AAS 25-124

DATA-DRIVEN SUMMARY OF MOTION IN AN EPHEMERIS MODEL OF CISLUNAR SPACE

Natasha Bosanac*

Spacecraft trajectory design and prediction in cislunar space will benefit from a comprehensive understanding of the solution space over a desired time horizon. This solution space consists of a diverse array of motions, with distinct geometries and itineraries. This paper uses a clustering-based framework to summarize these trajectories in a high-fidelity ephemeris model of cislunar space. First, trajectories are sampled across the Earth-Moon system in a geometry-based manner. Then, distributed clustering is used to discover groups of geometrically similar trajectories, each summarized by a representative member. The resulting data-driven summary is presented and analyzed in this paper.

INTRODUCTION

Trajectory design and prediction for spacecraft operating in cislunar space will benefit from a comprehensive understanding of the solution space over a desired time horizon. This multi-body gravitational system produces a chaotic solution space with a diverse array of motions, exhibiting a wide variety of geometries and itineraries. Existing dynamical systems techniques that are typically used to analyze a chaotic solution space in a lower-fidelity model may have limited applicability in higher-fidelity ephemeris models. For instance, some fundamental solutions, such as periodic and quasi-periodic orbits, or their finite-time equivalents that exist in lower-fidelity models may only be approximately retained in ephemeris models and rely on sufficient a priori knowledge of the solution space. Furthermore, these special types of solutions may not sufficiently capture the array of possible motions. As an alternative, simply sampling trajectories from an array of initial conditions across the 6D phase space may produce a diverse, high-dimensional, and extremely large amount of data. In the absence of any generalizable analytical expressions for categorizing these trajectories, manual analysis may be cumbersome and time-consuming.

The problem of automatically extracting insights from a large and high-dimensional dataset appears across a wide variety of disciplines. In these problems, clustering algorithms have been used to group similar data using a specified set of features and separate dissimilar data [1]. These data can even represent moving objects, producing a trajectory clustering problem [2]. After a dataset has been clustered, those groups may reveal fundamental modes or categories of data whereas representative members of each group may supply a digestible summary; these insights may drive knowledge discovery or support decision-making. Clustering has been useful in a wide variety

*Assistant Professor, Colorado Center for Astrodynamics Research, Smead Department of Aerospace Engineering Sciences, University of Colorado Boulder, 3775 Discovery Drive, Boulder, CO 80303.

of fields and applications including astronomy, medicine, studying driving routes, and air traffic management [2–6]. In astrodynamics, clustering methods have supported identifying bounded motions near distant retrograde orbits on a Poincaré map, grouping periodic orbits by a defined set of features, and extracting motion primitives that summarize trajectory arcs [7–9].

To automatically summarize a complex solution space in a multi-body system, Bosanac [10, 11] as well as Bonasera and Bosanac [12] have developed a clustering-based framework. The most recent implementation of this framework has been developed by Bosanac and is used in this paper. This approach was used to summarize the geometries exhibited by natural trajectories over 21 days in the circular restricted three-body problem [11]. This approach was then used by Bosanac and Joyner to summarize low-thrust trajectories governed by fixed thrust directions in the velocity-normal-conormal axes relative to the Moon [13].

The approach developed by Bosanac begins by describing a continuous trajectory using a finite-dimensional feature vector. First, each trajectory is sampled at a discrete set of states that are evenly spaced in the total absolute curvature, i.e., the curvature integrated along the arclength of the trajectory, along each half revolution. Each state is used to form two finite-dimensional feature vectors that summarize each continuous trajectory: 1) a shape-based feature vector that is composed of the local tangent unit vectors and 2) a position-based feature vector.

The feature vectors are used by a clustering algorithm to produce groupings of trajectories with a similar geometry [13]. When the dataset is diverse, Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) [14] is initially employed to construct coarse groupings using the shape-based feature vectors of entire trajectories. Then, a convoy detection approach is used to refine each coarsely constructed group by identifying trajectories that remain sufficiently close for their entire duration in each of the shape-based and then position-based feature vector spaces. The resulting clusters capture trajectories with a similar shape and path through the configuration space.

Using these definitions, the clustering-based framework presented by Bosanac begins by sampling the solution space to produce a sparse yet representative set of trajectories by strategically selecting initial conditions [13]. A uniform grid of position vectors is first defined. Then, at each position vector, velocity vectors are defined to produce maxima in the curvature. Because these velocity vectors exist in one or more one-parameter curves, they are discretely sampled. For each velocity vector along these curves, their associated trajectories are generated and summarized. At each position vector, the resulting set of velocity vectors are clustered via HDBSCAN using the shape-based feature vector. Representative members of these local groups of trajectories are added to the larger dataset that is summarized.

The trajectories sampled across the solution space are summarized via distributed clustering [11, 13]. First, the large dataset is partitioned into smaller subsets. This approach distributes millions of trajectories across hundreds of smaller datasets. Each dataset is independently clustered. These local clusters are then aggregated to form global clusters of similar trajectories that exist across distinct partitions. The result is a set of global clusters and their representative trajectories that summarize the geometries of a larger set of trajectories that exist across the solution space.

This paper focuses on leveraging the clustering-based framework developed by Bosanac to generate and examine a data-driven summary of the solution space for natural motion in a point mass ephemeris model of cislunar space. Prior work by Bosanac has applied an earlier version of this framework to a small subset of trajectories in an ephemeris model, offering only an early proof of

concept [11]. However, the approach used in this current paper produces a higher quality summary of the geometries of trajectories in an ephemeris model.

BACKGROUND

Reference Frames

Two reference frames are used for formulation and analysis. To formulate the equations of motion governing the spacecraft and generate solutions, a Moon-centered inertial frame is defined with an origin at the center of the Moon and axes $\hat{X}\hat{Y}\hat{Z}$ of the International Celestial Reference Frame (ICRF) [15, 16]. To analyze the trajectory of the spacecraft, a pulsating Earth-Moon rotating frame is employed. This definition uses the Earth-Moon barycenter as the origin and axes defined as follows: $\hat{x}$ points from the center of the Earth to the center of the Moon, $\hat{z}$ is aligned with the orbital angular momentum vector of the Earth-Moon system, and $\hat{y}$ completes the right-handed triad. These axes pulsate over time as they are scaled to ensure that the distance between the Earth and Moon is always equal to unity, even as the two celestial bodies travel on non-circular paths.

The state of the Moon relative to the Earth is used to express the axes of the Earth-Moon rotating frame in terms of the axes of the Moon-centered inertial frame. At an epoch t, the state vector of the Moon relative to the Earth, expressed in the axes of the ICRF, is defined as $\bar{X}_{E,L}(t) = [\bar{R}_{E,L}(t), \bar{R}'_{E,L}(t)]^T$, where the notation $(.)'$ indicates an observer in the inertial frame and the subscripts L and E indicate the Moon and Earth, respectively. Then, the axes are calculated at t as

$$\hat{x} = \frac{\bar{R}_{E,L}}{||\bar{R}_{E,L}||} \qquad \hat{z} = \frac{\bar{R}_{E,L} \times \bar{V}_{E,L}}{||\bar{R}_{E,L} \times \bar{V}_{E,L}||} \qquad \hat{y} = \hat{z} \times \hat{x} \tag{1}$$

to produce three column vectors. Their first time derivatives are equal to

$$\dot{\hat{x}}(t) = \frac{\bar{V}_{E,L}}{||\bar{R}_{E,L}||} - \hat{x}\frac{\hat{x} \cdot \bar{V}_{E,L}}{||\bar{R}_{E,L}||} \qquad \dot{\hat{z}}(t) \approx 0 \qquad \dot{\hat{y}}(t) = \hat{z} \times \dot{\hat{x}} \tag{2}$$

The assumption in the second expression is also used in GMAT [17]. The second time derivatives of these axes are calculated as

$$\ddot{\hat{x}}(t) = \frac{\bar{A}_{E,L}}{||\bar{R}_{E,L}||} - \frac{\bar{V}_{E,L}(\bar{V}_{E,L} \cdot \hat{x})}{||\bar{R}_{E,L}||^2} - \frac{\dot{\hat{x}}(\bar{V}_{E,L} \cdot \hat{x})}{||\bar{R}_{E,L}||} - \hat{x}\left(\frac{\bar{A}_{E,L} \cdot \hat{x}}{||\bar{R}_{E,L}||} + \frac{\bar{V}_{E,L} \cdot \dot{\hat{x}}}{||\bar{R}_{E,L}||} - \frac{(\bar{V}_{E,L}(t) \cdot \hat{x})^2}{||\bar{R}_{E,L}(t)||^2}\right) \tag{3}$$

$$\ddot{\hat{y}}(t) = \hat{z} \times \ddot{\hat{x}} \qquad \ddot{\hat{z}}(t) \approx 0 \tag{4}$$

Finally, the third time derivatives of these axes are equal to

$$\begin{aligned}\dddot{\hat{x}}(t) = {} & \frac{\bar{J}_{E,L}}{||\bar{R}_{E,L}||} - 2\frac{\bar{A}_{E,L}(\bar{V}_{E,L} \cdot \hat{x})}{||\bar{R}_{E,L}||^2} - \frac{\bar{V}_{E,L}(\bar{V}_{E,L} \cdot \dot{\hat{x}})}{||\bar{R}_{E,L}||^2} - \frac{\bar{V}_{E,L}(\bar{A}_{E,L} \cdot \hat{x})}{||\bar{R}_{E,L}||^2} + 2\frac{\bar{V}_{E,L}(\bar{V}_{E,L} \cdot \hat{x})^2}{||\bar{R}_{E,L}||^3} \\ & - \frac{\ddot{\hat{x}}(\bar{V}_{E,L} \cdot \hat{x})}{||\bar{R}_{E,L}||} - 2\frac{\dot{\hat{x}}(\bar{V}_{E,L} \cdot \dot{\hat{x}})}{||\bar{R}_{E,L}||} - \frac{\dot{\hat{x}}(\bar{A}_{E,L} \cdot \hat{x})}{||\bar{R}_{E,L}||} + \frac{\dot{\hat{x}}(\bar{V}_{E,L} \cdot \hat{x})^2}{||\bar{R}_{E,L}||^2} - \frac{\hat{x}(\bar{V}_{E,L} \cdot \ddot{\hat{x}})}{||\bar{R}_{E,L}||} - 2\frac{\ddot{\hat{x}}(\bar{A}_{E,L} \cdot \dot{\hat{x}})}{||\bar{R}_{E,L}||} \\ & - 3\frac{\hat{x}(\bar{V}_{E,L} \cdot \dot{\hat{x}})(\bar{V}_{E,L} \cdot \hat{x})}{||\bar{R}_{E,L}||^2} - \frac{\dot{\hat{x}}(\bar{A}_{E,L} \cdot \hat{x})}{||\bar{R}_{E,L}||} - \frac{\hat{x}(\bar{J}_{E,L} \cdot \hat{x})}{||\bar{R}_{E,L}||} - 3\frac{\hat{x}(\bar{V}_{E,L} \cdot \hat{x})(\bar{A}_{E,L} \cdot \hat{x})}{||\bar{R}_{E,L}||^2} \\ & + \frac{\dot{\hat{x}}(\bar{V}_{E,L} \cdot \hat{x})^2}{||\bar{R}_{E,L}||^2} - 2\frac{\hat{x}(\bar{V}_{E,L} \cdot \hat{x})^3}{||\bar{R}_{E,L}||^3}\end{aligned} \tag{5}$$

$$\dddot{\hat{y}}(t) = \hat{z} \times \dddot{\hat{x}} \qquad \dddot{\hat{z}}(t) \approx 0 \tag{6}$$

These expressions include the second time derivative (acceleration, $\bar{A}_{E,L} = \ddot{\bar{R}}_{E,L}$) and third time derivative (jerk, $\bar{J}_{E,L} = \dddot{\bar{R}}_{E,L}$) of the position vector of the Moon relative to the Earth. In this paper, these higher-order time derivatives are calculated using central finite differences of the velocity vector.

To support analysis, the state of the spacecraft is transformed from the Moon-centered inertial frame to the pulsating Earth-Moon rotating frame. Consider a spacecraft state vector that is expressed in the Moon-centered inertial frame as $\bar{X} = [X, Y, Z, \dot{X}, \dot{Y}, \dot{Z}]^T = [\bar{R}_{L,sc}, \bar{R}'_{L,sc}]^T$. In the Earth-Moon rotating frame, the spacecraft state vector is denoted as $\bar{x} = [x, y, z, \dot{x}, \dot{y}, \dot{z}]^T = [\bar{r}^T_{L,sc}, \dot{\bar{r}}^T_{L,sc}]^T$, where the notation $\dot{(.)}$ indicates a time derivative with an observer in the rotating frame and the subscript sc indicates the spacecraft. At an epoch t, the position vector is transformed from the inertial frame to the Earth-Moon rotating frame via the following expression:

$$\bar{r}_{B,sc}(t) = \left[{}^R C(t)^I\right] \bar{R}_{L,sc}(t) + \bar{r}_{B,L} \tag{7}$$

where the rotation matrix $[{}^R C(t)^I]$ equals

$$\left[{}^R C(t)^I\right] = \begin{bmatrix} \hat{x}^T(t) \\ \hat{y}^T(t) \\ \hat{z}^T(t) \end{bmatrix} \tag{8}$$

where the subscript B indicates the Earth-Moon barycenter. Time derivatives of this position vector are transformed from the inertial to rotating frame via the following transformations:

$$\bar{v}_{B,sc}(t) = \dot{\bar{r}}_{B,sc}(t) = \left[{}^R \dot{C}(t)^I\right] \bar{R}_{L,sc}(t) + \left[{}^R C(t)^I\right] \bar{V}_{L,sc}(t) \tag{9}$$

$$\bar{a}_{B,sc}(t) = \ddot{\bar{r}}_{B,sc}(t) = \left[{}^R \ddot{C}(t)^I\right] \bar{R}_{L,sc}(t) + 2\left[{}^R \dot{C}(t)^I\right] \bar{V}_{L,sc}(t) + \left[{}^R C(t)^I\right] \bar{A}_{L,sc} \tag{10}$$

$$\dddot{\bar{r}}_{B,sc}(t) = \left[{}^R \dddot{C}(t)^I\right] \bar{R}_{L,sc}(t) + 3\left[{}^R \ddot{C}(t)^I\right] \bar{V}_{L,sc}(t) + 3\left[{}^R \dot{C}(t)^I\right] \ddot{\bar{R}}_{L,sc}(t) + \left[{}^R C(t)^I\right] \dddot{\bar{R}}_{L,sc} \tag{11}$$

where the time derivatives of the rotation matrix are formed using the time derivatives of the axes of the rotating frame expressed in terms of inertial coordinates.

Ephemeris Model

A point mass ephemeris model is used to model the natural motion of a spacecraft within cislunar space. This dynamical model captures the point mass gravitational influences of the Moon, Earth, and Sun on a spacecraft with a comparatively negligible mass. The states of these celestial bodies are calculated using the DE440 lunar and planetary ephemerides along with the associated Spacecraft, Planet, Instrument, C-matrix, Events (SPICE) toolkit provided by NASA's Navigation and Ancilliary Information Facility [18, 19].

The path of the spacecraft is generated by numerically integrating the equations of motion from a specified initial condition and epoch. Mathematically, the equations of motion governing the spacecraft are written as

$$\ddot{\bar{R}}_{L,sc} = -GM_L\left(\frac{\bar{R}_{L,sc}}{R^3_{L,sc}}\right) + G\sum_{i=E,S} M_i\left(\frac{\bar{R}_{sc,i}}{R^3_{sc,i}} - \frac{\bar{R}_{L,i}}{R^3_{L,i}}\right) \tag{12}$$

where the subscript S indicates the Sun, G is the universal gravitational parameter, and M_i is the mass of body i [20]. Because the position vectors $\bar{R}_{i,j}$ describing the position of body j relative to i are time dependent, an initial epoch must also be associated with every solution in the point mass ephemeris model.

Curvature

The curvature along a trajectory captures the local geometry of the path. To define the concept of curvature, consider a nonlinear and continuous trajectory over the time interval $t \in [t_0, t_f]$ in a selected dynamical model. At any point along this trajectory, the state is described by the position vector $\bar{r}(t) = [x(t), y(t), z(t)]^T$ and velocity vector $\dot{\bar{r}}(t) = [\dot{x}(t), \dot{y}(t), \dot{z}(t)]^T$, whereas the acceleration vector $\ddot{\bar{r}}(t) = [\ddot{x}(t), \ddot{y}(t), \ddot{z}(t)]^T$ captures the dynamics. The distance traversed along this trajectory is defined as the arclength s and is equal to [21]

$$s = \int_{t_0}^{t_f} ds = \int_{t_0}^{t_f} \sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2} dt \tag{13}$$

At a selected location along the trajectory, the curvature $\kappa(t)$ reflects the rate at which the tangent vector orientation changes as a function of arc length [22]. As a result, this quantity reflects the extent to which the trajectory instantaneously deviates from a straight line. The nonnegative value of $\kappa(t)$ is calculated directly from the velocity and acceleration vectors as

$$\kappa(t) = \frac{||\dot{\bar{r}}(t) \times \ddot{\bar{r}}(t)||}{||\dot{\bar{r}}(t)||^3} \tag{14}$$

and possesses a singularity when $||\dot{\bar{r}}|| = 0$. The total absolute curvature is then defined as the integral of curvature along the arclength of the trajectory, essentially capturing the angle swept out by the trajectory relative to an evolving center of curvature and within the evolving osculating plane [22]. Mathematically, the total absolute curvature is equal to

$$\kappa_{tot}(t_0, t_f) = \int_{t_0}^{t_f} \kappa(s) ds = \int_{t_0}^{t_f} \kappa(s) \sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2} dt \tag{15}$$

Due to its definition, κ_{tot} monotonically increases by 2π with each revolution completed along the trajectory.

Clustering

This paper leverages two well-known clustering algorithms: the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm developed by Ester et al. [23] and the Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) algorithm developed by Campello, Moulavi, and Sander [14]. Both of these algorithms rely on discovering groups of data that lie in sufficiently dense regions within a feature vector space. However, DBSCAN is useful when the desired density threshold is known whereas HDBSCAN is useful when clustering a diverse dataset into groups with potentially distinct diversity, shape, and membership size.

DBSCAN DBSCAN discovers groups within a dataset as members with sufficient proximity and density. This clustering algorithm relies on the definition of a core point, which corresponds to a member with at least m_{pts} neighbors within a neighborhood of radius ϵ in a specified feature vector space [23]. Two core points with overlapping ϵ-sized neighborhoods that contain at least m_{pts}

neighbors are designated as density-connected [23]. A cluster is then formed via a set of core points where any two members can be reached through density-connected sequences of core points [23]. Clusters also contain border points, which are members of a dataset that lie in the m_{pts}-neighborhood of a core point but their own m_{pts}-neighborhoods have a larger radius than ϵ [23]. Noise points do not lie within the m_{pts}-neighborhood of any core points and, therefore, are not assigned to any clusters [23]. Accordingly, DBSCAN requires specifying the values of m_{pts} and ϵ to return a set of clusters and noise points.

HDBSCAN Although HDBSCAN uses a notion of density in discovering clusters, it transforms the dataset from the original feature vector space to a mutual reachability distance space. First, the i-th member of a dataset is assigned a core distance $d_{core}(\bar{f}_i)$, which corresponds to the distance to its m_{pts}-nearest neighbor [14]. This information is used to define the mutual reachability distance between the i-th and j-th members as

$$d_{reach}(\bar{f}_i, \bar{f}_j) = \max(d_{core}(\bar{f}_i), d_{core}(\bar{f}_j), d(\bar{f}_i, \bar{f}_j)) \tag{16}$$

where $d(\bar{f}_i, \bar{f}_j)$ is the distance between the feature vectors of the two members [14]. Accordingly, the mutual reachability distance between two members of a dataset captures both their separation in the feature vector space and the density via the size of their m_{pts}-neighborhoods.

HDBSCAN generates a hierarchy of all possible groupings of members in the mutual reachability distance space and selects the most stable set of clusters. To generate this hierarchy, a graph is constructed to reflect each member at its nodes and their mutual reachability distances as the edges. A minimum spanning tree is then constructed to summarize this graph [14]. Then, the cluster hierarchy is calculated by successively removing the edges with the highest weights and identifying the connected components [14]. Then, the clusters that maximize the stability, calculated using an excess of mass approach, are selected from the hierarchy if they possess at least m_{clmin} members [14]. Malzer and Baum have presented a modification that incorporates a minimum neighborhood size, ϵ_{merge}, for two members to be located in separate clusters [24]. Through this approach, HDBSCAN returns a set of cluster and a set of noise points. Because these clusters do not include border points, these members can be identified and relabeled in a post-processing step after clustering [25].

TECHNICAL APPROACH

Generating a Dataset

Generating Trajectories Trajectories that begin within the vicinity of the Moon are propagated from a fixed initial epoch and for a sufficient time interval to enable a diverse array of geometries to emerge. First, a constant initial epoch of January 1, 2025, 00:00.000 UTC is selected for all trajectories in this paper. Then, given a specified initial state, each trajectory is generated for a duration of up to 21 days in the point-mass ephemeris model via numerical integration in the Moon-centered inertial frame. However, the following early termination conditions are also employed: 1) impact with a spherical approximation of the Earth or Moon; and 2) reaching a maximum nondimensional distance from the Moon equal to 0.25, i.e., a quarter of the distance between the Earth and Moon. Due to this general definition of the duration of a trajectory, the entire dataset may exhibit a wide array of geometries.

Because the goal of this paper is to summarize trajectories by their geometry in the pulsating, Earth-Moon rotating frame, the total absolute curvature is calculated in the same frame during trajectory generation. At each step of the numerical integration process, the dimensional position,

velocity, and acceleration vectors are transformed from the Moon-centered inertial frame to the Earth-Moon pulsating rotating frame via Eq. 7 - 10. This process also requires calculating the rotation matrix between the two frames as well as its first and second time derivatives, which depend on the instantaneous state vector of the Moon relative to the Earth. The position, velocity, and acceleration vectors are then nondimensionalized using the instantaneous values of the characteristic quantities typically used for normalization in the circular restricted three-body problem (CR3BP). This information is used to evaluate the curvature via Eq. 14 at each instant of time and, therefore, calculate the total absolute curvature $\kappa(t_0, t_f)$ along each trajectory in the rotating frame. Accordingly, total absolute curvature is appended to the state vector during numerical integration. These frame transformations, required to calculate the total absolute curvature, result in extra matrix operations and calls to SPICE functions. These additional steps increase the computational time required to generate each trajectory, even in a fast implementation within C++.

Sampling Trajectories Each continuous trajectory is sampled at a set of discrete states that captures its geometry and supports generating a finite-dimensional description. To achieve these goals, states are sampled using the total absolute curvature of a trajectory, calculated in the pulsating, Earth-Moon rotating frame. Consider a trajectory with a total absolute curvature of $\kappa_{tot}(t_0, t_f)$ that completes up to $r = \lceil \kappa_{tot}(t_0, t_f)/\pi \rceil$ half-revolutions within the evolving osculating plane. This trajectory is first sampled at its initial condition. Then, for the first $r - 1$ half-revolutions, states are evenly sampled at every integer multiple of π/N_a in the total curvature. Then, during the r-th half-revolution, N_a states are evenly sampled in the remaining total absolute curvature, i.e., at every integer multiple of $(\kappa_{tot}(t_0, t_f) - (r-1)\pi)/N_a$. In this paper, $N_a = 3$ to balance limiting the number of sampled states with sufficiently capturing the geometry of the trajectory. This approach discretizes each trajectory into a total of $N_a r + 1$ states.

This sampling approach adapts to the geometry of each trajectory. Specifically, more samples are placed in regions of high geometric importance, i.e., when the curvature changes. Furthermore, the number of samples is determined by the number of revolutions completed by each trajectory, avoiding over- and under-sampling of trajectories. Finally, this approach can be automated without any further information due to the dependence of curvature on only time derivatives of the position vector. Together, these characteristics ensure an effective and automated geometry-aware sampling of each trajectory across a diverse dataset.

Describing Trajectories A continuous trajectory is described by two finite-dimensional feature vectors that are constructed using information at a discrete set of sampled states. To support a high-quality clustering result, a shape-based feature vector $\bar{f}_s$ and a position-based feature vector $\bar{f}_p$ are used at distinct steps of the clustering process. For a trajectory that is described by N_f states, the shape-based feature vector is defined as

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

where $\hat{T}_j = \bar{v}_j/v_j$ is calculated in the pulsating, Earth-Moon rotating frame at the j-th sampled state along the trajectory. This $3N_f$-dimensional vector possesses elements that are always within the range $[-1, 1]$, regardless of the region in which the spacecraft is located. Of course, there is a singularity when $v_j = 0$. Nevertheless, this feature vector supports identification of paths with a similar shape in the rotating frame. The position-based feature vector then supports separating trajectories with a similar shape that do not exist nearby in configuration space. This feature vector is defined as

$$\bar{f}_p = \left[\bar{r}_1, \bar{r}_2, ..., \bar{r}_{N_f-1}, \bar{r}_{N_f}\right] \tag{18}$$

where $\bar{r}_j$ is the nondimensional, position vector in the pulsating, Earth-Moon rotating frame at the j-th sampled state along the trajectory. The result is a $3N_f$-dimensional vector.

Clustering a Set of Trajectories

Coarse Shape-Based Clustering A diverse array of trajectories is first clustered to construct a coarse set of groups of trajectories with a sufficiently similar shape. These coarse grouping are generated using the shape-based feature vectors $\bar{f}_s$ of each entire trajectory. Furthermore, HDSBCAN is leveraged at this step to discover an unknown number of clusters with potentially distinct shapes, sizes, and densities. Each cluster includes a set of trajectories with shape-based feature vectors that create sufficiently dense groupings in the $3N_s$-dimensional space. This step produces a set of coarse, shape-based clusters shape-based clusters $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, ..., \mathcal{C}_n\}$ and a set of unlabeled, noise points $\mathcal{N}_c$. At this step, noise points that lie in the m_{pts}-neighborhood of any labeled trajectories in the shape-based feature vector space are assigned to the nearest cluster. Any remaining unlabeled trajectories are discarded at this stage.

During the coarse shape-based clustering step, the parameters governing HDBSCAN must be selected. First, m_{clmin} is selected to be small to prioritize discovering localized variations between trajectories, albeit at the expense of producing a larger number of clusters that are more sensitive to small density changes. Then, $m_{pts} = m_{clmin} - 1$ is selected to define each neighborhood to span the minimum size of a cluster. In addition, all feature vectors are compared using the Euclidean distance to support fast clustering [26]. Finally, $\epsilon_{merge} = 2\sqrt{N_s}\sin(\alpha/2)$ to support grouping together trajectories with N_s velocity unit vectors that are each an angle of α apart. This heuristic results in a generous grouping of trajectories that may potentially be similar, but can also result in trajectories with a few significantly different velocity unit vectors being grouped together incorrectly. Accordingly, cluster refinement is used to generate high-quality clusters from within each coarse, shape-based group.

Cluster Refinement A coarsely constructed group may be refined in a two-step process to identify high-quality clusters of trajectories that consistently follow a similar path through their entire duration. To implement this refinement step, the concept of a 'convoy' from trajectory clustering is employed. This term describes a set of moving objects that remain density-connected at each sampled state along their trajectories [2]. Implementation of this concept has involved clustering subsets of feature vectors, evaluated at the i-th state along all trajectories, and repeating this process n times for all n states along a trajectory [2]. Any trajectories that are consistently grouped together satisfy the density-connectedness criterion at every sampled state and, therefore, form a convoy or cluster of sufficiently similar paths.

The convoy detection approach motivates the cluster refinement process. Consider a set of coarsely grouped trajectories that all complete up to r revolutions, with a total curvature between $(r-1)\pi$ and $r\pi$. Each trajectory is segmented into half revolutions, i.e., the i-th segment encompasses the $(i-1)N_a + 1$-th to iN_a-th sampled states (or $iN_a + 1$-th when $i = r$ to capture the final state along the trajectory). A feature vector is then calculated for these sampled states. Next, the feature vectors for the i-th segment along all trajectories in the coarsely constructed set are clustered via DBSCAN using a specified neighborhood size that is selected as described in later paragraphs. This process is repeated for all r segments along the trajectory to produce r clustering results. Any trajectories that are consistently grouped together within these r clustering results produce one or more refined clusters. Any trajectories that are considered noise along any segment are unlabeled and discarded.

The cluster refinement process is applied in two steps. First, the coarsely constructed set of clusters are refined using the shape-based feature vector. This step ensures that trajectories appearing in the same cluster consistently possess a similar time evolution of the velocity unit vector and limits the impact of the curse of dimensionality in discovering high quality clusters. Then, each refined shape-based cluster is once again refined using the position-based feature vector. This step ensures that each cluster groups only trajectories with a similar shape and path through configuration space for their entire durations.

When using DBSCAN for cluster refinement, three governing parameters must be selected. First, the value of m_{pts} is selected to match the value used in the coarse clustering process. In addition, the Euclidean distance is once again used to compare two feature vectors. The value of ϵ is selected heuristically to approximate the size of the m_{pts} neighborhood of a trajectory that is sufficiently similar to other trajectories. Specifically, when DBSCAN is used for cluster refinement, $\epsilon_{merge} = m_{pts} \max(e, \epsilon_{threshold})$ where e is the median value of the distance from each trajectory to its nearest neighbor and $\epsilon_{threshold}$ is a reasonably small minimum threshold. Through this expression, ϵ_{merge} is approximated as m_{pts} multiplied by this median distance between nearest neighbors to provide a reasonable estimate and limit bias from any outliers in the group.

Cluster Representatives Each cluster is summarized using a single representative trajectory. This trajectory is selected as the medoid, which captures the member of the cluster that is most similar to all other members. This medoid of cluster C_k, composed of P_k members, is calculated as

$$\mathcal{T}_{med,k} = \text{argmin}_{\mathcal{T}_i \in \mathcal{C}_k} \left[\sum_{j=1, i \neq j}^{P_k} d(\bar{f}_i, \bar{f}_j) \right] \tag{19}$$

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.

Data-Driven Summarization Process

Step 1: Sample the Solution Space The first step of initial condition sampling involves selecting the position vectors. A uniform grid is defined in the configuration space of the Earth-Moon pulsating, rotating frame with a fixed grid spacing $\Delta_{pos} = 0.004$. This grid is defined within the following ranges: $x \in [-0.834, 1.156]$, $y \in [-0.2, 0.2]$, $z \in [-0.2, 0.2]$. At each position vector, the concept of the Jacobi constant from the CR3BP is leveraged to select viable initial conditions. Although the Jacobi constant is not a constant of motion in an ephemeris model, it supports consistently sampling the phase space. Given a selected value of Jacobi constant, $C_{J,des}$, only the position vectors that produce real-valued speeds in the rotating frame are retained.

Each initial condition is constrained to produce a maximum in the curvature in the Earth-Moon pulsating rotating frame. This condition supports identifying states in the rotating frame that are analogous to apses relative to meaningful reference points. Given a speed and initial position vector, each possible velocity vector is described by two spherical coordinates: an in-plane angle $\theta_{xy} \in [0, 360^\circ]$ measured from the x-axis and an out-of-plane angle $\theta_z \in [-90^\circ, 90^\circ]$. Using these definitions, the velocity vector is equal to

$$\bar{v} = \dot{x}\hat{x} + \dot{y}\hat{y} + \dot{z}\hat{z} = v\cos(\theta_{xy})\cos(\theta_z)\hat{x} + v\sin(\theta_{xy})\cos(\theta_z)\hat{y} + v\sin(\theta_z)\hat{z} \tag{20}$$

Each out-of-plane angle is sampled in increments of 2°. Then, the value of θ_{xy} is varied in increments of 1° and the associated velocity vectors are used to calculate the time derivative of the

curvature, $\dot{\kappa}(t_0)$. This calculation requires calculating the second and third time derivatives of the position vector in the rotating frame, thereby requiring transformations to/from the inertial frame and an evaluation of the equations of motion. Any two neighboring angles that produce values of $\dot{\kappa}(t_0)$ with opposite signs are used to identify an initial guess for a velocity vector that produces a maximum in the curvature. The x and y components of the velocity vector are iteratively updated using the fsolve toolbox in MATLAB to satisfy the following two conditions [27]:

$$\frac{d\kappa(\bar{x})}{dt} = 0, \qquad \frac{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}{v^2} - 1 = 0 \tag{21}$$

This process is repeated for every initial guess at every selected value of θ_z. The result is a set of one parameter curves formed by the velocity vector at each position and each speed.

At each position vector and speed, the array of discretely sampled velocity vectors that produce maxima in the curvature are summarized using localized, coarse clustering. Each initial state vector is transformed to the inertial frame at the selected initial epoch and propagated as described in the Background section. The feature vectors in Equations 17 and 18 are then calculated for each trajectory to reflect the shape and position evolution in the Earth-Moon rotating frame. The shape-based feature vectors are coarsely clustered using the approach described in the previous section with $m_{clmin} = 3$, $m_{pts} = 2$, and $\epsilon_{merge} = 0$. The result is a set of clusters summarizing the geometries of trajectories that emanate from the same position vector with a fixed speed at a curvature maximum. The representative members of each trajectory supply the initial conditions, trajectories, and feature vectors that are added to the dataset to be clusters. This process is repeated for all viable position vectors in the uniform grid.

Step 2: Cluster Individual Partitions of Trajectories All the trajectories contained within the large dataset are segmented to produce smaller partitions. Trajectories that complete the same number of half-revolutions in the Earth-Moon pulsating rotating frame are collected and then evenly distributed across N partitions of up to 10,000 members; this upper limit corresponds to a reasonable maximum dataset size to support fast clustering. Then, each partition contains only trajectories that have the same feature vector dimension and begin in a similar region of the configuration space. However, these trajectories may exhibit a diverse array of geometries.

Each partition is independently clustered. First, the trajectories in the i-th partition are coarsely clustered using the approach outlined in the previous section with $m_{clmin} = 5$, $m_{pts} = 4$, and $\epsilon_{merge} = 2\sin(10^\circ/2)\sqrt{N_s}$. Then, those clusters are refined following the procedure described earlier with $m_{clmin} = 5$, $m_{pts} = 4$, and the heuristically selected value of ϵ_{merge}. The result is a set of c_i local clusters of trajectories with a similar shape and path through the configuration space, $\mathcal{L}_i = \{\mathcal{L}_{i,1}, \mathcal{L}_{i,2}, ..., \mathcal{L}_{i,c_i}\}$. Any trajectories designated as noise at this stage are discarded.

Step 3: Aggregate Clusters to Form Global Summary Local clusters that correspond to sufficiently similar trajectories are merged across partitions to identify global clusters. This approach essentially implements distributed clustering: aggregating local clustering results to produce a global summary of a large dataset. This approach is implemented in two steps: 2) identification of candidate neighboring clusters and 2) assessment of each candidate pair to determine whether two local clusters should be merged.

To ensure computational efficiency, candidate neighboring clusters are identified using the representatives of local clusters. Specifically, for the i-th local cluster in a partition, its representative trajectory is compared to the representatives of all other clusters including trajectories with a total

absolute curvature of between $(r-1)\pi$ and $r\pi$. Then, the k nearest neighbors in each of the shape-based and position-based feature vector spaces are identified. Accordingly, up to $2k$ unique local clusters form the set of candidate neighbors to the i-th local clusters. In this paper, $k = 4$ is selected to balance computational efficiency with identifying several candidates.

Each pair of candidate neighboring clusters is examined to determine whether they should be merged. This merging decision process is performed by simultaneously clustering the feature vectors of their members using the refinement process outlined in the previous section with $m_{clmin} = 5$, $m_{pts} = 4$, and the heuristically selected value of ϵ_{merge}. If trajectories from each local cluster appear grouped together in each of the shape-based and position-based feature vector spaces, they are considered to follow sufficiently similar paths for their entire duration. As a result, these clusters are merged together. If either cluster possesses more than 10,000 members, the clusters are partitioned into smaller subsets and every combination of pairs from each candidate cluster are examined. Then, if any pair of partitions is considered to have sufficiently similar members, the entire clusters are merged together.

To calculate the global clusters from all the merging decisions, a graph is used. Each local cluster forms the nodes of the graph whereas edges are added between a pair of local clusters that should be merged. Each connected component of this graph forms a set of one or more local clusters that should be merged to form a global cluster. The result of this process is a set of P global clusters $\mathcal{G} = [\mathcal{G}_1, \mathcal{G}_2, ..., \mathcal{G}_P]$.

RESULTS

A global cluster summary is generated for natural motion in the point mass ephemeris model for trajectories that begin with a Jacobi constant of $C_J = 3.16$ in the lunar vicinity. The dataset consists of a total of 1,145,990 trajectories distributed over 123 partitions. Across all of these partitions, a total of 3,965 local clusters are generated. Following aggregation, there are 1,388 global clusters composed of 1,044,174 trajectories; 101,816 members of the full dataset are discarded as noise, typically due to lying in insufficiently dense regions of a feature vector space. Some global clusters are composed of as many as 34,995 members whereas others consist of as few as 5 members. In some cases, the smaller clusters may resemble other motions but may be too localized to be merged with other clusters.

To aid analysis of the motion types summarizing the selected clusters in the ephemeris model, a comparison to selected motions from the CR3BP is also performed. Specifically, the same process is applied to natural trajectories in the Earth-Moon CR3BP at a Jacobi constant of 3.16. These trajectories are grouped into 822 global clusters, composed of 586,365 members. Then, the global clusters from each of the ephemeris model and CR3BP are aggregated together following the procedure in Step 3 of the Data-Driven Summarization Approach, but with $k = 2$.

Selected global clusters from the ephemeris model as well as their associated motions in the lower-fidelity CR3BP are visualized. Specifically, representative members of selected clusters are displayed in Figures 1-5, categorized based on their motion types. In each figure, the representative members of trajectories in the ephemeris model are plotted in blue in three-dimensional space with their initial conditions highlighted by a blue circle marker; red colored representatives, in some figures, indicate trajectories from the CR3BP with a similar geometry. For both paths, transparent markers display selected initial conditions of other members of the cluster. A projection of the path of any representatives from the ephemeris model onto the xy-plane is drawn just below in gray to

provide perspective on a three-dimensional trajectory that is displayed via a two-dimensional figure. In addition, the Moon is located using a gray circle marker; this object is not scaled due to the use of a pulsating, rotating frame that would result in a time-varying radius in this view. Furthermore, the locations of L_1 and L_2 from the CR3BP are depicted as diamonds.

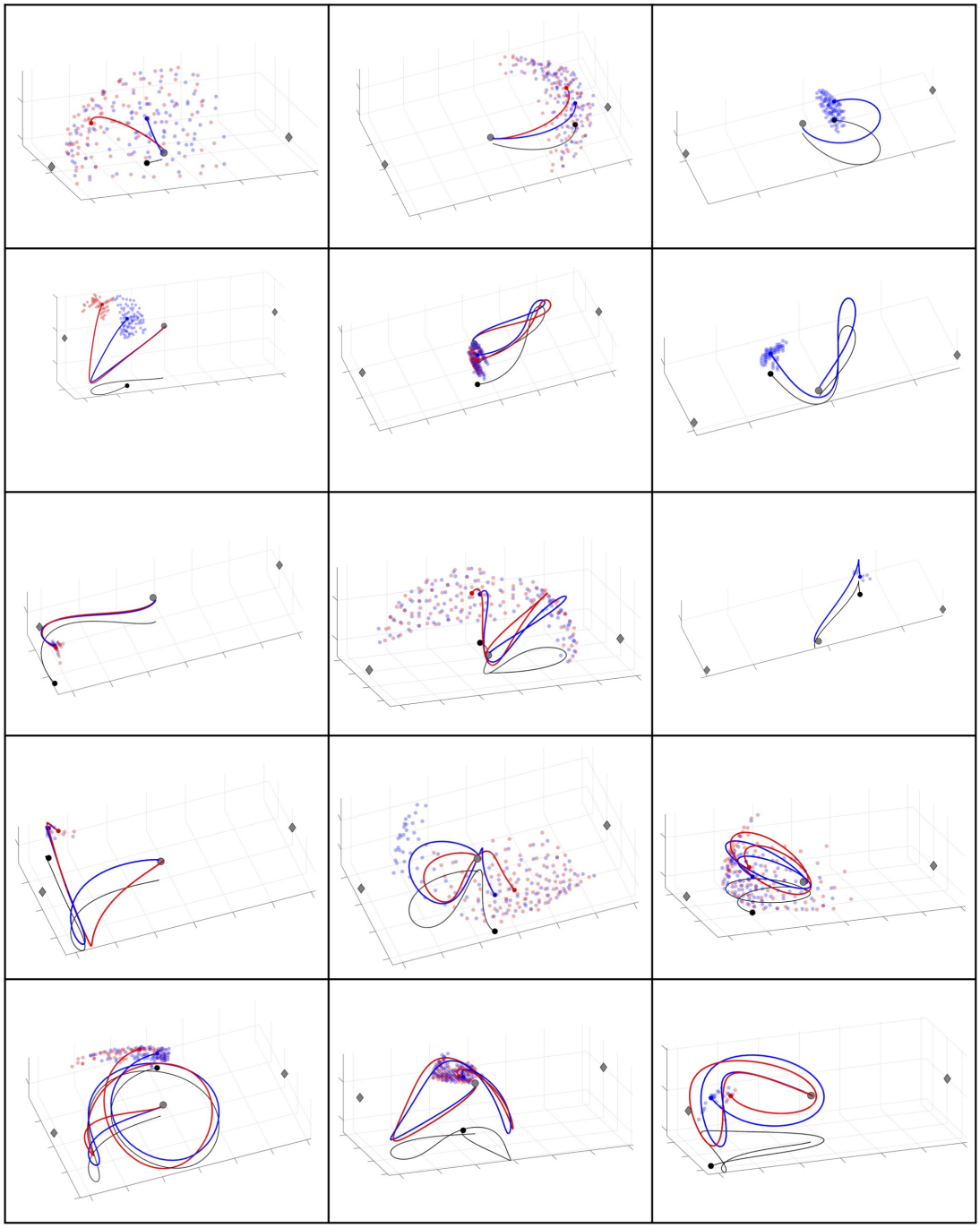

Figure 1. Selected representative trajectories associated with clusters that impact the Moon in ephemeris model (blue) and CR3BP (red) for comparison.

Analysis of Figures 1-5 reveals that the data-driven summary captures natural trajectories in the ephemeris model with a wide variety of geometries. Many of these geometries tend to be predicted by similar motions in the low-fidelity CR3BP. The subfigures which only display blue representatives indicate the trajectory was not associated with a similar motion from the CR3BP. These

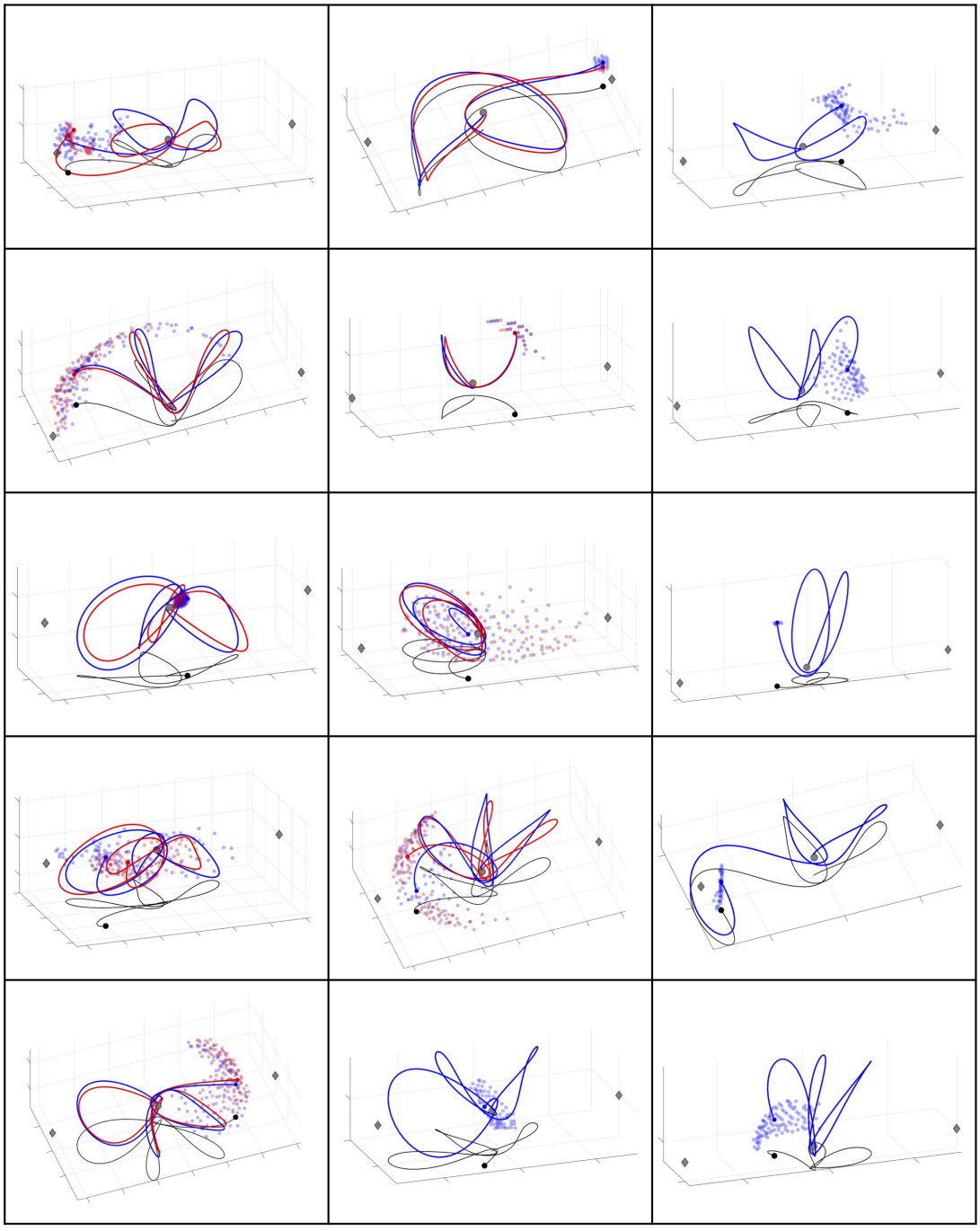

Figure 2. Selected representative trajectories associated with clusters that impact the Moon in ephemeris model (blue) and CR3BP (red) for comparison.

geometries are expected to be isolated due to: 1) representing a unique geometry in the ephemeris model, 2) lying sufficiently far from the similar motion in the CR3BP to not satisfy the density connectedness requirements during clustering based aggregation, 3) possessing a slightly different value of κ_{tot} that results in a different integer value for r compared to a similar motion from the

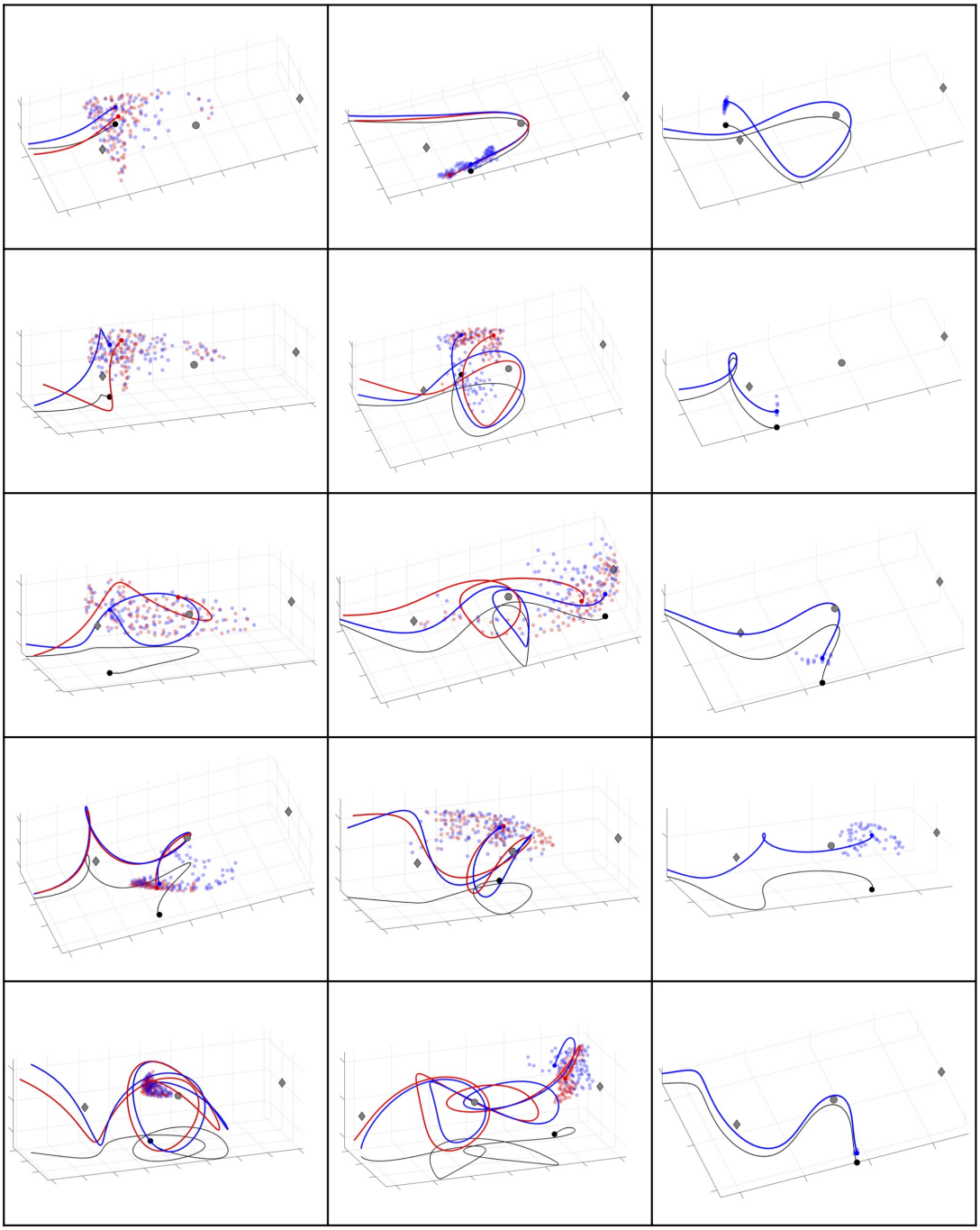

Figure 3. Selected representative trajectories associated with clusters that depart through the L_1 gateway in ephemeris model (blue) and CR3BP (red) for comparison.

CR3BP, 4) existing in a localized region of the phase space, 5) a similar motion may exist in the discarded points that were designated as noise during clustering, or 6) resembling a motion from the CR3BP at a slightly different energy level.

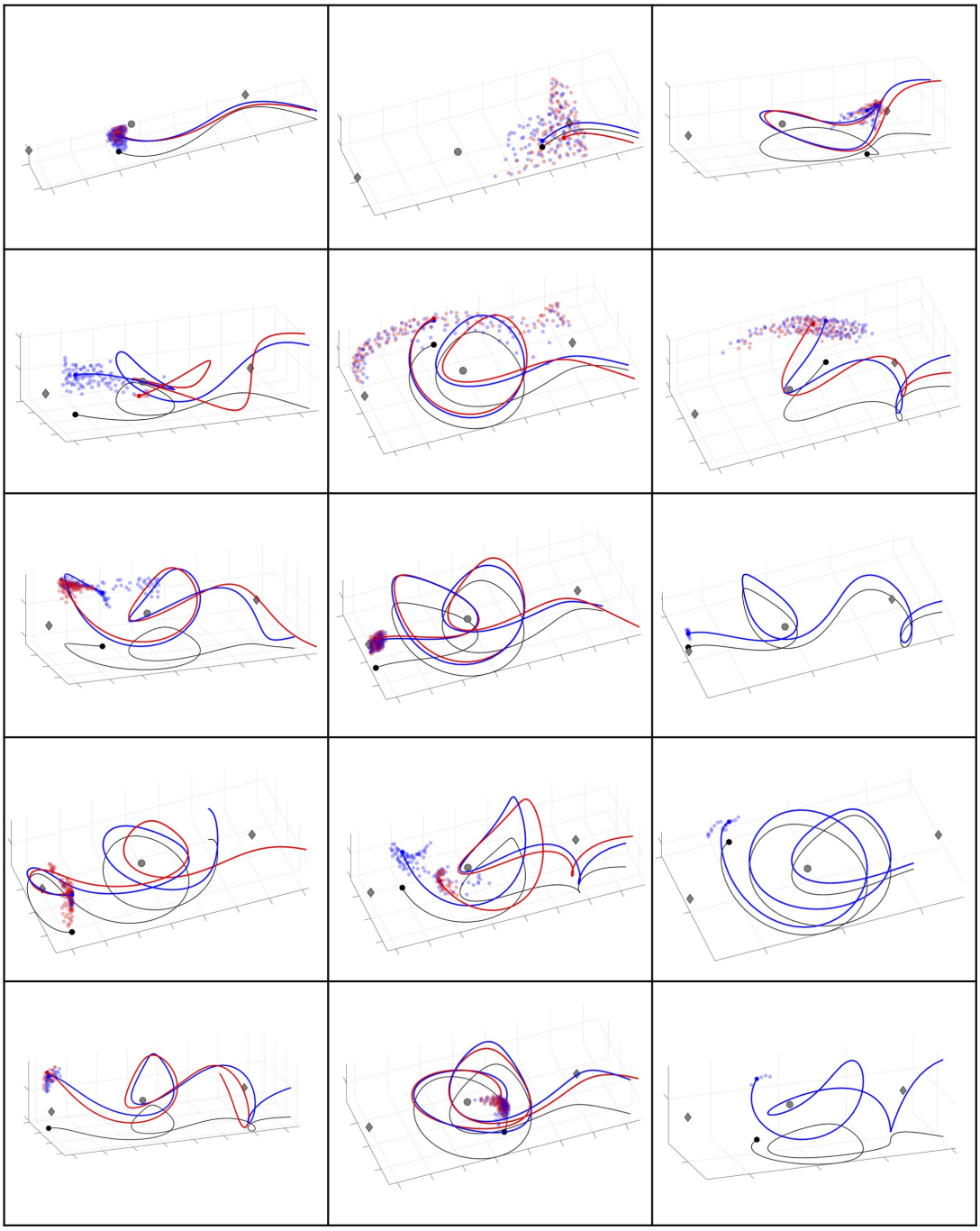

Figure 4. Selected representative trajectories associated with clusters that depart through the L_2 gateway in ephemeris model (blue) and CR3BP (red) for comparison.

CONCLUSIONS

This paper leveraged a data-driven framework to generate a high-resolution summary of natural trajectories in an Earth-Moon-Sun point mass ephemeris model. Each trajectory is generated from

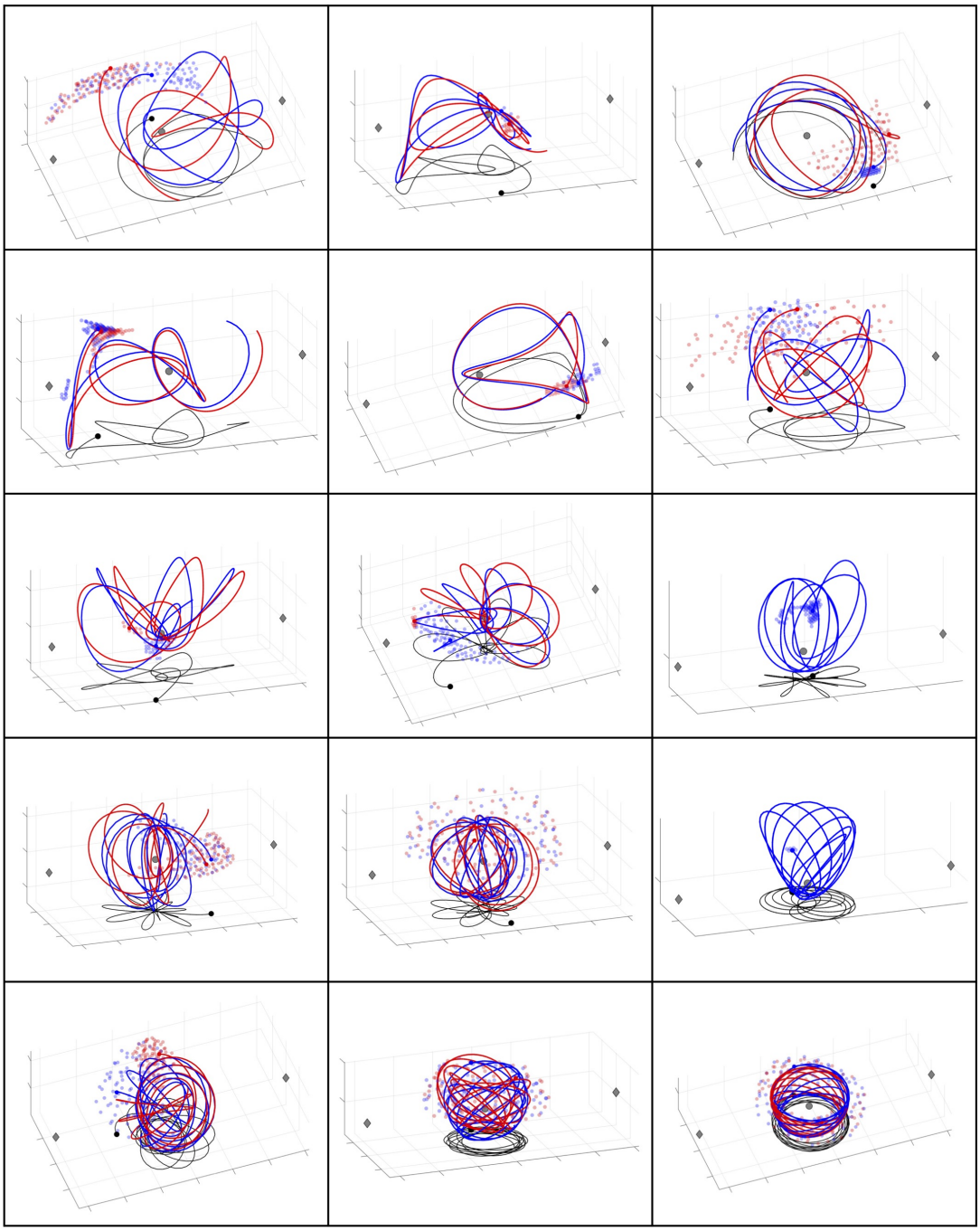

Figure 5. Selected representative trajectories associated with clusters that remain in the Moon's vicinity for 21 days in ephemeris model (blue) and CR3BP (red) for comparison.

a maximum in the curvature in the rotating frame at a fixed initial epoch on January 1, 2025 UTC for 21 days, or until impact with a primary. Then, each continuous trajectory is sampled using the geometry of the path to produce a discrete set of states. Each state is used to form finite-dimensional feature vectors that capture the shape of the trajectory and its path through the configuration space. Then, these trajectories are grouped using a distributed clustering approach that relies on HDBSCAN and DBSCAN. Each cluster captures a set of trajectories with a distinct geometry. Together, the clusters summarize the array of possible motions within this dataset in the point mass ephemeris model. Some of these clusters were presented, along with a comparison to the low-fidelity CR3BP.

ACKNOWLEDGMENT

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".

REFERENCES

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

[2] Y. Zheng and X. Zhou, *Computing with Spatial Trajectories*. New York, NY: Springer Science and Business Media, 2011, ch. 1, 5.

[3] I. Joncour, G. Duchêne, E. Moraux, and F. Motte, "Multiplicity and Clustering in Taurus Star Forming Region II. From Ultra-Wide Pairs to Dense NESTs," *Astronomy and Astrophysics*, vol. 620, no. A27, 2018.

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

[5] Z. Ivezić, A. Connolly, J. VanderPlas, and A. Gray, *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.

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

[7] N. Nakhjiri and B. F. Villac, "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.

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

[9] T. Smith and N. Bosanac, "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.

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

[11] N. Bosanac, "Data-Driven Summary of Natural Spacecraft Trajectories in the Earth-Moon System," in *2023 AAS/AIAA Astrodynamics Specialist Conference*, Big Sky, MT, Aug. 2023.

[12] S. Bonasera and N. Bosanac, "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.

[13] N. Bosanac and M. Joyner, "Data-Driven Summary of Continutous Thrust Trajectories in a Low-Fidelity Model of Cislunar Space," in *AAS/AIAA Astrodynamics Specialist Conference*, Broomfield, CO, Aug. 2024.

[14] R. Campello, D. Moulavi, and J. Sander, "Density-based clustering based on hierarchical density estimates," in *Advances in Knowledge Discovery and Data Mining*, J. Pei, V. Tseng, L. Cao, H. Motoda, and G. Xu, Eds., Heidelberg: Springer, Berlin, 2013.

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

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

[17] NASA Goddard Space Flight Center, *General Mission Analysis Tool Version R2020a: Mathematical Specifications*, 2020.

[18] R. Park, W. Folkner, J. Williams, and D. Boggs, "The JPL Planetary and Lunar Ephemerides DE440 and DE441," *The Astronomical Journal*, vol. 161, no. 3, 2021.

[19] C. Acton, "Ancillary Data Services of NASA's Navigation and Ancillary Information Facility," *Planetary and Space Science*, vol. 44, no. 1, pp. 65–70, 1996.

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

[21] K. Wardle, *Differential Geometry*. Mineola, NY: Dover Publications, Inc., 2008.

[22] N. Patrikalakis, T. Maekawa, and W. Cho, *Shape Interrogation for Computer Aided Design and Manufacturing*. 2009.

[23] M. Ester, H. Kriegel, J. Sander, and X. Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, AAAI Press, 1996.

[24] C. Malzer and M. Baum, "A Hybrid Approach To Hierarchical Density-Based Cluster Selection," in *2020 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, 2020, pp. 223–228.

[25] R. Campello, D. Moulavi, A. Zimek, and S. Sander, "Hierarchical Density Estimates for Data Clustering, Visualization, and Outlier Detection," *ACM Transactions on Knowledge Discovery from Data*, vol. 10, no. 1, pp. 1–51, 2015.

[26] L. McInnes, J. Healy, and S. Astels, "hdbscan: Hierarchical Density Based Clustering," *Journal of Open Source Software*, vol. 2, no. 11, 2017.

[27] The MathWorks Inc., *MATLAB version: 9.14 (R2023a)*, Natick, Massachusetts, United States, 2023. [Online]. Available: `https://www.mathworks.com`.