

Classification of Subgroups of Solar and Heliospheric Observatory (SOHO) Sungrazing Kreutz Comet Group by the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) Clustering Algorithm

Ulkar Karimova, Yu Yi[†]

Department of Astronomy and Space Science, Chungnam National University, Daejeon 34134, Korea

Sungrazing comets, known for their proximity to the Sun, are traditionally classified into broad groups like Kreutz, Marsden, Kracht, Meyer, and non-group comets. While existing methods successfully categorize these groups, finer distinctions within the Kreutz subgroup remain a challenge. In this study, we introduce an automated classification technique using the density-based spatial clustering of applications with noise (DBSCAN) algorithm to categorize sungrazing comets. Our method extends traditional classifications by finely categorizing the Kreutz subgroup into four distinct subgroups based on a comprehensive range of orbital parameters, providing critical insights into the origins and dynamics of these comets. Corroborative analyses validate the accuracy and effectiveness of our method, offering a more efficient framework for understanding the categorization of sungrazing comets.

Keywords: sungrazing comets, solar system origins, coronagraphic observations, density-based spatial clustering of applications with noise (DBSCAN) algorithm, comet categorization, Kreutz subgroups

1. INTRODUCTION

Comets, the celestial nomads of our cosmic neighborhood, have consistently captivated the attention and imagination of astronomers throughout the annals of human history. These enigmatic objects, composed of frozen materials, dust, and rocky constituents, hold the keys to deciphering the profound mysteries surrounding the inception of our solar system. As they venture close to the Sun, comets transform into transient celestial spectacles, manifesting sublimation processes that give rise to a beguiling luminous coma and an iconic tail, unflinchingly extending over vast distances, forever pointing away from the Sun, under the compelling influence of the solar wind (England 2002; Jones et al. 2018).

The understanding of comets, steeped in the pages of history, has undergone a remarkable evolution. Before the 1880s, the radiance of comets in close proximity to the Sun was attributed to a singular recurring sungrazing

comet. However, the seminal work of Kreutz and Kirkwood dismantled this long-standing theory, unveiling that these celestial vagabonds are, in truth, fragments born from past encounters with our Sun (Fig. 1).

More precise studies aimed at clarifying the classification of sungrazing comet groups have resulted in the categorization of sungrazers into four distinct groups. With the identification of the Meyer, Marsden, and Kracht groups, researchers embarked on a comparative analysis with the Kreutz group to discern their differences and similarities (Marsden 1967). One notable distinction lies in the trajectories and orbital dynamics of these comet groups. While Kreutz comets typically follow similar trajectories stemming from a single progenitor comet, other groups like Meyer, Marsden, and Kracht may exhibit more diverse orbital paths and characteristics, possibly influenced by variations in their origins or interactions within the solar system (Knight et al. 2010).

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 26 FEB 2024 Revised 02 MAR 2024 Accepted 04 MAR 2024

[†]Corresponding Author

Tel: +82-42-821-5468, E-mail: euyiyu@cnu.ac.kr

ORCID: <https://orcid.org/0000-0001-9348-454X>

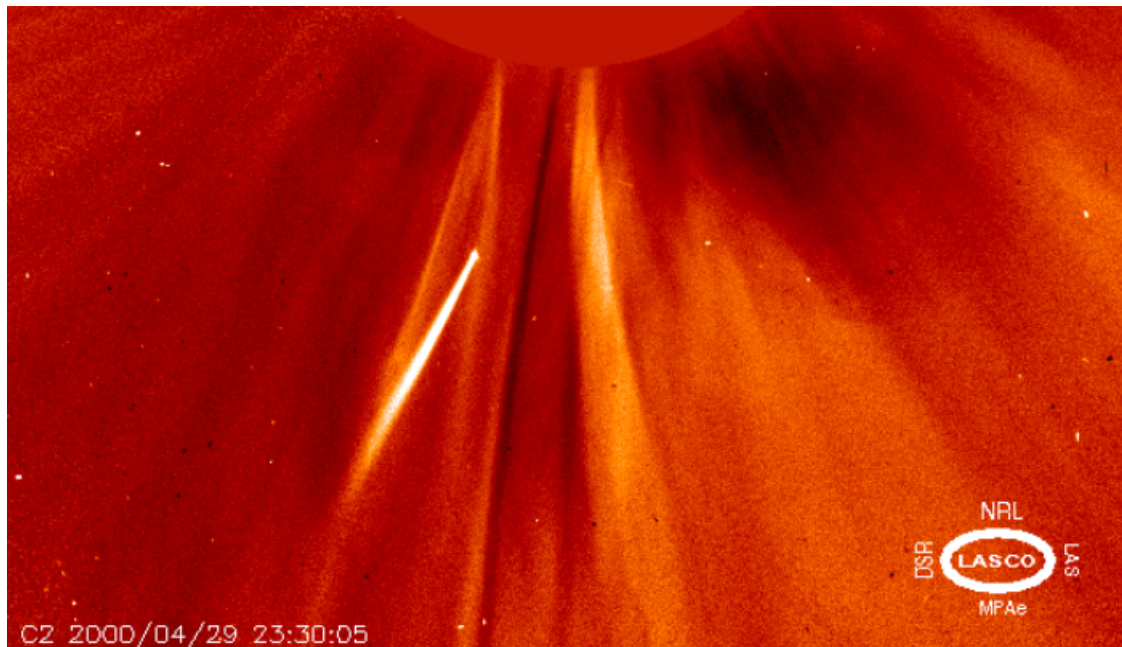


Fig. 1. This illustration depicts SOHO's comet, which was identified in real-time on April 28, 2000. Designated as C/2000 H2 (SOHO), the comet is captured in the image as observed through the LASCO C2 telescope on April 29, 2000. Adapted from SOHO (2020) with CC-BY-NC. Courtesy of SOHO/LASCO consortium. SOHO is a project of international cooperation between ESA and NASA. SOHO, Solar and Heliospheric Observatory; LASCO, The Large Angle and Spectrometric Coronagraph.

Initially, research efforts predominantly focused on the Kreutz group, driven by its abundance and visibility, particularly after the widespread adoption of coronagraphic observations post-1979. However, with the advent of technological advancements and the availability of more data, notably from the Solar and Heliospheric Observatory (SOHO)/The Large Angle and Spectrometric Coronagraph (LASCO) instrument in 1996, attention gradually shifted towards studying other comet groups such as Meyer, Marsden, and Kracht. Despite being investigated later, these groups were found to have fewer detected comets and were situated relatively farther from Earth compared to the Kreutz group. This realization prompted a renewed interest in the Kreutz group due to its proximity and higher detection rates (Lee et al. 2007).

Comprehensive investigations into the intricate features of comet structures and their orbital dynamics have fueled scientific inquiry (Hasegawa & Nakano 2001; Sekanina & Kracht 2013). Deciphering the resilience of sungrazing comets in the face of extreme conditions during perihelion passage is a matter of paramount importance. Variables such as size, distance, and composition (Marsden 2005), emerge as decisive factors influencing their endurance. Even the most imposing comets are not impervious to the perils of fragmentation, disintegration, or damage as they traverse the intense heat, radiation, and gravitational forces of their perilous journey

(Ohtsuka et al. 2003; Vokrouhlický et al. 2019).

Furthermore, the study of these celestial travelers offers invaluable insights into the very genesis of our solar system, notably with respect to the delivery of water to our terrestrial abode, Earth. Long-period comets, originating from the distant Oort cloud, traverse unique trajectories that unveil snapshots of the early narrative of our cosmic vicinity (Biermann et al. 1983; Whipple 2000; Rickman 2014). In addition, these comets serve as indispensable markers of the Sun's behavior, offering critical perspectives on solar flares and occasional encounters among themselves (Iseli et al. 2002; Jia et al. 2014). In light of their potential ramifications for Earth, the examination of interactions between these comets and the Sun stands as a matter of paramount consequence (Bzowski & Krolukowska 2005; Brown et al. 2011; Rasca et al. 2014; Hou et al. 2021; Fouchard et al. 2023).

Furthermore, scientists have ventured into the tantalizing exploration of potential connections between sungrazing comets, and the enigmatic celestial entity known as Planet X (Whitmire 2016; Batygin et al. 2019).

2. DATA AND METHODS

Commencing our study, we initiated the data collection phase by sourcing pertinent data from the SOHO webpage.

The data we acquired was selected specifically for the purpose of conducting in-depth observations and analyses of sungrazing comets.

2.1 Data Preparation

In the preparatory stages of our data analysis, we undertook various data preprocessing measures aimed at optimizing the dataset's quality and suitability for our research objectives. This process encompassed the removal of redundant or duplicate observations and comprehensive data cleansing to rectify any discrepancies or inaccuracies present in the raw data. Additionally, any potential outliers, which could introduce bias or distortion into our analyses, were meticulously identified and subsequently eliminated. Data transformation, if deemed necessary, was executed to normalize, or scale variables in alignment with the demands of our analysis.

2.1.1 Density-Based Spatial Clustering of Applications with Noise (DBSCAN) Parameters

The core of our analytical approach lies in the effective utilization of the density-based spatial clustering of applications with noise (DBSCAN) algorithm. The selection of DBSCAN parameters assumes critical importance in ensuring the precise categorization of the comet data. The epsilon value (ϵ), set at 0.15, determines the radius within which data points are considered as neighbors. When selecting the epsilon value (ϵ) for the DBSCAN algorithm, it's essential to find a balance. If the epsilon value is set too high, for instance, at 3, it would encompass a broader range, including a larger number of comets within the defined neighborhood. However, this approach also means incorporating more dissimilar parameters, potentially leading to less precise clustering results. Therefore, setting epsilon to 0.15, we strike a balance where we include enough comets while maintaining similarity among the parameters. This threshold ensures a comprehensive analysis by including enough comets while maintaining similarity among parameters. This balance is crucial for categorizing sungrazing comets effectively, capturing essential characteristics among clustered data points while minimizing outliers.

Furthermore, we established a minimum threshold of 5 comet instances as the requisite number of data points needed to form a dense region, contributing to the formation of distinct clusters. These parameters are of high significance, as they are instrumental in the identification of comet instances clustered based on their spatial density, thereby

facilitating the categorization of sungrazing comets into well-defined groupings.

2.1.2 Density-Based Spatial Clustering of Applications with Noise (DBSCAN) Overview

DBSCAN, an acronym for density-based spatial clustering of applications with noise, stands as a versatile and highly esteemed density-based clustering algorithm, renowned for its proficiency in uncovering clusters within intricate and multifaceted datasets. Unlike some other conventional clustering techniques, such as K-means, DBSCAN distinguishes itself by its unique feature: it does not necessitate a predetermined count of clusters. Furthermore, it excels in the identification of clusters characterized by varying shapes and structures. These exceptional capabilities hinge on the two pivotal parameters, namely, epsilon (ϵ) and the minimum number of data points (MinPts).

2.2 Parameters

Epsilon (ϵ): This parameter serves as a threshold, defining a radius that demarcates the ϵ -neighborhood, within which data points are considered neighbors. To merit inclusion within the same cluster, a data point must demonstrate at least MinPts data points within this ϵ -neighborhood. Essentially, ϵ establishes a proximity threshold, delineating the requisite closeness for data points to be deemed part of a shared cluster.

2.2.1 Minimum Data Points (MinPts)

The MinPts parameter stipulates the minimum count of data points necessary to constitute a dense region. A data point attains the status of a core point when its ϵ -neighborhood encompasses at least MinPts data points. Core points serve as a core point around which clusters coalesce.

2.2.2 Density-Based Spatial Clustering of Applications with Noise (DBSCAN) Process

DBSCAN systematically classifies data points into three distinct categories.

2.3 Core Points

These data points stand as the central pillars of clusters, demonstrating the presence of at least MinPts data points within their ϵ -neighborhoods.

Border Points: While not themselves core points, border

points reside within the ϵ -neighborhoods of core points, marking the periphery of clusters.

Noise Points: Noise points are data points devoid of core or border point attributes, evading association with any specific cluster. Typically, these points represent outliers within the dataset.

2.3.1 Computation of Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and Statistics for Grouping

Established categories (C-Established): These are the predefined or previously recognized categories of comets, typically based on traditional classification methods or expert knowledge.

Categories from DBSCAN (C-DBSCAN): These represent the categories or clusters identified by applying the DBSCAN algorithm to the comet data.

Jaccard similarity coefficient (J): The Jaccard similarity coefficient (J) between two sets A and B is defined as:

$$J(A, B) = \frac{A \cap B}{A \cup B} \quad (1)$$

In this context, A = C - Established and B = C - DBSCAN:

Therefore, the Jaccard similarity coefficient (J) between the established categories and the categories identified by DBSCAN is computed as:

$$J(\text{C-Established}, \text{C-DBSCAN}) = \frac{\text{C-Established} \cap \text{C-DBSCAN}}{\text{C-Established} \cup \text{C-DBSCAN}} \quad (2)$$

2.3.2 Statistical Analysis of Orbital Parameters using Jaccard Coefficient

After validating the categorization using the Jaccard coefficient, statistical analysis is performed on the orbital parameters of Kreutz comet subgroups.

This analysis provides insights into the variability of orbital parameters within each subgroup, further corroborating the clustering results. By assessing the consistency and variability of orbital parameters, the statistical analysis adds another layer of validation to the categorization process.

3. RESULTS

We have validated our comet categorization approach

based on the DBSCAN algorithm. Our validation process encompassed a meticulous comparison of our findings with previously established comet categories. This comprehensive verification involved an examination of the orbital parameter values associated with the categorized comets, assessing their alignment with the known comet categories.

- The DBSCAN method effectively demonstrates its prowess in grouping multivariable data within the sungrazing comet dataset.
- The Kreutz comet group was subdivided into four well-defined subgroups with unique orbital characteristics.
- Uniform clustering of perihelion positions across the Kreutz comet subgroups indicated their interconnected nature.
- Each subgroup (A, B, C) deviates by approximately ± 20 degrees to the left and right of the primary comet orbit.
- A minor away Group D was identified, indicative of slightly varying orbital parameters.

3.1 Heliocentric Ecliptic Mapping of All Groups

Building on the insights gained from our DBSCAN analysis and initial classifications, we embarked on an illuminating exploration of the heliocentric ecliptic map. This spatial representation provided valuable insights into following the successful application of the DBSCAN method to categorize comets into their respective groups, we initially presented Fig. 2, which provided a heliocentric ecliptic map illustrating the longitude (l) and latitude (b) of perihelion variables of comets.

The heliocentric ecliptic map presented in Fig. 3 delineates the four Kreutz subgroups distinctly, illustrating their precise locations on the map. This visual depiction enables clear differentiation between each subgroup, providing an accurate representation of their respective positions.

3.1.1 Scatter Matrix Validation

In our quest to validate and solidify the reliability of the DBSCAN method, we turned to the scatter plot matrix as a powerful tool. This visual representation, as displayed in Fig. 4, serves as a compelling means of discerning the distinctive subgroups within the Kreutz sungrazing comets, all accomplished post-DBSCAN application. The scatter plot matrix unveils the underlying structures and interrelationships between orbital parameters, offering a visual confirmation of the clustering efficacy achieved through the DBSCAN approach.

In this scatter matrix plot the ω ($^\circ$) represents the argument

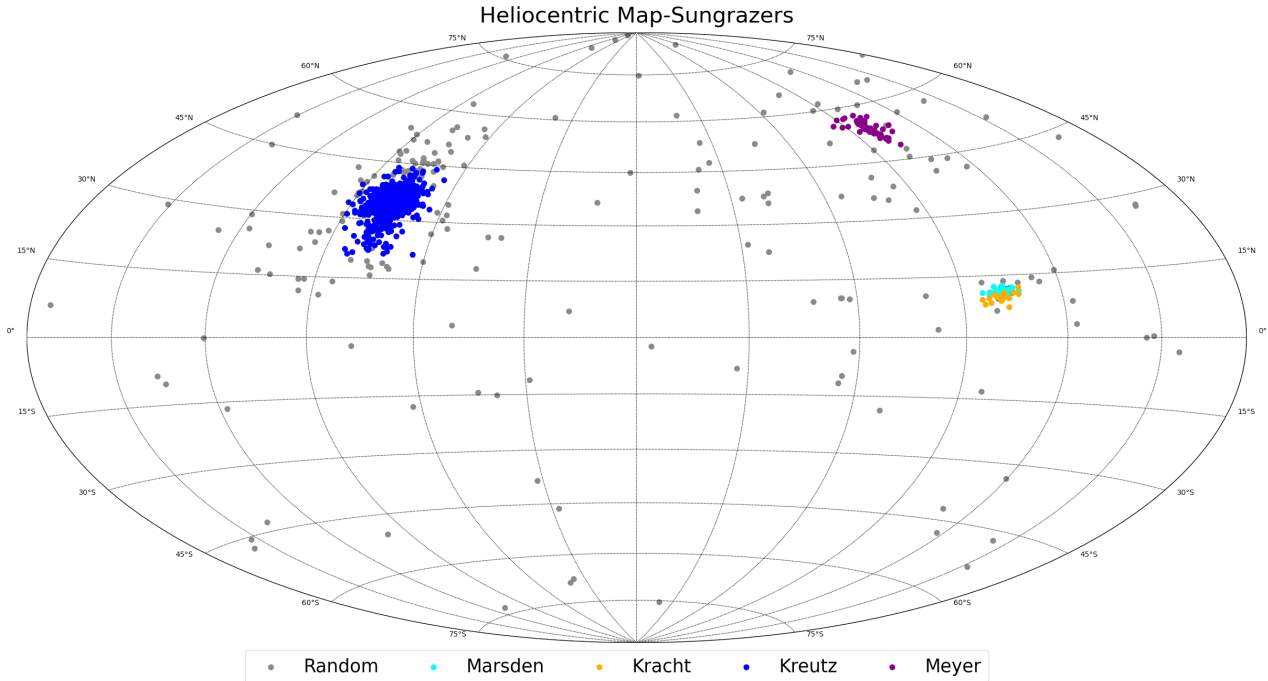


Fig. 2. Heliocentric ecliptic map for latitude and longitude of perihelion. Color code: cyan (Marsden), orange (Kracht), purple (Meyer), blue (Kreutz). The black-grey dots are random comets. The color-coded map illustrates the distribution of comet clusters, with perihelion directions represented by solid points. This map is generated using the Hammer projection in Python with data from Karimova & Yi (2023). Adapted from Karimova & Yi (2023) with permission of Springer Nature.

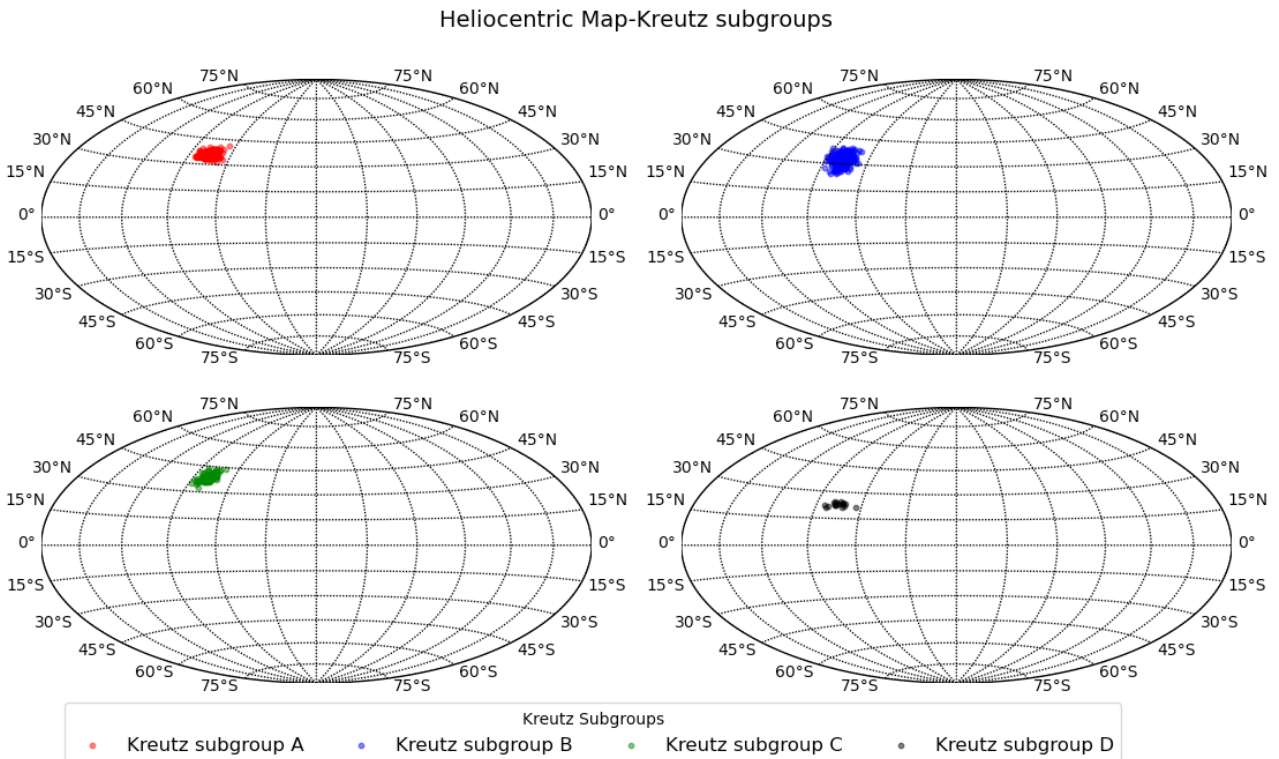


Fig. 3. Heliocentric ecliptic map for latitude and longitude of perihelion. Color code: red (Kreutz subgroup A), blue (Kreutz subgroup B), green (Kreutz subgroup C), black (Kreutz subgroup D). The color-coded map illustrates the distribution of comet clusters, with perihelion directions represented by solid points.

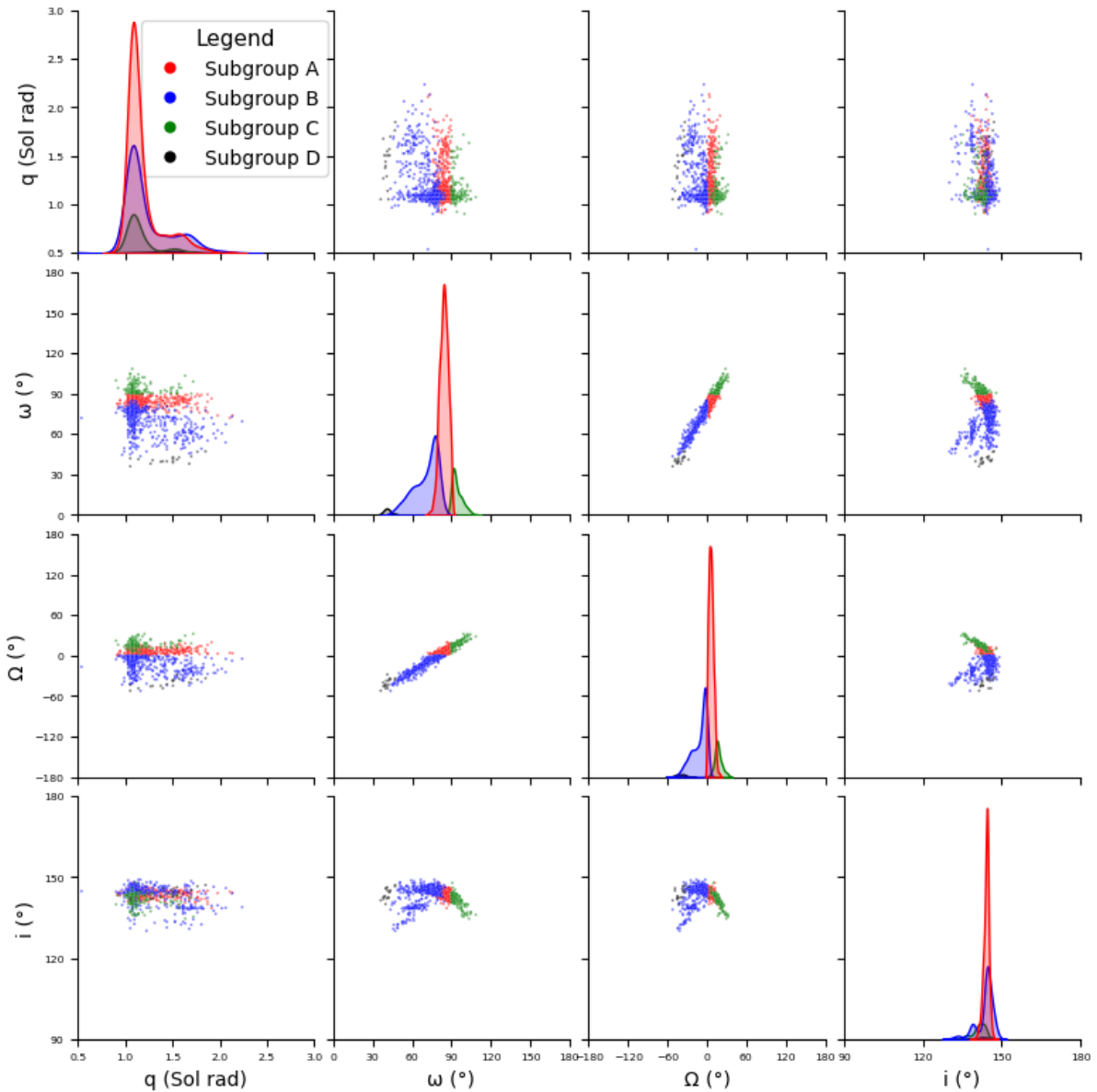


Fig. 4. Scatter matrix plot for determining the groupings of orbital parameters for four (4) Kreutz sungrazing comet subgroups - red (Kreutz subgroup A), blue (Kreutz subgroup B), green (Kreutz subgroup C), black (Kreutz subgroup D).

of perihelion, Ω ($^{\circ}$) is the longitude of the ascending node, i ($^{\circ}$) is the inclination of the orbit and the q stands for perihelion distance from the sun in astronomical units (AU).

Thus, through thorough analysis using the scatter plot matrix and mapping tool, we can confidently identify the presence of these subgroups within the Kreutz sungrazing comets. This visual confirmation further validates the precision and reliability of the DBSCAN methodology, affirming the accurate segmentation of comet data based on specific orbital characteristics. Such visual evidence significantly enhances our understanding of the dynamics

within this group.

Following the analysis of outcomes from employing the DBSCAN method and examining the Scatterplot Matrix, we calculated the mean orbital parameters for the four distinct Kreutz comet subgroups, as detailed in the subsequent table (Table 1).

3.1.2 The 3D Test of the Trajectories

The 3D visualization (Fig. 5) of the comet trajectories reveals distinct clustering among the three major comet subgroups (labeled A, B, and C), indicating their relatedness.

Table 1. The mean orbital parameters of Kreutz subgroups

Cluster	q (AU)	e	ω ($^\circ$)	Ω ($^\circ$)	i ($^\circ$)
Kreutz group (a)	0.00548	1	83.92831	5.883501	143.9945
Kreutz group (b)	0.005729	1	70.24084	348.2708	143.5961
Kreutz group (c)	0.005383	1	94.26993	17.00154	141.3661
Kreutz group (d)	0.006756	1	41.16722	319.4444	143.8644

AU, astronomical units.

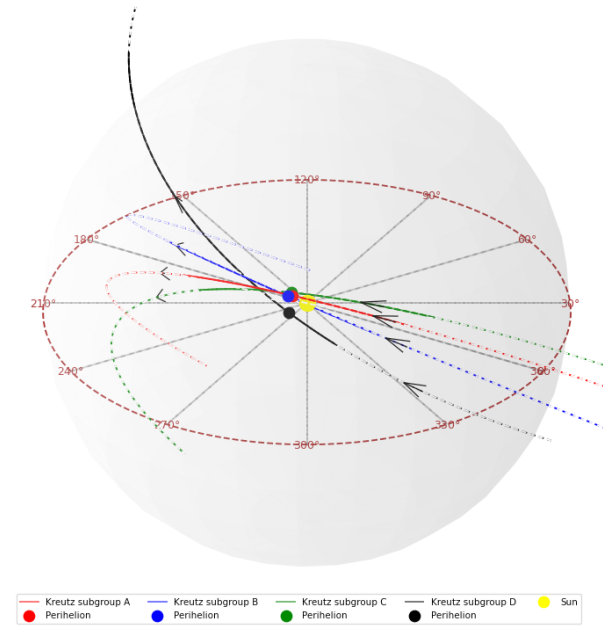


Fig. 5. 3D orbital trajectories of four Kreutz sungrazing comet groups (labeled labeled A-red, B-blue, C-green, and D-black) in a heliocentric ecliptic reference frame. Each group's trajectory is uniquely color-coded, providing a clear visual separation. Perihelion data is marked sequentially as a scatter dot corresponding to colors of trajectory. The trajectories are depicted with white dashed lines when ascending above the equator and after descending, aiding in the visualization of the comet's path.

However, the fourth subgroup (labeled D) appears slightly more isolated, suggesting potential variations in origin and destination trajectories compared to the other subgroups. This observation further corroborates the findings from the scatter matrix plot and the DBSCAN grouping methods, as the resulting trajectories align with the clusters identified through these techniques.

4. SUMMARY AND DISCUSSION

4.1 Validation and Discovery of Sungrazing Comet Groups & Subgroups

The application of the DBSCAN algorithm has unequivocally demonstrated its prowess in categorizing comets based on

orbital parameters, validating the method's precision and effectiveness in our classification efforts. In a significant breakthrough, our study utilized the DBSCAN algorithm to validate the existence of established comet groups, such as Kreutz, Meyer, Marsden, and Kracht.

4.2 In-Depth Kreutz Comet Exploration

Expanding our inquiry, we employed DBSCAN to analyze deeply the Kreutz comet group. This endeavor resulted in the subdivision of Kreutz into four distinct subgroups (A, B, C, D), each characterized by unique orbital signatures. Through this process, further reinforced by scatter plot matrix analysis, we underscored DBSCAN's capacity to partition complex celestial data into meaningful segments. A systematic observation within this section revealed that all Kreutz sungrazing comets consistently traverse their orbits in a clockwise direction, a phenomenon meriting further examination.

4.3 Orbital Variability within Kreutz Subgroups

Our investigation into the trajectories of the Kreutz comet group, utilizing advanced clustering methodologies, reveals noteworthy patterns. Subgroups A, B, and C showcase a high degree of orbital consistency, whereas subgroup D exhibits a discernible yet subtle deviation in its mean trajectory, particularly evident in the three-dimensional representations. This nuanced distinction in 3D orbital trajectories emphasizes the probable influence of argument of perihelion and perihelion distance variation within subgroup D, contributing to its unique orbital characteristics compared to the more uniform patterns observed in A, B, and C.

4.4 Destruction Dynamics of Kreutz Comet Group

Our investigation unveils a unique destruction pattern in the Kreutz comet group, occurring remarkably close to the Sun at limits of 0.8 to 1.1 solar radii, sometimes managing to travel further than that. Unlike conventional celestial bodies, their destruction process extends beyond established limits, necessitating further inquiry into the specific factors governing their resilience. The exceptional nature of the Kreutz comets, potentially could have been influenced by material composition or travel speed, underscores the need for focused research to unravel the intricate dynamics of cometary destruction in close proximity to the Sun.

ACKNOWLEDGMENTS

This work was supported by a research fund of Chungnam National University. The SOHO/LASCO data used here are produced by a consortium of the Naval Research Laboratory (Southwest, Washington DC, USA), Max-Planck-Institut für Sonnensystemforschung (MPS) (Göttingen, Germany), Laboratoire d'Astrophysique Marseille (LAM) (Marseille, France), and the University of Birmingham (UK). SOHO is a project of international cooperation between ESA and NASA.

ORCIDs

Ulkar Karimova <https://orcid.org/0000-0003-3396-2815>
Yu Yi <https://orcid.org/0000-0001-9348-454X>

REFERENCES

- Batygin K, Adams FC, Brown ME, Becker JC, The planet nine hypothesis, *Phys. Rep.* 805, 1-53 (2019). <https://doi.org/10.1016/j.physrep.2019.01.009>
- Biermann L, Huebner WF, Lüst R, Aphelion clustering of “new” comets: Star tracks through Oort’s cloud, *Proc. Natl. Acad. Sci. USA.* 80, 5151-5155 (1983). <https://doi.org/10.1073/pnas.80.16.5151>
- Brown JC, Potts HE, Porter LJ, Le Chat G, Mass loss, destruction and detection of sun-grazing and -impacting cometary nuclei, *Astron. Astrophys.* 535, A71 (2011). <https://doi.org/10.1051/0004-6361/201015660>
- Bzowski M, Krolikowska M, Sungrazing comets as source of pickup ions at Earth orbit and Ulysses, *Proceedings of the Solar Wind 11 / SOHO 16, Connecting Sun and Heliosphere Conference (ESA SP-592)*, Whistler, Canada, 12-17 Jun 2005.
- England KJ, Early sungrazer comets, *J. Br. Astron. Assoc.* 112, 13-28 (2002).
- Fouchard M, Higuchi A, Ito T, What long-period comets tell us about the Oort Cloud, *Astron. Astrophys.* 676, A104 (2023). <https://doi.org/10.1051/0004-6361/202243728>
- Hasegawa I, Nakano S, Possible Kreutz sungrazing comets found in historical records, *Publ. Astron. Soc. Jpn.* 53, 931-950 (2001). <https://doi.org/10.1093/pasj/53.5.931>
- Hou CP, He JS, Zhang L, Wang Y, Duan D, Dynamics of the charged particles released from a sun-grazing comet in the solar corona, *Earth Planet. Phys.* 5, 232-238 (2021). <https://doi.org/10.26464/epp2021023>
- Iseli M, Küppers M, Benz W, Bochsler P, Sungrazing comets: properties of nuclei and *in situ* detectability of cometary ions at 1 AU, *Icarus.* 155, 350-364 (2002). <https://doi.org/10.1006/icar.2001.6722>
- Jia YD, Russell CT, Liu W, Shou YS, Multi-fluid model of a sun-grazing comet in the rapidly ionizing, magnetized low corona, *Astrophys. J.* 796, 42 (2014). <https://doi.org/10.1088/0004-637X/796/1/42>
- Jones GH, Knight MM, Battams K, Boice DC, Brown J, et al., The science of sungrazers, sunskirters, and other near-Sun comets, *Space Sci. Rev.* 214, 20 (2018). <https://doi.org/10.1007/s11214-017-0446-5>
- Karimova U, Yi Y, SOHO Sungrazing comet groups classified by the scatterplot matrix, *J. Korean Phys. Soc.* 83, 733-740 (2023). <https://doi.org/10.1007/s40042-023-00935-2>
- Knight MM, A'Hearn MF, Biesecker DA, Faury G, Hamilton DP, et al., Photometric study of the Kreutz comets observed by SOHO from 1996 to 2005, *Astron. J.* 139, 926 (2010). <https://doi.org/10.1088/0004-6256/139/3/926>
- Lee SE, Yi Y, Kim YH, Brandt JC, Distribution of perihelia for SOHO sungrazing comets and the prospective groups, *J. Astron. Space Sci.* 24, 227-234 (2007). <https://doi.org/10.5140/JASS.2007.24.3.227>
- Marsden BG, Sungrazing comets, *Annu. Rev. Astron. Astrophys.* 43, 75-102 (2005). <https://doi.org/10.1146/annurev.astro.43.072103.150554>
- Marsden BG, The sungrazing comet group, *Astron. J.* 72, 1170 (1967). <https://doi.org/10.1086/110396>
- Ohtsuka K, Nakano S, Yoshikawa M, On the association among periodic comet 96P/Machholz, Arietids, the Marsden comet group, and the Kracht comet group, *Publ. Astron. Soc. Jpn.* 55, 321-324 (2003). <https://doi.org/10.1093/pasj/55.1.321>
- Rasca AP, Oran R, Horányi M, Mass loading of the solar wind by a sungrazing comet, *Geophys. Res. Lett.* 41, 5376-5381 (2014). <https://doi.org/10.1002/2014GL060990>
- Rickman H, The Oort Cloud and long-period comets, *Meteorit. Planet. Sci.* 49, 8-20 (2014). <https://doi.org/10.1111/maps.12080>
- Sekanina Z, Kracht R, Population of SOHO/STEREO Kreutz sungrazers and the arrival of comet C/2011 W3 (Lovejoy), *Astrophys. J.* 778, 24 (2013). <https://doi.org/10.1088/0004-637X/778/1/24>
- SOHO, SOHO observes 200TH comet (2020) [Internet], viewed 2023 Jul 20, available from: https://soho.nascom.nasa.gov/hotshots/2000_08_21/
- Vokrouhlický D, Nesvorný D, Dones L, Origin and evolution of long-period comets, *Astron. J.* 157, 181 (2019). <https://doi.org/10.3847/1538-3881/ab13aa>
- Whipple FL, Oort-Cloud and Kuiper-Belt comets, *Planet. Space Sci.* 48, 1011-1019 (2000). [https://doi.org/10.1016/S0032-0633\(00\)00070-2](https://doi.org/10.1016/S0032-0633(00)00070-2)
- Whitmire D, Periodic mass extinctions and the Planet X model reconsidered, *Mon. Not. R. Astron. Soc. Lett.* 455, L114-L117 (2016). <https://doi.org/10.1093/mnrasl/slv157>