# UNSUPERVISED LEARNING TO AID VISUALIZATION OF HIGHER-DIMENSIONAL POINCARÉ MAPS IN MULTI-BODY TRAJECTORY DESIGN

# Stefano Bonasera; Natasha Bosanac<sup>†</sup>

In spatial or nonautonomous models of multi-body systems, it is often challenging to explore the solution space via a higher-dimensional Poincaré map. In this paper, unsupervised learning techniques are used to improve analysis and visualization of Poincaré maps: clustering is used to group trajectories by their geometry, while manifold learning is used for both visualization, cluster correction and correlation. This approach is demonstrated using three examples: studying the persistence of solutions across various values of the independent variables in the elliptic restricted three-body problem and the point mass ephemeris models, and across models of increasing fidelity.

# **1 INTRODUCTION**

Poincaré mapping is often leveraged in rapid and informed trajectory design strategies within multi-body dynamical systems. When constructed appropriately, this technique reduces the complexity of visualization and analysis, revealing the underlying features of the solution space. For instance, trajectory designers explore the solution space in lower-fidelity models, such as the Circular Restricted Three-Body Problem (CR3BP), via Poincaré maps to support preliminary analysis: the crossings of a map may reveal patterns that enable mission orbit selection, identification of arcs with a desired behavior, or initial guess construction. This approach has been valuable in many applications, from aiding in the explanation of the natural motion of comets<sup>1</sup> to direct application in the trajectory design of the Lunar IceCube<sup>2</sup> and EQUULEUS CubeSat missions.<sup>3</sup>

The ease of use and information that may be extracted by a human trajectory designer analyzing a Poincaré map depends on the properties of the underlying dynamical model. When the dynamical model is nonautonomous, the trajectories are three-dimensional, or a constant of motion does not exist, each intersection of a trajectory with the surface of section possesses a multi-dimensional description. These map crossings are typically visualized via a projection onto a lower-dimensional space; the result is a nonunique set of map crossings, loss of information, data obscuration and, in some cases, a lack of patterns to reveal fundamental dynamical structures. Previous researchers have explored a variety of strategies to aid visualization and analysis of multi-dimensional data projected onto lower-dimensional Poincaré maps. Examples in the field of astrodynamics include the use of additional constraints to further reduce the dimensionality of the problem and the introduction of multivariate representations for each crossing.<sup>4,5</sup>

<sup>&</sup>lt;sup>\*</sup>Ph.D. Student, Ann and H.J. Smead Aerospace Engineering Sciences, 3775 Discovery Dr., ste-fano.bonasera@colorado.edu.

<sup>&</sup>lt;sup>†</sup>Assistant Professor, Ann and H.J. Smead Aerospace Engineering Sciences, 3775 Discovery Dr., natasha.bosanac@colorado.edu.

An alternative paradigm for using higher-dimensional Poincaré maps has recently been explored by Bosanac<sup>6</sup> and Bonasera and Bosanac.<sup>7</sup> These works pose the visualization and analysis of higher-dimensional data generated during Poincaré map construction as a Big Data problem. In the interdisciplinary field of Big Data, unsupervised learning techniques have been used to: discover groupings within the data without analytical expressions or labeling; perform a dimensionality reduction to visualize the data in a lower-dimensional space; and conduct knowledge discovery. Both clustering and dimension reduction algorithms have been leveraged in pursuit of these goals in a wide variety of applications. In Bosanac<sup>6</sup> and Bonasera and Bosanac,<sup>7</sup> we have previously used clustering to group the crossings on a Poincaré map according to the geometry of the associated trajectories and provide a summary of the solution space via a set of representative solutions. Specifically, we have leveraged Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN).<sup>8</sup> In Bosanac,<sup>6</sup> this clustering-based approach is presented and demonstrated in the planar CR3BP. In Bonasera and Bosanac,<sup>7</sup> the planar investigation is extended and applied to higher-dimensional datasets associated with spatial trajectories and nonautonomous systems.

This paper extends our previous work by leveraging manifold learning algorithms to enhance visualization and analysis of higher-dimensional Poincaré maps when used in combination with clustering. Manifold learning algorithms enable a multi-dimensional dataset to be projected onto an assumed nearby lower-dimensional manifold in an unsupervised manner. One example of a recent manifold learning algorithm is Uniform Manifold Approximation and Projection (UMAP).<sup>9</sup> UMAP constructs a lower-dimensional approximation of the higher-dimensional manifold associated with a dataset by minimizing the associated topological distance between them. UMAP has been successfully used to explore a wide variety of higher-dimensional datasets in nonlinear systems, including visualizing proteins in single cell biology,<sup>10</sup> investigating genetic structure in large cohorts<sup>11</sup> and classifying the origin of solar wind.<sup>12</sup> In this paper, we use both clustering and manifold learning to effectively analyze and visualize large higher-dimensional datasets associated with Poincaré maps capturing trajectories in multi-body systems. First, HDBSCAN is used to cluster the crossings on a map by the geometry of the associated trajectories. Then, we leverage UMAP to speed up the clustering process, enable visualization of higher dimensional Poincaré maps, refine the clustering output and correlate clusters of similar solutions across distinct maps.

This paper demonstrates the application of both clustering and dimension reduction to Poincaré maps by exploring the properties and persistence of solutions with distinct geometries across dynamical models of increasing fidelity. Rapid and informed trajectory design strategies often leverage models of lower fidelity for preliminary exploration of the solution space and initial guess construction. However, it is often challenging to define a model that is of sufficiently low fidelity to support rapid analysis and sufficiently high fidelity to predict the properties of the true solution space. Thus, this paper demonstrates the application of clustering and dimension reduction to identify fundamental solution geometries via maps constructed in the Sun-Earth CR3BP, the Sun-Earth Elliptic Restricted Three-Body Problem (ER3BP) and the Sun-Earth point mass ephemeris model. Clusters of solutions are associated between these maps to study their properties and persistence as the model fidelity is increased and parameters modified. This example demonstrates the value of using both clustering and dimension reduction to: 1) analyze and visualize the higher-dimensional data on a Poincaré map; and 2) conduct a preliminary exploration of the evolution of the solution space in models of increasing fidelity. As a result, this paper demonstrates the value of using techniques from the data mining community to reduce the complexity of analyzing and visualizing a complex solution space via Poincaré maps for trajectory design within multi-body systems.

### 2 BACKGROUND: DYNAMICAL MODELS

This paper examines the evolution of the solution space in the Sun-Earth system in three distinct dynamical models of increasing fidelity. The most fundamental dynamical model is the autonomous CR3BP, which assumes that the Sun and Earth follow circular orbits. Then, the ER3BP is used to incorporate a nonzero eccentricity for the orbits of the primaries. Finally, the highest fidelity dynamical model leveraged in this analysis is the point mass ephemeris model of the Sun-Earth system. In this section, each of these models is presented.

### 2.1 Circular Restricted Three-Body Problem

The CR3BP is used to model the motion of an assumed massless spacecraft under the gravitational influence of two point mass primaries, following circular orbits about their common barycenter.<sup>13</sup> Following introduction of these assumptions, three characteristic quantities are introduced to nondimensionalize mass, length and time:  $m^*$ , set equal to the sum of the masses of the primaries,  $P_1$  and  $P_2$ ;  $l^*$ , equal to the constant distance between  $P_1$  and  $P_2$ ; and  $t^*$ , producing a period of the two primaries equal to  $2\pi$ . An orthogonal frame  $(\hat{x}, \hat{y}, \hat{z})$ , rotating with the two primaries, is also introduced: the  $\hat{x}$ -axis is directed from  $P_1$  to  $P_2$ , the  $\hat{z}$ -axis is aligned with the orbital angular momentum of the system, and the  $\hat{y}$ -axis completes the right-handed triad. Using these definitions, the nondimensional state vector of the spacecraft relative to the system barycenter and in the rotating frame is written as  $\boldsymbol{x} = [x, y, z, \dot{x}, \dot{y}, \dot{z}]^T \in \mathbb{R}^6$ . Then, the nondimensional equations of motion for the spacecraft in the CR3BP and in the rotating frame are written as

$$\ddot{x} - 2\dot{y} = U_x, \qquad \ddot{y} + 2\dot{x} = U_y \qquad \ddot{z} = U_z \tag{1}$$

where  $\mu$  is the mass ratio, equal to  $\mu \approx 3.00348064 \times 10^{-6}$  in the Sun-Earth CR3BP, the pseudopotential function is  $U(\mathbf{x}) = (x^2 + y^2)/2 + (1 - \mu)/r_1 + \mu/r_2$ , and  $U_x$ ,  $U_y$  and  $U_z$  denote the partial derivatives of U with respect to x, y and z, respectively. The distances of the spacecraft from the two primaries are  $r_1 = \sqrt{(x + \mu)^2 + y^2 + z^2}$  and  $r_2 = \sqrt{(x - 1 + \mu)^2 + y^2 + z^2}$ , respectively. The CR3BP admits one integral of motion, labeled the Jacobi constant and equal to  $JC(\mathbf{x}) =$  $2U(\mathbf{x}) - \dot{x}^2 - \dot{y}^2 - \dot{z}^2$ .<sup>13</sup> At a fixed value of the Jacobi constant, the Zero Velocity Surfaces (ZVS) bound trajectories by separating allowable and forbidden regions of motion. The intersections of the ZVS with a plane are commonly referred to as the Zero Velocity Curves (ZVCs). Within the ZVS, a wide variety of solutions exist in the CR3BP: equilibrium points  $L_i$  for i = [1, 5], bounded solutions such as periodic and quasi-periodic orbits, and chaos.

### 2.2 Elliptic Restricted Three-Body Problem

The ER3BP is formulated similar to the CR3BP; however, the two primaries are assumed to follow elliptical orbits about the system barycenter. The distance between  $P_1$  and  $P_2$  is no longer constant as the primaries following conics with a nonzero eccentricity, equal to  $e_P \approx 0.0167$  in the Sun-Earth system. Thus, the value of the length characteristic quantity,  $l^*$ , is no longer constant. As a result, the primaries only appear at fixed locations over time in a pulsating, rotating frame. In addition, the true anomaly, f, of the primary system is used as an independent variable, rather than time, with f = 0 when the primaries are located at periapsis.<sup>13</sup> Using this configuration, the nondimensional equations of motion for the spacecraft in the ER3BP are written as:

$$x'' - 2y' = \omega_x, \qquad y'' + 2x' = \omega_y \qquad z'' = \omega_z \qquad t' = (1 - e_P^2)^{3/2} / (1 + e_P \cos f)^2$$
 (2)

where the prime (.)' indicates a derivative with respect to f and the pseudo-potential function is  $\omega(\mathbf{x}, f) = (U(\mathbf{x}) - z^2 e_P \cos f/2)/(1 + e_P \cos f)$ .<sup>13</sup> However, due to the explicit dependence of Eq. (2) on the time-like quantity f, an integral of motion no longer exists.<sup>14</sup>

#### 2.3 Point Mass Ephemeris Model

A point mass ephemeris dynamical model is often used as a higher-fidelity representation of the Sun-Earth system during initial trajectory construction activities. In this model,  $N_e$  bodies are each assumed to be spherically symmetric with a dimensional mass  $M_i$ . Body  $P_j$  is located in an inertial reference system with axes  $(\hat{X}, \hat{Y}, \hat{Z})$  and origin O by the position vector  $R_j =$  $X_j\hat{X} + Y_j\hat{Y} + Z_j\hat{Z}$ . The spacecraft,  $P_3$ , is located at the coordinates (X, Y, Z); its relative position vector with respect to body  $P_j$  is written as  $R_{j,3} = (X - X_j)\hat{X} + (Y - Y_j)\hat{Y} + (Z - Z_j)\hat{Z}$ . The point mass ephemeris model is formulated by expressing the acceleration acting on body  $P_3$  in a  $P_j$ -centered J2000 inertial coordinate system due to the gravitational influence of all  $N_e$  bodies. In this paper, the mass of the spacecraft is assumed to be negligible in comparison to each of the  $N_e$ bodies.<sup>2</sup> The resulting dimensional equations of motion for a spacecraft in the point mass ephemeris model and in the  $P_j$ -centered J2000 inertial coordinate system are written as:

$$\begin{split} \ddot{X}_{j,s/c} &= -\frac{GM_j}{R_{j,s/c}^3} X_{j,s/c} + G \sum_{\substack{i=1\\i \neq j}}^{N_e} M_i \left( \frac{X_{s/c,i}}{R_{s/c,i}^3} - \frac{X_{j,i}}{R_{j,i}^3} \right) \\ \ddot{Y}_{j,s/c} &= -\frac{GM_j}{R_{j,s/c}^3} Y_{j,s/c} + G \sum_{\substack{i=1\\i \neq j}}^{N_e} M_i \left( \frac{Y_{s/c,i}}{R_{s/c,i}^3} - \frac{Y_{j,i}}{R_{j,i}^3} \right) \\ \ddot{Z}_{j,s/c} &= -\frac{GM_j}{R_{j,s/c}^3} Z_{j,s/c} + G \sum_{\substack{i=1\\i \neq j}}^{N_e} M_i \left( \frac{Z_{s/c,i}}{R_{s/c,i}^3} - \frac{Z_{j,i}}{R_{j,i}^3} \right) \end{split}$$
(3)

where G is the universal gravitational constant. Once an initial epoch and spacecraft state are defined, the states of all bodies included in the point mass ephemeris model are retrieved using the Jet Propulsion Laboratory DE421 ephemerides via the SPICE toolkit.<sup>15</sup> The same ephemerides are leveraged when transforming the spacecraft state into the rotating frame.<sup>16</sup>

# 3 BACKGROUND: POINCARÉ MAPS

Poincaré mapping is a tool from dynamical systems theory that is often leveraged to reduce the dimension of continuous trajectories via their intersections with a surface of section, facilitating the analysis of complex solution spaces. A Poincaré map is constructed by first defining a surface of section that is transverse to the flow. Each continuous arc is then reduced to a sequence of discrete points, corresponding to subsequent intersections of this hyperplane. These intersections are then represented via a Poincaré map.

This paper focuses on prograde periapsis maps for planar trajectories in the Sun-Earth system, constructed using the same set of initial conditions across different dynamical models and system parameters. To supply a high-level overview of the process for constructing a prograde periapsis map, consider its implementation in the planar CR3BP at a fixed Jacobi constant; additional details appear in Bosanac and Bonasera and Bosanac.<sup>6,7</sup> First, a periapsis surface of section, relative to the

Earth in the Sun-Earth CR3BP, is defined as:<sup>17</sup>

$$(x - 1 + \mu)\dot{x} + y\dot{y} + z\dot{z} = 0 \quad \cup \quad (x - 1 + \mu)\ddot{x} + y\ddot{y} + z\ddot{z} + \dot{x} + \dot{y} + \dot{z} > 0 \tag{4}$$

Initial conditions are seeded as prograde perigees in the vicinity of the secondary, defined within the boundaries of the  $L_1$  and  $L_2$  gateways and the ZVS. By fixing the Jacobi constant, the velocity magnitude at each feasible combination of configuration space variables is  $v = \sqrt{2U - JC}$ . If v is a real number, the prograde perigee condition enables direct calculation of the velocity vector associated with an initial position vector that is located in the plane of the primaries and corresponds to an initially prograde direction of motion. Each initial perigee then possesses the form  $x_{IC} = [x, y, z = 0, \dot{x}, \dot{y}, \dot{z} = 0]^T$ . For each initial prograde perigee, a trajectory is propagated until: 1) a specified number of subsequent perigees occurs; 2) the trajectory exits the Earth vicinity through either the  $L_1$  or  $L_2$  gateways; or 3) the trajectory passes within a nondimensional distance of  $10^{-5}$  to the Earth. In the CR3BP, the map crossings associated with the perigee surface of section are uniquely represented on a two-dimensional projection such as (x, y). However, when a similar initial condition set is used to construct a set of trajectories in the ER3BP or the point mass ephemeris model, the resulting map crossings are higher-dimensional. Under these conditions, it may be challenging to employ these maps for rapid and informed trajectory design strategies. Thus, data mining techniques, such as clustering and manifold learning, are employed to effectively explore the available solution space and reduce the complexity of the analysis.

# 4 UNSUPERVISED LEARNING

This work leverages unsupervised learning to organize the trajectories associated with Poincaré map crossings into clusters based on their geometry. A dataset is first constructed by describing each member with a feature vector; in this paper, the feature vector summarizes the trajectory associated with each map crossings. A selected clustering algorithm then divides the trajectory set into groups: arcs in the same cluster are considered similar, while arcs in different sets are assumed dissimilar. Among the variety of available unsupervised techniques for clustering, this paper leverages the Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) algorithm, developed by Campello, Moulavi and Sander, following the approach outlined by Bosanac and Bonasera and Bosanac.<sup>6–8</sup> This section contains only a brief high-level overview of the input parameters. For more details about the clustering procedure, see Bosanac and Bonasera and Bosanac.<sup>6,7</sup> Each cluster, identified by HDBSCAN, is then summarized by a unique, representative trajectory. Learning manifold algorithms offer a computationally efficient unsupervised approach for the representation of higher-dimensional data that is clustered by HDBSCAN. Among the variety of existing algorithms, the Uniform Manifold Approximation and Projection (UMAP), developed by McInnes, Healy and Melville is leveraged.<sup>9</sup> UMAP enables an unsupervised dimension reduction, preserving the structure of the data in this projection, even in datasets generated from nonlinear models. This study combines HDBSCAN and UMAP to obtain a precise final clustering result and to correlate clusters across different datasets. In this section, an overview of HDBSCAN is presented, together with the definition of the feature and the approach used to define a cluster representative. An overview of UMAP is then presented.

### 4.1 Clustering via HDBSCAN

HDBSCAN is a density-based hierarchical clustering algorithm developed by Campello, Moulavi and Sander.<sup>8</sup> This algorithm takes as an input the dataset  $[T] = {t_1, t_2, ..., t_N}$ , composed of

N members  $t_i \in \mathbb{R}^M$ . Then, HDBSCAN clusters data in sufficiently dense-regions of the *M*-dimensional space, assigning each datapoint  $t_i$  to a cluster or labeling it as a noise point. Two input parameters govern the clustering algorithm:  $m_{pts}$  and  $m_{clSize}$ . The first,  $m_{pts}$ , influences the computed distances  $d(\cdot, \cdot)$  between two datapoints by changing the size of the local neighborhood used to assess density; the second,  $m_{clSize}$ , sets the minimum number of datapoints that may be considered to form a single cluster. In this analysis, the HDBSCAN algorithm is accessed via the hdbscan clustering library in Python.<sup>18</sup>

### 4.2 Feature Vector Definition

Unsupervised learning algorithms, such as HDBSCAN and UMAP, require that each continuous trajectory arc is summarized by a finite-dimensional feature vector. The feature vector is designed to balance the fidelity of the trajectory representation with computational performance, while avoiding the well-known curse of dimensionality.<sup>6</sup> The work presented in this paper follows the approach developed by Bosanac and Bonasera and Bosanac by using a geometry-based representation for each trajectory. Each continuous arc is summarized by sequence of apses, i.e., periapses and apoapses, calculated with respect to the secondary.<sup>6,7</sup> The feature vector is defined as  $\mathbf{t}_i = [\mathbf{s}_{i,0}, \mathbf{s}_{i,1}, \dots, \mathbf{s}_{i,k}, \dots, \mathbf{s}_{i,N_a}]^T \in \mathbb{R}^{5N_a}$ , with  $N_a$  representing the maximum number of apses and  $\mathbf{s}_{i,k}$  formulated as

$$\mathbf{s}_{i,k} = [\tau_{i,k}, x_{i,k}, y_{i,k}, \dot{x}_{i,k}, \dot{y}_{i,k}]^T \in \mathbb{R}^5$$
(5)

where  $(x_{i,k}, y_{i,k}, \dot{x}_{i,k}, \dot{y}_{i,k})$  represents the state components of the k-th apse in the rotating frame, while  $\tau_{i,k}$  is the time normalized by the total propagation time along that trajectory. Terminating conditions are also employed, including impacting one of the primaries or departing through the  $L_1$  or  $L_2$  gateways, to produce some trajectories that complete fewer than  $N_a$  apses. In this case, the feature vector is filled with placeholder vectors as  $s_{i,P} = \pm [0, 10, 0, 0, 0]^T$ . The positive placeholder is used when the trajectory ends prior to reaching an apoapsis, while the negative value is used when the trajectory terminates prior to reaching a periapsis.<sup>6</sup> After evaluating each feature vector, the dataset is normalized to mitigate the potential for ill-conditioning. A component-wise normalization is applied such that each component of the feature vector is in the range [-1, 1].

#### 4.3 Cluster Representatives

Following clustering, each group of trajectories is described by a single, representative arc and associated map crossing. Following the approach of Bosanac and Bonasera and Bosanac, this representative member of a cluster is defined as the medoid.<sup>6,7</sup> A medoid, or clustroid, is defined as the member of a cluster that is most similar to the other members of the same cluster.<sup>19</sup> The medoid  $t_{med}^{(j)}$  of cluster  $C_j = \{t_1^{(j)}, t_2^{(j)}, \ldots, t_{M_j}^{(j)}\}$ , with cardinality  $|C_j| = M_j$ , is calculated as

$$\boldsymbol{t}_{med}^{(j)} = \operatorname{argmin}_{\boldsymbol{t}_{k}^{(j)} \in C_{j}} \sum_{i=1, i \neq k}^{M_{j}} d\left(\boldsymbol{t}_{i}^{(j)}, \, \boldsymbol{t}_{k}^{(j)}\right)$$
(6)

To improve the computational efficiency of calculating the medoid, a soft-clustering modification of HDBSCAN, available in the *hdbscan* clustering library in Python, is employed: only those members of a cluster that possess a unitary probability for belonging to that cluster are used in the summation above. This approach excludes members of a cluster that tend to lie at the boundary of a cluster from this computation of the medoid.

### 4.4 Dimension Reduction via UMAP

The Uniform Manifold Approximation and Projection (UMAP) algorithm, developed by McInnes et al., offers one approach for dimensionality reduction via manifold learning.<sup>9</sup> UMAP is founded on the premise that the majority of the high-dimensional information describing each member of a dataset is intrinsically redundant; thus, UMAP focuses on revealing the latent variables associated with a dataset.<sup>9</sup> The algorithm leverages ideas from algebraic and fuzzy topology to support the dimensionality reduction process. In fact, UMAP constructs a projection of the dataset onto a low-dimensional manifold that is topologically similar to the high-dimensional description of the data. First, the input dataset is assumed to be uniformly distributed on a high-dimensional manifold. Then, UMAP calculates a local Riemannian metric on the high-dimensional manifold that is assumed to be locally-connected using fuzzy simplicial sets. This process produces a fuzzy topological structure which is representative of the higher-dimensional dataset. To generate a topological equivalent in the low-dimensional space, UMAP initializes a candidate low-dimensional representation using spectral embedding techniques. With a fixed high-dimensional topological representation, UMAP optimizes the low-dimensional projections by minimizing the cross-entropy between the 1-simplicies of the two representations. The dimension of the projection space  $\mathbb{R}^n$  is user-defined. The optimization problem then employs stochastic gradient descent for computational efficiency; however, for reproducibility, the user can fix the random seed generator for a relatively small additional computational cost. This approach produces a useful lower-dimensional embedding for the high-dimensional dataset.

UMAP is governed by a variety of fundamental input parameters that impact the computed embedding. Among the variety of input parameters, this work focuses on three fundamental quantities:  $n_{neigh}$ ,  $m_{dist}$  and  $n_{comp}$ . The first,  $n_{neigh} \in \mathbb{N}^+$ , supplies a measure of the local versus global structure of the dataset used in calculating the lower-dimensional projection; a low value produces a final embedding that is focused on the local structure. The second input parameter,  $m_{dist} \in [0, 1]$ , balances the density level of the projections; a low value favors a condensed final embedding. The third parameter,  $n_{comp} \in \mathbb{N}^+$ , sets the dimension of the lower-dimensional Euclidean manifold. This work employs UMAP to visually investigate the relative association between clusters associated with distinct datasets and to correct the clustering output. To aid visualization of the projected dataset,  $n_{comp} = 3$ . Due to the large size of the computed datasets, high values for  $n_{neigh}$  and low values for  $m_{dist}$  are selected to generate a global and condensed projection of the dataset.

# 5 OVERVIEW OF TECHNICAL APPROACH

This section supplies an overview of the data-driven procedure used to track the persistence of trajectories across distinct dynamical models and values of the independent variable. This procedure is inspired by distributed data mining. First, perigee maps are clustered using HDBSCAN for the same initial condition set, propagated using different dynamics. These maps are clustered individually to reduce the computational complexity of processing a large dataset. Map crossings that are designated as noise points but exist close to a cluster are reassigned using UMAP. Each cluster, within each dataset, is then sampled and aggregated to produce a global summary. In an approach that is motivated by Bosanac and Bonasera and Bosanac, clusters in this global summary are then correlated.<sup>6,7</sup> However, in this particular paper, UMAP is leveraged to perform this correlation process. Clusters that are correlated across distinct maps correspond to trajectories that persist in their general geometry across various dynamical models, governed by distinct independent variables. Each of these steps is described within this section.

### 5.1 Clustering Individual Perigee Maps with HDBSCAN

The first step in studying the evolution and persistence of solution geometries across multiple dynamical models is to construct a set of perigee maps across each model and for various parameters. For each perigee map, the same set of initial conditions is employed, defined using planar and prograde perigees that possess a Jacobi constant of JC = 3.00088 in the CR3BP, similar to the one investigated by Bosanac and Bonasera and Bosanac.<sup>6,7</sup> Once this set of initial conditions is defined in the CR3BP, the same initial states are used to define propagation in the higher fidelity dynamical models. To construct the initial condition set, planar position vectors are seeded within the vicinity of the Earth: this paper uses up to 401 *x*-coordinates between the locations of the Sun-Earth  $L_1$  and  $L_2$  points in the CR3BP and up to 401 locations of the *y*-coordinate in [-0.01, 0.01], while the *z*-coordinate is set equal to 0. For each position vector, the associated speed is calculated from the Jacobi constant as  $v = \sqrt{2U - JC}$  in the CR3BP. If this expression produces a real value, the position vector lies within the ZVS and is added to the initial condition set. Then, the periapsis condition is used to calculate the velocity unit vector and, therefore, produce the complete state at the initial condition.

For the same set of initial conditions, multiple datasets are constructed for each dynamical model and selected values of dependent variables, where appropriate. For the dataset describing trajectories in the CR3BP, each prograde perigee is propagated in the CR3BP until: a total of 7 apses with respect to the Earth occurs; the spacecraft distance from the secondary falls below  $r_2 < 10^{-5}$ ; or the trajectory departs the Earth vicinity through either the  $L_1$  or the  $L_2$  gateways. Each continuous trajectory is then discretized via the geometry-based feature vector outlined in Equation (5). If a trajectory admits fewer than 7 apses, a placeholder vector is leveraged to produce the same dimension for each feature vector. Following this procedure, the complete dataset [T] is composed of |T| = 31544 map crossings. To generate a similar dataset in the Sun-Earth ER3BP, the same set of initial conditions are propagated using the ER3BP equations of motion in Equation (2). However, this procedure is repeated multiple times for various initial values of the  $f_0$ , the true anomaly of the primary system. To generate a dataset in the point mass ephemeris model, the equations of motion from Equation (3) govern propagation of each initial condition. Due to the forms of these equations, each initial condition that is originally expressed in the rotating frame is transformed into the Earth-centered inertial coordinate system for a specified initial modified Julian date (MJD). Once the apses along the trajectory, calculated with respect to the Earth, are recorded, the state at each apsis is transformed into the rotating frame at the associated epoch. This procedure is repeated multiple times for various values of the initial modified Julian date, which influences the ephemerides of the Sun and Earth.

Each dataset, formed using the same initial condition set but generated using different dynamical models and system parameters, is input to the HDBSCAN clustering algorithm. All dataset are clustered using the same set of input parameters, selected as  $m_{pts} = 200$  and  $m_{clSize} = 100$ . This clustering procedure is first applied to the dataset associated with planar prograde perigees in the Sun-Earth CR3BP at a Jacobi constant of JC = 3.00088, generated for up to 7 apses. The resulting 13 clusters of map crossings is displayed in dimensional coordinates in the rotating frame in Figure 1(a), with each cluster identified by a unique shade of red or blue and labeled; HDBSCAN identifies 6.23% of the dataset as noise, indicated via black points, for the specified combination of feature vector definition and input parameters. Of course, alternate feature vector definitions and input parameters may yield a different number and form of clusters. The equilibrium points are displayed as red diamonds in this figure, while the zero velocity curves outline the gray shaded



Figure 1. Clustered perigee maps near the Earth, constructed for the same set of initial conditions in: (a) the Sun-Earth CR3BP at JC = 3.00088 and (b) the Sun-Earth ER3BP with  $f_0 = \pi/2$ .

forbidden regions. As an additional example, Figure 1(b) displays the clusters of map crossings for trajectories generated in the ER3BP with an initial true anomaly of  $f_0 = \pi/2$  and the same set of input parameters. Note that, in this figure and similar maps constructed in higher-fidelity models, the equilibrium points and ZVCs calculated in the CR3BP are overlaid only to supply perspective. Analysis of Figure 1(b) reveals that this map, constructed for this specified initial true anomaly in the ER3BP, admits similar clusters to those obtained for the CR3BP in Figure 1(a), with some slight shifts and distortion. It is possible that some of the clusters in Figure 1(b) correspond to trajectories with a similar geometry to those captured on the clustered map in Figure 1(a); the goal of this paper is to correlate these clusters of similar solutions across distinct maps.

## 5.2 Refining Noise Points with UMAP

Although the HDBSCAN clustering algorithm is well-suited to discover clusters of distinct densities within a dataset, it may produce noise points near the boundaries of clusters or in the sensitive regions close to the Earth; thus, UMAP is used to robustly assign noise points to clusters of similar solutions. Specifically, the entire dataset is projected onto the lower-dimensional manifold calculated by UMAP and noise points that lie close to a cluster are relabeled. As an example of this procedure, consider the perigee map constructed in the ER3BP with  $f_0 = \pi/2$ , as displayed in Figure 1(b). The high-dimensional dataset [T] that describes the trajectories associated with each map crossing in this dynamical model are projected onto a three-dimensional space via UMAP using the input parameters  $n_{neigh} = 200$  and  $m_{dist} = 0.0$ : this parameterization produces a lowerdimensional representation that focuses on the global structure of the dataset. Figure 2(a) displays the dataset projected onto the resulting three-dimensional space, with points colored according on the associated cluster, consistent with Figure 1(b). Analysis of this figure reveals that UMAP produces a lower-dimensional embedding that separates the dataset in a similar manner to the groupings discovered by HDBSCAN, despite not being provided with these labels. However, there are a few differences. First, clusters such as cluster 8 are split into sub-groups by UMAP and appear in distinct regions of the embedding. Second, some larger clusters, such as clusters 12 and 14, are bounded



Figure 2. Noise correction applied to the map in Figure 1(b), composed of trajectories generated in the ER3BP with  $f_0 = \pi/2$ : (a) UMAP projection of the pre-processed map and (b) post-processed map.

by noise points, depicted in black. Since the lower dimensional embedding constructed by UMAP preserves the structure of the data, but not density, this projection is used to refine the clustering result and reassign these noise points that appear to coincide with clusters. For each noise point identified by HDBSCAN, the distance to a maximum of 2000 randomly-selected points is calculated for each cluster using the coordinates in the lower-dimensional embedding. A noise point is reassigned to a cluster if the distance to any member in that cluster is lower than 0.5. Of course, this is a user-selected parameter; yet, performing this distance-based reassignment strategy in the lower-dimensional representation is more robust than performing it in the full *M*-dimensional feature vector space. The results of this noise reassignment approach, applied to the clustered map in Figure 1(b), are depicted in Figure 2(b): the fraction of noise within the dataset decreases from 11.46% following application of HDBSCAN to 2.18% after noise reassignment. Visual inspection of Figure 2(a) suggests that this result is reasonable. Thus, this approach is applied to each clustered map. The resulting maps are then used to perform cluster correlation and associate clusters of similar solutions across distinct dynamical models and values of the independent variable.

### 5.3 Correlating Clusters Across Maps with UMAP

UMAP is leveraged to construct a relative robust procedure for correlating clusters of geometrically similar solutions across distinct maps. Performing this correlation step in the full Mdimensional space in a straightforward and robust manner would be challenging due to the curse of dimensionality and the sensitivity of states within distinct regions of the phase space. However, the lower-dimensional embedding constructed by UMAP preserves relative distances between members of a dataset, locating similar members in nearby regions and separating dissimilar members. Since the minimum distance between two datapoint on the UMAP projection is an input parameter to the algorithm, it is relatively straightforward to construct a distance-based cluster correlation procedure that is implemented in the lower-dimensional space and governed by a single value of a distance-based threshold. This procedure is implemented as follows:

• Generate and cluster each map: D distinct datasets,  $\{[T]_1, [T]_2, \dots, [T]_D\}$ , are constructed

for each perigee map, generated for each dynamical model and set of system parameters. For each map  $[T]_i$ , with  $i \in [1, D]$ , HDBSCAN used with  $m_{pts} = 200$  and  $m_{clSize} = 100$  to obtain an initial clustering result of  $G_i$  clusters. Then, noise reassignment is performed using UMAP as described in Section 5.2.

- Sample clusters in each map to produce a global cluster summary: Each cluster  $[S]_i^j$ , with  $j \in \mathbb{N}^+$ , of each map described by the dataset  $[T]_i$ , is sampled to produce a subset of  $n_{sub} = 300$  points. The sampled map crossings from all clusters across all maps are aggregated to produce a single reduced global dataset [P]. These subsets of cluster members are identified by leveraging the soft-clustering modification to HDBSCAN, available in the hdbscan clustering library in Python: only members of  $[S]_i^j$  with a probability p > 0.8 of being assigned to  $[S]_i^j$  are used to populate [P]. If the number of members that satisfies this condition for cluster  $[S]_i^j$  is greater than a user-specified maximum value  $n_{sub}$ , these points are sampled to contribute no more than  $n_{sub}$  members to the global dataset. Once this procedure has been implemented for all clusters across all maps, the reduced dataset [P] is populated with a total of  $[Q]^j = \sum_{i=1}^D G_i$  clusters, each represented by up to  $n_{sub}$  members.
- Construct a lower-dimensional embedding of the global dataset: The global dataset [P] is renormalized in a component-wise manner and then processed with UMAP to obtain a threedimensional representation of the higher-dimensional dataset. This procedure is implemented using  $n_{neigh} = 100$  and  $m_{dist} = 0.0$  to enhance the global structure of the projected dataset.
- Correlate clusters across distinct maps: The lower-dimensional projection of the global dataset [P] is analyzed to automatically correlate clusters of similar trajectories. First, the distance of the projected representations of each member of  $[Q]^j$  from members of all clusters is calculated, excluding points from the same original map  $[T]_i$ . The minimum values of these distances are recorded for each member and used to compute an average minimum distance between two clusters  $[Q]^j$  and  $[Q]^k$  in the global dataset using the lower-dimensional embedding. If the minimum average distance between clusters  $[Q]^j$  and  $[Q]^k$  is lower than a threshold value  $t_{avg}$ , the two clusters are correlated and are assigned the same cluster ID.

When two clusters are correlated, they are assumed to be composed of map crossings that produce trajectories with a similar geometry, despite existing in different dynamical models and at different values of the independent variable. In fact, UMAP is not provided with any information about the generating model and its parameters; it is only provided a summary of the geometry of the associated trajectories, via a sequence of apses. As a result, key information about the evolution and persistence of trajectories of a specific geometry may be inferred in a robust, data-driven manner.

## 6 RESULTS: STUDYING CLUSTER PERSISTENCE ACROSS DYNAMICAL MODELS

The presented data-driven approach for clustering and correlating trajectories that exist across distinct dynamical models is demonstrated in the Sun-Earth system. Several perigee maps are constructed for the same set of initial conditions with trajectories generated in each of the CR3BP, ER3BP and point mass ephemeris models. Each perigee map is clustered individually via HDB-SCAN and some noise points reassigned using UMAP. Then, clusters in each map are correlated to study their persistence across models and values of the independent variables. In this section, this procedure is demonstrated using three examples that study the evolution of trajectories, described

by their geometry, across: 1) various values of the initial true anomaly in the ER3BP; 2) various initial epochs in the point mass ephemeris model; and 3) across dynamical models.

### 6.1 Cluster Persistence Across Initial True Anomalies in the ER3BP

The map crossings associated with several perigee maps in the ER3BP, each constructed for various values of the initial true anomaly, are clustered individually and then combined in a data-driven approach to form a single global clustering result. Consider 10 perigee maps constructed for the following distinct values of the initial true anomaly of the primaries:  $f_0 = [0, \pi/6, 5\pi/24, \pi/4, 7\pi/24, \pi/4, \pi/24, \pi/4, \pi/24, \pi/4, \pi/24, \pi/24, \pi/4, \pi/24, \pi$  $\pi/3, \pi/2, 13\pi/24, 7\pi/12, 2\pi/3$ ]. Each of these perigee maps is clustered individually using HDB-SCAN with the selected input parameters, as described in Section 5.1, and noise points reassigned as appropriate, as described in Section 5.2. These individually clustered datasets are then processed to form a global cluster summary as described in Section 5.3. During this procedure, clusters with a minimum average distance of  $t_{avg} = 1.5$  are considered sufficiently close in the lower-dimensional embedding to be correlated; of course, this is a tunable parameter that is selected iteratively and may influence the results. Nevertheless, following this procedure with the selected minimum average distance for cluster correlation produces a global clustering of 26 unique clusters across the perigee maps constructed at the specified values of the initial true anomaly in the ER3BP. Figure 3 displays a subset of the generated maps, constructed with distinct values of the initial true anomaly for the primaries. In these figures, clusters are labeled and uniquely colored in shades of red and blue; clusters that are correlated across multiple perigee maps are indicated with the same color and cluster ID. The gray arrow in this figure indicates the increasing value of  $f_0$  used to generate the trajectories associated with each map crossing. Recall that the equilibrium points and ZVCs are calculated in the CR3BP, but displayed in these figures only for reference. To support analysis of these results, the representatives of clusters 4 to 7 and 17 to 20 are plotted in Figure 4, colored in shades of blue based on the originating initial system anomaly: darker shades are associated with lower values of  $f_0$ . Each trajectory starts from the associated green marker. In Figure 4, the Earth is represented as a grey circle, while  $L_1$  and  $L_2$  are indicated by red diamonds.

The persistence of trajectories that are generated in the ER3BP and admit a similar geometry across various values of the initial true anomaly is studied using the specific, fixed set of initial conditions used in this paper. The first two maps, corresponding to  $f_0 = 0$  and  $f_0 = \pi/4$ , admit clusters of similar solutions that do not shift or distort significantly as the true anomaly is set equal to these two values and for this specific set of initial conditions. Clusters 0 to 9 exist within a narrow region of the configuration space, while clusters 10 and 11 encompass a significant region of the configuration space. Note that an alternative feature vector definition or input parameter selection may produce more or fewer clusters. Two small white lobes appear close to the Earth in these first two perigee maps. These lobes represent trajectories that naturally depart the Earth vicinity before completing at least one apse: for this reason, these trajectories are excluded from the dataset.

As the initial true anomaly is increased to  $f_0 = \pi/2$ , new clusters of solutions emerge. Analysis of the third map at  $f_0 = \pi/2$  reveals that new clusters appear in the regions of the configuration space previously encompassed by clusters 10 and 11 at lower values of the initial true anomaly. Some of these clusters, including clusters 17, 19 and 20 depart the Earth vicinity through either the  $L_1$  and  $L_2$  gateways before completing 7 apses with respect to the secondary. As displayed in Figure 4, trajectories in these clusters are similar to those governed by the stable manifolds associated with the  $L_1$  and  $L_2$  Lyapunov orbits in the Sun-Earth CR3BP. Other clusters, such as 4 to 7, maintain their overall shape and size between these first three perigee maps.



Figure 3. Clustered perigee maps for various values of the initial true anomaly in the Sun-Earth ER3BP. Clusters of similar solutions are correlated across maps.



Figure 4. Selected cluster representatives from the maps in Figure 3; darker trajectories are associated with maps constructed at lower values of the initial true anomaly.

When the initial true anomaly increases beyond  $f_0 = \pi/2$ , the solution space associated with the perigee maps in Figure 3 admits more significant changes. For instance, clusters of trajectories that depart the Earth vicinity after 5 apses, such as clusters 17 and 20, encompass a larger region of the configuration space. A new group, cluster 18, also emerges in a nearly anti-symmetric configuration to cluster 19: as depicted in Figure 4, trajectories in cluster 19 resemble the stable manifold associated with an  $L_1$  Lyapunov orbit in the Sun-Earth CR3BP, while trajectories in cluster 18 are similar to those governed by an  $L_2$  Lyapunov orbit in the Sun-Earth CR3BP. Interestingly, cluster 18 emerges at a higher value of the true anomaly than cluster 19. Clusters 17 to 20 appear to encompass a wider region of the configuration space, clusters 10 and 11 shrink while clusters 2 to 7 remain relatively unchanged.

The last two maps, at the bottom of Figure 3, possess a significantly different cluster configuration to those constructed at low values of the initial true anomaly. First, clusters 17 to 20, encompass a larger region of the configuration space in the  $f_0 = 7\pi/12$  map. When the initial true anomaly is increased further to  $f_0 = 2\pi/3$ , clusters 18 and 19 each split clusters 23 and 24, respectively, into two regions on the map. As a result, new clusters emerge, namely clusters 22 and 25. Clusters 2 to 7, however, remain relatively unchanged through these increases in the initial true anomaly. In addition, the central white lobes that correspond to trajectories departing the Earth vicinity before completing one additional apsis encompass a larger region of the configuration space.

From a global perspective, this data-driven approach to assessing the persistence of trajectories across various values of the initial true anomaly supplies valuable insights into the solution space. Two distinct behaviors are evident. First, the size and shape of the clusters associated with escaping trajectories are significantly impacted by changes in the initial true anomaly, for the single, fixed initial condition set examined in this paper. Conversely, the regions of the configuration space encompassed by clusters 2 to 7 remain relatively unaffected by these changes in the initial true anomaly is a work in progress and would likely provide a more thorough understanding of the persistence of trajectories with a specific geometry in the ER3BP as this quantity is changed.

### 6.2 Cluster Persistence Across Epochs in the Point Mass Ephemeris Model

The map crossings associated with several perigee maps in a point mass ephemeris model of the Sun-Earth system, each constructed for various values of the initial epoch, are clustered individually. Recall that only the gravitational influence of the Sun and the Earth are included in this initial analysis. Consider 8 initial epochs  $t_0 = [29020, 29023, 29025, 29028, 29030, 29033, 29035, 2940]$ MJD, spanning 20 days, from Jun 19, 2020 at 12:00pm to Jul 9, 2020 at 12:00pm UTC. This particular time frame and discretization is selected consistent with the sensitivity of the solution space in the point mass ephemeris model to changes in the initial epoch. Using the same initial condition set employed in the CR3BP and ER3BP, each prograde perigee is propagated forward in time in the point mass ephemeris model of the Sun-Earth system. Since this dynamical model includes the ephemerides of the Sun and Earth in their true orbits, a planar initial condition results in trajectories that eventually extend slightly out of the plane of the primaries. The out-of-plane components  $[z, \dot{z}]$  do not appear in the feature vector in eq. (5) and are not considered in the presented implementation. However, this is a specific setup due to the presented planar map: indeed, including the out-of-plane components would extensively separate solutions reporting a modest out-of-plane displacement, leading to a different cluster differentiation of the maps and a subsequent erroneous final UMAP output.



Figure 5. Clustered perigee maps for various values of initial epoch in the Sun-Earth point mass ephemeris model. Clusters of similar solutions are correlated across maps.

After each map is individually clustered and noise points reassigned where appropriate, UMAP is employed, with  $t_{avg} = 1$ , to correlate clusters of similar solutions. Figure 5 displays the resulting global cluster assignments across 6 of these perigee maps, labeled with their associated initial epoch, in the point mass ephemeris model of the Sun-Earth system; the gray arrow indicates an increasing initial epoch. Each cluster is uniquely labeled and colored in shades of red and blue; the equilibrium points are displayed as red diamonds and the ZVCs bound the gray forbidden regions. Similar to the previous example, clusters of trajectories with a similar geometry and the same initial conditions exist across multiple initial epochs and encompass different regions of the configuration space. Figure 6 displays the representative solutions for several clusters at each initial epoch over which the cluster exists, colored in shades of blue based on the initial date; darker shades are associated with low values of  $t_0$ . Each trajectory begins at the green marker. In this figure, the Earth is represented as a gray circle, while  $L_1$  and  $L_2$  are indicated by red diamonds.



Figure 6. Selected cluster representatives from the maps in Figure 5; darker trajectories are associated with maps constructed at lower values of the initial epoch.

Several clusters exist across the entire range of initial epochs, but appear to shift and change in their shape. For instance, clusters 13 and 14 encompass a significant range of map crossings near the center of each perigee map; at the boundaries of these clusters, clusters 7 to 10 also tend to persist throughout the range of epochs. Clusters 1 and 2 occupy a significant region of the perigee map at lower initial epochs, and gradually shrink as the epoch increases; these clusters are composed of trajectories that naturally depart the Earth vicinity after completing 5 apses.

Some clusters only appear across a limit range of initial epochs, under the assumptions used to create the initial condition set for each perigee map. For instance, consider cluster 21, which corresponds to trajectories that naturally depart through the  $L_1$  gateway after one revolution, similar to the trajectories that lie inside the stable manifold of  $L_1$  orbits in the CR3BP. This particular cluster emerges at an epoch of  $t_0 = 29028$  MJD. However, trajectories exhibiting this kind of motion are present in the  $t_0 = 29025$  MJD perigee map, although the number of trajectories is lower than the threshold of  $m_{clSize} = 200$ ; thus these trajectories are not associated by HDBSCAN to a cluster. Another example is cluster 22 that exists at epochs including and after  $t_0 = 29033$  MJD. This cluster represents trajectories that naturally depart through the  $L_2$  gateway. These two types of trajectories, captured within clusters 21 and 22, encompass a wider region of the configuration space across the initial condition set as the initial epoch is increased; thus, a wider variety of trajectories lead to departure through the  $L_2$  gateway.

This set of perigee maps demonstrates the existence and persistence of clusters of trajectories with a similar geometry as the initial epoch in the point mass ephemeris model evolves. The size, shape and existence of a few clusters across the initial epoch range remains relatively unchanged, particularly for clusters 7 to 10 which correspond to bounded prograde motion. However, the existence and persistence of trajectories that leave the Earth vicinity, such as those contained within clusters 1, 2, 21 and 22, depends strongly on the initial epoch. Some trajectories pose challenges for the clustering framework under the selected initial condition discretization, input parameters, and feature vector definition; this is especially true for trajectories associated with the map crossings near Sun-Earth  $L_1$  and  $L_2$  points, at the boundaries of central oval-shaped map and in proximity of the Earth. In fact, clusters 0, 5, 6, 9, 18, 17 and 23 tend to be constituted by underrepresented trajectories, rapidly changing in time throughout the different map representations. Their existence, as well as their time evolution, is an effect of the employed discretization for the initial map grid and the subsequent UMAP correlation. A finer grid would better partition and correlate this structure, at the expense of a significant adjunct computational effort; such analysis is an ongoing effort.

### 6.3 Cluster Persistence Across Distinct Dynamical Models

In this final example, clusters of trajectories, generated in distinct dynamical models from the same initial condition set, are correlated to study their persistence as the underlying dynamics evolve. This example analysis uses three datasets: the perigee map generated and clustered in the CR3BP, a map generated in the ER3BP and a map generated in the Sun-Earth point mass ephemeris model, all using the same set of initial conditions. Following implementation of the cluster correlation process with  $t_{avq} = 1$ , the resulting maps are displayed in Figure 7, labeled by the dynamical model and the value of the input parameter. Note that the bottom left map, is generated in the ER3BP using  $f_0 = 7\pi/12$ , while the ephemeris map at the bottom right is constructed with  $t_0 = 29035$ MJD. Simultaneous analysis of these three maps reveals useful insights. For example, clusters associated with escaping motion are successfully correlated: clusters 0 and 1 group trajectories naturally departing the Earth vicinity before 3 apses. Similarly, clusters 4 and 5 group trajectories naturally departing after 5 apses. Other clusters of solutions, which exist across different true anomalies and initial epochs, for the ER3BP and ephemeris models, respectively, are again linked and numbered, e.g., clusters 2 and 3, 6 and 8. Interestingly, these same solutions also exist across each of the three dynamical models. Other solutions, such as those in clusters number 10 and 11 also appear to exist across different dynamical models.

At negative values of y, some solution geometries persist between the CR3BP and ER3BP, but not in the ephemeris model. For instance, consider clusters 12 and 13 which exist in the CR3BP and ER3BP. In the ephemeris map, cluster 17 occupies the part of the configuration space associated with clusters 12 and 13 in the CR3BP and ER3BP. To explore this observation further, Figure 8 displays the UMAP projections of the reduced dataset [P], colored according to the final clustering



Figure 7. Clustered perigee maps for the same initial conditions in the: (top) CR3BP, (bottom-left) ER3BP at  $f_0 = 7\pi/12$ , (bottom-right) Sun-Earth point mass ephemeris model at  $t_0 = 29035$  MJD. Clusters of similar solutions are correlated across maps.



Figure 8. UMAP projection of the global reduced dataset in Figure 7.

and consistent with Figure 7. The points associated with each of these three clusters are labelled with the associated cluster ID and appears in the zoomed-in view. While many solutions appear condensed and well-grouped in a specific region of the projected space, cluster 17, in the lower part of the figure, is split in two halves. Two zoomed-in regions in the neighborhood of clusters 12 and 13 are displayed in the right part of the figure, highlighting the splitting of cluster 17. The adopted color scheme associates the light-blue color to cluster 12, the darker red shade to cluster 13 and the lighter red shade to cluster 17. Clearly, cluster 17 can be linked to both cluster 12 and 13, i.e. its representative members are located in a neighborhood of these two clusters. However, the current cluster correlation approach does not make this decision, likely due to the selected value of  $t_{avg}$ . A larger value may potentially produce a different result; investigation into this result is ongoing. Additional ongoing work is focusing on exploring whether similar results appear as the feature vector definition and aggregation process is modified.

## 7 CONCLUDING REMARKS

In this paper, we use clustering to group trajectories based on geometrical similarity and study their evolution throughout progressively higher-fidelity dynamical models. In particular, we focus on trajectories associated with the crossings of a planar periapsis map near the Earth and generated in three distinct dynamical models of the Sun-Earth system: 1) the circular restricted three-body problem (CR3BP); 2) the elliptic restricted three-body problem (ER3BP), governed by the initial true anomaly of the primary system as an independent variable and 3) the point mass ephemeris model governed by the initial epoch. First, the initial conditions leveraged to construct the CR3BP map

are used across all dynamical models and several values of the independent variables in the ER3BP and point mass ephemeris model. These maps are first clustered individually using HDBSCAN and projected onto a three-dimensional reduced space with UMAP, in a completely unsupervised manner. The UMAP projection is used to correct the clustering result from HDBSCAN and decrease the amount of generated noise for each map. Then, the clusters in each map are reduced to a smaller set of representative members. The resulting global summary of all clusters across all maps is projected onto a three-dimensional space calculated by UMAP. This process enables clusters of similar trajectories to be correlated between different maps in a computationally efficient manner. Three examples are used to demonstrate this approach: 1) periapsis maps at different initial true anomalies in the ER3BP; 2) periapsis maps at different initial epochs in the Sun-Earth point mass ephemeris model; and 3) periapsis maps generated in each of the CR3BP, ER3BP and point mass ephemeris models. The first two scenarios enable an analysis of the influence of the independent variable on the proprieties of the solution space near the Earth in each dynamical model. The third example supports analysis of the persistence of trajectories with a specific geometry as the fidelity of the dynamical model is increased.

The procedure and analysis presented in this paper supply two fundamental contributions. First, the developed data-driven approach successfully correlates trajectories in progressively higher-fidelity dynamical models to support analysis of their persistence and existence. The resulting insight is useful for astrodynamicists selecting low-fidelity dynamical models for rapid initial guess construction in multi-body systems. Furthermore, this insight supplies the trajectory designer with a mechanism for understanding how the design space evolves as the model fidelity is increased. Second, UMAP and HDBSCAN, dimension reduction and clustering techniques in the field of unsupervised learning, are successfully used to improve the analysis and visualization of higher-dimensional Poincaré maps in a computationally efficient manner. In fact, UMAP is used in this work as an effective post-processing mechanism, to either improve the clustering output or correlate data across different datasets. The combined use of HDBSCAN and UMAP also enables a reduction of large datasets into smaller set of representative solutions for simplified analysis.

### 8 ACKNOWLEDGMENT

This work was completed at the University of Colorado Boulder under NASA Grant 80NSSC18K1536.

### REFERENCES

- W. S. Koon, M. W. Lo, J. E. Marsden, and S. D. Ross, "Resonance and Capture of Jupiter Comets," *Celestial Mechanics and Dynamical Astronomy*, Vol. 81, No. 27, September 2001, pp. 27 – 38, 10.1023/A:1013398801813.
- [2] N. Bosanac, A. Cox, K. Howell, and D. Folta, "Trajectory Design for a Cislunar CubeSat Leveraging Dynamical Systems Techniques: The Lunar IceCube Mission," *Acta Astronautica*, Vol. 144, March 2018, pp. 283–296, 10.1016/j.actaastro.2017.12.025.
- [3] K. Oshima, S. Campagnola, C. H. Yam, Y. Kayama, Y. Kawakatsu, N. Ozaki, Q. Verspieren, K. Kakihara, K. Oguri, and R. Funase, "EQUULEUS Mission Analysis: Design of the Transfer Phase," 31st International Symposium on Space Technology and Science, Matsuyama, Japan, June 2017.
- [4] A. Haapala, *Trajectory Design in the Spatial Circular Restricted Three-Body Problem Exploiting Higher-Dimensional Poincare Maps.* PhD thesis, Purdue University, West Lafayette, IN, 2014.
- [5] G. Gómez, W. Koon, M. Lo, J. Marsden, J. Masdemont, and S. Ross, "Connecting Orbits and Invariant Manifolds in the Spatial Restricted Three-Body Problem," *Nonlinearity*, Vol. 17, 2004, pp. 1571–1606.
- [6] N. Bosanac, "Data Mining Approach to Using Poincaré Maps in Multi-Body Trajectory Design Strategies," *Journal of Guidance, Control and Dynamics*, Vol. 43, No. 6, 2020, pp. 1190–1200, 10.2514/1.G004857.

- [7] S. Bonasera and N. Bosanac, "Applications of Clustering to Higher-Dimensional Poincaré Maps in Multi-Body Systems," *30th AIAA/AAS Space Flight Mechanics Meeting*, Orlando, FL, January 2020.
- [8] R. Campello, D. Moulavi, and J. Sander, *Density-Based Clustering Based on Hierarchical Density Estimates*. Berlin, Heidelberg: Springer, 2013, 10.1145/2733381.
- [9] McInnes, L., Healy, J., Melville, K., "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction," arXiv:1802.03426, (9 February 2018).
- [10] Cao J.; Spielmann, M.; Qiu, X.; Huang, X.; Ibrahim, D.M.; Hill, A.J.; Zhang, F.; Mundlos, S.; Christiansen, L.; Steemers, F.J.; Trapnell, C.; Shendure, J., "The Single-Cell Transcriptional Landscape of Mammalian Organogenesis," *Nature*, Vol. 566, 2019, pp. 496 – 502.
- [11] Diaz-Papkovicha, A; Anderson-Trocmé, L.; Ben-Eghan, C.; Gravel, S., "UMAP Reveals Cryptic Population Structure and Phenotype Heterogeneity in Large Genomic Cohorts," *PLos Genetics*, Vol. 15, No. 11, 2019, 10.1371/journal.pgen.1008432.
- [12] Bloch, T.; Watt, C.; Owens, M.; McInnes, L; Macneil, A.R., "Data-Driven Classification of Coronal Hole and Streamer Belt Solar Wind," *Solar Physics*, Vol. 295, No. 41, 2020.
- [13] V. Szebehely, Theory of Orbits: The Restricted Problem of Three Bodies. London, UK: Academic Press, 1967.
- [14] S. Campagnola, M. Lo, and P. Newton, "Subregions of Motion and Elliptic Halo Orbits in the Elliptic Restricted Three-Body Problem," *Advances in the Astronautical Sciences*, Vol. 130, 2007, pp. 1541– 1556.
- [15] "SPICE Toolkit," https://naif.jpl.nasa.gov/naif/aboutspice.html, 2016. Accessed July 2020.
- [16] T. Pavlak, Trajectory Design and Orbit Maintenance Strategies in Multi-body Dynamical Regimes. PhD thesis, School of Aeronautics Astronautics, Purdue University, West Lafayette, IN, May 2013.
- [17] B. Villac and D. Scheeres, "On the Concept of Periapsis in Hill's Problem," *Celestial Mechanics and Dynamical Astronomy*, Vol. 90, 2004, pp. 165–178, 10.1007/s10569-004-0405-9.
- [18] L. McInnes, J. Healy, and S. Astels, "hdbscan: Hierarchical Density Based Clustering," *The Journal of Open Source Software*, Vol. 2, 2017.
- [19] P. Cichosz, *Data Mining Algorithms: Explained Using R*. West Sussex, United Kingdom: John Wiley and Sons, 2015.