Statistical equivalence of metrics for meteor dynamical association

Eloy Peña-Asensio [a,b,*], Juan Miguel Sánchez-Lozano [c]

[a] *Department of Aerospace Science and Technology, Politecnico di Milano, Via La Masa 34, Milano 20156, Lombardia, Italy*
[b] *Departament de Química, Universitat Autònoma de Barcelona, Carrer dels Til·lers, Bellaterra 08193, Catalonia, Spain*
[c] *Centro Universitario de la Defensa, Universidad Politécnica de Cartagena, C/ Coronel López Peña S/N, Base Aérea de San Javier, Santiago de la Ribera 30720, Spain*

Abstract

Meteor showers, originating as a result of the activity of comets or the disruption of large objects, provide a unique window into the composition and dynamics of our Solar System. While modern meteor detection networks have amassed extensive data, distinguishing sporadic meteors from those belonging to specific meteor showers remains challenging. In this study, we statistically evaluate and compare four orbital similarity criteria within five-dimensional parameter space (D_{SH}, D_D, D_H, and ϱ_2) to study dynamical associations using the already classified meteors (manually by a human) in CAMS database as a benchmark. In addition, we assess various distance metrics typically used in Machine Learning with two different vectors: ORBIT, grounded in heliocentric orbital elements, and GEO, predicated on geocentric observational parameters. To estimate their degree of correlation and efficacy, the Kendall rank correlation coefficient and the Top-k accuracy are employed. The statistical equivalence of the Top-1 results is examined using the Kolmogorov–Smirnov test and the percentage of Top-1 agreement is calculated on an event-by-event basis. Additionally, we compute the optimal cut-offs for all methods for distinguishing sporadic background events. Our findings demonstrate the superior performance of the sEuclidean metric in conjunction with the GEO vector. Within the scope of D-criteria, D_{SH} emerged as the preeminent metric, closely followed by ϱ_2. The Bray-Curtis metric displayed an advantage compared to the other distance metrics when paired with the ORBIT vector for Top-k accuracy, however, the Cityblock metric is more effective when considering the sporadic background. ϱ_2 stands out as the most equivalence to the distance metrics when utilizing the GEO vector and the most compatible with GEO and ORBIT simultaneously, whereas D_D aligns more closely when using the ORBIT vector. The stark contrast in D_D's behavior compared to other D-criteria highlights potential inequivalence. Our results suggest that geocentric features provide a more robust basis than orbital elements for meteor dynamical association. Most distance metrics associated with the GEO vector surpass the D-criteria when differentiating the meteoroid background. Accuracy displayed a dependence on solar longitude with a pronounced decrease around 180° matching an apparent increase in the meteoroid background activity, tentatively associated with the transition from the Perseids to the Orionids. Considering lately identified meteor showers, ∼27% of meteors in CAMS would have different associations. This work unveils that Machine Learning distance metrics can rival or even exceed the performance of tailored orbital similarity criteria for meteor dynamical association.
© 2024 COSPAR. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Keywords: Meteor; Meteoroids; Meteor showers; Meteoroid streams; Statistics

1. Introduction

Within the expanse of our planetary system, remnants from its formation provide glimpses into the early stages of our cosmic neighborhood (Bottke et al., 2002; Walker and Cameron, 2006). Among these remnants, comets

* Corresponding author at: Department of Aerospace Science and Technology, Politecnico di Milano, Via La Masa 34, Milano 20156, Lombardia, Italy.
E-mail addresses: eloy.pena@polimi.it, eloy.pena@uab.cat (E. Peña-Asensio).

emerge as witnesses to the dramatic events that shaped our nearby environment. These celestial bodies can undergo processes of disruption due to various factors such as volatile sublimation when approaching the Sun, tidal forces, or impacts with other bodies. According to the theory of formation and evolution of small bodies of the Solar System (Whipple, 1951; Bredikhin, 1954; Plavec, 1954; Hughes, 1986; Babadzhanov and Obrubov, 1992), meteoroid streams are formed mainly as a result of the activity of comets or the ejection of meteoroids from cometary nuclei with various initial velocities (Chapman, 2010; Tóth et al., 2011; Gritsevich et al., 2012). Meteoroids exhibit a diverse composition, including rock, metal, or a combination of both, and span a wide range of sizes, from micrometerscale grains to larger objects up to one meter in diameter (Trigo-Rodríguez and Llorca, 2006; Trigo-Rodríguez and Llorca, 2007; Koschny and Borovicka, 2017). Despite their heterogeneous characteristics, these meteoroids share a common origin, derived from a parent body, which imparts certain similarities among them.

Additionally, though less common, asteroids can also generate meteoroid streams as a result of catastrophic impact events. Some associations have been found, such as the case of the potentially hazardous asteroid (3200) Phaethon (1983 TB), whose origin could be the nucleus of an extinct comet (Zhong-Yi et al., 2020), and the Geminids meteor shower. Multiple studies have confirmed the high probability that the Geminids are dynamically associated with such asteroid (Whipple, 1983; Gustafson, 1989; Williams and Wu, 1993). However, as they traverse the space, the influence of planetary perturbations and nongravitational forces gradually renders them indistinguishable from the background population (Olsson-Steel, 1986; Bottke et al., 2000; Pauls and Gladman, 2005; Brož, 2006; Koschny et al., 2019).

Eventually, the journey of meteoroids brings them into intersecting paths with the Earth's orbit, leading to captivating interactions with our planet (Brown et al., 2002; Murad and Williams, 2002; Gritsevich, 2009; Trigo-Rodríguez, 2022). As these meteoroids penetrate the Earth's atmosphere, they experience a dramatic transformation fueled by the intense heat generated through air molecule friction. The high-speed entry produces enormous amounts of heat, causing the outer layers of the meteoroids to rapidly vaporize (Popova et al., 2019). This process, known as ablation, leads to the formation of a glowing plasma sheath surrounding the meteoroid (Ceplecha et al., 1998; Silber et al., 2018). The energy released during atmospheric aerobraking causes the visible phenomenon known as a meteor, which is called a fireball or bolide if its magnitude surpasses that of the planet Venus. When a meteoroid stream intersects the Earth's path periodically, it gives rise to the phenomenon of meteor showers (Jenniskens, 1994; Jenniskens, 1998; Jenniskens, 2006; Vaubaillon et al., 2019; Jenniskens, 2023). The meteors within them share common features, including their time of occurrence, apparent origin in the sky, known as the radiant, and their geocentric impact velocity, as well as their orbital elements in an equivalent manner.

Thanks to the existence of meteor detection networks (Jacchia and Whipple, 1956; Ceplecha, 1957; Bland, 2004; Trigo-Rodriguez et al., 2005; Weryk et al., 2007; Jenniskens et al., 2011; Kornoš et al., 2014; Gritsevich et al., 2014; Colas et al., 2015; SonotaCo, 2016; Gardiol et al., 2016; Devillepoix et al., 2020; Colas et al., 2020; Boaca et al., 2022; Borovička et al., 2022), researchers now have access to an abundance of videos and images that play a key role in improving our understanding of the distribution and characteristics of meteoroid streams. The availability of this extensive dataset has revolutionized our capacity to investigate and analyze incoming extraterrestrial material. Through processing these videos and extracting relevant information, we can determine essential parameters such as trajectory, velocity, brightness, and other physical properties of meteors (Ceplecha, 1987; Borovicka, 1990; Vida et al., 2020; Peña-Asensio et al., 2021; Peña-Asensio et al., 2024). This valuable information significantly contributes to our knowledge about meteor showers, assisting in the identification and tracking of potentially hazardous complexes that may pose a threat to Earth in the near term (Voloshchuk and Kashcheev, 1996; Halliday, 1987; Borovička et al., 2015; Trigo-Rodríguez et al., 2017; Trigo-Rodríguez and Blum, 2022).

In this regard, accurate classification of sporadic meteors from those belonging to the same stream is of utmost importance, as the latter provides valuable insights into the population density of future impactors (Wiegert and Brown, 2004; Porubčan et al., 2004; Jopek and Williams, 2013; Dumitru et al., 2017; Jenniskens, 2017; Vaubaillon et al., 2019; Peña-Asensio et al., 2022; Peña-Asensio et al., 2023). Determining the point at which a meteor shower transitions from a cohesive entity to a collection of unrelated meteoroids (sporadic background), or establishing the criteria to accurately associate meteors with a specific shower, poses a significant challenge. To tackle the issue of orbital dynamical association, multiple endeavors have been undertaken to define similarity criteria or D-criteria. These criteria aim to effectively differentiate between events that are associated with a specific meteoroid stream and those that are unrelated to other objects or swarms. Ultimately, analyzing the impact features can aid in associating meteorites with their parent bodies (Carbognani and Fenucci, 2023).

In this study, we assess the rank correlation, efficacy, and equivalence of four five-dimensional similarity criteria designed for quantifying dynamical associations between meteor orbits, as well as various distance metrics with two different vectors (one shared with the D-criteria). The evaluation is conducted using a comprehensive meteor database and extends to exploring alternative metrics for orbit association, as well as computing the optimal thresholds for each method. The objective is to elucidate the statistical strengths, limitations, and similarities of each approach, thereby providing a robust framework for future

research in meteor associations with parent bodies or meteoroid streams.

In Section 2, we detail the database utilized and the methodology applied. Section 3 presents our findings, and Section 4 provides a summary of the key outcomes of our study.

2. Data and procedures

The methodology presented herein is designed to analyze multiple meteor dynamical association approaches by comparing five-dimensional orbital similarity criteria and various vector-based distance metrics typically used in Machine Learning. For the latter, we use as a vector (1) the same parameters utilized by the similarity criteria defined by some heliocentric orbital elements, which we termed as ORBIT, and (2) the four-dimensional vector proposed by Sugar et al. (2017) and named here as GEO. It should be noted that while the term "metrics" may be appropriate to describe the D-criteria to a certain extent, in this work, we use the term "metrics" exclusively to refer to vector-based distance metrics, which are further explained.

This section is subdivided into different subsections. SubSection 2.1 elaborates on the data sources utilized. SubSection 2.2 presents D-criteria for comparing the orbital elements of two orbits. In SubSection 2.3, we introduce the two vectors that will be used along with the distance metrics. In SubSection 2.4 we explain the theoretical background used for calculating the rank correlations, comparing the performances with the Top-k accuracy method, and estimating the equivalence with the Kolmogorov–Smirnov test and Top-1 event-by-event agreement. Finally, in SubSection 2.5, we detail our strategy to determine the optimal thresholds for distinguishing between sporadic background and meteor showers. All implementations of the statistical analyses were conducted utilizing the *SciPy* library (Virtanen et al., 2020).

2.1. Databases

CAMS, short for the Cameras for All-Sky Meteor Surveillance project (Jenniskens et al., 2011), is an international initiative sponsored by NASA and managed by the Carl Sagan Center within the SETI Institute, located in California, USA. Its primary objective is to monitor and map meteor activity through nighttime optical video surveillance, employing triangulation techniques. It annually records an average of half a million meteor orbits, although the publication of this data stopped in 2016. The last release was the Meteoroid Orbit Database v3.0, which includes 471,582 events registered since 2010.

While there are other automated meteor detection networks, CAMS stands out as the primary and most widely recognized repository of meteor data. Nevertheless, it was noted that its performance in accurately detecting fast

meteors falls short in comparison to its detection of slower meteors (Koseki, 2017; Koseki, 2022). To address this issue, we implement a filtering mechanism to exclude lower-quality detections and to reduce spurious data, requiring a minimum convergence angle of 15 degrees between cameras, ensuring an estimated velocity error of no more than 10% of the nominal value, not allowing hyperbolic orbits, and selecting perihelion distance compatible with impacts on the Earth.

Certainly, we rely on the classification provided by CAMS as a ground truth, which may not be infallible. However, the classification within this database did not utilize any formal dissimilarity criteria. Instead, it depended on human visual clustering within sun-centered ecliptic longitude-latitude representations, with clusters manually delineated using specific coordinates and geocentric velocity limits (Jenniskens et al., 2018). Our analyses proceeds under the presumption that the CAMS classification is accurate, a premise that, regardless, serves our primary objective of assessing the equivalence between metrics and D-criteria.

For identifying meteoroid streams responsible for meteor showers, we use the V.2 list of all known showers from the IAU Meteor Data Center, updated in January 2024 (Jopek and Jenniskens, 2011; Jopek and Kanuchová, 2013; Jopek and Kaňuchová, 2017; Jenniskens et al., 2020). To facilitate the association of these meteor showers with entries in the CAMS database, we employ the IAU numeral code. This list includes 1484 entries, 956 corresponding to unique meteor showers. To ensure a direct comparison of association performances, we filter both CAMS and IAU meteor shower datasets to include only identical, unique meteor showers.

2.2. Orbital similarity criteria

Orbital elements such as inclination i, eccentricity e, longitude of the ascending node Ω, perihelion distance q, and argument of the perihelion ω allow us to determine the path of any moving object following a Keplerian trajectory in our Solar System. Likewise, it is possible to look for the connection between a meteor shower and its parent body (or any two objects) through the similarities of their orbits.

This search approach is not recent. The first attempts focused on measuring the degree of similarity between orbits were designed in the second half of the last century, they were so-called D-criteria. The first D-criteria was introduced by Southworth and Hawkins (1963):

$$D_{SH}^2 = (e_B - e_A)^2 + (q_B - q_A)^2 + \left(2\sin\frac{I_{AB}}{2}\right)^2$$
$$+ \left(\frac{e_B + e_A}{2}\right)^2\left(2\sin\frac{\pi_{BA}}{2}\right)^2, \tag{1}$$

where other concepts of geometry come into play such as the angles between their respective perihelion points (π_{BA}) and between the inclinations of the orbits (I_{AB}).

Drummond (1981) not only defined the angle between the perihelion points on each orbit (θ_{BA}) by adding both the ecliptic longitude (λ) and the perihelion latitude (β), but also weighted the terms e and q to provide a metric in which each term contributed equally to the overall sum. In this way, a new variant of the D_{SH} criterion, named D_D in honor of its creator, was developed:

$$D_D^2 = \left(\frac{e_B - e_A}{e_B + e_A}\right)^2 + \left(\frac{q_B - q_A}{q_B + q_A}\right)^2 + \left(\frac{I_{AB}}{\pi}\right)^2$$
$$+ \left(\frac{e_B + e_A}{2}\right)^2 \left(\frac{\theta_{BA}}{\pi}\right)^2, \tag{2}$$

A decade later, Jopek (1993) carried out a random perturbation model of several orbits, ignoring i, Ω, and ω, to analyze the D_{SH} and D_D criteria. He found dependency relationships of q and e values for the reference orbit; q in the case of D_{SH} and e for the criterion D_D. To reduce these dependency relationships between orbital parameters, Jopek proposed a new similarity criterion, D_H, defined by:

$$D_H^2 = (e_B - e_A)^2 + \left(\frac{q_B - q_A}{q_B + q_A}\right)^2 + \left(2\sin\frac{I_{AB}}{2}\right)^2$$
$$+ \left(\frac{e_B + e_A}{2}\right)^2 \left(2\sin\frac{\pi_{BA}}{2}\right)^2. \tag{3}$$

Note that these D-criteria cannot be categorized mathematically as metrics due to their violation of the triangle inequality (Kholshevnikov et al., 2016). Instead, they are more appropriately defined as quasimetrics, as they adhere to a relaxed form of the triangle inequality (Milanov et al., 2019). Contemporary functions, such as ϱ_2, enable the precise quantification of orbital similarity through consistent mathematical formulations:

$$\varrho_2^2 = (1 + e_1^2)p_1 + (1 + e_2^2)p_2 - 2$$
$$\times \sqrt{p_1 p_2}(\cos I + e_1 e_2 \cos P) \tag{4}$$

with

$$p = a(1 - e^2), \tag{5}$$
$$\cos I = \cos i_1 \cos i_2 + \sin i_1 \sin i_2 \cos(\Omega_1 - \Omega_2), \tag{6}$$
$$\cos P = \sin i_1 \sin i_2 \sin \omega_1 \sin \omega_2 + \cos \omega_1 \cos \omega_2 \cos(\Omega_1 - \Omega_2)$$
$$+ \cos i_1 \cos i_2 \sin \omega_1 \sin \omega_2 \cos(\Omega_1 - \Omega_2) \tag{7}$$
$$+ (\cos i_2 \cos \omega_1 \sin \omega_2 - \cos i_1 \sin \omega_1 \cos \omega_2) \sin(\Omega_1 - \Omega_2).$$

The limit values of such D-criteria, also called thresholds, cut-off levels, or upper limits, determine whether two objects may be associated. Being, for example, A and B a meteor and meteor shower respectively, if the distance $D(A, B)$ between A and B is greater than this limit value, the association must be discarded. The smaller this distance is, the greater the possibility that there is a dynamical similarity between two objects, and, therefore, the meteoroid belongs to the meteoroid stream.

Some studies on the suitability of these criteria have already been carried out. For example, Galligan (2001) explored the performance of four similarity functions in the near-ecliptic region–D_{SH}, D_D, D_H, and D_N (Valsecchi et al., 1999)–, resulting in D_N criterion being the most stable in the case of the lack of a priori information on orbital inclination regimes, while D_{SH}, which is based on meteor shower dispersion theoretical models, is more suitable with very different cut-off levels. However, D_N has not been adopted in our approach due to its less straightforward application from the standard parameters provided in meteor databases. Likewise, Moorhead (2015) analyzed such cut-off values to determine a chosen acceptable false-positive rate and distinguish which showers are significant within a set of sporadic meteors. Jenniskens (2008) and Rudawska et al. (2015) introduced the four-dimensional metrics D_B and D_X, respectively. However, to maintain consistency within the parameter space domain analyzed in this study, we opt not to include these criteria.

Through these values, it has been possible to associate meteor showers with parent bodies such as the 109P (1862) III Swift–Tuttle comet and the Perseid meteor shower, whose first connection data from the late 19th century when Schiaparelli calculated the orbits of the Perseids and discovered their strong similarity to that of this comet. Involved on this connection, (Sokolova et al., 2014) calculated the cut-off level of D_{SH} resulting in $D_{SH} \leqslant 0.2$. Literature provides more classical examples such as the April Lyrids, whose extremely small value of the D_D criterion ($D_D = 0.009$) suggests that such meteors showers have indeed come from comet Thatcher (Arter and Williams, 1997). Other recent examples are the case of a fireball detected in the night sky over Kyoto whose likely parent, with $D_{SH} = 0.0079$, could be the binary near-Earth asteroid (164121) 2003 YT1 (Kasuga et al., 2020); the binary asteroid 2000 UG11 associated with Andromedids ($D_{SH} = 0.183$ and $D_H = 0.176$) and the asteroid (4179) Toutatis, with values of $D_{SH} = 0.180$ and $D_H = 0.175$, that postulate it associated with October Capricornids (Dumitru et al., 2017); the meteor shower June epsilon Ophiuchids, whose values in three D-criteria ($D_{SH} = 0.05$, $D_D = 0.03$ and $D_H = 0.06$) confirm that is likely to originate from comet 300P/Catalina (Matlovic et al., 2020); or the recently observed fall and recovery of the Traspena meteorite is posited to be linked with the potentially hazardous asteroid 1989 QF (Minos), exhibiting $\varrho_2 = 0.1059$ (Andrade et al., 2023). We note the absence of a cut-off estimate works for ϱ_2, unlike the traditional D-criteria.

Although the cases mentioned above demonstrate the usefulness of the similarity criteria, some limitations confirm the need to investigate these metrics. For example, Galligan (2001) found that, for the case of the D_{SH} criterion, it is necessary to use different upper limits depending on the orbital inclination angle of the stream. In fact, Sokolova et al. (2014), intending to improve the reliability of identification of the observed objects, recommends analyzing the D_{SH} threshold values independently for each meteoroid complex. Following that approach, the study of comparison of four similarity criteria carried out by Rudawska et al. (2012) confirmed the difficulty in obtaining one specific value of threshold that would fit all cases,

reaching the conclusion that the ideal threshold depends on the cluster analysis method, the meteors shower, and the sample; this latter statement is also seconded by Jopek and Bronikowska (2017). Ye (2018) also pointed out that the traditional D-criteria may not necessarily reflect a shared origin of two objects due to the orbital evolution influenced by planetary perturbations.

In short, these studies are clear examples of the need to analyze the effectiveness and equivalence of the different approaches to establish dynamical associations of meteors.

2.3. Meteor vectors and distance metrics

In the preceding section, we discussed five-dimensional D-criteria for associating meteors with meteor showers. While these approaches are widely used, they are not without limitations. It is an active research topic for which there is no consensus on either criteria or thresholds. To search for alternatives and compare their performance, we introduce two meteor vectors –ORBIT and GEO– to evaluate multiple Machine Learning distance metrics in meteor-shower association.

The ORBIT vector focuses simply on the same five heliocentric orbital elements that are used by the above-mentioned orbital similarity criteria, which allows for a more direct comparison of the effectiveness:

$$ORBIT = \begin{bmatrix} q \\ e \\ i/\pi \\ \frac{\sin(\omega)+1}{4} \\ \frac{\cos(\omega)+1}{4} \\ \frac{\sin(\Omega)+1}{4} \\ \frac{\cos(\Omega)+1}{4} \end{bmatrix}. \tag{8}$$

Note that the database has been filtered to minimize spurious events, ensuring the inclusion of only non–hyperbolic orbits ($0 < e \leqslant 1$) that intersect Earth's orbit, specifically with $0 < q \leqslant 1$ au. The inclination, when normalized by 180°, spans the range [0, 1]. For the circular components, ω and Ω, which range from [-1, 1], we normalize them to [0, 1] and assign half the weight to each circular component. Utilizing sine and cosine functions for the circular angle ω and Ω, we effectively account for the shortest circular distance between angles, ensuring that 358° is recognized as 4° away from 2°, rather than 356°. Consequently, all five independent parameters are normalized and weighted equally, constructing a five-dimensional space vector.

The GEO vector is based mainly on geocentric observable parameters and was proposed by Sugar et al. (2017). This six-component vector (but in four-dimensional space as it has only four independent parameters) inherently addresses the issue of longitude wrapping. It normalizes the six components to ensure that each variable contributes equally. The vector's initial two components represent the meteor's position, as the meteoroid intersects the Earth's

orbit. The subsequent three components define the unit vector opposite to the meteor's velocity direction. The final component represents the magnitude of the geocentric velocity, normalized by the maximum velocity allowed for the study population:

$$GEO = \begin{bmatrix} \cos(\lambda_\odot) \\ \sin(\lambda_\odot) \\ \sin(\lambda_g - \lambda_\odot)\cos(\beta_g) \\ \cos(\lambda_g - \lambda_\odot)\cos(\beta_g) \\ \sin(\beta_g) \\ v_g/72 \end{bmatrix}. \tag{9}$$

In this vector, v_g represents the geocentric velocity in kilometers per second, $\lambda_\odot$ is the solar longitude, β_g is the geocentric ecliptic latitude of the radiant, and $\lambda_g - \lambda_\odot$ being the Sun-centered ecliptic longitude of the radiant. All components span the range [-1, 1], except for the element related to velocity, which varies between [0, 1]. Given that velocity measurements are subject to the greatest degree of error, the authors allowed a reduced weight for the velocity.

Although the D-criteria are theoretically five-dimensional, the orbits of the meteors are constrained by having impacted the Earth, virtually reducing the dimensionality by one. Consequently, this dimensionality reduction enables a comparison between the performances of the GEO and ORBIT vectors.

In the quest to develop a robust methodology for associating meteors with their parent meteor showers, we explore various distance metrics typically used in Machine Learning that can quantify the similarity between the previously defined vectors. In Table 1, we introduce the distance metrics that are employed in this study.

2.4. Statistical analysis

2.4.1. Rank correlations

We select the Kendall rank correlation coefficient (τ) to measure the ordinal association between the distance metrics and D-criteria. Mathematically, it is defined as:

$$\tau = \frac{2}{n(n-1)} \sum_{i<j} sgn(x_i - x_j) sgn(y_i - y_j), \tag{10}$$

where $(x_1, y_1), \ldots, (x_n, y_n)$ are a set of samples of the variables.

τ is notable for its ability to measure the strength and direction of the relationship between two variables without requiring them to be on the same scale. Unlike parametric correlations like Pearson's, which assume linear relationships and normal distribution of data, Kendall's approach is based on the ranking of data points, assessing concordance and discordance in their relative ordering across two datasets. It focuses on rank rather than absolute values obviates the need for identical scales between datasets. Consequently, we can employ it to compare the results of

Table 1
Summary of distance metrics.

Metric Name	Formula	Brief Explanation
Euclidean	$\sqrt{\sum_i (u_i - v_i)^2}$	Square root of sum of squared differences
sEuclidean	$\sqrt{\sum_i \frac{(u_i - v_i)^2}{V_i^2}}$	Normalized Euclidean by variance V_i
Cityblock (Manhattan)	$\sum_i |u_i - v_i|$	Sum of absolute differences
Cosine Similarity	$1 - \frac{u \cdot v}{\|u\| \cdot \|v\|}$	Angle between vectors
Canberra	$\sum_i \frac{|u_i - v_i|}{|u_i| + |v_i|}$	Weighted Cityblock distance
Bray-Curtis	$\frac{\sum_i |u_i - v_i|}{\sum_i |u_i + v_i|}$	Normalized weighted Cityblock
Chebyshev	$\max_i |u_i - v_i|$	Maximum absolute difference

the D-criteria and the distance metric without applying any normalization. We use the asymptotic method to compute Kendall's tau, which provides an efficient and scalable approximation suitable for large datasets and handles ties effectively.

The process is as follows. For each meteor in the dataset, we first compute its similarity/closeness to every meteor shower based on predefined D-criteria and distance metrics (both for GEO and ORBIT vectors). These calculations yield two separate sets of rankings for every meteor: one set derived from the D-criteria and another from the distance metrics. Each set sorts all meteor showers from the most to the least similar to the meteor in question. Once we obtain these rankings, the τ is computed for each meteor, comparing the two sets of rankings to ascertain the degree of ordinal classification. For more information into the Kendall rank correlation coefficient applied here refer to Kendall (1938); Fenwick (1994); Hollander et al. (2013).

2.4.2. Top-k accuracy

The heart of the present study centers on the evaluation of the classification accuracy of various D-criteria and distance metrics. To address this challenge, a unified methodology is imperative for the consistent application of statistical tests across all approaches under consideration. Despite the diversity in metrics and D-criteria, they converge on a singular objective: to quantify the association between a meteor and its corresponding meteor shower. As such, the Top-k accuracy is employed as a standardizing criterion to compare the overall accuracy among the various methods (Xia et al., 2009).

It quantifies the frequency with which the correct label class is included among the first k predicted labels. In the specific context, these labels denote the meteor showers associated with individual meteoroid impacts as classified by CAMS. For each meteor in the dataset, the similarities and distances are calculated in relation to all reference meteor showers. These values are subsequently sorted in ascending order to generate a ranked list. A successful classification in the Top-1 category occurs when the meteor shower with the minimum similarity or distance aligns with the meteor shower associated with the meteor in the CAM dataset. Similarly, a Top-5 success is recorded if the associated meteor shower is among the top five labels in the ranked list, and this extends analogously for other values of k.

In the present study, multiple tests encompassing Top-1, Top-5, and Top-10 accuracy are performed to evaluate the efficacy of D-criteria and distance metrics in associating a meteor with its originating meteor shower. This multi-tiered approach enables both a precise assessment of the top prediction (Top-1) and an evaluation of the model's capacity to identify a broader set of correct associations (Top-5 and Top-10). While one might assume that the Top-1 accuracy is paramount for meteor association, it is important to consider the significance of conducting Top-5 and Top-10 analyses. These extended evaluations yield insights into the efficacy of various ranking methodologies, going beyond mere concurrence with CAMS classifications. These analyses aid in contrasting the variability in rankings produced by different metrics. It is distinct when two metrics diverge at the Top-1 level yet converge within the Top-5, compared to a scenario where they diverge up to the Top-10.

2.4.3. Kolmogorov–Smirnov Test

The Kolmogorov–Smirnov test (K-S test) serves as a robust, non-parametric statistical method designed to assess the goodness-of-fit and equivalence of continuous, one-dimensional probability distributions. The test is particularly advantageous due to its distribution-free nature, making it applicable to datasets without the assumption of any specific distribution. The K-S test is employed in two primary contexts: the one-sample K-S test and the two-sample K-S test. The two-sample K-S aims to compare two empirical distributions and to determine if the two

samples come from the same distribution. The K-S statistic D is:

$$D = \sup_x |F_{1,n}(x) - F_{2,m}(x)| \qquad (11)$$

where $F_{1,n}(x)$ and $F_{2,m}(x)$ are the empirical distribution functions of the two samples of sizes n and m, respectively. Here we follow the treatment explained in Hodges (1958).

When applying the K-S test to Top-1 test results, interpreting the results sheds light on the comparative distributions of accuracy between classification methods. Failing to reject the null hypothesis H_0 indicates no statistically significant difference in accuracy distributions, but it does not affirm equivalence in method performance. Conversely, rejecting H_0 suggests a statistically significant difference, supporting the alternative hypothesis H_1 that the samples originate from distinct distributions. This outcome implies that H_0 does not adequately explain the observed data, with the decision to reject based on the significance level α, set here at 0.05 for 95% confidence.

2.4.4. Top-1 Agreement

Consider two classifiers tested on a dataset consisting of two equally sized classes. The first classifier might excel in identifying Class A but fail to recognize Class B, whereas the second classifier achieves the opposite, accurately identifying Class B while mistaking instances of Class A. Despite both classifiers reporting an overall accuracy of 50%, their distinct performance on the individual classes reveals a divergent understanding and representation of the underlying patterns in the data. This example underscores the necessity of applying another test, as (1) Kendall's correlation assesses whether the order of rankings is similar between two sets of observations and (2) K-S is specifically focused on the shape of accuracy distributions rather than precise values.

For this reason, we calculate as well the percentage of Top-1 coincidence between distance metrics and D-vectors on an event-by-event basis, which provides a direct measure of agreement on the most preferred classification outcome, capturing the extent to which different approaches concur on the single best classification. This straightforward metric offers an immediate sense of the hit-and-miss between approaches. A heatmap is an optimal visualization tool for showcasing the pairwise agreement between classification metrics, using a rectangular matrix to highlight the magnitude of their coincidences.

2.5. Differentiating the sporadic background

The last part of our work deals with the effective discrimination of the sporadic background from meteor events that are associated with specific showers. We calculate the Top-1 accuracy values across the entire (non-filtered) database and construct the Receiver Operating Characteristic (ROC) curves for each D-criteria and distance metric, utilizing both the GEO and ORBIT vectors, using binary labels from CAMS (0: sporadic; 1: associated). The ROC curve represents the diagnostic ability of a binary classifier system as its discrimination threshold is varied. Using the ROC curve output, it is possible to quantify the optimal threshold that maximizes the classifier's performance with Youden's J statistic (Youden, 1950; Schisterman et al., 2005):

$$J = \frac{TP}{TP + FN} + \frac{TN}{TN + FP} - 1, \qquad (12)$$

where TP represents the true positives, FN the false positives, TN the true negatives, and FP the false positives.

To synthesize the overall performance of each classification method in differentiating the sporadic background, we utilize the Matthews Correlation Coefficient, usually denoted by MCC or ϕ (Matthews, 1975). The ϕ offers a measure of the quality of binary classifications, encapsulating sensitivity, specificity, and the balance between them. It ranges from -1 (total disagreement between prediction and observation) to 1 (perfect prediction), with 0 denoting random guessing. The ϕ is defined as:

$$\phi = \frac{(TP \cdot TN) - (FP \cdot FN)}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}}. \qquad (13)$$

3. Results

Similar to Section 2, where we detailed the database and methodology in distinct subsections, the results section is also organized into subsections for clarity and depth. SubSection 3.1 examines the dataset, SubSection 3.2 presents the rank correlation estimations, SubSection 3.3 reports on the accuracy results, SubSection 3.4 explores the equivalence between distance metrics and D-criteria, SubSection 3.5 offers the level of coincidence between approaches for the Top-1 tests, and finally SubSection 3.6 provides optimal cut-offs and false positive rates.

3.1. Population analysis

Within the extensive CAMS database, 24.6% of its entries can be directly linked to a distinct meteor shower. In contrast, 75.4% of the data points are categorized as sporadic events, implying they are part of the broader meteoroid background rather than specific meteor showers. After applying the filters mentioned in Section 2, the database reduces its number to account for 102,680 orbits.

The number of unique meteor shower classifications is somewhat constrained, amounting to 376 distinct categories. A total of 80% of these classified meteor showers have been observed more than 10 times. A quarter of them, or 25%, boasts over 100 individual recorded meteor events. An even smaller fraction, 5%, can claim over 1000 meteor instances. Four of the meteor showers stand out due to their frequent documentation: the Perseids, Orionids,

Geminids, and Southern Taurids (enumerated in descending order based on their observation frequency). Meteors belonging to these showers have been observed more than 10,000 times.

Regarding the IAU meteor shower database, after filtering it reduces its number to 724, having 355 unique IDs shared with the CAMS database. Note that ~30% are duplicate entries, corresponding to distinct values for the same meteor shower estimated in different studies.

A key aspect of our analysis of associations is the parameter of solar longitude, that correlates meteor activity with Earth's specific orbital locations. Such a correlation is instrumental in discerning patterns and understanding recurring meteoritic phenomena. To visually represent this correlation, Fig. 1 offers a histogram that plots impacting meteoroid classifications (sporadic or associated) based on solar longitude. The most active meteor showers are annotated. It can be observed an apparent concentration of the meteoroid background activity toward 180° of solar longitude.

3.2. Degree of rank correlation

For each of the showers listed in the IAU database, we compute the similarity/closeness between the shower and each meteor in the CAMS database using the D-criteria and all distance metric combinations. We then calculate the Kendall rank correlation between each D-criterion and each vector-metric combination.

Fig. 2 displays the Kendall rank correlation between the evaluated D-criteria and distance metrics. Each column corresponds to a particular distance metric, and the plots are color-coded by D-vectors. The box plots encapsulate the quartile distribution of the samples, where each sample denotes the rank correlation between the D-criteria and distance metrics for a meteor with all meteor showers. The calculation is performed for each meteor against all meteor showers, a process executed separately for both the GEO and ORBIT vectors. Points lying outside the whiskers of the box plots are classified as outliers, positioned more than 1.5 times the interquartile range away from the median (Q2, depicted by the box's central line). A homogeneous dataset would result in a compact interquartile range, with the median equidistant from the box's extremes (Q1 and Q3), indicating symmetry. The span from the plot to each whisker indicates the data's variability or spread, suggesting a more concentrated distribution if the span is shorter and greater dispersion if it is extended.

The different figures reveal particular features in the Kendall rank correlation between D-criteria and distance metrics, as delineated by the employment of GEO and ORBIT vectors. The sEuclidean metric paired with the GEO vector consistently demonstrates the highest median correlation across all D-criteria, indicating a robust ordinal association. In contrast, the ORBIT vector presents a distinctive landscape. D_D criterion, when evaluated with ORBIT vectors, achieves the highest correlation values. ϱ_2 criterion exhibits considerable variability in correlation, as evidenced by notably wide box plots for some distance metrics when using the ORBIT vectors. This behavior starkly contrasts with the other D-criteria, pointing to ϱ_2 unique response to the parameters captured by ORBIT vectors. While the GEO vector is characterized by a greater number of lower outliers, indicating instances of significantly divergent rankings, ORBIT vectors show fewer upper outliers. The results show a general tendency for the median correlation values to either be randomly centered or skewed across both vectors and all metrics. This variability suggests that no singular pattern of correlation prevails universally. Additionally, the maximum whisker extension observed with the Cosine distance metric, specifically when paired with the GEO vector and D_D criterion, signals instances of high variability or dispersion in the degree of correlation.
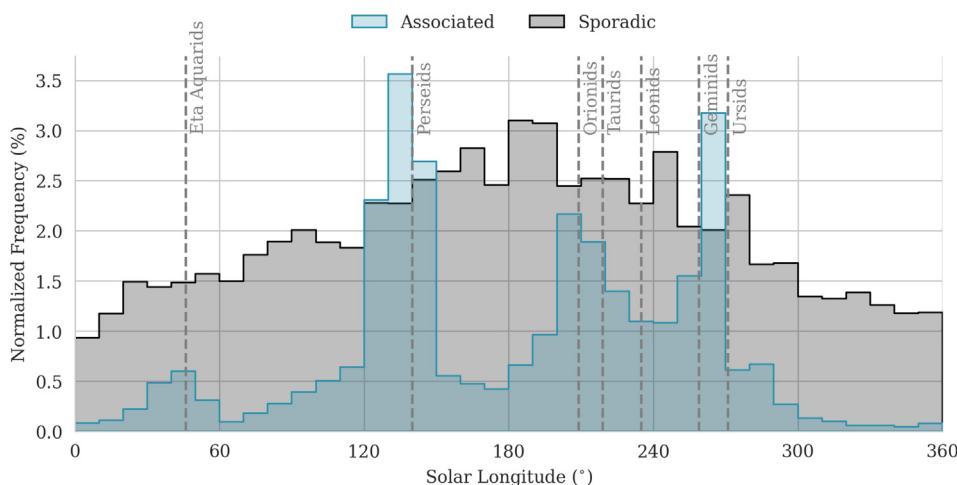

Fig. 1. Histogram of CAMS database as a function of the solar longitude. Sporadic and associated meteors are depicted.
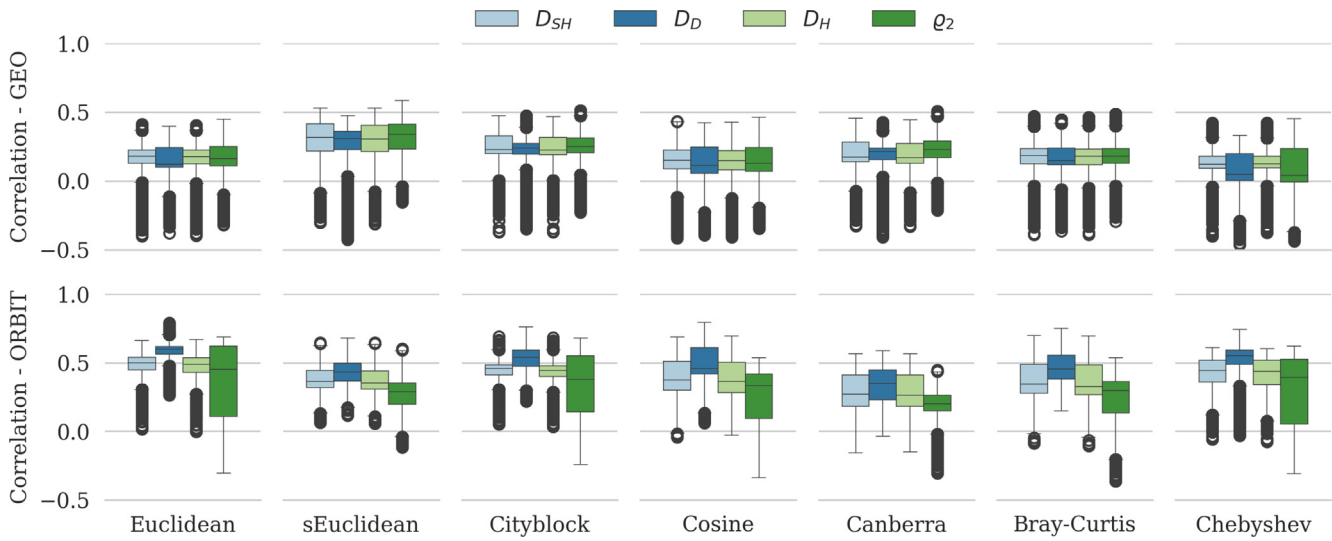
Fig. 2. Kendall rank correlation between D-criteria and metric distances for associated meteors in CAMS database. Each column corresponds to a unique vector (ORBIT or GEO). Each sample symbolizes the rank correlation between the similarity criteria and the distance metrics of each of the meteors from the CAMS database concerning the distinct meteor showers. Outlier values surpass 1.5 times the interquartile range of the median.

Table 2
Top-k accuracies of D-criteria for associated meteors in CAMS database.

Top-k	D_{SH} (%)	D_D (%)	D_H (%)	ϱ_2 (%)
1	**86.23**	80.07	85.83	85.56
5	94.90	93.60	94.99	**95.67**
10	97.50	96.63	97.60	**97.93**

3.3. Accuracy of best choices

Using both the D-criteria and the employed meteor vectors and distance metrics, each meteor is juxtaposed with all showers, as detailed in Tables 2,3. This approach facilitates representing, in percentages, instances where the associated shower from the CAMS database in terms of distance and similarity aligns (Top-1) ranks among the five showers with the most minimal similarity and standardized Euclidean distance values (Top-5) or falls within the top ten showers (Top-10).

The optimal D-criterion for achieving Top-1 accuracy is D_{SH} (86.23%), whereas ϱ_2 excels in both Top-5 (95.67%) and Top-10 (97.93%) categories and have a good accuracy in Top-1 (85.56%). Conversely, D_D ranks as the least effec-

tive across all evaluations, markedly lagging behind the others, which exhibit comparable performances. The sEuclidean when combined with the GEO vector demonstrates superior performance (87.06%) over the other D-criteria in achieving Top-1 accuracy and overly the rest of the distance metrics in all Top-k tests. When paired with the ORBIT vector, the Bray-Curtis metric delivers the highest overall accuracy (including the D_D criterion in all tests), except for Top-5 accuracy, where the Euclidean metric slightly outperforms it. Across the distance metrics evaluated, the GEO vector is found to yield better outcomes than the ORBIT vector. The Chebyshev metric exhibits the worst results with the GEO vector, while the sEuclidean

Table 4
Mean accuracies and standard deviations for Top-k tests across the D-criteria and distance metrics with GEO and ORBIT vectors.

Test	All (%)	D-criteria (%)	GEO (%)	ORBIT (%)
Top-1	83.7 ± 2.5	84.2 ± 2.5	**85.1 ± 1.1**	81.8 ± 2.5
Top-5	93.6 ± 1.3	**94.8 ± 0.7**	93.4 ± 1.2	93.2 ± 1.1
Top-10	96.2 ± 1.0	**97.4 ± 0.5**	95.9 ± 0.9	95.9 ± 0.8

Table 3
Top-1, Top-5, and Top-10 accuracies of distance metrics for associated meteors in CAMS database.

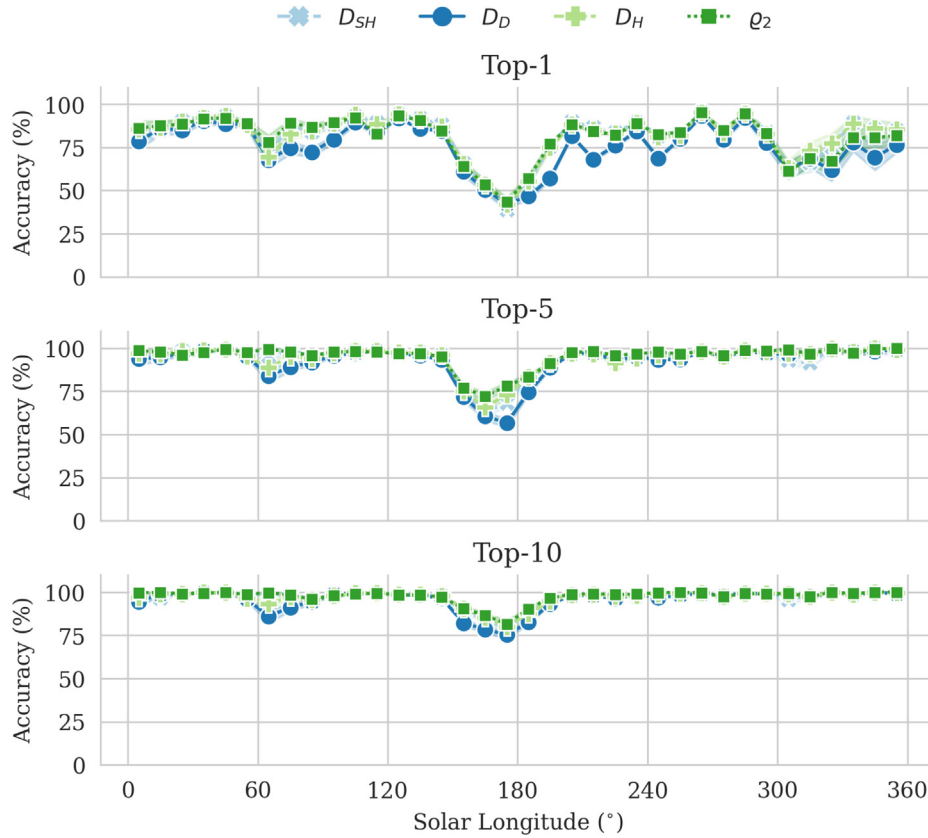|	Euclidean	sEuclidean	Cityblock	Cosine	Canberra	Bray-Curtis	Chebyshev
				Top-1 (%)			
GEO	84.68	**87.06**	84.79	84.63	85.89	84.96	83.41
ORBIT	83.67	77.53	83.92	83.83	79.64	**83.96**	80.11
				Top-5 (%)			
GEO	92.99	**95.03**	93.72	92.95	94.35	93.89	90.94
ORBIT	**94.17**	92.68	94.05	93.88	91.06	94.10	92.29
				Top-10 (%)			
GEO	95.56	**97.06**	96.16	95.52	96.66	96.23	94.05
ORBIT	96.37	96.25	96.57	96.19	95.08	**96.61**	94.45

Fig. 3. Top-k accuracies along solar longitude of the D-criteria for associated meteors in CAMS database.
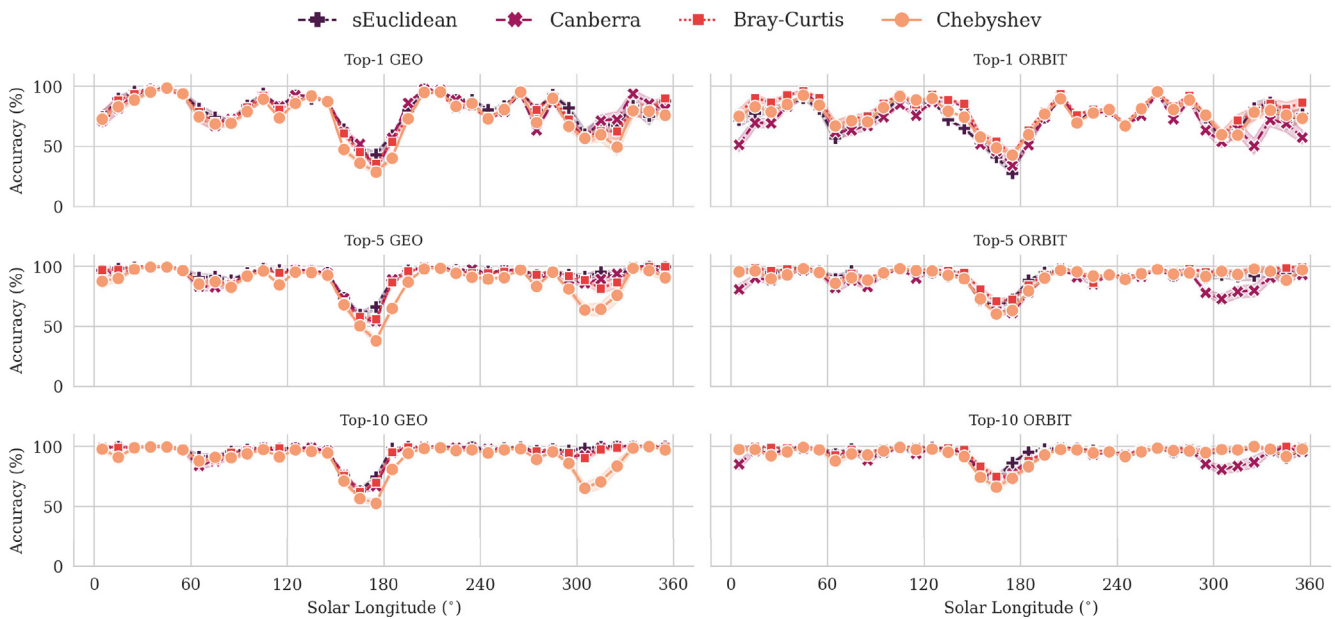

Fig. 4. Top-k accuracies along solar longitude of different distance metrics for associated meteors in CAMS database.

and Canberra present the lower performances for the ORBIT vector. Table 4 shows the mean accuracies. The distance metrics combined with the GEO vectors offer the best overall accuracy for Top-1, while the D-criteria outstrip in Top-5 and Top-10.

Fig. 3 illustrates the variation in Top-k test accuracy as a function of solar longitude across different D-criteria. Similarly, Fig. 4 displays the Top-k results for the sEuclidean, Canberra, Bray-Curtis, and Chebyshev distance metrics. Across all evaluations, the results are of the same

order of magnitude. A distinct pattern emerges: for Top-1, the accuracy variation is irregular, whereas, for Top-10, it tends towards uniformity, except for a notable decrease (up to 50%) around 180° of solar longitude. Visually, the lower performance of D_D is prominent, and D_{SH} and ϱ_2 excel, especially at solar longitudes between 170° and 220°, as well as around 70° in Top-5 and Top-10, and 350° in Top-1 (with a sudden increase of the accuracy of D_H). Conversely, the performances of the distance metrics generally follow the same trend, albeit less uniformly in the Top-10 distribution. Besides the common peak at 180°, it is observed that they struggle to associate meteors at around 310°, where Chebyshev (with GEO vector) and Canberra (with ORBIT vector) exhibit remarkably lower performances.

The trend of minimum accuracy in meteor association is pinpointed at 180° solar longitude, aligning with an apparent increase in the meteoroid background activity, as depicted in Fig. 1. This time frame also bridges the Perseids and Orionids, meteor showers renowned for their high activity and velocities above 60 km/s, expecting, in consequence, a diffuseness of their parameters. Instrumental constraints correlate meteoroid velocity with measurement inaccuracies (Hajduková and Kornoš, 2020). As a result, high-velocity meteoroids are more challenging to accurately characterize. This is depicted in Fig. 5, showcasing a concentration of high apparent velocities within this specific solar longitude range. It is conceivable that these meteoroids were once part of such swarms but have lost their orbital affinity due to temporal decoherence, making

many of them challenging to distinguish. Furthermore, the increased activity during these periods, characterized by similar velocities, may have influenced the association process conducted by CAMS.

3.4. Statistical qquivalence

Fig. 6 displays classification outcomes labeled as H_0 or H_1, corresponding to the hypothesis tested for each data comparison for the Top-1 accuracy results from the metric distance and the D-criteria. Labels are determined based on p-values: instances where the p-value is less than 0.05 are marked as H_1, indicating the rejection of the null hypothesis (H_0) in favor of the alternative (H_1), suggesting a statistically significant difference between the compared distributions. Conversely, instances with a p-value greater than or equal to 0.05 retain the H_0 label, indicating insufficient evidence to reject the null hypothesis, thus suggesting no statistically significant difference between distributions under examination.

The figure reveals a distinct pattern in the distribution of hypothesis testing results, particularly when evaluating the D_D criterion with the GEO vector. Contrary to the other D-criteria, which generally do not reject the null hypothesis H_0 when paired with the GEO vector, D_D stands out by predominantly rejecting H_0 (indicated by H_1), suggesting differences in distributions. This trend is reversed for the ORBIT vector, where D_D results in non-rejection of H_0, except for one the Canberra metric. This behavior is markedly different from the other criteria tested with the ORBIT
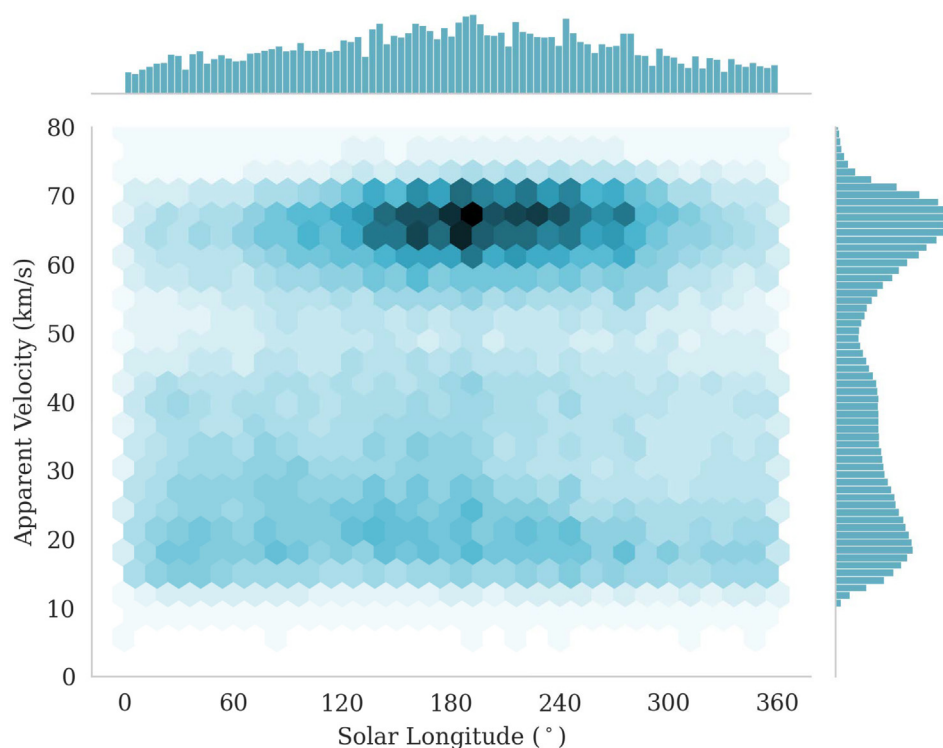

Fig. 5. 2D-histogram of sporadic meteor apparent velocities and solar longitudes at impact in the CAMS database. Darker colors denote higher density.
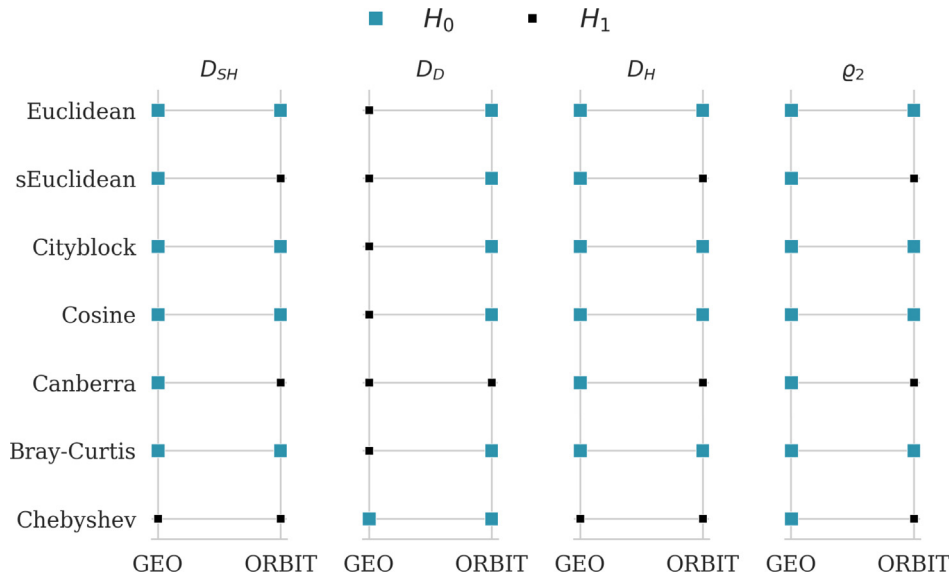
Fig. 6. K-S test comparing Top-1 accuracies of distance metrics and D-criteria with a 95% level of confidence for associated meteors in CAMS database. H_0 indicates no statistically significant difference between distributions, while H_1 indicates a significant difference between the compared distributions.

vector, which rejects H_0 for the same three metrics (sEuclidean, Canberra, and Chebyshev). ϱ_2 appears the most likely compatible with both vectors at the same time.

The consistent failure to reject H_0 with GEO vector for all distance metrics under the ϱ_2 criterion does not confirm the distributions being identical but rather indicates the test lacked sufficient evidence to demonstrate statistical differences. This outcome positions ϱ_2 as the D-criterion that is most plausibly comparable to the distance metrics in terms of meteor association when using the GEO vector. Also, the ϱ_2 metric exhibits the highest probability of being compatible with both vectors simultaneously.

3.5. Event-by-event agreement

The heatmap on Fig. 7 visualizes the agreement level between various D-vectors and distance metrics, showcasing their comparative analysis for the Top-1 results across the two meteor vectors. Each cell represents the percentage of Top-1 coincidence between pairs, with GEO-related comparisons highlighted in shades of blue for intuitive analysis, and ORBIT-related comparisons in shades of red, enabling a clear distinction between the two meteor vectors used. The diagonal, intentionally left blank, separates GEO and ORBIT results for a dual analysis within a single visual representation. The cross-accuracies of D-criteria, independent of GEO or ORBIT vectors, are outlined with a black frame in the figure's top left corner.

Analyzing the heatmap reveals that the D-vector D_{SH} has a strong event-by-event alignment with D_H for Top-1 (97.43%), indicating these criteria frequently concur on their top classifications. This is closely followed by ϱ_2 and D_H (94.69%). Within the GEO vector, Euclidean and Cosine (99.66%), along with Cityblock and Bray-Curtis (99.15%), show the highest levels of coincidence in Top-1.

The sEuclidean metric generally shows good agreement for the GEO vector across various metrics and D-criteria (∼88%), except when paired with D_D (83.37%). For the ORBIT vector, Cityblock and Bray-Curtis (98.99%), as well as Euclidean and Cosine (95.62%), exhibit the highest values. There is better alignment between D_D and the ORBIT vector (reaching ∼86% with various distance metrics) than seen with the GEO vector.

The consistency observed in the heatmap resonates with the findings from Kendall's correlation and the K-S test. These statistical measures support the identified patterns of agreement and discrepancy among the classifiers, providing robustness to the analysis and confirming the reliability of these patterns.

3.6. Thresholds and confusion matrices

Table 5 presents the evaluation of D-criteria and distance metrics within the CAMS database, considering both sporadic and associated meteor events, where optimal thresholds and the effectiveness of different methods are encapsulated. The standout performer among D-criteria is D_{SH}, distinctly outshining others with a ϕ of 0.6400. Conversely, D_D emerges as the least effective.

When using the GEO vector, the sEuclidean metric takes precedence, exhibiting the highest overall accuracy and a ϕ value of 0.6464, closely followed by Cityblock and Bray-Curtis metrics. The scenario shifts when transitioning to the ORBIT vector, where Cityblock edges out as the frontrunner, albeit with Bray-Curtis and Euclidean not far behind, suggesting a competitive field with closely matched performances. The sEuclidean metric with ORBIT vector does not mirror its GEO vector success, hinting at vector-specific behavior that influences metric efficacy.
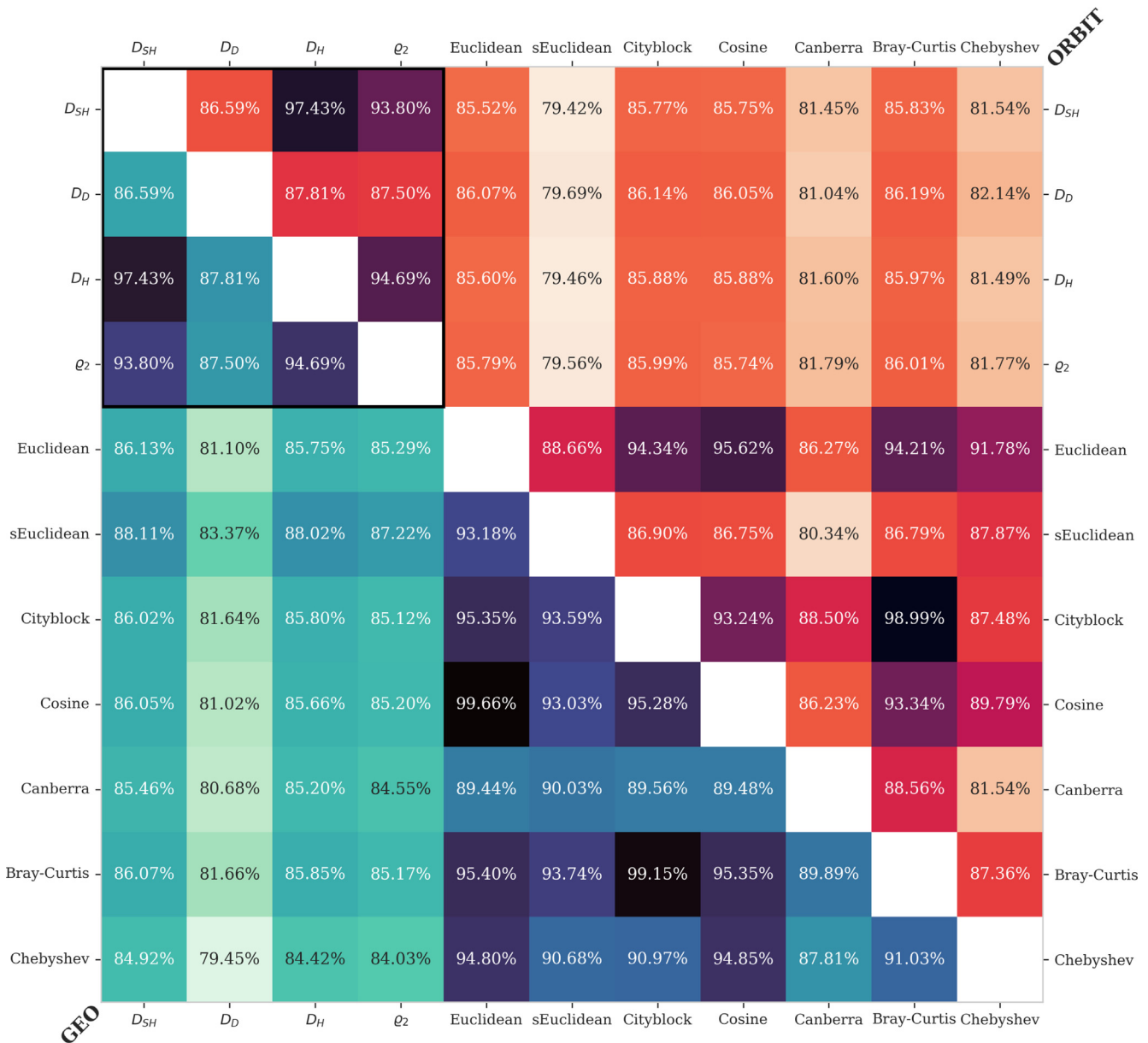
Fig. 7. Heatmap of cross-coincidence between D-vectors and distance metrics using GEO (lower triangle, blue colormap) and ORBIT (upper triangle, red colormap) vectors of Top-1 accuracies for associated meteors in CAMS database. D-vector's own cross-coincidences are highlighted within a black rectangle in the top left corner.

Cityblock, while not outperforming other distance metrics in replicating CAMS' associated Top-1 classifications, excelled in more effectively distinguishing the sporadic background on average. Except for D_{SH}, all distance metrics applied to the GEO vector–aside from Cosine–surpass the rest of the D-criteria in terms of the ϕ. Interestingly, despite generally lower performance with the ORBIT vector, several distance metrics still exceed some D-criteria performances. Cityblock, in particular, scores relatively close to achieving the superior results of D_{SH} and sEuclidean.

Additionally, the observed thresholds for traditional D-criteria (D_{SH}, D_D, and D_H) align perfectly with values documented in the scientific literature, reinforcing the validity of our findings.

As an additional note to our findings, it is noteworthy that upon incorporating the complete list of meteor showers—not limited to those used within the CAMS database—an average of 27% of the meteor classified (Top-1) by all D-criteria and distance metrics would align better with newly recognized meteor showers. In future efforts, we aim to do a comparative analysis by testing other databases such as GMN (Vida et al., 2021) and EDMOND (Kornoš et al., 2014).

4. Conclusions

This study undertook a statistical evaluation of four orbital similarity criteria (or D-criteria) within a five-dimensional parameter space to probe the dynamical asso-

Table 5
Threshold, accuracies, and Matthews correlation coefficients for different D-criteria and distance metrics in the CAMS database taking into account the sporadic and associated events.

Method	Cut-off	TP (%)	FN (%)	FP (%)	TN (%)	Acc. (%)	ϕ
			D-criteria				
D_{SH}	0.170	59.51	8.62	7.21	24.66	84.17	**0.6400**
D_D	0.083	58.34	9.78	8.35	23.52	81.87	0.5877
D_H	0.176	58.81	9.32	7.76	24.12	82.92	0.6122
ϱ_2	0.174	59.56	8.56	7.97	23.91	83.47	0.6213
			Distance metric with the GEO vector				
Euclidean	0.138	58.75	9.37	7.13	24.75	83.50	0.6279
sEuclidean	0.370	59.25	8.88	6.78	25.10	84.34	**0.6464**
Cityblock	0.252	59.45	8.67	7.20	24.67	84.13	0.6393
Cosine	0.004	57.85	10.27	6.97	24.91	82.76	0.6154
Canberra	0.380	59.98	8.15	8.22	23.65	83.63	0.6228
Bray-Curtis	0.037	59.80	8.33	7.48	24.40	84.19	0.6387
Chebyshev	0.089	61.01	7.11	8.83	23.05	84.06	0.6282
			Distance metric with the ORBIT vector				
Euclidean	0.086	60.23	7.89	8.01	23.86	84.10	0.6335
sEuclidean	0.542	59.80	8.32	9.71	22.17	81.97	0.5803
Cityblock	0.173	59.63	8.50	7.46	24.42	84.04	**0.6359**
Cosine	0.002	57.47	10.65	6.94	24.94	82.41	0.6093
Canberra	0.431	55.69	12.43	8.13	23.74	79.44	0.5454
Bray-Curtis	0.031	59.31	8.81	7.29	24.58	83.90	0.6343
Chebyshev	0.061	59.55	8.57	8.54	23.34	82.89	0.6062

ciations within meteor data. Utilizing the extensive data compiled by the CAMS network, we have not only relied on D-criteria (D_{SH}, D_D, D_H, and ϱ_2) but also ventured into distance metrics commonly applied in Machine Learning (Euclidean, sEuclidean, Cityblock, Cosine, Canberra, Bray-Curtis, and Chebyshev), investigated across two distinctive meteor vectors. One vector termed ORBIT, based on heliocentric orbital elements, is essentially shared with the D-criteria, and the other one, GEO, based on geocentric observational parameters, was proposed by Sugar et al. (2017). Our methodology hinged on the Kendall rank correlation coefficient and Top-k accuracy tests to assess the correlation and performance of these criteria and metrics. We also applied the Kolmogorov–Smirnov test and computed the level of coincidence of individual Top-1 results for discerning the statistical equivalence of the different approaches. Finally, we calculated the optimal thresholds and evaluated their performances in distinguishing the sporadic background from the meteor showers.

Our key findings can be summarized as follows:

- The sEuclidean metric paired with the GEO vector demonstrates superior performances than the D-criteria and the other distance metrics in achieving Top-1 accuracy (87.06%).
- Regarding the D-criteria, the D_{SH} criterion holds the upper hand in achieving Top-1 accuracy (86.23%), while ϱ_2 maintains dominance in both the Top-5 (95.67%) and Top-10 (97.93%) categories (surpassed by D_{SH} in Top-1 accuracy by 0.67%).
- The Bray-Curtis metric, allied with the ORBIT vector, demonstrated a consistent edge over other distance metrics, outperforming the D_D criterion across all Top-k

tests (83.96%, 94.10%, and 96.61%, in increasing order of k) and only slightly beaten by the Euclidean metric in Top-5 accuracy by a negligible difference (0.07%).
- D_D exhibits an opposite trend to the other D-criteria when evaluating its equivalence against distance metrics with the GEO vector.
- Among the D-criteria, ϱ_2 appears as the most likely similar to the distance metrics with the GEO vector, being also the most compatible with both GEO and ORBIT vectors at the same time.
- In general terms, the D-criteria and the metric distances provide similar accuracies in Top-k tests (83.7±2.5%, 93.6±1.3%, 96.2±1.0%, in ascending order of k), with the D_D and the metric Chebyshev performing worse.
- The mean highest accuracies are achieved with the GEO (85.1±1.1%, 93.4±1.2%, and 95.9±0.9%) rather than the ORBIT vector (81.8±2.5%, 93.2±1.1%, and 95.9±0.8%) within the Top-1, Top-5, and Top-10 tests, and surpassing also the D-criteria in Top-1 (84.4±2.5%, 94.8±0.7%, and 97.4±0.5%). This suggests that geocentric parameters provide a more robust basis than orbital elements for meteor dynamical association.
- We observed moderate solar longitude-dependent deviations and a common significant decrease in accuracy around 180° of solar longitude. We tentatively linked these features to heightened meteoroid background activity and the interface with two of the most active, high-velocity meteor showers: the Perseids and the Orionids.
- The sEuclidean metric achieves the highest overall accuracy (84.34%) and ϕ of 0.6464 for GEO vector applications, excelling in distinguishing the sporadic

background. It is closely followed by Cityblock ($\phi = 0.6393$ and 84.13% accuracy) and Bray-Curtis metrics ($\phi = 0.6387$ and 84.19% accuracy).

- Among D-criteria, D_{SH} distinguishes itself with a ϕ of 0.6400, translating to an 84.17% accuracy rate in separating the background, while D_D emerges as the least effective, with a ϕ of 0.5877 and an accuracy of 81.87%.
- In the ORBIT vector context, Cityblock takes the lead ($\phi = 0.6359$ and 84.04% accuracy), closely challenged by Bray-Curtis ($\phi = 0.6343$ and 83.90% accuracy) and Euclidean ($\phi = 0.6335$ and 84.10% accuracy).
- Excluding Cosine, all distance metrics associated with the GEO vector surpass the D-criteria in ϕ when differentiating the meteoroid background.
- Despite the ORBIT vector's generally lower performance, various distance metrics still exceed certain D-criteria in effectiveness.
- Optimal cut-offs for all D-criteria and distance metrics are provided, founded on the CAMS database classification.
- Based on these approaches, $\sim$27% of associated meteors in CAMS would align with showers identified after the database's release.
- Future research will concentrate on studying effectiveness, equivalences, and thresholds for a synthetic impacting population, exploring the performance and specific attributes of the methods for each individual meteor shower.

The work culminates in the significant revelation that Machine Learning distance metrics can rival or even outperform the specifically tailored orbital similarity criteria for meteor dynamical association. This opens up new pathways for the use of computational techniques in the field of meteor science, offering an opportunity to refine our approaches to classifying meteor showers and sporadic meteors alike.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

References

Andrade, M., Docobo, J.Á., García-Guinea, J., et al., 2023. The Traspena meteorite: heliocentric orbit, atmospheric trajectory, strewn field, and petrography of a new L5 ordinary chondrite. Mon. Not. R. Astron. Soc. 518 (3), 3850–3876. https://doi.org/10.1093/mnras/stac2911.

Arter, T.R., Williams, I.P., 1997. The mean orbit of the April Lyrids. Mon. Not. R. Astron. Soc. 289 (3), 721–728. https://doi.org/10.1093/mnras/289.3.721.

Babadzhanov, P.B., Obrubov, Y.V., 1992. Evolution of short-period meteoroid streams. Celest. Mech. Dynam. Astron. 54, 111–127. https://doi.org/10.1007/BF00049547.

Bland, P.A., 2004. The desert fireball network. Astron. Geophys. 45 (5), 5–20.

Boaca, I., Gritsevich, M., Birlan, M., et al., 2022. Characterization of the Fireballs Detected by All-sky Cameras in Romania. Astrophys. J. 936 (2), 150. https://doi.org/10.3847/1538-4357/ac8542.

Borovicka, J., 1990. The comparison of two methods of determining meteor trajectories from photographs. Bull. Astron. Inst. Czechoslovakia 41, 391.

Borovička, J., Spurný, P., Brown, P., 2015. Small Near-Earth Asteroids as a Source of Meteorites. https://doi.org/10.2458/azu_ua-press_9780816532131-ch014.

Borovička, J., Spurný, P., Shrbený, L., et al., 2022. Data on 824 fireballs observed by the digital cameras of the European Fireball Network in 2017–2018. I. Description of the network, data reduction procedures, and the catalog. Astron. Astrophys. 667, A157. https://doi.org/10.1051/0004-6361/202244184, arXiv:2209.11186.

Bottke, J., William, F., Rubincam, D.P., Burns, J.A., 2000. Dynamical evolution of main belt meteoroids: numerical simulations incorporating planetary perturbations and Yarkovsky Thermal Forces. Icarus 145 (2), 301–331. https://doi.org/10.1006/icar.2000.6361.

Bottke, W.F., Morbidelli, A., Jedicke, R., et al., 2002. Debiased orbital and absolute magnitude distribution of the near-earth objects. Icarus 156 (2), 399–433. https://doi.org/10.1006/icar.2001.6788.

Bredikhin, F.A., 1954. Etyudy o meteorakh (Etudes on Meteors). USSR Acad. Sci..

Brož, M., 2006. Yarkovsky Effect and the Dynamics of the Solar System Ph.D. thesis. Charles University in Prague.

Brown, P., Spalding, R.E., ReVelle, D.O., et al., 2002. The flux of small near-Earth objects colliding with the Earth. Nature 420 (6913), 294–296. https://doi.org/10.1038/nature01238.

Carbognani, A., Fenucci, M., 2023. Identifying parent bodies of meteorites among near-Earth asteroids. MNRAS 525 (2), 1705–1725. https://doi.org/10.1093/mnras/stad2382, arXiv:2308.01931.

Ceplecha, Z., 1957. Photographic Geminids 1955. Bull. Astron. Inst. Czechoslovakia 8, 51.

Ceplecha, Z., 1987. Geometric, dynamic, orbital and photometric data on meteoroids from photographic fireball networks. Bull. Astron. Inst. Czechoslovakia 38, 222.

Ceplecha, Z., Borovička, J., Elford, W.G., et al., 1998. Meteor phenomena and bodies. Space Sci. Rev. 84, 327–471. https://doi.org/10.1023/A:1005069928850.

Chapman, C.R., 2010. Asteroids: Stripped on passing by Earth. Nature 463 (7279), 305–306. https://doi.org/10.1038/463305a.

Colas, F., Zanda, B., Bouley, S., et al., 2020. FRIPON: a worldwide network to track incoming meteoroids. Astron. Astrophys. 644, A53. https://doi.org/10.1051/0004-6361/202038649, arXiv:2012.00616.

Colas, F., Zanda, B., Bouley, S., et al., 2015. FRIPON, the French fireball network. In European Planetary Science Congress (pp. EPSC2015–800).

Devillepoix, H., Cupák, M., Bland, P., et al., 2020. A global fireball observatory. Planet. Space Sci. 191, 105036. https://doi.org/10.1016/j.pss.2020.105036.

Drummond, J.D., 1981. A test of comet and meteor shower associations. Icarus 45 (3), 545–553. https://doi.org/10.1016/0019-1035(81)90020-8.

Dumitru, B.A., Birlan, M., Popescu, M., et al., 2017. Association between meteor showers and asteroids using multivariate criteria. Astron. Astrophys. 607, A5. https://doi.org/10.1051/0004-6361/201730813.

Fenwick, P.M., 1994. A new data structure for cumulative frequency tables. Software: Pract. Exp. 24 (3), 327–336. https://doi.org/10.1002/spe.4380240306.

Galligan, D.P., 2001. Performance of the D-criteria in recovery of meteoroid stream orbits in a radar data set. MNRAS 327 (2), 623–628. https://doi.org/10.1046/j.1365-8711.2001.04858.x.

Gardiol, D., Cellino, A., & Di Martino, M. (2016). PRISMA, Italian network for meteors and atmospheric studies. In A. Roggemans, & P. Roggemans (Eds.), International Meteor Conference Egmond, the Netherlands, 2–5 June 2016 (p. 76).

Gritsevich, M., Lyytinen, E., Moilanen, J., et al., 2014. First meteorite recovery based on observations by the Finnish Fireball Network. In: Rault, J.L., Roggemans, P. (Eds.), Proceedings of the International Meteor Conference, Giron, France, 18-21 September 2014, pp. 162–169.

Gritsevich, M.I., 2009. Determination of parameters of meteor bodies based on flight observational data. Adv. Space Res. 44 (3), 323–334. https://doi.org/10.1016/j.asr.2009.03.030.

Gritsevich, M.I., Stulov, V.P., Turchak, L.I., 2012. Consequences of collisions of natural cosmic bodies with the Earth's atmosphere and surface. Cosm. Res. 50 (1), 56–64. https://doi.org/10.1134/S0010952512010017.

Gustafson, B.A.S., 1989. Geminid meteoroids traced to cometary activity on Phaethon. Astron. Astrophys. 225 (2), 533–540.

Hajduková, M., Kornoš, L., 2020. The influence of meteor measurement errors on the heliocentric orbits of meteoroids. Planet. Space Sci. 190, 104965. https://doi.org/10.1016/j.pss.2020.104965.

Halliday, I., 1987. Detection of a meteorite "stream": Observations of a second meteorite fall from the orbit of the Innisfree chondrite. Icarus 69 (3), 550–556. https://doi.org/10.1016/0019-1035(87)90023-6.

Hodges, J.L., 1958. The significance probability of the smirnov two-sample test. Arkiv for Matematik 3 (5), 469–486. https://doi.org/10.1007/BF02589501.

Hollander, M., Wolfe, D.A., Chicken, E., 2013. Nonparametric Statistical Methods. Wiley Series in Probability and Statistics, 3rd ed. Wiley-Blackwell, Hoboken, NJ.

Hughes, D.W., 1986. Hollow meteoroid streams. Nature 319, 718.

Jacchia, L.G., Whipple, F.L., 1956. The harvard photographic meteor programme. Vistas Astron. 2, 982–994. https://doi.org/10.1016/0083-6656(56)90021-6.

Jenniskens, P., 1994. Meteor stream activity I. The annual streams. Astronomy & Astrophysics 287, 990–1013.

Jenniskens, P., 1998. On the dynamics of meteoroid streams. Earth, Planets and Space 50, 555–567. https://doi.org/10.1186/BF03352149.

Jenniskens, P., 2006. Meteor Showers and their Parent Comets.

Jenniskens, P., 2008. Meteoroid streams that trace to candidate dormant comets. Icarus 194 (1), 13–22. https://doi.org/10.1016/j.icarus.2007.09.016.

Jenniskens, P., 2017. Meteor showers in review. Planet. Space Sci. 143, 116–124. https://doi.org/10.1016/j.pss.2017.01.008.

Jenniskens, P., 2023. Atlas of Earth's Meteor Showers, 1st ed. Elsevier, Amsterdam.

Jenniskens, P., Baggaley, J., Crumpton, I., et al., 2018. A survey of southern hemisphere meteor showers. Planet. Space Sci. 154, 21–29. https://doi.org/10.1016/j.pss.2018.02.013.

Jenniskens, P., Gural, P.S., Dynneson, L., et al., 2011. CAMS: Cameras for Allsky Meteor Surveillance to establish minor meteor showers. Icarus 216 (1), 40–61. https://doi.org/10.1016/j.icarus.2011.08.012.

Jenniskens, P., Jopek, T.J., Janches, D., et al., 2020. On removing showers from the IAU Working List of Meteor Showers. Planet. Space Sci. 182, 104821. https://doi.org/10.1016/j.pss.2019.104821.

Jopek, T.J., 1993. Remarks on the Meteor Orbital Similarity D-Criterion. Icarus 106 (2), 603–607. https://doi.org/10.1006/icar.1993.1195.

Jopek, T.J., Bronikowska, M., 2017. Probability of coincidental similarity among the orbits of small bodies – i. pairing. Planet. Space Sci. 143, 43–52. https://doi.org/10.1016/j.pss.2016.12.004.

Jopek, T.J., Jenniskens, P.M., 2011. The Working Group on Meteor Showers Nomenclature: A History, Current Status and a Call for Contributions. In W.J. Cooke, D.E. Moser, B.F. Hardin, & D. Janches (Eds.), Meteoroids: The Smallest Solar System Bodies (pp. 7–13).

Jopek, T.J., Kanuchová, Z., 2013. The meteoroids 2013. In: Proceedings of the Astronomical Conference held at AM University, Poznan, Poland (pp. 353–364). University Press.

Jopek, T.J., Kaňuchová, Z., 2017. IAU Meteor Data Center-the shower database: A status report. Planetary and Space Science, 143, 3–6. doi:10.1016/j.pss.2016.11.003. arXiv:1607.00661.

Jopek, T.J., Williams, I.P., 2013. Stream and sporadic meteoroids associated with near-Earth objects. Mon. Not. R. Astron. Soc. 430 (3), 2377–2389. https://doi.org/10.1093/mnras/stt057.

Kasuga, T., Sato, M., Ueda, M., et al., 2020. A fireball and potentially hazardous binary near-earth asteroid (164121) 2003 yt1. Astron. J. 159 (2), 47. https://doi.org/10.3847/1538-3881/ab4e1b.

Kendall, M.G., 1938. A new measure of rank correlation. Biometrika 30 (1/2), 81. https://doi.org/10.2307/2332226.

Kholshevnikov, K.V., Kokhirova, G.I., Babadzhanov, P.B., et al., 2016. Metrics in the space of orbits and their application to searching for celestial objects of common origin. Mon. Not. R. Astron. Soc. 462 (2), 2275–2283. https://doi.org/10.1093/mnras/stw1712.

Kornoš, L., Koukal, J., Piffl, R., et al., 2014. EDMOND Meteor Database. In: Gyssens, M., Roggemans, P., Zoladek, P. (Eds.), Proceedings of the International Meteor Conference, Poznan, Poland, 22-25 August 2013, pp. 23–25.

Koschny, D., Borovicka, J., 2017. Definitions of terms in meteor astronomy. WGN, Journal of the International Meteor Organization 45 (5), 91–92.

Koschny, D., Soja, R.H., Engrand, C., et al., 2019. Interplanetary Dust, Meteoroids, Meteors and Meteorites. Space Sci. Rev. 215 (4), 34. https://doi.org/10.1007/s11214-019-0597-7.

Koseki, M., 2017. SonotaCo network and CAMS. eMeteorNews 2 (4), 105–108.

Koseki, M. (2022). SonotaCo net vs CAMS vs EDMOND vs GMN in the case of the chi Cygnids (CCY#0757). eMeteorNews, 7(6), 381–389.

Matlovic, P., Kornos, L., Kovácová, M., et al., 2020. Characterization of the june epsilon ophiuchids meteoroid stream and the comet 300p/catalina. Astron. Astrophys. 636, A122. https://doi.org/10.1051/0004-6361/202037727.

Matthews, B., 1975. Comparison of the predicted and observed secondary structure of t4 phage lysozyme. Biochimica et Biophysica Acta (BBA) - Protein. Structure 405 (2), 442–451. https://doi.org/10.1016/0005-2795(75)90109-9.

Milanov, D.V., Milanova, Y.V., Kholshevnikov, K.V., 2019. Relaxed triangle inequality for the orbital similarity criterion by Southworth and Hawkins and its variants. Celestial Mechanics and Dynamical Astronomy 131 (1), 5. https://doi.org/10.1007/s10569-019-9884-6.

Moorhead, A.V., 2015. Performance of D-criteria in isolating meteor showers from the sporadic background in an optical data set. Mon. Not. R. Astron. Soc. 455 (4), 4329–4338. https://doi.org/10.1093/mnras/stv2610.

Murad, E., Williams, I.P., 2002. Meteors in the Earth's Atmosphere.

Olsson-Steel, D., 1986. The origin of the sporadic meteoroid component. MNRAS 219, 47–73. https://doi.org/10.1093/mnras/219.1.47.

Pauls, A., Gladman, B., 2005. Decoherence time scales for meteoroid streams. Meteoritics & Planetary Science 40, 1241–1256. https://doi.org/10.1111/j.1945-5100.2005.tb00186.x.

Peña-Asensio, E., Trigo-Rodríguez, J.M., Gritsevich, M., et al., 2021. Accurate 3D fireball trajectory and orbit calculation using the 3D-FIRETOC automatic Python code. Mon. Not. R. Astron. Soc. 504 (4), 4829–4840. https://doi.org/10.1093/mnras/stab999, arXiv:2103.13758.

Peña-Asensio, E., Trigo-Rodríguez, J.M., Rimola, A., 2022. Orbital Characterization of Superbolides Observed from Space: Dynamical Association with Near-Earth Objects, Meteoroid Streams, and Iden-

tification of Hyperbolic Meteoroids. The Astronomical Journal 164 (3), 76. https://doi.org/10.3847/1538-3881/ac75d2, arXiv:2206.03115.

Peña-Asensio, E., Trigo-Rodríguez, J.M., Rimola, A., et al., 2023. Identifying meteorite droppers among the population of bright 'sporadic' bolides imaged by the Spanish Meteor Network during the spring of 2022. Mon. Not. R. Astron. Soc. 520 (4), 5173–5182. https://doi.org/10.1093/mnras/stad102, arXiv:2301.03515.

Peña-Asensio, E., Visuri, J., Trigo-Rodríguez, J.M., et al., 2024. Oort cloud perturbations as a source of hyperbolic Earth impactors. Icarus 408, 115844. https://doi.org/10.1016/j.icarus.2023.115844, arXiv:2310.12673.

Plavec, M., 1954. On the relations between minor planets and meteor streams. Bull. Astron. Inst. Czechoslovakia 5, 38.

Popova, O., Borovička, J., & Campbell-Brown, M.D. (2019). Modelling the Entry of Meteoroids. In G.O. Ryabova, D.J. Asher, & M.J. Campbell-Brown (Eds.), Meteoroids: Sources of Meteors on Earth and Beyond (p. 9).

Porubčan, V., Williams, I.P., Kornoš, L., 2004. Associations Between Asteroids and Meteoroid Streams. Earth Moon and Planets 95 (1–4), 697–712. https://doi.org/10.1007/s11038-005-2243-5.

Rudawska, R., Matlovič, P., Tóth, J., et al., 2015. Independent identification of meteor showers in EDMOND database. Planet. Space Sci. 118, 38–47. https://doi.org/10.1016/j.pss.2015.07.011, arXiv:1406.6598.

Rudawska, R., Vaubaillon, J., Atreya, P., 2012. Association of individual meteors with their parent bodies. Astron. Astrophys. 541. https://doi.org/10.1051/0004-6361/201118762.

Schisterman, E.F., Perkins, N.J., Liu, A., et al., 2005. Optimal cut-point and its corresponding youden index to discriminate individuals using pooled blood samples. Epidemiology 16 (1), 73–81. https://doi.org/10.1097/01.ede.0000147512.81966.ba, URL: https://doi.org/10.1097/01.ede.0000147512.81966.ba.

Silber, E.A., Boslough, M., Hocking, W.K., et al., 2018. Physics of meteor generated shock waves in the Earth's atmosphere - A review. Adv. Space Res. 62 (3), 489–532. https://doi.org/10.1016/j.asr.2018.05.010, arXiv:1805.07842.

Sokolova, M.G., Nefedyev, Y.A., Varaksina, N.Y., 2014. Asteroid and comet hazard: Identification problem of observed space objects with the parental bodies. Adv. Space Res. 54 (11), 2415–2418. https://doi.org/10.1016/j.asr.2014.08.002.

SonotaCo, 2016. Observation error propagation on video meteor orbit determination. WGN, Journal of the International Meteor Organization 44 (2), 42–45.

Southworth, R.B., Hawkins, G.S., 1963. Statistics of meteor streams. Smithsonian Contributions to Astrophysics 7, 261–285.

Sugar, G., Moorhead, A., Brown, P., et al., 2017. Meteor shower detection with density-based clustering. Meteoritics and Planetary Science 52 (6), 1048–1059. https://doi.org/10.1111/maps.12856, arXiv:1702.02656.

Tóth, J., Vereš, P., Kornoš, L., 2011. Tidal disruption of NEAs - a case of Příbram meteorite. MNRAS 415 (2), 1527–1533. https://doi.org/10.1111/j.1365-2966.2011.18799.x, arXiv:1103.4257.

Trigo-Rodriguez, J., Castro-Tirado, A., Llorca, J., et al., 2005. The development of the spanish fireball network using a new all-sky ccd systems. Earth, Moon, and Planets 95 (1–4), 553–567.

Trigo-Rodríguez, J.M., 2022. Asteroid impact risk: Impact hazard from asteroids and comets.

Trigo-Rodríguez, J.M., Blum, J., 2022. Learning about comets from the study of mass distributions and fluxes of meteoroid streams. MNRAS 512 (2), 2277–2289. https://doi.org/10.1093/mnras/stab2827, arXiv:2109.14428.

Trigo-Rodríguez, J.M., Gritsevich, M., Palme, H., Eds., 2017. Dynamic Sources of Contemporary Hazard from Meteoroids and Small

Asteroids volume 46 of Astrophysics and Space Science Proceedings. doi:10.1007/978-3-319-46179-3_2.

Trigo-Rodríguez, J.M., Llorca, J., 2006. The strength of cometary meteoroids: clues to the structure and evolution of comets. MNRAS 372 (2), 655–660. https://doi.org/10.1111/j.1365-2966.2006.10843.x.

Trigo-Rodríguez, J.M., Llorca, J., 2007. Erratum: The strength of cometary meteoroids: clues to the structure and evolution of comets. MNRAS 375 (1). https://doi.org/10.1111/j.1365-2966.2006.11310.x, 415–415.

Valsecchi, G.B., Jopek, T.J., Froeschlé, C., 1999. Meteoroid stream identification: a new approach — I. Theory. Mon. Not. R. Astron. Soc. 304 (4), 743–750. https://doi.org/10.1046/j.1365-8711.1999.02264.x.

Vaubaillon, J., Neslušan, L., Sekhar, A. et al. (2019). From parent body to meteor shower: The dynamics of meteoroid streams.

Vida, D., Gural, P.S., Brown, P.G., et al., 2020. Estimating trajectories of meteors: an observational Monte Carlo approach - I. Theory. Mon. Not. R. Astron. Soc. 491 (2), 2688–2705. https://doi.org/10.1093/mnras/stz3160, arXiv:1911.02979.

Vida, D., Šegon, D., Gural, P.S., et al., 2021. The Global Meteor Network - Methodology and first results. Mon. Not. R. Astron. Soc. 506 (4), 5046–5074. https://doi.org/10.1093/mnras/stab2008, arXiv:2107.12335.

Virtanen, P., Gommers, R., Oliphant, T.E., et al., 2020. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nat. Methods 17, 261–272. https://doi.org/10.1038/s41592-019-0686-2.

Voloshchuk, Y.I., Kashcheev, B.L., 1996. The Meteor Complex near the Earth's Orbit: Sporadic Background, Streams, and Associations. II.A Comparative Analysis of Some Statistical Characteristics of Stream and Sporadic Meteoric Bodies. Sol. Syst. Res. 30 (6), 480.

Walker, R.M., & Cameron, A.G.W. (2006). URL: doi: 10.2307/j.ctv1v7zdmm. doi:10.2307/j.ctv1v7zdmm.

Weryk, R.J., Brown, P.G., Domokos, A., et al., 2007. The southern ontario all-sky meteor camera network. Earth, Moon, and Planets 102 (1–4), 241–246. https://doi.org/10.1007/s11038-007-9183-1.

Whipple, F., 1983. TB and the Geminid Meteors. International Astronomical Union Circular 3881, 1.

Whipple, F.L., 1951. A Comet Model. II. Physical Relations for Comets and Meteors. Astrophys. J. 113, 464. https://doi.org/10.1086/145416.

Wiegert, P., Brown, P., 2004. The problem of linking minor meteor showers to their parent bodies: initial considerations. Earth Moon and Planets 95 (1–4), 19–26. https://doi.org/10.1007/s11038-005-4342-8.

Williams, I.P., Wu, Z., 1993. The Geminid meteor stream and asteroid 3200 Phaethon. Mon. Not. R. Astron. Soc. 262 (1), 231–248. https://doi.org/10.1093/mnras/262.1.231.

Xia, F., Liu, T.-Y., Li, H., 2009. Statistical consistency of top-k ranking. In: Bengio, Y., Schuurmans, D., Lafferty, J., Williams, C., Culotta Inc, A. (Eds.), Advances in Neural Information Processing Systems. Curran Associates, volume 22, URL: https://proceedings.neurips.cc/paper_files/paper/2009/file/ba2fd310dcaa8781a9a652a31baf3c68-Paper.pdf.

Ye, Q.-Z., 2018. Meteor showers from active asteroids and dormant comets in near-Earth space: A review. Planet. Space Sci. 164, 7–12. https://doi.org/10.1016/j.pss.2018.04.018, arXiv:1804.07245.

Youden, W.J., 1950. Index for rating diagnostic tests. Cancer 3 (1), 32–35. https://doi.org/10.1002/1097-0142(1950)3:1<32::aid-cncr2820030106>3.0.co;2-3.

Zhong-Yi, L., Fumi, Y., Yi-Chuan, L., et al., 2020. A study of photometric observations of (3200) phaethon from lulin observatory. Planet. Space Sci. 194, 105114. https://doi.org/10.1016/j.pss.2020.105114.