o}\times\mathbf{u}_\rho \tag{3}$$

$$\mathbf{h}_2 = \mathbf{u}_\rho\times(\dot{\alpha}\cos(\delta)\mathbf{u}_\alpha + \dot{\delta}\mathbf{u}_\delta) \tag{4}$$

$$\mathbf{h}_3 = \mathbf{u}_\rho\times\dot{o} + o\times(\dot{\alpha}\cos(\delta)\mathbf{u}_\alpha + \dot{\delta}\mathbf{u}_\delta) \tag{5}$$

$$\mathbf{h}_4 = o\times\dot{o} \tag{6}$$

Similarly, Eqs. (7) through (8) provide the functions utilized in Eq. (2).

$$\gamma_1 = \rho\mathbf{r}\times\mathbf{u}_\alpha\cos(\delta) \tag{7}$$

$$\gamma_2 = \rho\mathbf{r}\times\mathbf{u}_\delta \tag{8}$$

$$\gamma_3 = \mathbf{r}\times\dot{o} + \dot{\rho}o\times\mathbf{u}_\rho \tag{9}$$

The unit vectors $\mathbf{u}_\rho, \mathbf{u}_\alpha, \mathbf{u}_\delta$ are functions of the right ascension and declination angles.

$$\mathbf{u}_\rho = \begin{bmatrix}\cos\alpha\cos\delta & \sin\alpha\cos\delta & \sin\delta\end{bmatrix}^T \tag{10}$$

$$\mathbf{u}_\alpha = \begin{bmatrix}-\sin\alpha & \cos\alpha & 0\end{bmatrix}^T \tag{11}$$

$$\mathbf{u}_\alpha = \begin{bmatrix}-\cos\alpha\sin\delta & -\sin\alpha\sin\delta & \cos\delta\end{bmatrix}^T \tag{12}$$

Similar to the angular momentum derivation, the energy equations derived in terms of the measurements. Where the first equation is a function of range and range-rate, while the other version is in terms of the right ascension and declination rates.

$$\varepsilon(\rho,\dot{\rho}) = \dot{\rho}^2 + \omega_1\dot{\rho}^2 + \omega_2\rho^2 + \omega_3\rho + \omega_4 - \frac{2\mu}{\sqrt{\rho^2+\omega_5+\omega_0}+\frac{u}{a_{max}}} \tag{13}$$

$$\varepsilon(\dot{\alpha},\dot{\delta}) = z_1\dot{\alpha}^2 + z_2\dot{\delta}^2 + 2z_3\dot{\alpha} + 2z_4\dot{\delta} + z5 - \frac{u}{a_{max}} \tag{14}$$

The associated variables for both of the energy equations are shown below:

$$\omega_0 = |o|^2; \quad \omega_1 = 2(\dot{o}^2\cdot\mathbf{u}_\rho) \tag{15}$$

$$\omega_2 = \dot{\alpha}^2 \cos^2\delta + \dot{\delta}^2; \quad \omega_3 = 2\dot{\alpha}\cos\delta(\dot{o}\cdot\mathbf{u}_\alpha) + 2\dot{\delta}(\dot{o}\cdot\mathbf{u}_\delta) \tag{16}$$

$$\omega_4 = |\dot{o}|^2; \quad \omega_5 = 2(o^2\cdot\mathbf{u}_\rho) \tag{17}$$

$$z_1 = \rho^2\cos^2\delta/2; \quad z_2 = \rho^2/2 \tag{18}$$

$$z_3 = (\rho\dot{\mathbf{r}}\cdot\mathbf{u}_\alpha)\cos\delta/2; \quad z_4 = (\rho\dot{\mathbf{r}}\cdot\mathbf{u}_\delta)/2 \tag{19}$$

$$z_5 = \frac{1}{2}\left(\dot{\rho}^2 + \omega_1\dot{\rho} + \omega_4 - \frac{2\mu}{\sqrt{\rho^2+\omega_5\rho+\omega_0}}\right) \tag{20}$$

Given a set of asynchronous optical and radar measurements, the Newton-Rapshon optimization then simply sets the angular momentum and energy equations equal to each other and solve for the remaining variable to maintain a conserved orbit solution. Specifically, we set $\mathbf{h}(\rho_2,\dot{\rho}_2) = \mathbf{h}(\dot{\alpha}_1,\dot{\delta}_1)$. Similarly, the energy equation provides $\varepsilon(\rho_2,\dot{\rho}_2) = \varepsilon(\dot{\alpha}_1,\dot{\delta}_1)$. Accordingly, Eqs. (21) and (22) show the conservation of angular momentum and energy in terms of asynchronous measurements.

$$\begin{aligned}\mathbf{h}_1\dot{\rho} + \mathbf{h}_2\rho^2 + \mathbf{h}_3\rho + \mathbf{h}_4 \\ -\gamma_1\dot{\alpha} - \gamma_2\dot{\delta} - \gamma_3 = 0\end{aligned} \tag{21}$$

$$\begin{aligned}\dot{\rho}^2 + \omega_1\dot{\rho}^2 + \omega_2\rho^2 + \omega_3\rho + \omega_4 - \frac{2\mu}{\sqrt{\rho^2+\omega_5\rho+\omega_0}} \\ -z_1\dot{\alpha}^2 - z_2\dot{\delta}^2 - 2z_3\dot{\alpha} - 2z_4\dot{\delta} - z5 = 0\end{aligned} \tag{22}$$

3. CONVEXITY WITH INFORMATION DIVERGENCE

Information divergence techniques focus on the measure of distances between two PDFs. There are various types of distances including but not limited to Kullback-Leibler, Rényi or the α-divergence, Bhattacharyya and Burbea-Rao ([1], [4], [3], and [11]). These are widely used in the field of image processing and portions of machine learning. In our application, the information divergence metric plays a role in determining the difference between the prior and implicitly updated PDFs. However, since these PDFs are non-Gaussian in nature, we take the step to utilize Gaussian Mixture Models in order to study the distributions on a component level. This is based on the convexity property which is shown from the Bregmann divergence as in [7], which the type of minimization problem was posed in [2] by Amari for the α-divergence. Specifically, we leverage the MATLAB *fitgmdist* and *cluster* sub-routines to fit a Gaussian Mixture Model (GMM) to the implicitly updated PDF. This enables us to explore the convexity of two particular information divergences, the Burbea-Rao and Bhattacharyya divergences, which both require Gaussian PDFs (i.e. mean and covariance) as input. As a result, we are able to explicitly measure consensus between PDFs by minimizing the distance between the three decentralized set of implicitly updated PDFs. Once consensus is determined, we employed a centroid calculation to obtain a fused Gaussian representation of adjacent GMM components from the three decentralized implicitly updated PDFs. This is further explored by analyzing some the sensitivities as for the resulting distribution and utilizing various GMM component combinations towards centroid representation.

3.1 Burbea-Rao Centroid

The first information distance metric we studied for this paper is called the Bregmann divergence $B_F(p,q)$ where p and q are distributions represented by $\mathcal{N}(\mu_p,\Sigma_p)$ and $\mathcal{N}(\mu_q,\Sigma_q)$, respectively. The utilization of the Bregmann divergance helps in providing an equivalent expression to assemble the Burbea-Rao divergence. This allows us to portray the Burbea-Rao divergence with convex properties. Eq. (23) provides an expression $F(x)$ that is strictly convex and differentiable.

$$B_F(x,p) = F(x) - F(p) - \langle x-p, \nabla F(p)\rangle \tag{23}$$

Nielsen proposes the barycenter technique in finding the minimzation of the information-theoretic distances between various distributions. Eq. (24) provides the optimization. The $d(x,p_i)$ represents the divergence measure to be used without the loss of generality.

$$(OPT) : \min_{x} \sum_{i=1}^{n} w_i d(x, p_i) \tag{24}$$

Burbea-Rao divergence is described in the form provided by Eq. (25), such that it is formulated with the function $F(\bullet)$, which is convex.

$$BR_F(p,q) = \frac{F(p)+F(q)}{2} - F\left(\frac{p+q}{2}\right) \geq 0 \tag{25}$$

To obtain the centroid of the Burbea-Rao divergence the divergence equation requires further generalizations, where Nielsen introduces a positive weight $\alpha \in (0,1)$

$$BR_F^{\alpha}(p,q) = \alpha F(p) + (1-\alpha)F(q) - F(\alpha + (1-\alpha)q) \tag{26}$$

Therefore, by applying Eq. (24) to the generalized form of the Burbea-Rao divergence in Eq. (26), the centroid or the minimum distance to a particular point is described by the following equation:

$$(OPT) : c = \arg\min_{x} \sum_{i=1}^{n} w_i BR_F^{(\alpha_i)}(x, p_i) \tag{27}$$

Eq. (27) can be re-written with the energy function presented in Eq. (28) where it is specifically decomposed into a sum of convex $(\sum_{i=1}^{n} w_i\alpha_i)F(c)$ and concave portions $-\sum_{i=1}^{n} w_i F(\alpha_i c + (1-\alpha_i)p_i)$.

$$E(c) = \left(\sum_{i=1}^{n} w_i\alpha_i\right) F(c) - \sum_{i=1}^{n} w_i F\left(\alpha_i c + (1-\alpha_i)p_i\right) \tag{28}$$

A popular method to solve this convex-concave optimization problem is the Convex-ConCave Procedure or CCCP as described in [12]. An initial position for the centroid c_0 is chosen and then iterated to converge to a centroid location. A possible starting point is $c_0 = \sum_{i=1}^{n} w_i p_i$. The CCCP method for this centroid utilizes the following set of equations (29) and (30). Where Eq. (30) provides the next point of iteration c_{t+1}.

$$\nabla F(c_{t+1}) = \frac{1}{\sum_{i=1}^{n} w_i\alpha_i} \sum_{i=1}^{n} w_i\alpha_i \nabla F\left(\alpha_i c_t + (1-\alpha_i)p_i\right) \tag{29}$$

$$c_{t+1} = \nabla F^{-1}\left(\frac{1}{\sum_{i=1}^{n} w_i\alpha_i} \sum_{i=1}^{n} w_i\alpha_i \nabla F\left(\alpha_i c_t + (1-\alpha_i)p_i\right)\right) \tag{30}$$

The process to develop both the distance measure and centroid requires a definition for $F(\bullet)$ and $\nabla F(\bullet)$. From the definitions provided in [12] and [10] the mean $\lambda_v = \mu$ and covariance $\lambda_M = \Sigma$ are translated into new definitions θ_v and θ_M. Then, $\theta = (\theta_v, \theta_M)$ is utilized to form the convex function $F_\theta(\theta)$. Finally, the distance measure is then based on the well known Eq. (31) for a multivariate normal distribution (MVN) with $\lambda := (\lambda_v, \lambda_M)$.

$$p_\lambda(x;\lambda) := \frac{1}{(2\pi)^{d/2}\sqrt{|\lambda_M|}} \exp\left(-\frac{1}{2}(x-\lambda_v)^T \lambda_M^{-1}(x-\lambda_v)\right) \tag{31}$$

In the canonical form of the exponential family of equations, the probability density in Eq. (31) is defined with $F_\theta(\theta)$ as follows:

$$p_\theta(x;\theta) := \exp\left(\langle t(x), \theta\rangle - F_\theta(\theta)\right) = p_\lambda(x;\lambda(\theta)) \tag{32}$$

where Eq. (33) shows the definition utilized in Eq. (32). Furthermore (34) provides the inverse definition

$$\theta = (\theta_v, \theta_M) = \left(\Sigma^{-1}\mu, -\frac{1}{2}\Sigma^{-1}\right) \tag{33}$$

$$\theta(\lambda) = (\lambda_v^{-1}\lambda_v, -\frac{1}{2}\lambda_M^{-1}) \tag{34}$$

Now $F_\theta(\theta)$ is defined as:

$$F_\theta(\theta) = \frac{1}{2}\left(d\log\pi - \log|\theta_M| + \frac{1}{2}\theta_v^T\theta_M^{-1}\theta_v\right) \tag{35}$$

Taking the gradient of the convex function $\nabla F_\theta(\theta)$ yields the following definition $\eta = (\eta_v, \eta_M)$.

$$\nabla F_\theta(\theta) = \eta = (\eta_v, \eta_M) \tag{36}$$

where

$$\eta_v = \lambda_v = \mu \tag{37}$$

$$\eta_M = -\lambda_M - \lambda_v\lambda_v^T = -\Sigma - \mu\mu^T \tag{38}$$

Eqs. (37) and (38) provide a definition in the MVN format in order to perform the computations for each iteration within the CCCP. Algorithm 1 provides the formulation of the CCCP utilizing the definitions from [6].

Algorithm 1: Burbea-Rao Centroid

Result: Centroid:μ_c and Σ_c
Inputs: $\mu_{i=1,\ldots,n}, \Sigma_{i=1:n}$;
Initialization: $\theta_{vc} = \mu_0, \theta_M = \Sigma_0, k = 1, \varepsilon$;
while $\Delta\eta_v \leq \varepsilon$ **do**
  for *i = 1:n* **do**
    Compute the term: $\nabla F(\alpha_i c_t + (1-\alpha_i)\theta_i)$;
    $\theta_{vk(i)} = \alpha_i\theta_{vc} + (1-\alpha_i)\theta_{v(i)}$;
    $\theta_{Mk(i)} = \alpha_i\theta_{Mc} + (1-\alpha_i)\theta_{M(i)}$;
    $\theta_{vc} = \theta_{vk(i)}$;
    $\theta_{Mc} = \theta_{Mk(i)}$;
  end
  $\eta_v = \sum_{i=1}^n \left(w_i\alpha_i\theta_{vk(i)}\right)$;
  $\eta_M = \sum_{i=1}^n \left(w_i\alpha_i\theta_{Mk(i)}\right)$;
  Compute the $\nabla F(\bullet)^{-1}$;
  $\lambda_v(k) = \eta_v$;
  $\lambda_M(k) = -\eta_M - \eta_v\eta_v^T$;
  $\Delta\eta_v = ||\lambda_v(k) - \lambda_v(k-1)||$;
  $k = k+1$
end
$\mu_c = \lambda_v(k)$;
$\Sigma_c = \lambda_M(k)$;

Algorithm 2: Bhattacharyya Centroid

Result: Centroid:μ_c and Σ_c
Inputs: $\mu_{i=1,\ldots,n}, \Sigma_{i=1:n}$;
Initialization: $\mu_c = \mu_0, \Sigma_c = \Sigma_0, k = 1, \varepsilon, \zeta = 0^{(m\times n)}, \xi = 0^{(m\times 1)}$;
while $\Delta\eta_v \leq \varepsilon$ **do**
  for *i = 1:n* **do**
    Compute the term: ζ, ξ, A;
    $U_i = (\Sigma_c + \Sigma_i)^{-1}$;
    $\zeta = \zeta + U_i + U_i^T$;
    $\xi = \xi + (U_i + U_i^T)\mu_i$;
    $A = A + 2U_i^T - U_i^T(\mu_c - \mu_i)(\mu_c - \mu_i)^T U_i^T$
  end
  Compute the update to $\mu_{c_{(k+1)}}$:;
    $\mu_c = \zeta^{-1}\xi$;
  Compute the B term:;
    $B = A + A^T - diag(A)$;
  Compute the update to $\Sigma_{(k+1)}$:;
    $\Sigma_c = 2n(B + diag(B))^{-1}$;
  Compute criteria for exit:;
    $\Delta\mu_c = ||\mu_c(k) - \mu_c(k-1)||$;
  Reset Values for next iteration:;
    $\zeta = 0^{(m\times n)}, \xi = 0^{(m\times 1)}$;
  $k = k+1$
end
μ_c;
Σ_c;

Now we have an algorithm that enables us to address the posed consensus optimization problem where we explicitly compute the information distance between two or more decentralized sources of information. The algorithm also enabled us to utilize sensor weights to represent the performance of a sensor. However, for this study we kept the weight at unity and controlled the sensor performance through different values of measurement noise.

3.2 Bhattacharyya Centroid

From [7] Nielsen provides a description of the Bhattacharyya distances or divergences as part of the Burbea-Rao divergence presented in the previous subsection. The algorithm used for the Bhattacharyya centroid is shown in the previous subsection as well. For arbitrary probability distributions in Eqs. (32) and (31) the overlap is measured using the Bhattacharyya coefficient as seen in [7] and shown in Eq. (40). This coefficient ranges from 0 and 1, $0 \leq C(p,q) \leq 1$.

$$C(p,q) = \int \sqrt{p(x)q(x)} \tag{39}$$

The Bhattacharyya measure utilizes $\sqrt{p(x)}$ and $\sqrt{q(x)}$ as unit vectors and employs a dot product to obtain the cosine of the angle between them. The distance measure is obtained by taking the negative natural logarithm as shown below.

$$B(p,q) = -\ln C(p,q) \tag{40}$$

Nielsen considers the generalization of the Bhattacharyya coefficients and divergences as Chernoff divergences by B_α. Eq. (41) below links the Bhattacharyya definition and the Chernoff coefficient $\int_x p^\alpha(x)q^{1-\alpha}(x)dx$.

$$B_\alpha(p,q) = -\ln \int_x p^\alpha(x)q^{1-\alpha}(x)dx = -\ln C_\alpha(p,q) \tag{41}$$

Now applying the probability distributions in Eqs. (32) and (31) yields a weighted asymmetric Brubea-Rao divergence or in short the Burbea-Rao divergence.

$$B_\alpha(p_F(x;\theta_p, p_F(x;\theta_{\theta_q})) \tag{42}$$

where,

$$BR_F^{(\alpha)}(\theta_p,\theta_q) = \alpha F(\theta_p + (1-\alpha)F(\theta_q) - F(\alpha\theta_p + (1-\alpha)\theta_q) \tag{43}$$

From these formulations the Bhattacharyya/Burbea-Rao distance between arbitrary distributions p and q is provided in Eq. (44).

$$\begin{aligned} BR_F(\lambda_p,\lambda_q) = &\frac{1}{8}(\mu_p-\mu_q)^T \left(\frac{\Sigma_p+\Sigma_q}{2}\right)^{-1}(\mu_p-\mu_q) \\ &+ \frac{1}{2}\ln \frac{\det \frac{\Sigma_p+\Sigma_q}{2}}{\sqrt{\det\Sigma_p \det\Sigma_q}} \end{aligned} \tag{44}$$

To compute the Bhattacharyya centroid, Nielsen utilizes Eq. (44) in order to produce an energy equation to be optimized $L(c)$ such that Eq. (45) is the sum over the distances between the centroid and the arbitrary set of distributions $i = 1,...,n$.

$$\begin{aligned} L(c) = &\sum_{i=1}^{n} \frac{1}{8}(\mu_c-\mu_i)^T \left(\frac{\Sigma_c+\Sigma_i}{2}\right)^{-1}(\mu_c-\mu_i) \\ &+ \frac{1}{2}\log \frac{\det \frac{\Sigma_c+\Sigma_i}{2}}{\sqrt{\det\Sigma_c \det\Sigma_i}} \end{aligned} \tag{45}$$

Computing the the Bhattacharyya distance in direct form requires developing the matrix differential to Eq. (45) with respect to μ_c in order to start minimizing the function. This computation is represented by Eq. (46).

$$\frac{\partial L}{\partial \mu_c} = \sum_{i=1}^{n} \left[U_i + U_i^T\right][\mu_c - \mu_i] \tag{46}$$

where,

$$U_i = (\Sigma_c + \Sigma_i)^{-1} \tag{47}$$

In order to estimate the centroid $\mu_{c_{(t)}}$ an iterative approach is executed such that one can take Eq. (46) to provide the subsequent point $\mu_{c_{(t+1)}}$. Eq. (48) provides the update for the mean $\mu_{c_{(t+1)}}$.

$$\mu_{c_{(t+1)}} = \left[\sum_{i=1}^{n} [U_i + U_i^T]\right]^{-1} \left[\sum_{i=1}^{n} [U_i + U_i^T]\mu_i\right] \tag{48}$$

Next, matrix differentials in Eq. (49) with respect to Σ_c are utilized to compute the centroid covariance $\Sigma_{c_{(t+1)}}$.

$$\begin{aligned} \frac{\partial L}{\partial \Sigma_c} = &\sum_{i=1}^{n} -U_i^T(\mu_c - \mu_i)(\mu_c - \mu_i)^T U_i^T \\ &+ 2\sum_{i=1}^{n} U_i^T - \sum_{i=1}^{n} \Sigma_c^{-T} \end{aligned} \tag{49}$$

Since for this process Σ_c is symmetric, an easier estimation is applied as follows.

$$\frac{dL}{d\Sigma_c} = \frac{\partial L}{\partial \Sigma_c} + \left(\frac{\partial L}{\partial \Sigma_c}\right)^T - diag\left(\frac{\partial L}{\partial \Sigma_c}\right) \tag{50}$$

Finally, the A matrix is formed in Eq. (51):

$$A = \sum_{i=1}^{n} 2U_i^T - U_i^T(\mu_c - \mu_i)(\mu_c - \mu_i)^T U_i^T \tag{51}$$

Since Σ_c is symmetric

$$B = A + A^T - diag(A) \tag{52}$$

we can solve for the updated iteration step for Σ_{c+t} as follows:

$$\Sigma_c^{t+1} = 2n\left[(B^{(t)}) + diag(B^{(t)})\right]^{-1} \tag{53}$$

3.3 Weighted Sided Gaussian KLD-Centroids

This section approaches another set of centroids provided by Schnitzer et al. [9], which provides information theoretic centroids based on the Kullback-Leibler divergence (KLD or KL). The optimization problems are posed in a similar manner to the Bhattacharyya and Burbea-Rao centroids. Eqs. (54) to (56) provide the minimization of the divergences (left-sided, right-sided and symmetric) upon the selection of the multiple distributions.

$$c_L = \arg\min_c \frac{1}{n}\sum_{i=1}^{n} D(c||x_i) \tag{54}$$

$$c_R = \arg\min_c \frac{1}{n}\sum_{i=1}^{n} D(x_i||c) \tag{55}$$

$$c_S = \arg\min_c \frac{1}{n}\sum_{i=1}^{n} \frac{D(x_i||c) + D(c||x_i)}{2} \tag{56}$$

The λ_i in the equations below is a weighted term, where the sum is equal to 1 ($\sum_{i=1}^{n} \lambda_i = 1$). The set of algorithms become dependent on the number of distributions. Hence, for each sided and symmetric centroid is based on the weight and the sum for the means and covariances. Equation (57) and (58) represent the righ-sided KL centroid.

$$\mu_{c_R} = \sum_{i=1}^{n} \lambda_i \mu_i \tag{57}$$

$$\Sigma_{c_R} = \sum_{i=1}^{n} \lambda_i \left(\mu_i \times \mu_i^T + \Sigma_i\right) - \mu_{c_R} \times \mu_{c_R}^T \tag{58}$$

The procedure for the left-sided centroid is slightly different, such that the Σ_{c_L} is needed first in order to compute the left-sided mean centroid. Eqs. (59) and (60) provide the centroid equations.

$$\mu_{c_L} = \Sigma_{c_L} \times \sum_{i=1}^{n} \lambda_i \left(\Sigma - i^{-1} \times \mu_i\right) \tag{59}$$

$$\Sigma_{c_L} = \left(\sum_{i=1}^{n} \lambda_i \Sigma_i^{-1}\right)^{-1} \tag{60}$$

The symmetric equations rely on the right-sided and left-sided centroids from the previous equations. The resultant equations are shown in (61) and (62).

$$\mu_{c_S} = \frac{1}{2}(\mu_{c_L} + \mu_{c_R}) \tag{61}$$

$$\Sigma_{c_S} = \frac{1}{2} \sum_{i=\{L.R\}}^{n} (\mu_{c_i} \times \mu_{c_i}^T + \Sigma_{c_i}) - \mu_{c_S} \times \mu_{c_S}^T \tag{62}$$

Practically speaking, the method from Schnitzer et al. is convenient because it does not rely on convergence, but rather is dependent on the set of distributions to analyze. This paper utilizes these methods in order to further understand the consensus or closeness between multiple decentralized implicitly updated PDFs. It it important to note that Nielsen also provides similar definitions in [8]. However, the right and left side definitions are different due to the notation used between the two authors.

4. SENSOR UPDATE AND INFORMATION DIVERGENCE CENTROID RESULTS

This work utilizes the observations made from site Toodyay as shown in Table 1 and the update is provided by the Otago, Maui and Moorook sites. The latitude longitude and altitude are shown in the 1, where Toodyay's notation S_1 represents measurements taken at t_1. While, the sites that provide update S_2, provide an update 30 minutes later at t_2. From these sites, the set of solutions obtained from angular momentum and energy provides three major distributions as shown in the left plot of Fig. 1.

Table 1: List of sites for the sensor updates. Site, latitude, longitude and altitude

Sensor Site	Lat. Long. (deg)	Alt. (m)
S_1 Toodyay	-31.52384°, 116.3362°	263.9
S_2 Otago	20.454°, 124.5545°	250
S_2 Maui	-45.87456°, 170.503388°	150
S_2 Moorook	-34.269178°, 140.341894°	35.2

This paper does not focus on the ambiguity issues developed from the solution space, where typically one would see two major set of distributions per site update from conservation equations. The distribution at the center of this analysis contains the truth, such that the vacant distribution is intuitively ruled out due to the motion of the object. Where, Fig. 1 depicts the positions of the particles and the division to the GMM composed of 6 sub-clusters for the distribution mode containing the truth state. The number of GMM clusters has no particular rule, where this number provides a good starting point to investigate the information divergence metrics on centroids.

Table 2 provides the input values for simulations, depicting the state for each site and the uncertainties associated with each site σ_α,σ_δ for optical and σ_ρ, $\sigma_{\dot{\rho}}$ for radar measurements. The set of simulations was done using 5,000 particles, for which only about half remained within the major cluster associated with the truth.

Table 2: Input Values for the simulation: Site Position,Velocities and Uncertainties

Input	Value
S_1 (km) $[x,y,z]'$	$[-5058, 2024, -3306]'$
S_1 (km/s) $[v_x, v_y, v_z]'$	$[-0.1476, -0.3683, 0.0002834]'$
S_1 (arc-sec) $[\sigma_\alpha, \sigma_\delta]'$	$[2.0, 2.0]'$
S_2 (Otago) (km) $[x,y,z]'$	$[-3775, -2368, -4548]'$
S_2 (Otago) (km/s) $[v_x, v_y, v_z]'$	$[0.1727, -0.2746, 0.0003346]'$
S_2 (Otago) (m,cm/s) $[\sigma_\rho, \sigma_{\dot{\rho}}]'$	$[30, 3.0]'$
S_2 (Maui) (km) $[x,y,z]'$	$[-2486, -5420, -2255]'$
S_2 (Maui) (km/s) $[v_x, v_y, v_z]'$	$[0.3953, -0.1816, 0.0007643]'$
S_2 (Maui) (m,cm/s) $[\sigma_\rho, \sigma_{\dot{\rho}}]'$	$[60, 6.0]'$
S_2 (Moor.) (km) $[x,y,z]'$	$[-5280, -184, -3561]'$
S_2 (Moor.) (km/s) $[v_x, v_y, v_z]'$	$[0.01342, -0.3845, 0.00002738]'$
S_2(Moor.) (m,cm/s) $[\sigma_\rho, \sigma_{\dot{\rho}}]'$	$[30, 3.0]'$

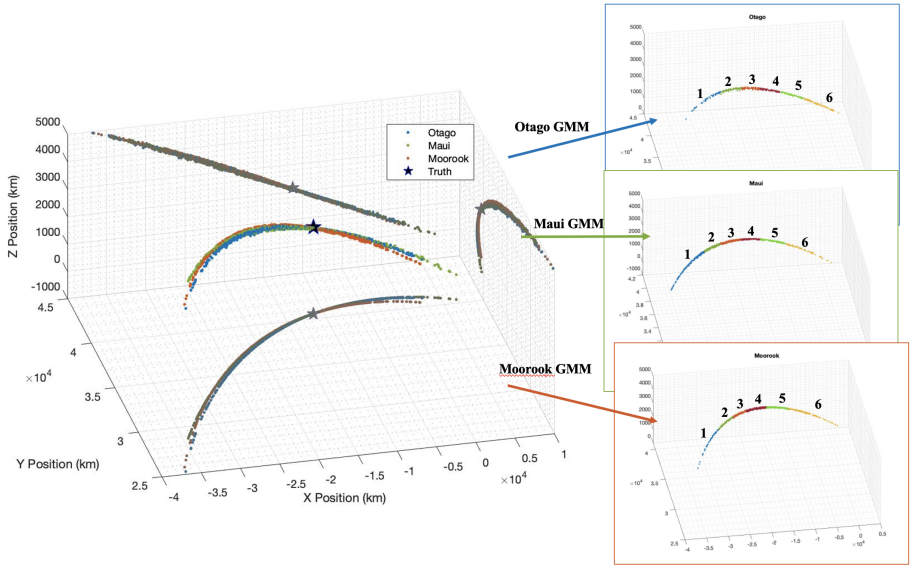

Fig. 1: Otago, Maui and Moorook Position distributions to 6 Gaussian Mixture Models

Each site update provides 6 sub-clusters to analyze and apply the divergence centroids. A first analysis is conducted along the numbered sub-clusters in an ascending order fashion. Fig. 2 depicts the selection of a sub-cluster no. 4 from Fig. 1, where each of the site updates distributions to compute the information divergence centroids. Fig. 3 depicts the computations for sub-cluster no. 5. From these results the centroid distributions carry more information from the fusion of GMMs distributions than with information divergence. In Fig. 2 and Fig. 3, the sensitivity due to each divergence metric is shown, where the Burbea-Rao and Bhattacharyya are seemingly less sensitive than most of the Kullback-Leibler right-sided and symmetrical divergences. It is important to note that the Burbea-Rao and the Kullback-Leibler left-sided divergence depict the same mean centroid and covariance. From the set of centroid distributions, the sensitivity varies across the different metrics. The set mean centroid appears to be within the same vicinity and relevant to the GMMs distributions.

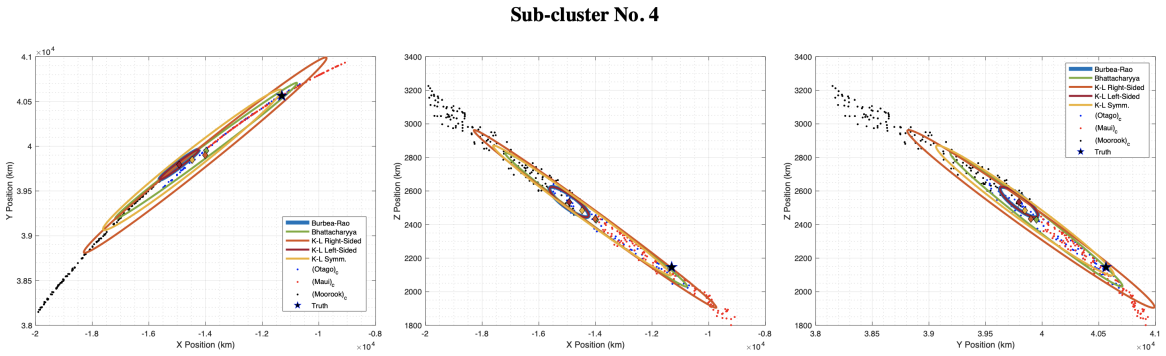

Fig. 2: Sub-cluster No. 4 shown with divergence centroids placed on top of the particles. Noting the computed centroid distribution for each metric and its center denoted by the colored diamonds.

Fig. 3 depicts the sensitivity due to the non-linear behavior of the distribution. The Burbea-Rao and Kullback-Leibler divergence portray a smaller covariance ellipse relative to the other metrics. Fig. 4 depicts this characteristic of sensitivity, where the result is relatively smaller distributions.

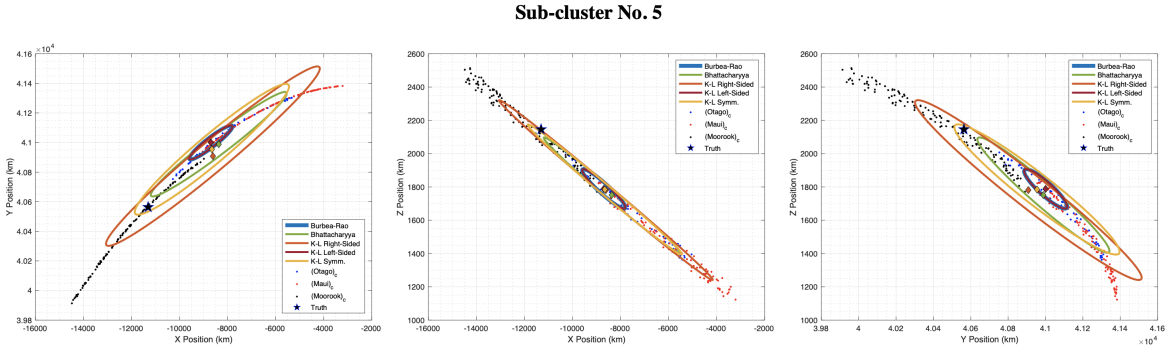

Fig. 3: Sub-cluster No. 5 shown with divergence centroids placed on top of the particles. Noting the computed centroid distribution for each metric and its center denoted by the colored diamonds. The Brubea-Rao centroid distribution overlaps the KL-Left-Sided.

In Fig. 4 shows that over the set of clusters, the Burbea-Rao and KL-left sided distribution provides a relatively more symmetrical covariance than other metrics across the set of clusters. The symmetrical characteristic may provide further insight into the sensitivity due to non-linearity. Such that Burbea-Rao and KL-left sided centroid distributions are useful to identify overlapping regions from the contributing distributions.

The utility of the metrics becomes useful when determining the sub-clusters that reside closest to the truth it becomes useful to create centroid distributions in a non-sequential order, which allows the identification of an intersection between the distributions. We analyze various sample groups from the sub-clusters and measure the closeness of the grouping using the Bhattacharyya distance, which presents a method in measuring consensus between sources.

For the first sample group, chosen in an intuitive manner due to the GMMs presented in Fig. 1 on the plots on the right. Where, the Otago sub-cluster 1 (sc. 1) would be grouped with Maui sc. 1 and Moorook sc. 1. The analysis that follows aims to measure consensus through how close the GMMs within that group to the computed centroid. This is done with all the centroid metrics in order to see potential differences and sensitivities. Fig. 5 on the top plot presents the maximum Bhattacharyya distance from all three sites GMMs for that group. The bottom plot in Fig. 5 presents the mimimum Bhattacharyya distance from the centroid to each of the GMMs. The maximum and minimum aim to present the worst case scenario and the best case scenario. In this case the best group with the lowest overall distance is Group 1. This is an interesting result, where this suggesting a better consensus present in the first region of particles. However, this does not help determine where the truth might be present in the sub-clusters.

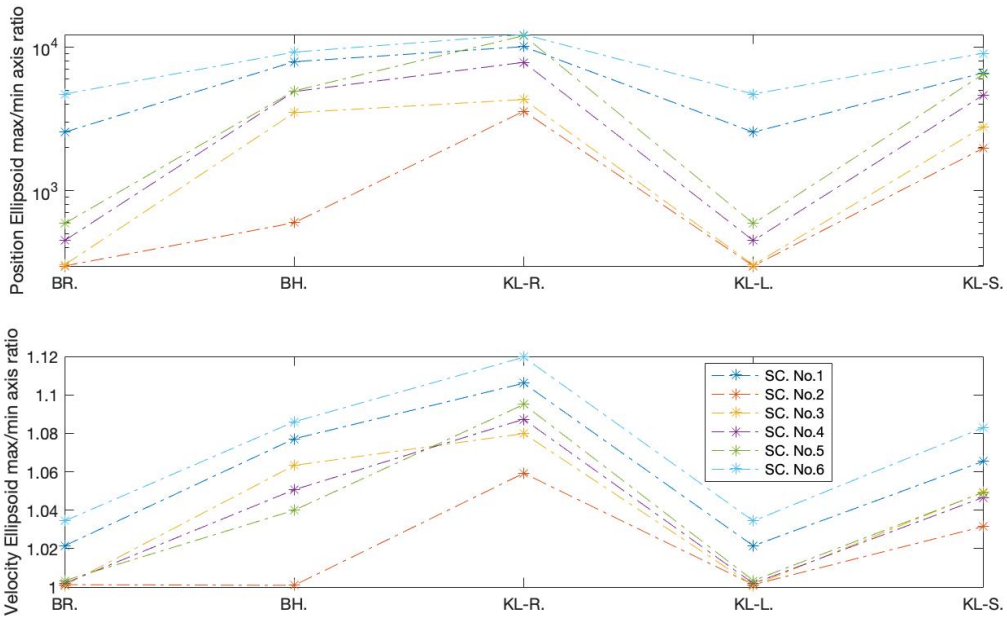

Fig. 4: The top plot provides the position centroid covariance ellipsoid ratios between largest to smallest axis for each of the sub-cluster in sequential order. The bottom depicts similar information for the velocity centroid covariance. BR.: Burbea-Rao, BH.: Bhattacharyya, KL-R: KL-right sided, KL-L: KL-left sided, KL-S: KL-Symmetrical.

In order to see if this works in determining a better consensus around the truth, a different sgroup is constructed as depicted in Fig. 6 Group 1 now represents Otago sc. 2, Maui sc. 2 and Moorook sc. 2. Similarly, for sc. 3 for all sites in Group 2 and different for Group 3 such that this represents the set of sub-clusters that is the closest to the truth. Group3 consists of Otago sc. 4, Maui sc. 4 and Moorook sc. 5 from Fig. 1 order. This particular Group 3 shows values that are the lowest distances between all groups, which hints at a better consensus around the truth. To test further the set of groups around the truth another group set is formed.

Fig. 7 provides anew set of groups, where Group 2 and Group 4 show another set of groupings in a different combination from the one discussed above. Group 3 in this case carried over from the previous group set provides further comparison with (sc. 4, sc. 4, sc. 5) combination in the Otago, Maui, Moorook order. The comparison here shows that the combination (sc. 4, sc. 4, sc. 5) demonstrates the lowest Bhattacharyya distance between the GMMs of the group and the centroid. The various centroid metrics show the same result, where this combination depicts a minima in this case. From these results one can start formulating a problem towards the search of a minima in the statistical distance after applying the information divergence centroid to groupings.

There are certainly some sensitivities due to each measure as discussed earlier, where the centroid metrics produce various distributions. The KL-right, KL-symmetrical set of centroid measures appear to be less sensitive across, when looking at Fig. 5, Fig. 6, and Fig. 7. Where the maximum and minimum values stay within a certain margin, while the Burbea-Rao, Bhattacharyya and KL-left provide greater sensitivity across. This might provide greater help in determining a minima. In Fig. 5 when observing Bhattacharyya, Burbea-Rao and KL-left centroid maximum and minimum values it can be easily seen that the minima is presented in the range between Group 3 to Group 5. Where in that group set the grouping is homogeneous (Group 3: sc. 3, sc. 3, sc. 3) to (Group 5: sc. 5, sc. 5, sc. 5).

Table 3 provides the values that are protrayed in the maximum and minimum figures, such that some of the minima can be difficult to distinguish on the scales. From this table it is shown that the sc. 4, sc. 4, sc. 5 has the carries the least amount of distance and thus better grouping. This leads to the centroid application towards consensus and obtain a set of particles that is closer to the truth.

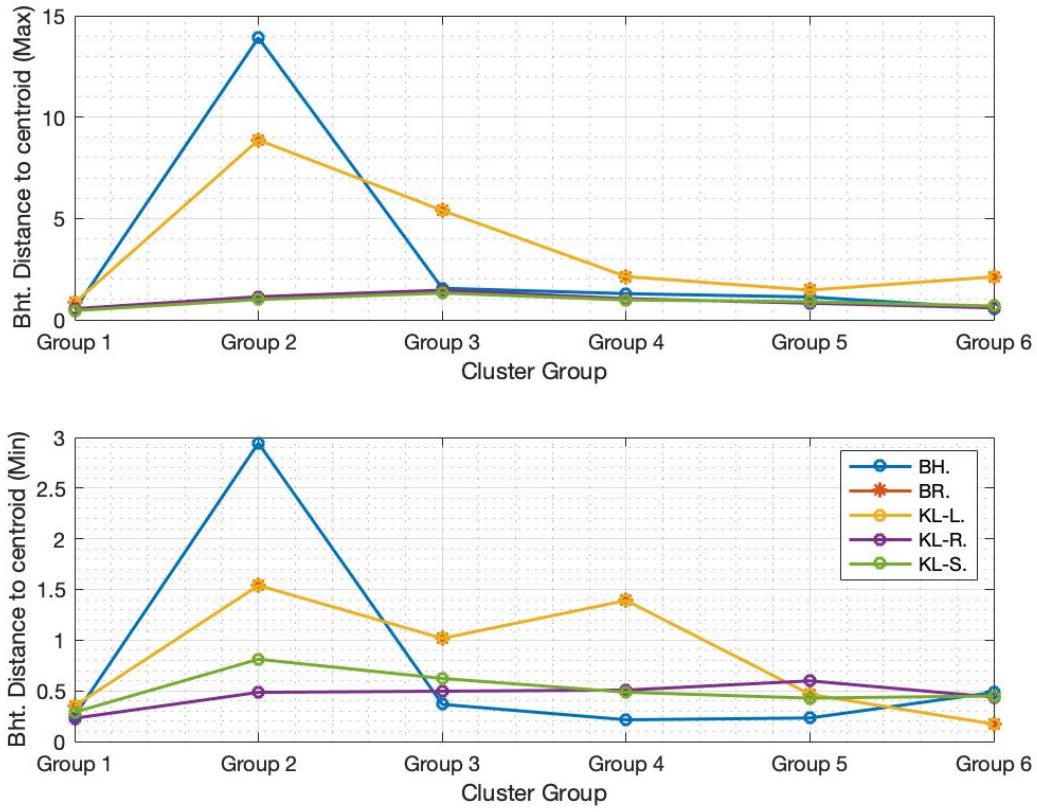

Fig. 5: The top plot depicts the maximum Bhattacharyya distance from the centroid metric to the GMM group. Bottom plot depicts similar information with minimum Bhattacharyya distance to the GMM group. Note that in this groups the set is homogeneous with (sc. 1, sc.1, sc. 1) on the Otago, Maui, Moorook order for Group 1 and (sc. 2, sc. 2, sc. 2) for Group 2 and so on for each group in this set. The Bhattacharyya metric is presented in blue, Burbea-Rao in red, KL-Left Sided in yellow, KL-Right sided in purple and KL-symmetric in green.

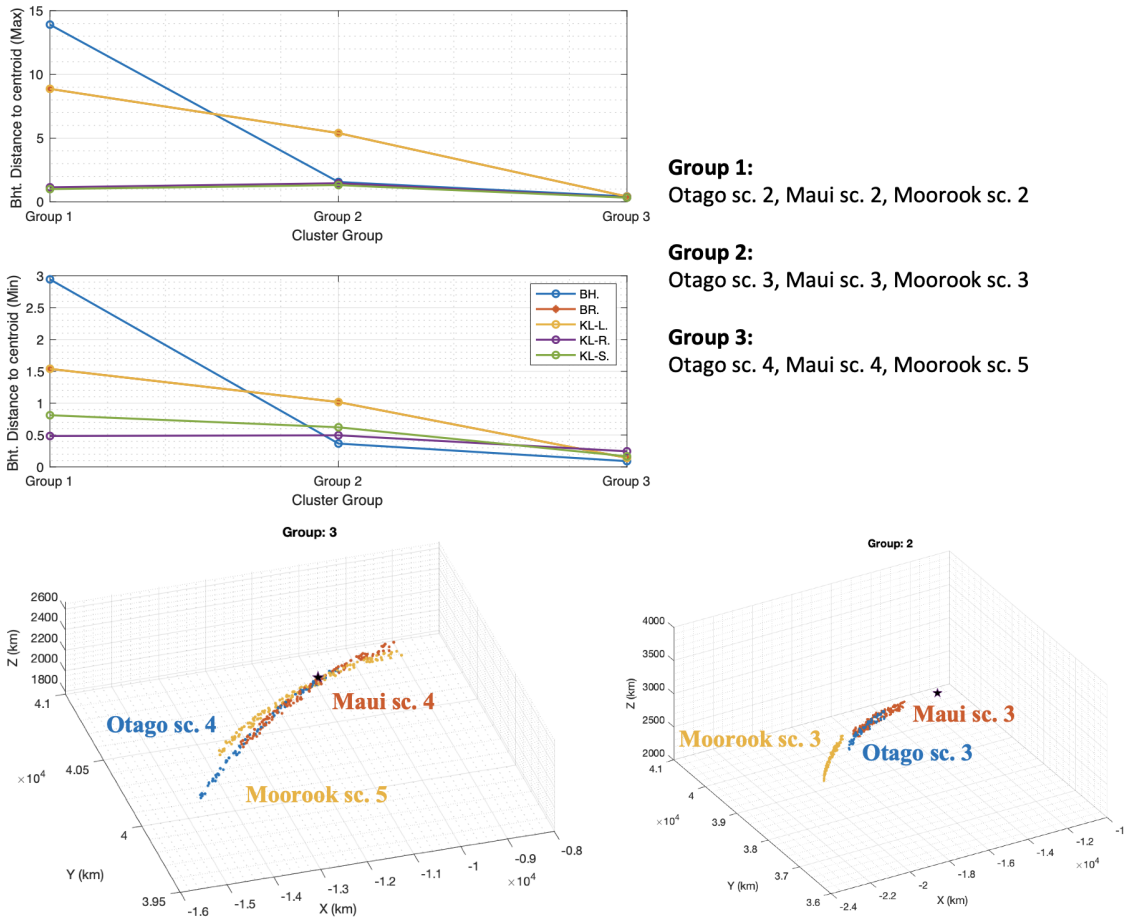

Fig. 6: Plots depicting the maximum and minimum Bhattacharyya distance from centroid to each GMM component within each group. Bottom clusters depict where the truth lies in the set of cluster groups denoted by the star. Group set to depict the decrease in Bhattacharyya distance.

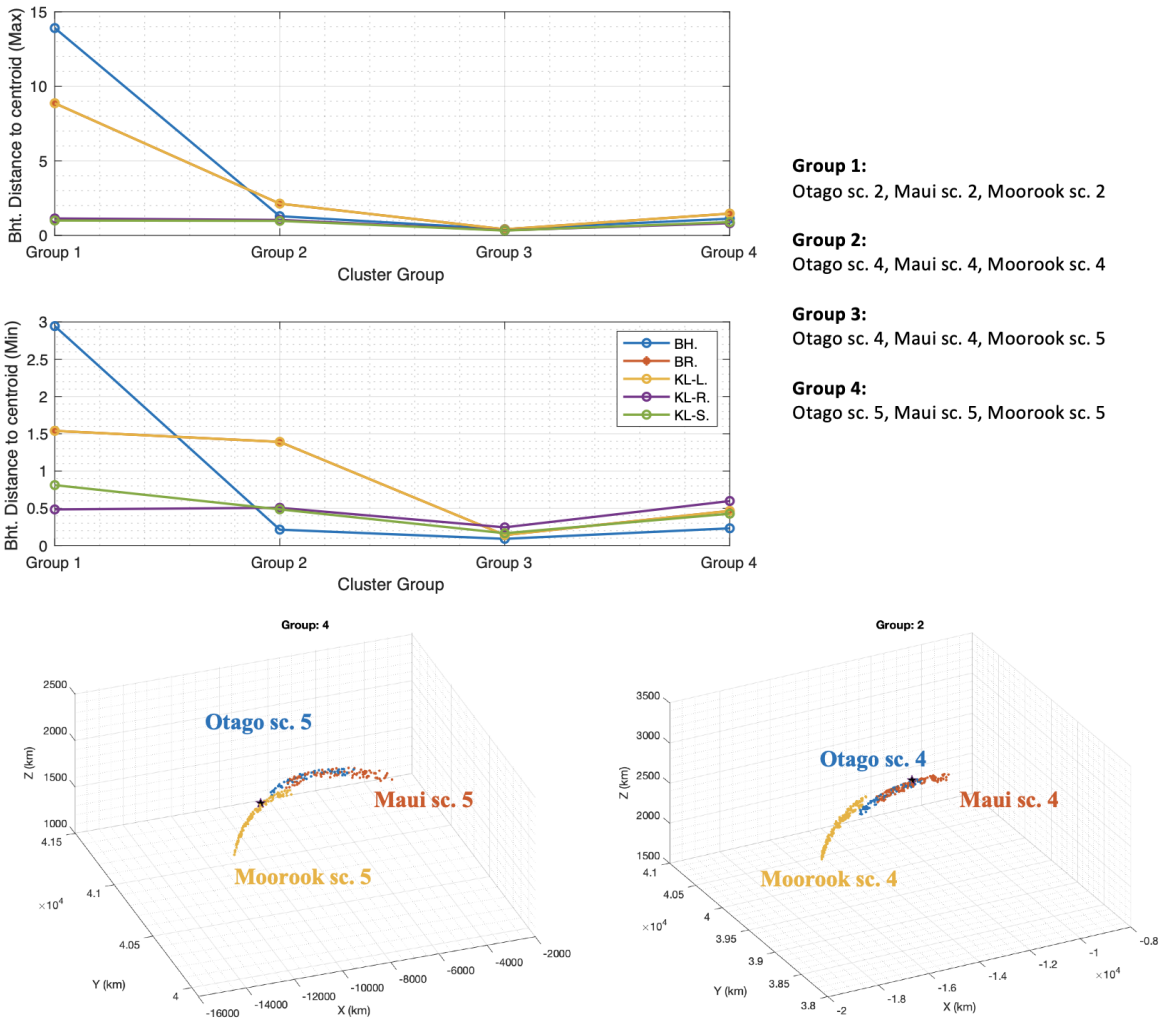

Fig. 7: Set of groups chosen to depict the minimum in Bhattacharyya distances to the centroids for each group. Noting that the a minimum is shown for group made up of Otago sc. 4, Maui sc. 4 and Moorook sc. 5.

Table 3: Sample of Bhattacharyya centroid metric for the Bhattacharyya distance values to the set of groups. The highlighted row shows the lowest maximum and minimum.

Grouping (Otago, Maui, Moorook)	Bht. Distance to centroid			max.	min.
1,1,1	0.2524	0.2764	0.4956	0.4956	0.2524
3,3,3	0.3666	0.6941	1.5587	1.5587	0.3666
4,4,4	0.2143	0.7213	1.2914	1.2914	0.2143
4,4,5	0.4238	0.0906	0.1643	0.4238	0.0906
5,5,4	1.1572	1.326	1.8386	1.8386	1.1572
5,5,5	0.2322	0.6205	1.1262	1.1262	0.2322

5. CONCLUSION AND IMPROVEMENTS

The results presented in this paper establish an initial analysis for methods of applying consensus. The decentralized update approach from a first optical measurement to three separate radar updates generates set of particles, which were be broken up using GMM algorithm. The GMM splitting performed by Matlab does not imply the best algorithm for this case. Further research must be attributed into implementing GMM splitting algorithm forming a new set of groups based on the PDF mode analyzed. The case presented focuses on the consensus between the three separate set of particles and attempt to fuse the information to yield a region closer to the truth. We utilize the convexity of information divergence in order to provide this fusion of information carrying distribution information, which is then carried to obtain a measure of distribution grouping. The results suggest that a minimum Bhattacharyya distance from a selected group of GMM components and its computed distribution centroid carried the region closest to the truth. However, this requires additional set of simulations with increased number of particles as some of the results hinted the presence other local minima that does not necessarily represent region nearest to the truth. For object's orbit yielded promising results, however there still remains to explore a vast set of orbits and sensor locations to fully understand and test the approach to consensus.

6. REFERENCES

[1] Syed Mumtaz Ali and Samuel D Silvey. A general class of coefficients of divergence of one distribution from another. *Journal of the Royal Statistical Society: Series B (Methodological)*, 28(1):131–142, 1966.

[2] Shun-ichi Amari. Integration of stochastic models by minimizing α-divergence. *Neural computation*, 19(10):2780–2796, 2007.

[3] Jacob Burbea and C Radhakrishna Rao. Entropy differential metric, distance and divergence measures in probability spaces: A unified approach. *Journal of Multivariate Analysis*, 12(4):575–596, 1982.

[4] Imre Csiszár. Information-type measures of difference of probability distributions and indirect observation. *studia scientiarum Mathematicarum Hungarica*, 2:229–318, 1967.

[5] White House. Space policy directive-3, national space traffic management policy. *US Government*, 18(06), 2018.

[6] Frank Nielsen. On the jensen–shannon symmetrization of distances relying on abstract means. *Entropy*, 21(5):485, 2019.

[7] Frank Nielsen and Sylvain Boltz. The burbea-rao and bhattacharyya centroids. *IEEE Transactions on Information Theory*, 57(8):5455–5466, 2011.

[8] Frank Nielsen and Richard Nock. Sided and symmetrized bregman centroids. *IEEE transactions on Information Theory*, 55(6):2882–2904, 2009.

[9] Dominik Schnitzer, Arthur Flexer, Gerhard Widmer, and Martin Gasser. Islands of gaussians: The self organizing map and gaussian music similarity features. In *ISMIR*, pages 327–332. Citeseer, 2010.

[10] Bharath K Sriperumbudur and Gert RG Lanckriet. On the convergence of the concave-convex procedure. In *Nips*, volume 9, pages 1759–1767. Citeseer, 2009.

[11] Tim Van Erven and Peter Harremos. Rényi divergence and kullback-leibler divergence. *IEEE Transactions on Information Theory*, 60(7):3797–3820, 2014.

[12] Alan L Yuille and Anand Rangarajan. The concave-convex procedure. *Neural computation*, 15(4):915–936, 2003.

[13] Waqar H Zaidi, Weston R Faber, Islam I Hussein, Michael Mercurio, Christopher WT Roscoe, Matthew P Wilkins, and Paul W Schumacher Jr. Debris object orbit initialization using the probabilistic admissible region with asynchronous heterogeneous observations. In *Advanced Maui Optical and Space Surveillance (AMOS) Technologies Conference*, page 126, 2017.