AAS 23-107

DATA-DRIVEN SUMMARY OF NATURAL SPACECRAFT TRAJECTORIES IN THE EARTH-MOON SYSTEM

Natasha Bosanac*

Trajectory designers and ground operators may encounter situations where digestible summaries of a wide array of possible spacecraft motions within cislunar space could be valuable for decision-making. However, the diversity of trajectories within this chaotic gravitational environment renders manual analysis a time-consuming and overwhelming task. To address this challenge, this paper uses a clustering-based framework to summarize the distinct geometries exhibited by a large set of spacecraft trajectories with short to medium flight times in the Earth-Moon system. This approach is applied to trajectories generated in both the circular restricted three-body problem and an ephemeris model of cislunar space.

INTRODUCTION

In the near future, an increasing number of spacecraft are expected to operate in cislunar space. Accordingly, trajectory designers will need to design mission orbits and transfers to achieve a wide variety of objectives. Simultaneously, analysts will need to regularly predict the possible future motions of space objects for space domain awareness and collision avoidance. Both of these astrodynamics tasks require an initial understanding of the wide array of spacecraft motions that are possible within the complex and chaotic dynamical environment of cislunar space.

In the Earth-Moon system, a low-fidelity dynamical model such as the circular restricted three-body problem (CR3BP) is often used to gain preliminary insight into the solution space. This model is autonomous when formulated in an Earth-Moon rotating frame and admits fundamental solutions governing the flow throughout the system.[1] Some fundamental solutions and comparable finite-time solutions may also exist in other approximate dynamical models such as the elliptic restricted three-body problem and the bicircular restricted four-body problem. It is in these lower-fidelity models that designers and analysts use dynamical systems techniques to rapidly gain valuable insights into a wide array of bounded motions and natural transport mechanisms.[2] However, in higher fidelity models such as an ephemeris model, these fundamental solutions and/or comparable finite-time solutions no longer exist and only some may be approximately retained. Directly visualizing and extracting meaningful insight from a wide array of epoch-dependent motions that are generated in an ephemeris model may be time-consuming and overwhelming, even when using dynamical systems techniques such as Poincaré mapping.

---

*Assistant Professor, Colorado Center for Astrodynamics Research, Smead Department of Aerospace Engineering Sciences, University of Colorado Boulder, Boulder, CO 80303.

DISTRIBUTION A: Approved for public release; distribution is unlimited. Public Affairs release approval #AFRL-2023-3637. The views expressed are those of the author and do not reflect the official guidance or position of the United States Government, the Department of Defense, or of the United States Air Force.

A wide variety of technical disciplines encounter a similar problem: large amounts of high-dimensional information may be generated through simulations or experiments and may be difficult or time-consuming for a human to analyze or extract meaningful insights. In these 'big data' problems, data mining techniques have proven valuable. For instance, clustering algorithms construct an unsupervised grouping of similar data in a finite-dimensional feature space while also separating dissimilar data.[5] Clustering has also been used to group trajectories that are either sampled to produce a time sequence description or characterized by a model.[32] Dimension reduction algorithms may be used to further reduce the dimension of the description.[20] When a clustering algorithm is applied to a set of data, the resulting clusters may supply a digestible summary that supports further analysis or decision-making. Classification schemes, a form of supervised learning, may then be used to associate new or unlabeled data with an existing cluster.[5] These data mining techniques have been used for analysis and knowledge discovery in various technical disciplines such as astronomy, medicine, studying driving routes, and air traffic management.[8–11,32] In astrodynamics, these techniques have been used to detect regions of bounded motions near distant retrograde orbits on a Poincaré map, group periodic orbits with similar characteristics, and extract motion primitive sets that summarize families of periodic orbits and their hyperbolic invariant manifolds.[12–15]

Bosanac as well as Bonasera and Bosanac have previously developed a clustering framework for extracting a summary of spacecraft trajectories in the Sun-Earth CR3BP.[16–18] First, a large set of trajectories were generated from initial perigees at a single energy level for a few revolutions around the Earth. In a curve-based discretization scheme, each trajectory was then sampled at several apses to construct a finite-dimensional feature vector.[16] To accommodate a large dataset, these trajectories were separated into partitions. Each partition was then independently clustered using Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) algorithm to extract groups of trajectories with a similar geometry.[16] The clustering results from each partition were sampled and aggregated in a binary tree structure that employs both clustering and dimension reduction via manifold learning.[17] The result was a global cluster summary that supplied a digestible and data-driven summary of the geometries of a wide array of trajectories that exist at a specific energy level in the Sun-Earth CR3BP while also supporting association to natural transport mechanisms.[17] This paper leverages and builds upon this prior work.

To gain insight into some of the possible motions of a spacecraft operating in cislunar space, this paper uses a clustering framework to extract a summary of natural spacecraft trajectories at a single energy level. Specifically, a wide variety of trajectories are generated from prograde perilunes to span flight times of up to 21 days in each of the Earth-Moon CR3BP and an ephemeris model of cislunar space. To capture the shape of nonlinear trajectories that visit various regions of the Earth-Moon system, the concept of curvature from differential geometry is employed to introduce an updated approach to sampling and describing each trajectory using a finite-dimensional feature vector. Then, HDBSCAN is used along with the cluster aggregation procedure presented by Bonasera and Bosanac to group these feature vectors and, therefore, construct clusters of geometrically similar trajectories.[17] This framework is used to extract a digestible summary of the trajectories generated at an energy level where both the L_1 and L_2 gateways open and, therefore, the solution space is diverse. These summaries are constructed in each of the two models of the Earth-Moon system. Accordingly, this paper offers the following original contributions: introducing a new trajectory sampling and description scheme, applying this framework to the Earth-Moon system, and constructing a data-driven summary of spatial trajectories in an ephemeris model.

DISTRIBUTION A: Approved for public release; distribution is unlimited. Public Affairs release approval #AFRL-2023-3637.

DYNAMICAL MODELS

Circular Restricted Three-Body Problem

The circular restricted three-body problem (CR3BP) is used to approximate the motion of a spacecraft in cislunar space. In this model, the Earth and Moon are modeled as spherically symmetric masses following circular orbits around their barycenter. In addition, the spacecraft is assumed to possess a comparatively negligible mass.[1]

To describe the state, time, and mass parameters of this system, a normalization scheme is often employed.[1,2] Length quantities are normalized by the assumed constant distance between the Earth and Moon whereas mass quantities are normalized by the total mass of the system. Time quantities, however, are normalized to produce a mean motion of the Earth and Moon that is equal to unity.

To formulate an autonomous dynamical system and facilitate visualization, the equations of motion are specified in a rotating frame. In the Earth-Moon rotating frame, the origin is located at the barycenter of the system and the three axes $\hat{x}\hat{y}\hat{z}$ are defined as follows: $\hat{x}$ is directed from the Earth to Moon, $\hat{z}$ is aligned with the orbital angular momentum vector of the Earth-Moon system, and $\hat{y}$ completes the orthogonal, right-handed triad. In this rotating frame, the nondimensional state of the spacecraft is specified as $\bar{x} = [x, y, z, \dot{x}, \dot{y}, \dot{z}]^T$ where $\dot{(.)}$ indicates a time derivative with respect to an observer in the rotating frame. The nondimensional equations of motion governing the spacecraft are then written in this rotating frame as

$$\ddot{x} - 2\dot{y} = \frac{\partial U^*}{\partial x}, \qquad \ddot{y} + 2\dot{x} = \frac{\partial U^*}{\partial y}, \qquad \ddot{z} = \frac{\partial U^*}{\partial z} \tag{1}$$

where μ is the mass ratio comparing the Moon's mass to the total mass of the system, $r_1 = \sqrt{(x+\mu)^2 + y^2 + z^2}$, $r_2 = \sqrt{(x-1+\mu)^2 + y^2 + z^2}$, and the pseudo-potential function is $U^* = (x^2 + y^2)/2 + (1-\mu)/r_1 + \mu/r_2$.[1] These equations of motion produce one integral of motion, the Jacobi constant, that is equal to $C_J = 2U^* - \dot{x}^2 - \dot{y}^2 - \dot{z}^2$. At a single value of C_J, zero velocity surfaces capture the set of position vectors that correspond to a zero velocity vector and bound the allowable motions of a spacecraft in the CR3BP.[2] Within these zero velocity surfaces, the chaotic solution space can be composed of a diverse array of trajectories.

Ephemeris Model

An ephemeris model supplies a higher-fidelity representation of the dynamical environment in cislunar space. In this paper, this ephemeris model is constructed to include only the point mass gravitational influences of the Earth, Moon, and Sun on the spacecraft. This level of fidelity captures the effects of the most dominant gravitational bodies traveling on their true, non-circular paths while avoiding the significant increases in computational time associated with higher-order gravity models or parameter dependencies in solar radiation pressure models, for example.

In the ephemeris model, the state of the spacecraft is described using two reference frames. The Moon-centered inertial frame is defined using the center of the Moon as the origin and the axes $\hat{X}\hat{Y}\hat{Z}$ of the International Celestial Reference Frame (ICRF).[21,22] This frame is used to formulate the equations of motion and numerically integrate the path of the spacecraft. In addition, the pulsating Earth-Moon rotating frame is defined using the same origin and mathematical definition of the axes $\hat{x}\hat{y}\hat{z}$ as the rotating frame used in the CR3BP. However, this frame is labeled as 'pulsating' because the scale is defined using the instantaneous distance between the Earth and Moon to ensure both celestial bodies possess fixed locations along the $\hat{x}$-axis.

DISTRIBUTION A: Approved for public release; distribution is unlimited. Public Affairs release approval #AFRL-2023-3637.

The equations of motion for the spacecraft are written in nondimensional form in the Moon inertial frame. Consistent with the CR3BP, the spacecraft is assumed to possess a negligible mass compared to the three celestial bodies. Length, mass, and time quantities are also normalized using the same scheme as the CR3BP. However, the values for the characteristic quantities used in normalization are time-dependent.[25] The state of the spacecraft is then expressed in the Moon-centered inertial frame as $\bar{X} = [X, Y, Z, \dot{X}, \dot{Y}, \dot{Z}]^T = [\bar{R}_{L,sc}^T, \bar{V}_{L,sc}^T]^T$. Using these definitions, the nondimensional equations of motion for the spacecraft are written as

$$\ddot{\bar{R}}_{L,sc} = -GM_L \left(\frac{\bar{R}_{L,sc}}{R_{L,sc}^3} \right) + G \sum_{i=E,S} M_i \left(\frac{\bar{R}_{sc,i}}{R_{sc,i}^3} - \frac{\bar{R}_{L,i}}{R_{L,i}^3} \right) \tag{2}$$

where subscripts L, E, S, and sc indicate the Moon, Earth, Sun, and spacecraft; M_i is the mass of body i; $\bar{R}_{i,j}$ locates body j relative to body i in the inertial frame; and, in this expression, $\dot{()}$ indicates a derivative with respect to time for an observer in the inertial frame.[26] To evaluate the position vectors of the celestial bodies, the DE440 lunar and planetary ephemerides, provided by NASA's Navigation and Ancilliary Information Facility, are used along with the SPICE toolkit.[3,4]

Although the state of the spacecraft is integrated in the inertial frame, a frame transformation is employed to describe and visualize each trajectory in the pulsating Earth-Moon rotating frame. At an epoch t, the transformation from a position vector $\bar{R}_{L,sc}(t)$ in the Moon-centered inertial frame to a position vector $\bar{r}_{B,sc}(t)$ in the pulsating rotating frame is calculated as

$$\bar{r}_{B,sc}(t) = \left[{}^R C(t)^I \right] \bar{R}_{L,sc}(t) + \bar{r}_{B,L} \tag{3}$$

where $\bar{r}_{i,j}$ locates body j relative to body i in the rotating frame and the subscript B corresponds to the Earth-Moon barycenter.[21,24] In this expression, the rotation matrix $[{}^R C(t)^I]$ is equal to

$$\left[{}^R C(t)^I \right] = \begin{bmatrix} \hat{x}^T(t) \\ \hat{y}^T(t) \\ \hat{z}^T(t) \end{bmatrix} \tag{4}$$

where $\hat{x}, \hat{y}, \hat{z}$ are column vectors of the axes of the Earth-Moon rotating frame expressed in the Moon inertial frame.[21,24] These unit vectors are calculated as

$$\hat{x}(t) = \frac{\bar{R}_{E,L}(t)}{||\bar{R}_{E,L}(t)||} \qquad \hat{y}(t) = \hat{z}(t) \times \hat{x}(t) \qquad \hat{z}(t) = \frac{\bar{R}_{E,L}(t) \times \bar{V}_{E,L}(t)}{||\bar{R}_{E,L}(t) \times \bar{V}_{E,L}(t)||} \tag{5}$$

where $\bar{R}_{E,L}(t)$ and $\bar{V}_{E,L}(t)$ are the position and velocity vectors of the Moon relative to the Earth in the Moon-inertial frame at the associated epoch.[21,24] This transformation may be inverted to transform a position vector from the rotating frame to the inertial frame.

To transform the spacecraft velocity and acceleration vectors from the Moon inertial frame to the pulsating Earth-Moon rotating frame, the time derivatives of the axes of the rotating frame are used. Specifically, the transformation for the velocity vector is written as

$$\bar{v}_{B,sc}(t) = \left[{}^R \dot{C}(t)^I \right] \bar{R}_{L,sc}(t) + \left[{}^R C(t)^I \right] \bar{V}_{L,sc}(t) \tag{6}$$

assuming that the Moon's location is instantaneously fixed along the $\hat{x}$ axis. The transformation for the acceleration vector from $\bar{A}_{L,sc}$ in the inertial frame to $\bar{a}_{B,sc}(t)$ in the rotating frame is equal to

$$\bar{a}_{B,sc}(t) = \left[{}^R \ddot{C}(t)^I \right] \bar{R}_{L,sc}(t) + 2 \left[{}^R \dot{C}(t)^I \right] \bar{V}_{L,sc}(t) + \left[{}^R C(t)^I \right] \bar{A}_{L,sc} \tag{7}$$

DISTRIBUTION A: Approved for public release; distribution is unlimited. Public Affairs release approval #AFRL-2023-3637.

To calculate the first and second time derivatives of the rotation matrix, the associated derivatives of the $\hat{x}\hat{y}\hat{z}$ axes are also calculated. The first derivatives are equal to[21,24]

$$\dot{\hat{x}}(t) = \frac{\bar{V}_{L,sc}(t)}{||\bar{R}_{L,sc}(t)||} - \hat{x}\frac{\hat{x} \cdot \bar{V}_{L,sc}(t)}{||\bar{R}_{L,sc}(t)||} \qquad \dot{\hat{z}}(t) \approx 0 \qquad \dot{\hat{y}}(t) = \hat{z} \times \dot{\hat{x}} \tag{8}$$

The second derivatives of the axes of the rotating frame are equal to

$$\ddot{\hat{x}}(t) = \frac{\bar{A}_{L,sc}}{||\bar{R}_{L,sc}||} - \frac{\bar{V}_{L,sc}(\bar{V}_{L,sc} \cdot \hat{x})}{||\bar{R}_{L,sc}||^2} - \frac{\dot{\hat{x}}(\bar{V}_{L,sc} \cdot \hat{x})}{||\bar{R}_{L,sc}||} - \hat{x}\left(\frac{\bar{A}_{L,sc} \cdot \hat{x}}{||\bar{R}_{L,sc}||} + \frac{\bar{V}_{L,sc} \cdot \dot{\hat{x}}}{||\bar{R}_{L,sc}||} - \frac{(\bar{V}_{L,sc}(t) \cdot \hat{x})^2}{||\bar{R}_{L,sc}(t)||^2}\right) \tag{9}$$

$$\ddot{\hat{y}}(t) = \hat{z} \times \ddot{\hat{x}} \tag{10}$$

$$\ddot{\hat{z}}(t) \approx 0 \tag{11}$$

The assumption that $\dot{\hat{z}}(t) \approx 0$ is also employed in NASA Goddard Space Flight Center's General Mission Analysis Tool.[23]

CLUSTERING

Clustering algorithms perform an unsupervised grouping of members of a dataset based on their similarity in a specified feature space. These algorithms are often categorized as partition-based, grid-based, density-based, hierarchical, and/or model-based.[27] Furthermore, hard clustering algorithms uniquely assign each member to a group whereas soft or fuzzy clustering algorithms assign each member a probability of membership in each group.[27] The type of clustering algorithm employed for exploratory data analysis in a specific application may depend on the type of data, available computational resources, and a priori knowledge of the structure of the dataset.

Hierarchical Density-Based Spatial Clustering of Applications with Noise

To cluster a diverse dataset describing nonlinear trajectories, this paper employs the Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) algorithm that was developed by Campello, Moulavi, and Sander.[19] HDBSCAN is used in this paper because it does not require a prior knowledge of the number of clusters; can discover clusters of various shapes, densities, and distances; and can label members of a dataset that lie in insufficiently dense regions as noise points. Bosanac as well as Bonasera and Bosanac have demonstrated that this algorithm can extract the distinct geometries of trajectories near the Earth in the Sun-Earth CR3BP.[16,17]

HDBSCAN is a hierarchical and density-based clustering algorithm that labels the N members of a dataset, each described by an M-dimensional feature vector $\bar{f}_i$ for $i = [1, N]$, as belonging to a specific cluster or as a noise point.[19] First, HDBSCAN calculates the core distance of each member, defined as the distance to its $(m_{pts} - 1)$-th nearest neighbor in the M-dimensional feature space. Then, HDBSCAN calculates the mutual reachability distance between each pair of members of the dataset. When calculated between the i-th and j-th member, this mutual reachability distance is defined as $d_{reach}(\bar{f}_i, \bar{f}_j) = \max(d_{core}(\bar{f}_i), d_{core}(\bar{f}_j), d(\bar{f}_i, \bar{f}_j))$ where $d(\bar{f}_i, \bar{f}_j)$ is the distance between the feature vectors of the two members, assessed using a specified distance measure. The mutual reachability distance between all pairs of members of the dataset is used to form a weighted graph: the nodes are the members of the dataset and the edge weights are the associated mutual reachability distances. This graph is summarized by a minimum spanning tree (MST), with a self-loop added to each node. From this extended MST, a hierarchical representation of possible clusters

DISTRIBUTION A: Approved for public release; distribution is unlimited. Public Affairs release approval #AFRL-2023-3637.

is extracted. Using this cluster hierarchy, groupings that are sufficiently stable as a function of distance and possess at least m_{clmin} members are selected as clusters; members that do not satisfy this criteria are labeled as noise points. A modification presented by Malzer and Baum merges clusters that are separated by a distance that falls below a specified threshold, ϵ_{merge}.[29] In this paper, HDBSCAN is accessed in Python using the *hdbscan* library developed by McInnes, Healy, and Astels.[28] For a more detailed discussion of the mathematical foundations of HDBSCAN, see Campello, Moulavi, and Sander 2013.[19]

Distributed Clustering

Distributed clustering enables the problem of grouping the members of a large dataset to be solved in a computationally tractable manner. Common approaches to distributed clustering tend to follow four general steps.[27] First, the dataset is partitioned into multiple, smaller datasets that are independently clustered to produce local clustering results. Then, these local clustering results are subsampled. The summarized representations of each local clustering result are aggregated to form a global cluster summary, composed of a subset of representative members from the original, large dataset. The resulting summary can be returned to each local clustering result to either improve the original clusters or label any data that does not appear in the global cluster summary.

Trajectory Clustering

Trajectory clustering is a form of time-series clustering that is applied to the evolution of continuous or discrete variables describing the movement of an object over time.[30] Common approaches to describing these trajectories involve 1) constructing a lower dimensional model of each trajectory or 2) summarizing each trajectory using a sequence of relevant variables that evolve over time. Models of trajectories may describe its shape via a bounding object, construct a function approximation, or perform a frequency transformation. Alternatively, describing each trajectory by a sequence of coordinates or other relevant variables involves selecting the resolution and sampling scheme to ensure the time sequences sufficiently represents the trajectory without excess storage requirements.[32] If the trajectory is discretized into a time sequence, there are two common approaches to clustering this higher-dimensional dataset. One approach discretizes the trajectory into subsegments, clusters those subsegments, and identifies each trajectory by its sequence of subsegments; this is the approach used by the well-known TRACLUS algorithm.[31] Another approach simply clusters the higher-dimensional description of the trajectory using a desired distance measure.

Curse of Dimensionality

When the members of a dataset are described in a high-dimensional feature space, the clustering process can suffer from the well-known 'curse of dimensionality'.[27] The most relevant consequences to the implementation in this paper include 1) the decreased relative importance of subsequent dimensions when calculating the distance between two objects using a Euclidean distance, 2) selecting the most important features to describe a dataset to limit the total number of dimensions, and 3) discovering sufficiently dense regions in a high-dimensional space when using density-based clustering algorithms. Unfortunately, there are no generalizable solutions to the 'curse of dimensionality'. One common approach, dimension reduction, constructs an embedding of a dataset onto a lower-dimensional space. However, this approach can also alter the groupings discovered by a clustering algorithm and must be used with caution.[33] In this paper, dimension reduction is not employed due to its tendency to require parameter tuning to avoid extracting either a small number

DISTRIBUTION A: Approved for public release; distribution is unlimited. Public Affairs release approval #AFRL-2023-3637.

of excessively large clusters or an excessive number of small clusters. Accordingly, the clustering of a dataset in a high-dimensional feature space via HDBSCAN may produce some noise points in regions of insufficient density or near the boundaries of some clusters.

TECHNICAL APPROACH

To generate a data-driven summary of trajectories that begin near the Moon in the Earth-Moon system, this paper uses a framework that builds upon prior work by Bosanac as well as Bonasera and Bosanac.[16,17] This framework leverages distributed clustering to group trajectories by their geometry in an unsupervised manner. However, to accommodate trajectories that may depart the Moon vicinity, a new trajectory summarization and description approach is presented. An overview of the entire framework is presented in this section with brief examples.

Step 1: Generate and Partition Trajectory Dataset

Trajectories are generated from initial conditions that are defined using apses relative to the Moon and partitioned into smaller subsets. In this paper, the focus is on prograde initial conditions that begin between the L_1 and L_2 gateways and lie within the zero velocity surfaces that are defined in the CR3BP at a specified Jacobi constant, $C_{J,d}$. Although there are a variety of options for initial condition definition, apses offer an intuitive interpretation as well as a straightforward calculation approach across a larger dataset. To support a distributed clustering approach that reduces time and computational expense, these apses are generated in smaller partitions. Within the i-th partition, a grid of initial conditions is defined at fixed values of the z-coordinate z_i and out-of-plane angle of the velocity vector θ_i. The definition of these initial conditions follows the approach presented by Bonasera and Bosanac.[17] Then, trajectories are generated from each initial condition over a specified time horizon by numerical integration in either the CR3BP or the ephemeris model.

Within the i-th partition, a grid of initial conditions is defined in the Earth-Moon rotating frame. This grid specifies combinations of N_x and N_y nondimensional x and y coordinates in the Earth-Moon rotating frame that are evenly spaced within the following ranges: $x \in [x_{min}, x_{max}]$, $y \in [y_{min}, y_{max}]$. For the j-th combination of these two position variables along with the specified value of z_i, the position vector $\bar{r}_{i,j}$ is defined. At the specified value of $C_{J,d}$, the Jacobi constant expression is rearranged to calculate the speed as $v_{i,j} = \sqrt{2U^*(\bar{r}_{i,j}) - C_{J,d}}$. If $v_{i,j}$ is real-valued, the spacecraft lies within the zero velocity surfaces associated with $C_{J,d}$ in the CR3BP at that location and the position vector is used to compute an initial condition.

For each viable position vector $\bar{r}_{i,j}$ within the i-th partition, the velocity vector $\bar{v}_{i,j}$ is calculated to produce a prograde apse relative to the Moon with a Jacobi constant of $C_{J,d}$. This velocity vector must satisfy the following mathematical expression in the rotating frame:

$$(x_{i,j} - 1 + \mu)\dot{x}_{i,j} + y_{i,j}\dot{y}_{i,j} + z_{i,j}\dot{z}_{i,j} = 0 \tag{12}$$

However, the solution to this expression lies in a two-dimensional nullspace. Accordingly, the out-of-plane angle of the velocity vector θ_i serves as an additional constraint. This angle is defined using two basis vectors for the nullspace of Equation 12: $\hat{u}_{1,i,j}$ lies in the xy-plane and produces a specific angular momentum unit vector with a positive z-component whereas $\hat{u}_{2,i,j}$ is perpendicular to $\hat{u}_{1,i,j}$ with a positive z-component. Thus, for a specified value of θ_i, the velocity vector is calculated as:

$$\bar{v}_{i,j} = v_{i,j} \frac{\cos(\theta_i)\hat{u}_{1,i,j} + \sin(\theta_i)\hat{u}_{2,i,j}}{||\cos(\theta_i)\hat{u}_{1,i,j} + \sin(\theta_i)\hat{u}_{2,i,j}||} \tag{13}$$

DISTRIBUTION A: Approved for public release; distribution is unlimited. Public Affairs release approval #AFRL-2023-3637.

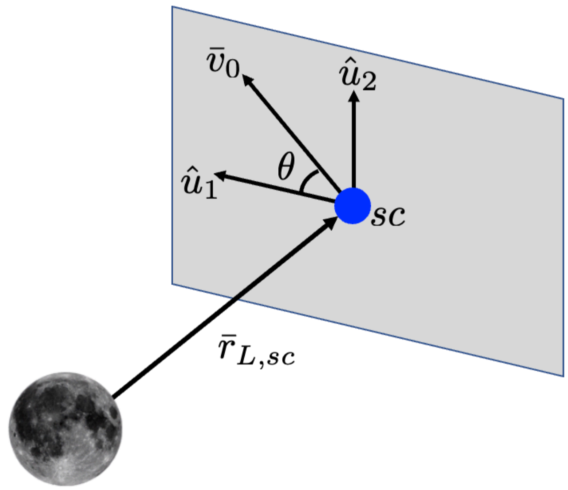

Figure 1. Definition of velocity vector for an apse using out-of-plane velocity angle θ. Moon image credit: nasa.gov, not shown to scale.

This definition is displayed graphically in Figure 1. If the resulting velocity vector produces a specific angular momentum vector relative to the Moon with a positive z-component, the state vector supplies a valid prograde apse in the i-th partition.

Within the i-th partition, trajectories are propagated in a desired dynamical model from each valid initial condition. In this paper, trajectories are generated until one of the following termination conditions is satisfied: a maximum integration time of 21 days, selected as a time interval of interest when studying observed objects in cislunar space domain awareness; or passing within the equatorial radius of the Earth or Moon. Note that these termination conditions enable the examined trajectories to extend beyond the Moon vicinity and potentially visit various regions of the Earth-Moon system. For trajectories generated in the CR3BP, numerical integration is performed directly from the state vector for the j-th trajectory, labeled $\bar{x}_{i,j} = [\bar{r}_{i,j}, \bar{v}_{i,j}]^T$, using Equation 1. However, for trajectories generated in the ephemeris model, the equations of motion are specified in an inertial frame. Thus, the initial state vector $\bar{x}_{i,j}$ is dimensionalized using the instantaneous characteristic quantities at a fixed initial epoch t_0, transformed into the Moon-centered inertial frame at t_0, and then nondimensionalized using the same constant characteristic quantities as in the CR3BP. The resulting nondimensional and inertial state vector $\bar{X}_{i,j} = [\bar{R}_{i,j}^T, \bar{V}_{i,j}^T]^T$ is then numerically integrated forward in time from t_0 using Equation 2 until the termination conditions are met.

Step 2: Summarize each Trajectory

To effectively describe the shape of a continuous trajectory by a sequence of discrete states, a trajectory sampling scheme is defined using concepts from differential geometry. A continuous, curved trajectory $\bar{r}(t) = (x(t), y(t), z(t))$ that is generated over a time interval $t \in [t_0, t_f]$ possesses an arclength s that is calculated as[34]

$$s = \int_{t_0}^{t_f} ds = \int_{t_0}^{t_f} \sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2} dt \tag{14}$$

After parameterizing the trajectory by its arclength, i.e., $\bar{r}(s)$, the tangent $\hat{T}$, normal $\hat{N}$, and binormal $\hat{B}$ vectors may be defined at any location along the curve. Both $\hat{T}$ and $\hat{N}$ lie in the osculating

DISTRIBUTION A: Approved for public release; distribution is unlimited. Public Affairs release approval #AFRL-2023-3637.

plane, defined as the plane that passes through three sequential points as the arclength between them approaches an infinitesimally small value.[35] Within this plane, the $\hat{T}$ vector is tangent to the path whereas the $\hat{N}$ vector is pointed towards the center of curvature and along the derivative of $\hat{T}$ with respect to the arclength.[34] Then, the $\hat{B}$ vector completes the orthogonal, right-handed triad.

Using an arclength parameterization and the $\hat{T}\hat{N}\hat{B}$ basis vectors, two parameters describe how a trajectory curves in three-dimensional space: the curvature $\kappa(t)$, which captures the deviation from a straight line in the osculating plane, and the torsion $\tau(t)$, reflecting the change in orientation of the osculating plane. For a position vector $\bar{r}(t)$ with velocity vector $\dot{\bar{r}}$ and acceleration vector $\ddot{\bar{r}}$, the curvature $\kappa(t)$ is mathematically calculated as

$$\kappa(t) = \frac{||\dot{\bar{r}} \times \ddot{\bar{r}}||}{||\dot{\bar{r}}||^3} \tag{15}$$

The curvature is also equal to the rate of change of the angle swept out by the tangent vector with respect to arclength.[35] Integrating $\kappa(t)$ along a curved trajectory produces the total absolute curvature, equal to[35]

$$\kappa_{tot}(t_0, t_f) = \int_{t_0}^{t_f} \kappa(s)ds = \int_{t_0}^{t_f} \kappa(s)\sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}dt \tag{16}$$

essentially capturing the cumulative angle traced out within the osculating plane as the path revolves around a moving center of curvature.

In the field of computer graphics, curves and surfaces have been discretized to capture their shape by sampling points using the total absolute curvature; this approach lies within a broader class of shape interrogation methods.[35] In regions of high curvature, where the path significantly deviates from a straight line, this quantity increases more rapidly. However, unlike the curvature $\kappa(t)$ that may possess a wide range of values along a single trajectory, the total absolute curvature increases monotically and increases by approximately 2π with each revolution. Accordingly, the value of the total curvature along a curve offers insight meaningful locations that capture its shape.

In this paper, each trajectory is discretized into a fixed number of states N_s that are equally distributed in the total absolute curvature, calculated in the rotating frame. As an example, consider two planar trajectories generated in the Earth-Moon CR3BP at $C_{J,d} = 3.165$ and displayed in black in Figure 2. In this figure, the Moon is located by a gray circle, not to scale, and the L_1 and L_2 equilibrium points are located by red diamonds. The blue circles indicate $N_s = 30$ states that are equally-distributed along each trajectory as a function of its total absolute curve whereas the purple circle locates the initial condition. In Figure 2a), the trajectory performs over three revolutions around the Moon. The total curvature along this trajectory is equal to approximately $3.37(2\pi)$ with regions of high curvature located near apolune and perilune as well as the curve near L_1. Accordingly, the total curvature changes rapidly in these regions and some of the sampled states tend to be concentrated in their vicinity. Along the trajectory plotted in Figure 2b), however, each revolution around the Moon possesses a much smaller difference between the perilune and apolune distances. Consistent with this geometry, the total curvature does not change rapidly and the sampled states appear more evenly distributed along the arclength of the trajectory.

For a sufficiently large value of N_s, states can be well-sampled along the entire trajectory but concentrated near regions of high curvature, e.g., near apses. This approach supplies a fixed-length and curve-based representation of a trajectory that both supports the use of a simple Euclidean distance metric and minimizes the loss of information about the shape of the continuous trajectory due

DISTRIBUTION A: Approved for public release; distribution is unlimited. Public Affairs release approval #AFRL-2023-3637.

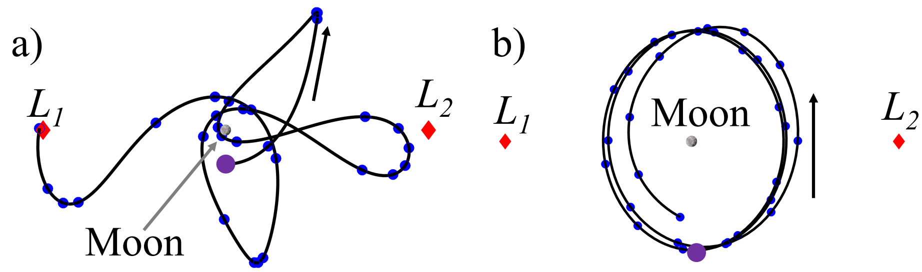

Figure 2. Examples of $N_s = 30$ states evenly distributed along two planar trajectories, generated in the CR3BP at $C_J = 3.165$, as a function of the total absolute curvature in the Earth-Moon rotating frame.

to discretization. Unlike an apse-based discretization, this approach also does not rely on determining one or more meaningful reference point/s for calculating the apse as a trajectory visits multiple regions of a system. Furthermore, this approach has been observed to adequately represent a curved trajectory with lower values of N_s than equally distributing the states along a trajectory in time. Note that the torsion is not currently incorporated into the discretization scheme used in this paper due to its dependency on the third time derivative of the position vector in the rotating frame, which is more mathematically intensive to calculate along a trajectory that is generated in the ephemeris model; this extension is an avenue of ongoing work.

Each discretized trajectory is then described by a finite-dimensional feature vector. This feature vector serves as an input to a clustering algorithm and must, therefore, reflect characteristics of the dataset that are of interest to the analyst. In this paper, the goal is to group trajectories by their geometry. However, the geometry of a nonlinear trajectory in three-dimensional space cannot be analytically defined. Furthermore, rapidly computing meaningful clusters using the distance between two trajectories in the configuration space may be challenging when the trajectory encompasses regions of distinct sensitivities or extends far from its initial condition. Accordingly, a shape-based feature vector is defined using the tangent vector expressed in the rotating frame at each of the N_s points that are equally distributed in the total absolute curvature along a trajectory. With this definition, the tangent vector captures the local shape using quantities that remain within the range $[-1, 1]$. Furthermore, the trajectory discretization scheme indirectly supplies global shape information to the feature vector. Mathematically, the feature vector for the i-th trajectory is defined as

$$\bar{f}_i = \left[\hat{T}_1, \hat{T}_2, ..., \hat{T}_{N_s-1}, \hat{T}_{N_s}\right] \tag{17}$$

to produce a $(3N_s)$-dimensional vector. However, it is important to note that there may be a wide variety of alternative, meaningful descriptions of a trajectory that produce distinct groupings of trajectories during clustering for various applications.

Step 3: Cluster Individual Partitions

The trajectories contained within each partition are independently clustered via HDBSCAN with a fixed set of governing parameters. First, the Euclidean distance is selected as the distance metric used to compare two trajectories via the difference between their feature vectors. This distance metric supports fast clustering with a reasonable computational load; however, it does require the use of feature vectors with a fixed length across the whole dataset, consistent with the definitions in Step

DISTRIBUTION A: Approved for public release; distribution is unlimited. Public Affairs release approval #AFRL-2023-3637.

2. In addition, HDBSCAN is governed by three parameters, selected to be fixed when clustering each partition: the minimum cluster size $m_{clmin} = 50$; the number of neighbors used in distance computations $m_{pts} = 2$; and the distance threshold for merging clusters $\epsilon_{merge} = 0.3$. The values of these three parameters are selected through observation of their impact on the resulting clusters across multiple partitions, balancing reducing the number of trajectories assigned as noise points by HDBSCAN, producing a reasonable number of clusters, avoiding the existence of many small clusters in high density regions, and avoiding producing too few clusters that do not sufficiently differentiate trajectories of distinct geometries. Using these governing parameters, HDBSCAN is used to group trajectories in the i-th partition into a set of clusters $\mathcal{C}_i$ and noise points $\mathcal{N}_i$ that produce the i-th local cluster summary.

To demonstrate the process of individually clustering each partition to produce groups of geometrically similar trajectories, consider a set of trajectories generated at $C_{J,d} = 3.165$ with $\theta = 0°$ in the Earth-Moon CR3BP. Three partitions of trajectories are constructed at $z_1 \approx -0.0017$, $z_2 \approx -0.0257$, and $z_3 \approx -0.0977$ with $N_x = N_y = 200$ for $x \in [0.836, 1.156]$ and $y \in [-0.12, 0.12]$. Within this grid, only a subset of the position vectors produce valid initial perilunes at a Jacobi constant of $C_{J,d}$. The feature vectors describing the trajectories within each partition are independently clustered using HDBSCAN with the selected governing parameters. The resulting clusters are displayed in Figure 3 for each partition. In this figure, the initial perilunes of each trajectory are projected onto the xy-plane of the Earth-Moon rotating frame and uniquely colored according to their cluster assignment using shades of red and blue. Note that some colors may be repeated across both subfigures, but the clusters are, at this step, independent. In these figures, black points indicate noise points. Each partition in Figure 3 is grouped into the following number of clusters and noise points: a) $\mathcal{C}_1$ is composed of 44 clusters of 28,535 trajectories and $\mathcal{N}_1$ includes 2,149 noise points; b) $\mathcal{C}_2$ is composed of 33 clusters of 25,753 trajectories and $\mathcal{N}_2$ includes 3,425 noise points; and c) $\mathcal{C}_3$ is composed of 11 clusters of 4,571 trajectories and $\mathcal{N}_3$ includes 77 noise points. In each case, the noise points tend to lie predominantly at either the cluster boundaries or sensitive regions, i.e., near the Moon. However, because the goal is to extract the fundamental geometries exhibited by the generated trajectories, rather than precisely determine the regions of the phase space that similar trajectories encompass, this level of noise is deemed acceptable. Nevertheless, reducing the number of trajectories designated as noise across each partition is an avenue of ongoing work.

Within each partition, the clustering process successfully groups trajectories of similar geometry and separates trajectories of distinct geometries. To demonstrate this observation, consider the clustering result plotted in Figure 3 a) for a partition of trajectories corresponding to $z_1 \approx -0.0017$,

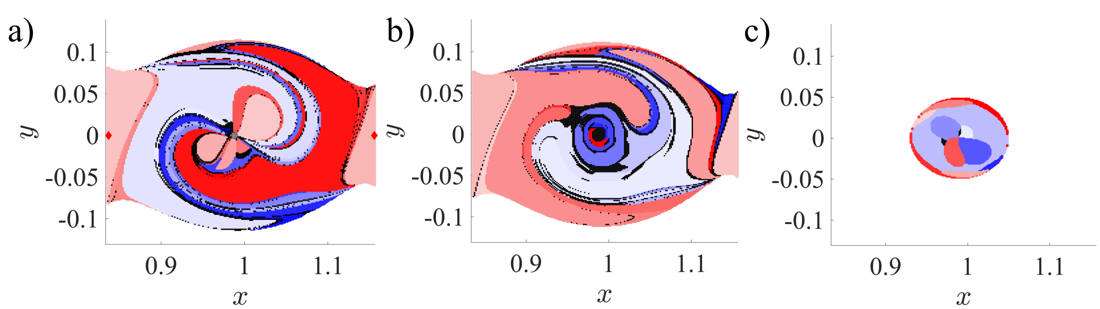

Figure 3. Examples of clustering results for three partitions at a) $z_1 \approx -0.0017$, b) $z_2 \approx -0.0257$, and c) $z_3 \approx -0.0977$.

DISTRIBUTION A: Approved for public release; distribution is unlimited. Public Affairs release approval #AFRL-2023-3637.

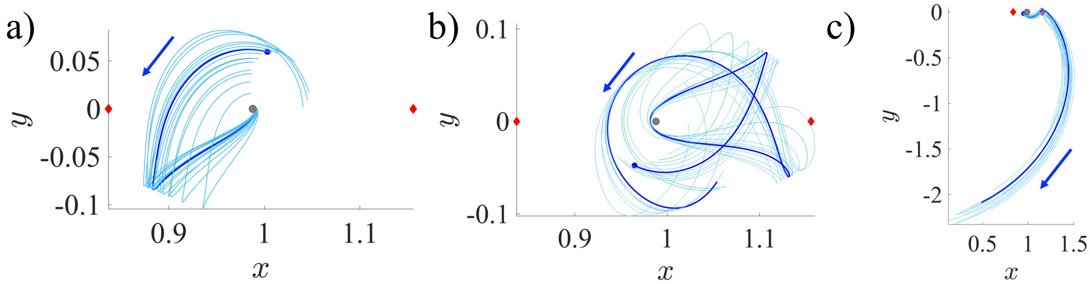

Figure 4. Examples of three clusters within a partition with $z_1 \approx -0.0017$, $\theta_1 = 0°$, and $C_{J,d} = 3.165$.

$\theta_1 = 0°$ and $C_{J,d} = 3.165$ and generated in the Earth-Moon CR3BP. Figure 4 displays a subset of trajectories in 3 of the 44 clusters in $\mathcal{C}_1$ generated for this partition and projected onto the xy-plane of the Earth-Moon rotating frame. Within each subfigure, one trajectory is highlighted in thick blue with a circle locating the initial condition. In these figures, the Moon is located by a gray circle (not to scale) and two red diamonds locate L_1 and L_2. Several additional members, sampled across each cluster, are colored in light blue. As demonstrated in these examples, each group of trajectories possesses a similar geometry in the three-dimensional configuration space. Furthermore, across these clusters, their geometry is distinct. Note that in some cases, trajectories with a similar shape that begin and end in distinct regions of the solution space may be grouped in the same cluster. This is not unexpected as the feature vector only captures the shape rather than the location of each sampled state along the trajectory. Ongoing work includes performing a subsequent cluster refinement step to further separate any subgroups of trajectory in distinct locations of the configuration space.

Step 4: Aggregate Clusters Across Partitions

To ensure that clustering a large dataset is computationally feasible, a cluster aggregation procedure is employed. This aggregation process is similar to the procedure presented by Bonasera and Bosanac and is performed in the following two phases:[17]

1. The local cluster summaries of partitions with the same values of the out-of-plane angle of the velocity vector are aggregated to produce an intermediate θ-cluster summary
2. The θ-cluster summaries are aggregated to produce a global cluster summary, $\mathcal{G}$

Within each phase, clusters are aggregating by subsampling previous clustering results and grouping their reduced set of members via HDBSCAN. The cluster aggregation process in each phase follows a binary tree structure as described in this subsection.

To reduce the computational load of cluster aggregation, each set of clusters and noise points that are identified after each application of HDBSCAN is subsampled. Specifically, each cluster with more than $N_{a,min} = 200$ members is subsampled to retain every $N_{a,samp} = 2$ members in the ordered list of cluster members. For the noise trajectories, every $N_{a,noise} = 2$ trajectories in the ordered list are retained in case they are later clustered with trajectories from other partitions.

Cluster aggregation in each phase follows a binary tree structure. This procedure is graphically depicted in Figure 5. First, the local cluster summary of each partition is subsampled. Then, the reduced set of members of each neighboring pair of partitions for a fixed value of θ is input to HDBSCAN to produce an intermediate cluster summary. At the next level of cluster aggregation,

DISTRIBUTION A: Approved for public release; distribution is unlimited. Public Affairs release approval #AFRL-2023-3637.

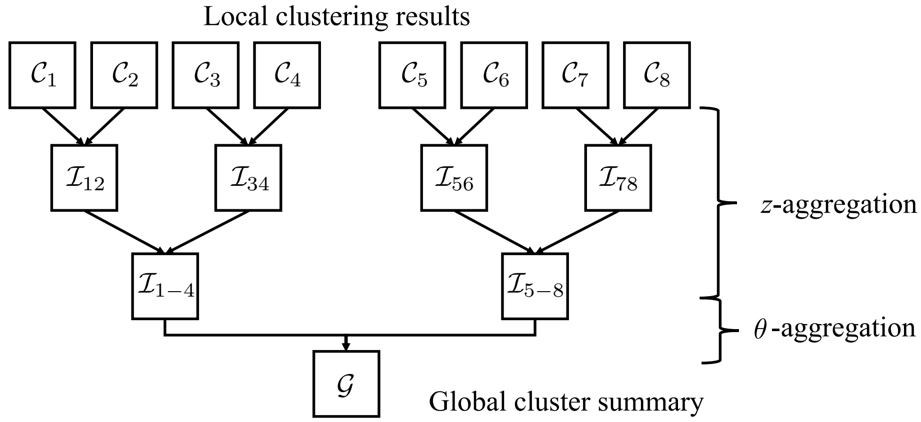

Figure 5. Conceptual depiction of cluster aggregation process.

each neighboring pair of intermediate cluster summaries is subsampled and input to HDBSCAN to produce a new intermediate cluster summary. When an odd number of cluster summaries exist in any step of the cluster aggregation process, however, that summary proceeds to the next step without subsampling. This process, labeled z-aggregation, continues until there is one θ-cluster summary. This z-aggregation process is performed independently for each value of θ. Next, the θ cluster summaries are aggregated using the same binary tree structure to produce one global cluster summary $\mathcal{G}$. At the final step of generating the global clustering result, ϵ_{merge} is increased to 1 to avoid the discovery of smaller, dense clusters that exist within a broader region of similar solutions. Overall, this approach enables geometrically similar trajectories across distinct partitions to be combined into a single cluster, regions with more information to be split into multiple clusters when geometric differences are discovered, and mitigation of the impact of a previous poor clustering result.

Step 5: Extract Representative Trajectories from Each Cluster

To facilitate visualization and analysis, each cluster is described by a single representative trajectory, similar to the approach presented by Bosanac as well as Bonasera and Bosanac.[16,17] This representative trajectory is extracted as the medoid, which is defined as the member of a cluster that is most similar to the other members.[36] Mathematically, the trajectory $\mathcal{T}_{med,k}$ that is the medoid of cluster C_k is calculated using the feature vectors of its P_k members as

$$\mathcal{T}_{med,k} = \text{argmin}_{\mathcal{T}_i \in \mathcal{C}_k} \left[\sum_{j=1, i \neq j}^{P_k} d(\bar{f}_i, \bar{f}_j) \right] \tag{18}$$

where $d(\bar{f}_i, \bar{f}_j)$ is the Euclidean distance between the feature vectors of the i-th and j-th trajectories and $\mathcal{T}_i$ is the i-th trajectory. Examples of cluster representatives appear as thick blue curves for the three clusters displayed earlier in Figure 4.

RESULTS

Summarizing Trajectories in the Earth-Moon CR3BP

The clustering based framework is used to generate a data-driven summary of natural spacecraft trajectories at $C_{J,d} = 3.165$ in the Earth-Moon CR3BP. The initial conditions are gener-

DISTRIBUTION A: Approved for public release; distribution is unlimited. Public Affairs release approval #AFRL-2023-3637.

ated using the following grid of xy-coordinates: $N_x = N_y = 200$ for $x \in [0.836, 1.156]$ and $y \in [-0.12, 0.12]$. These initial conditions are generated in partitions with 64 values of $z \in [-0.108, 0.108]$ and 19 values of $\theta = [-89°, -80°, ..., -10°, 0°, 10°, ..., 80°, 89°]$. Within each partition, there are between 36 and 30,684 trajectories for a total of 23,824,964 trajectories. Each trajectory is discretized into $N_s = 30$ states that are equally distributed along the trajectory as a function of the total absolute curvature. During the cluster aggregation process, the following governing parameters are supplied to HDBSCAN: $m_{clmin} = 50$, $m_{pts} = 2$, $\epsilon_{merge} = 0.3$. At the final step of θ-aggregation, ϵ_{merge} is increased to a value of 1 to avoid the detection of many smaller clusters of trajectories with a similar geometry.

Using the clustering framework to group the feature vectors associated with the generated trajectories in a distributed manner produces a global clustering result with 218 clusters composed of 51,235 trajectories. The representative trajectories of these clusters are displayed in Figures 6- 8 in the Earth-Moon rotating frame. In these figures, the L_1 and L_2 equilibrium points are located by red diamonds whereas the Moon and, where applicable, the Earth are displayed as gray circles that are not to scale. The initial condition along each trajectory is located by a circle marker. Note that some subfigures feature multiple representatives of similar geometry in distinct colors.

Analysis of Figures 6-8 reveals the extraction of a wide array of trajectories with distinct geometries. First, note that in some cases, paths that are symmetric about the xy-plane in the Earth-Moon rotating frame appear to be separated. Where possible these trajectories are located in the same subfigure. Further analysis is required to identify the reason that these trajectories are separated into distinct clusters, particularly those that are expected to encompass trajectories located in the xy-plane with $\dot{z} = 0$. In other cases, trajectories that have a similar geometry but distinct phasing are also separated. In any case, the top two rows of Figure 6 display trajectories that depart the Moon vicinity through the L_1 or L_2 gateways after performing a distinct number of revolutions. A majority of the remaining representative trajectories within this figure perform a few revolutions around the Moon with a distinct geometry, apsidal rotation, and extension predominantly above or below the xy-plane. Alternatively, the seventh and eighth rows of Figure 7 feature representatives that impact the Moon after a high inclination and high eccentricity revolution around the Moon. The final two rows of Figures 7 and 8 also include trajectories with a close pass to the Moon but tracing out a tulip-shaped path with high inclination and high eccentricity revolutions.

Summarizing Trajectories in an Ephemeris Model

The clustering based framework is used to summarize a set of natural spacecraft trajectories in the point mass ephemeris model of cislunar space. The initial conditions are generated using the following grid of xy-coordinates: $N_x = N_y = 100$ for $x \in [0.836, 1.156]$ and $y \in [-0.12, 0.12]$. These initial conditions are generated with 16 values of $z \in [-0.108, -0.0.0035]$ and 19 values of $\theta = [-89°, -80°, ... - 20°, -10°, 0°, 10°, 20°, ..., 80°, 89°]$. Note that this example uses a coarser grid of initial conditions than in the previous example in the CR3BP as well as only half the possible range of values of z due to the increased computational time associated with generating trajectories in an ephemeris model. Next, the velocity vector is defined to produce a Jacobi constant of $C_{J,d} = 3.165$ in the CR3BP. Although the Jacobi constant is not constant in the ephemeris model, this approach supports a proof of concept. Within each partition, there are between 10 and 7,630 trajectories for a total of 1,455,172 trajectories. Each trajectory is discretized into $N_s = 30$ states that are equally distributed along the trajectory as a function of the total absolute curvature, calculated in the pulsating Earth-Moon rotating frame.

DISTRIBUTION A: Approved for public release; distribution is unlimited. Public Affairs release approval #AFRL-2023-3637.

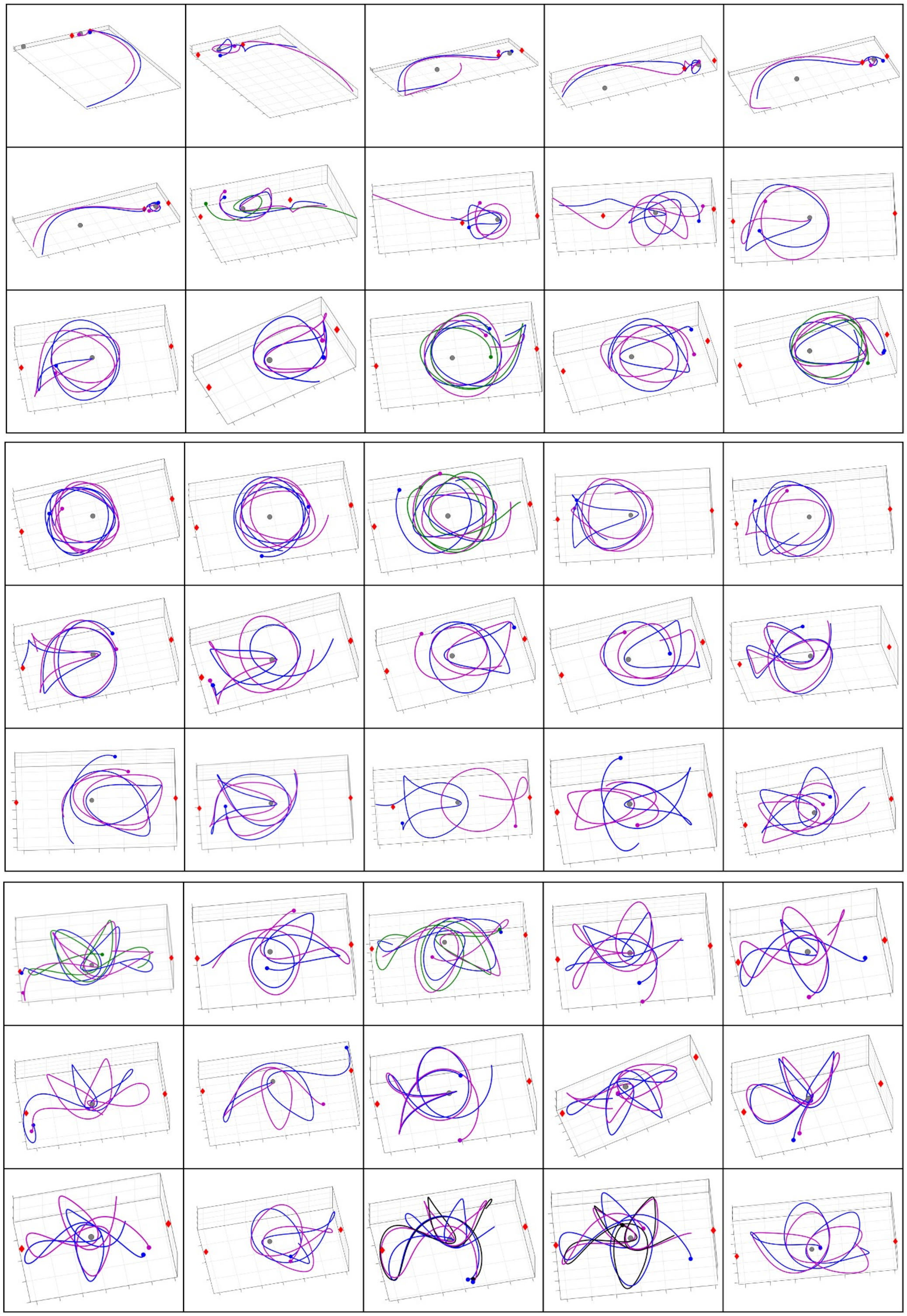

Figure 6. Representative, spatial trajectories of the 218 clusters in the global cluster summary constructed at $C_{J,d} = 3.165$ in the Earth-Moon CR3BP.

DISTRIBUTION A: Approved for public release; distribution is unlimited. Public Affairs release approval #AFRL-2023-3637.

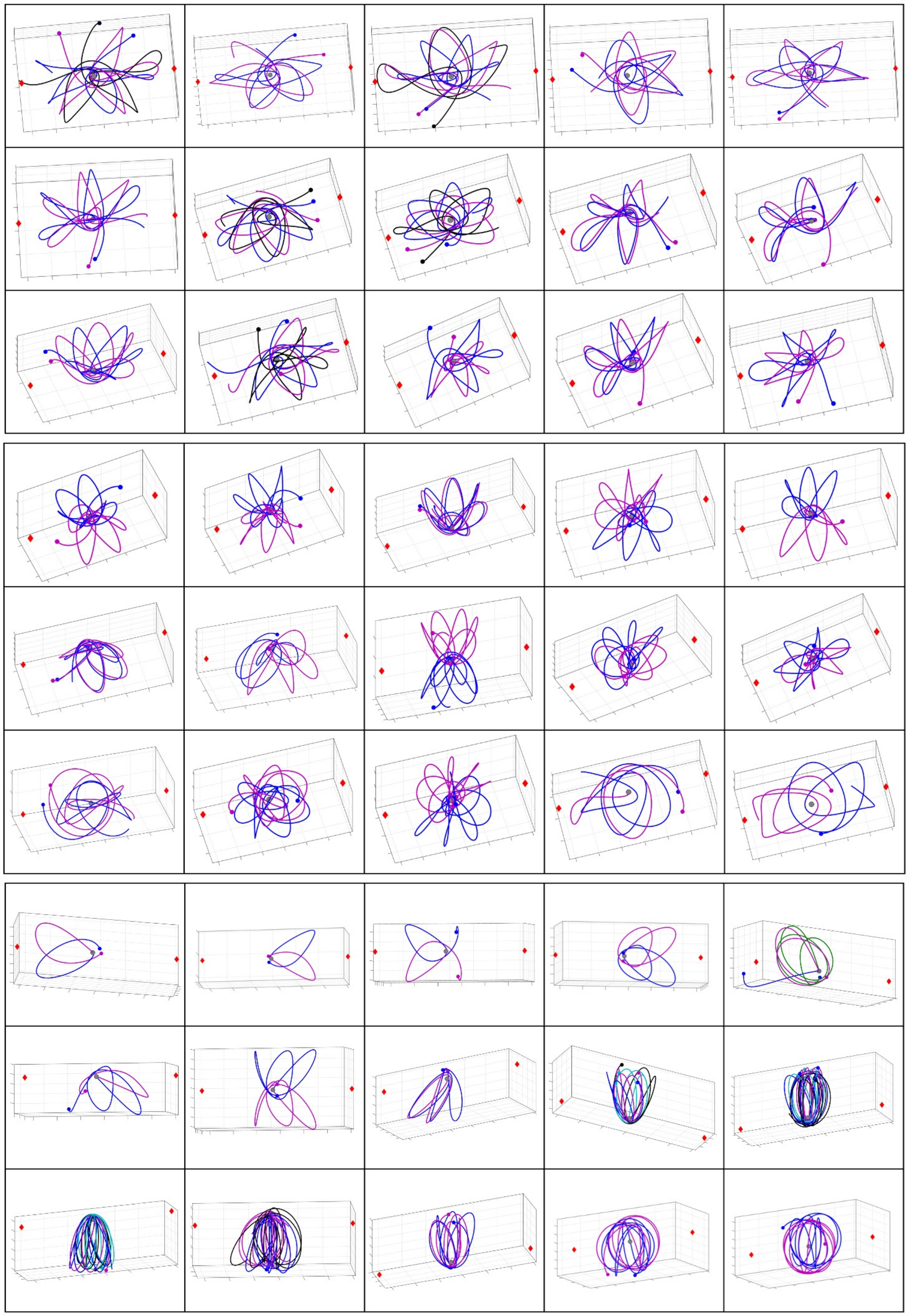

Figure 7. Representative, spatial trajectories of the 218 clusters in the global cluster summary constructed at $C_{J,d} = 3.165$ in the Earth-Moon CR3BP.

DISTRIBUTION A: Approved for public release; distribution is unlimited. Public Affairs release approval #AFRL-2023-3637.

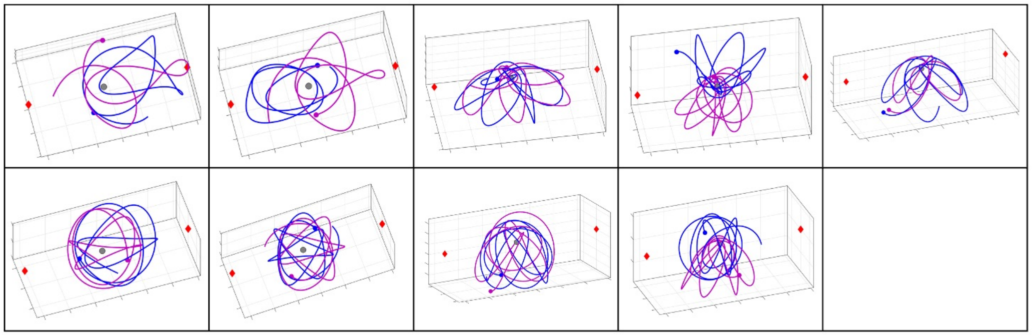

Figure 8. Representative, spatial trajectories of the 218 clusters in the global cluster summary constructed at $C_{J,d} = 3.165$ in the Earth-Moon CR3BP.

The clustering framework is used to group the feature vectors associated with the large array of trajectories in a distributed manner to produce a global clustering result. During the cluster aggregation process, the following governing parameters are supplied to HDBSCAN: $m_{clmin} = 10$, $m_{pts} = 2$, $\epsilon_{merge} = 1$. These parameters are updated due to the coarser grid used to define the initial conditions in this example. At the final step of θ-aggregation, ϵ_{merge} is increased to a value of 2 to avoid the detection of many smaller clusters.

Following cluster aggregation, the global clustering result is composed of 52 clusters of 13,325 trajectories. The representative trajectories of these clusters are displayed in Figure 9 in the pulsating Earth-Moon rotating frame. In these figures, the L_1 and L_2 equilibrium points are located by red diamonds whereas the Moon and, where applicable, the Earth are displayed as gray circles that are not to scale. The initial condition along each trajectory is located by a circle marker. Note that some subfigures feature multiple representatives of similar geometry in distinct colors.

Analysis of Figure 9 reveals the recovery of an array of trajectories with distinct geometries. The top row features trajectories that depart through the L_1 or L_2 gateways after a distinct number of revolutions. Because the initial conditions in this case only encompass negative values of z, this clustering result does not feature 2 clusters of geometrically similar trajectories that exist above or below the xy plane. The second and third rows of Figure 9 feature cluster representatives of distinct geometry that perform at least one revolution around the Moon before either 21 days has elapsed or lunar impact. The fourth row includes cluster representatives that impact the Moon after completing a high inclination and high eccentricity revolution. The final row features trajectories that trace out a tulip shape via high inclination and high eccentricity revolutions with a maximum extension above or below the xy-plane.

Conclusions

A clustering-based framework is used to extract a data-driven summary of a diverse set of trajectories in the Earth-Moon system. First, these trajectories are generated from prograde apses at a single energy level for up to 21 days. Then, these trajectories are discretized by evenly sampling states as a function of the total absolute curvature. The tangent vector at each of these states is used to form a feature vector that effectively captures the shape of the trajectory. These feature vectors, computed for all trajectories, are grouped into partitions based on their initial z-coordinate and out-of-plane velocity angle. Trajectories in each partition are clustered individually using HDBSCAN and their clustering results are aggregated by subsampling and reapplying the clustering process in

DISTRIBUTION A: Approved for public release; distribution is unlimited. Public Affairs release approval #AFRL-2023-3637.

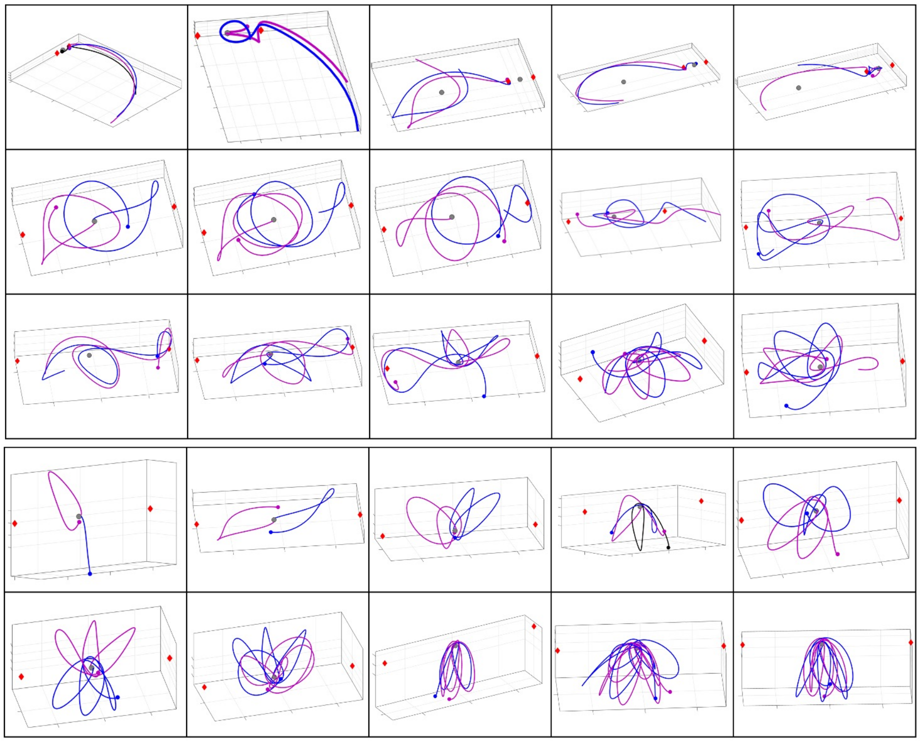

Figure 9. Representative, spatial trajectories of the 52 clusters in the global cluster summary constructed at $C_{J,d} = 3.165$ in the point mass ephemeris model of the Earth-Moon system.

a binary tree structure. The resulting global cluster summary is then analyzed using representative trajectories of each cluster.

This data-driven framework is used to extract a summary of natural spacecraft trajectories that begin from apses at a single energy level where the L_1 and L_2 gateways are open in the Earth-Moon system. This summary is constructed for trajectories that are generated in two models of distinct fidelity: the Earth-Moon CR3BP and a point mass ephemeris model of the Earth, Moon, and Sun. In each case, the global cluster summary captures trajectories with distinct geometries, performing distinct revolutions around the Moon, and traveling through various regions of the system. Comparing these two global cluster summaries supports analysis of the impact of model fidelity on the characteristics of trajectories generated from the same set of initial conditions.

ACKNOWLEDGMENTS

This work was performed at the University of Colorado Boulder and supported by funding from the Air Force Research Laboratory through a cooperative agreement titled "Multi-Domain Awareness, Decision, and Exploitation".

DISTRIBUTION A: Approved for public release; distribution is unlimited. Public Affairs release approval #AFRL-2023-3637.

REFERENCES

[1] Szebehely, V., *Theory of Orbits: The Restricted Problem of Three Bodies*, Academic Press, London, 1967.

[2] Koon, W. S., Lo, M. W., Marsden, J. E., Ross, S. D., *Dynamical Systems, the Three Body Problem and Space Mission Design,* Marsden Books, New York, 2011.

[3] Park, R.S., Folkner, W.M., Williams, J.G., Boggs, D.H., "The JPL Planetary and Lunar Ephemerides DE440 and DE441," *The Astronomical Journal*, Vol. 161, No. 3, 2021.

[4] Acton, C.H., "Ancillary Data Services of NASA's Navigation and Ancillary Information Facility," *Planetary and Space Science*, Vol. 44, No. 1, pp. 65-70, 1996. DOI 10.1016/0032-0633(95)00107-7

[5] Han, J., and Kamber, M., *Data Mining: Concepts and Techniques, 2nd ed.*, Proquest EBook Central: Elsevier Science and Technology, New York, 2006, Chap. 7.

[6] Zheng,Y.; Zhou,X., *Computing with Spatial Trajectories,* Springer Science & Business Media, New York, 2011, Chaps. 1, 5.

[7] Aggarwal, C. C., and Reddy, C. K., *Data Clustering: Algorithms and Applications*, Chapman and Hall CRC Press, Boca Raton, FL, 2018, Chaps. 2.1.2, 3, 9.2.

[8] Ivezić, Z., Connolly, A.J., VanderPlas, J.T., and Gray, A., *Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data, Updated Edition*, Princeton University Press, 2019, Ch. 9.4.

[9] Joncour, I., Duchêne, G., Moraux, E., and Motte, F., "Multiplicity and Clustering in Taurus Star Forming Region II. From Ultra-Wide Pairs to Dense NESTs," *Astronomy and Astrophysics*, Vol. 620, Paper A27, 2018.

[10] McLachlan, G. J., "Cluster Analysis and Related Techniques in Medical Research," *Statistics Methods in Medical Research*, Vol. 1, No. 1, pp. 27–48, 1992.

[11] Gallego, C. E. V., Gómez Comendador, V. F., Saez Nieto, F. J., and Martinez, M. G., "Discussion of Density-Based Clustering Methods Applied for Automated Identification of Airspace Flows," *37th Digital Avionics Systems Conference*, IEEE Publ., Piscataway, NJ, 2018.

[12] Hadjighasem, A., Karrasch, D., Teramoto, H., and Haller, G., "Spectral- Clustering Approach to Lagrangian Vortex Detection," *Physical Review E*, Vol. 93, No. 6, 2016.

[13] Nakhjiri, N. and Villac, B. F. "Automated Stable Region Generation, Detection, and Representation for Applications to Mission Design," *Celestial Mechanics and Dynamical Astronomy*, Vol. 123, No. 1, pp. 63–83, 2015.

[14] Villac, B., Anderson, R., and Pini, A. "Computer Aided Ballistic Orbit Classification Around Small Bodies," *Journal of Astronautical Sciences*, Vol. 63, No. 3, pp. 175–205, 2016.

[15] Smith, T.R; Bosanac, N, "Constructing Motion Primitive Sets to Summarize Periodic Orbit Families and Hyperbolic Invariant Manifolds in a Multi-Body Systems," *Celestial Mechanics and Dynamical Astronomy*, Vol. 134, No. 7, 2022.

[16] Bosanac, N., "Data-Mining Approach to Poincaré Maps in Multi-Body Trajectory Design," *Journal of Guidance, Control, and Dynamics*, Vol. 43, No. 6, 2020.

[17] Bonasera, S, Bosanac, N, "Applying Data Mining Techniques to Higher-Dimensional Poincaré Maps in the Circular Restricted Three-Body Problem," *Celestial Mechanics and Dynamical Astronomy*, Vol. 133, No. 51, 2021.

[18] Bonasera, S., Bosanac, N, "Unsupervised Learning to Aid Visualization of Higher-Dimensional Poincarè Maps in Multi-Body Trajectory Design," *2020 AAS/AIAA Astrodynamics Specialist Virtual Conference*, August 2020.

[19] Campello, R., Moulavi, D., and Sander, J., "Density-Based Clustering Based on Hierarchical Density Estimates," In: Pei, J., Tseng, V.S., Cao, L., Motoda, H., Xu, G. (eds) *Advances in Knowledge Discovery and Data Mining,* Springer, Berlin, Heidelberg, 2013.

[20] Lee, J.A., and Verleysen, M., 2007, *Nonlinear Dimension Reduction*, Springer Science + Business Media, New York, NY.

[21] Folta, D., Bosanac, N., Elliott, I.L., Mann, L., Mesarch, R., Rosales, J., "Astrodynamics Convention and Modeling Reference for Lunar, Cislunar, and Libration Point Orbits (Version 1.1)", NASA/TP–20220014814, 2022.

[22] Petit,G., Luzum, B. (eds), *IERS Conventions (2010)*, International Earth Rotation and Reference Systems Service, Technical Note 36, 2010.

[23] General Mission Analysis Tool Version R2020a: Mathematical Specifications.

[24] AGI Documentation Team and Subject Matter Experts, "Rotating Libration Point Coordinate System Technical Note," July 2020. Online: `https://help.agi.com/stk/index.htm#gator/eq-rlp.htm`, Last Accessed: 27 May 2021.

DISTRIBUTION A: Approved for public release; distribution is unlimited. Public Affairs release approval #AFRL-2023-3637.

[25] Pavlak, T, “Trajectory Design and Orbit Maintenance Strategies in Multi-Body Dynamical Regimes,” Ph.D. Dissertation, Purdue University, West Lafayette, IN, 2013.

[26] Vallado, D.A., *Fundamentals of Astrodynamics and Applications: Fourth Edition*, Microcosm Press, Hawthorne, CA, 2013.

[27] Aggarwal, C. C., and Reddy, C. K., *Data Clustering: Algorithms and Applications*, Chapman and Hall CRC Press, Boca Raton, FL, 2018, Chaps. 2.1.2, 3, 9.2.

[28] McInnes, L., Healy, J., and Astels, S. “hdbscan: Hierarchical Density Based Clustering,” Journal of Open Source Software, Vol. 2 No. 11, 2017.

[29] Malzer, C., Baum, M., “A Hybrid Approach To Hierarchical Density-Based Cluster Selection,” *2020 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, Karlsruhe, Germany,pp. 223-228, 2020.

[30] Kisilevich, S., Mansmann, F., Nanni, M., Rinzivillo, S. ,“Spatio-temporal clustering,” In: Maimon, O., Rokach, L. (eds) *Data Mining and Knowledge Discovery Handbook*, Springer, Boston, MA. 2009.

[31] Lee, J.G., Han, J., Whang, K.Y., “Trajectory Clustering: A Partition-and-Group Framework,” *SIGMOD ’07: Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data* , pp. 593–604, 2007.

[32] Zheng,Y.; Zhou,X., *Computing with Spatial Trajectories,* Springer Science & Business Media, New York, 2011, Chaps. 1, 5.

[33] Wenskovitch, J., Crandell, I., Ramakrishnan, N., House, L., Leman, S., and North, C., “Towards a Systematic Combination of Dimension Reduction and Clustering in Visual Analytics,” IEEE Transactions on Visualization and Computer Graphics, Vol. 24, No. 1, pp. 131–141, 2018.

[34] Wardle, K.L., *Differential Geometry*, Dover Publications, Inc., Mineola, NY, 2008.

[35] Patrikalakis, N.M., Maekawa, T., Cho, W., *Shape Interrogation for Computer Aided Design and Manufacturing*, E-Book, 2009.

[36] Cichosz, P., *Data Mining Algorithms: Explained Using R.* John Wiley and Sons, West Sussex, United Kingdom, 2015.

DISTRIBUTION A: Approved for public release; distribution is unlimited. Public Affairs release approval #AFRL-2023-3637.