10-2-2023

Using Machine Learning to Predict Hypervelocity Fragment Propagation of Space Debris Collisions

Katharine Larsen

Riccardo Bevilacqua

IAC-23-A6.2.3

Using Machine Learning to Predict Hypervelocity Fragment Propagation of Space Debris Collisions

Katharine E. Larsen[a]*, Riccardo Bevilacqua[b]

[a] *Department of Aerospace Engineering, Embry-Riddle Aeronautical University, 1 Aerospace Blvd, Daytona Beach, Florida, USA 32114*, larsenk2@my.erau.edu
[b] *Department of Aerospace Engineering, Embry-Riddle Aeronautical University, 1 Aerospace Blvd, Daytona Beach, Florida, USA 32114*, bevilacr@erau.edu
* Corresponding Author

Abstract

The future of spaceflight is threatened by the increasing amount of space debris, especially in the near-Earth environment. To continue operations, accurate characterization of hypervelocity fragment propagation following collisions and explosions is imperative. While large debris particles can be tracked by current methods, small particles are often missed. This paper presents a method to estimate fragment fly-out properties, such as fragment, velocity, and mass distributions, using machine learning. Previous work was performed on terrestrial data and associated simulations representing space debris collisions. The fragmentation of high-velocity fragmentation can be modeled by terrestrial fragmentation tests, such as static detonations. Recently, stereoscopic imaging techniques have become an addition to static arena testing. Collecting data with this method provides position vector and mass information faster and more accurately than previous manual-collection methods. Additionally, there is limited space debris data of similar accuracy on individual fragments. Therefore, this imaging technique was used as the primary collection method for the previous research data. Now, two-line element (TLE) sets for Iridium 33 are also used. Machine learning methodologies are leveraged to predict fragmentation fly-out from the collision event with Cosmos 2251. First, gaussian mixture models (GMMs) are used to model the probability distribution of the particles for a given desired characteristic at Julian dates following the event. Once this training data is generated, regression techniques can be used to predict these characteristics. K-nearest neighbor (K-NN) regressors are used to estimate the spatial distribution of the satellite fragments. Monte Carlo simulations are then used to validate the results, finding that this technique accurately estimates the total number of fragments expected to intersect a region of interest at a given time. Following this work, the same technique can be used to estimate the velocity and mass distributions of the debris. This information can then be used to estimate the kinetic energy of the particle and classify it to avoid future debris collisions.
Keywords: fragmentation, space debris, machine learning

1. Introduction

For the future of space operations, accurate characterization of hypervelocity fragment propagation following satellite break-up, including collisions and explosions, is imperative. Debris fragments are often tracked by organizations, such as the United States Space Surveillance Network, using ground-based radars. However, the data is often sparsely collected as collection is limited to location, maximum altitude, and debris size. Therefore, proper characterization of debris particles through modeling is still necessary to avoid collisions and debris creation in the future.

One of the most prominent existing models, widely used as a reference for others, is NASA's Standard Satellite Breakup Model (SSBM), or Standard Breakup Model (SBM), based on both ground-based impact experiments and data from one on-orbit collision [1]. Models following, such as NASA's EVOLVE 4.0, a one-dimensional, low earth orbit (LEO) model, implement this model [2]. NASA's LEO to GEO (geostationary orbit) Environmental Debris model (LEGEND) improves upon these previous EVOLVE models by expanding information into three-dimensional GEO orbits [1,3,4].

All current models must be updated regularly to consider new debris and environmental information. For example, DebriSat, a ground-based experiment used to model catastrophic breakup of payloads in the LEO environment, has been used to update the NASA's SBM with more current satellite qualities such as structure, material, and technology [5,6].

Currently, general perturbation element sets, distributed by the North American Aerospace Defense Command (NORAD), for artificial satellite and debris data can be accessed from open-source satellite catalogs as two-line element (TLE) sets. Common resources include Space-Track.org and CelesTrak [7,8]. TLE sets are one of the current methods used to maintain satellite and debris data. As they are collected as general perturbation elements, the proper model must be selected to propagate the orbital elements as position and velocity

[9,10]. The accuracy for propagated TLE sets should be evaluated thoroughly as it varies, partially due to the sparse updates. However, different methods can be used to appropriately propagate TLE sets over a longer time range.

1.1 Previous Work

Formed in the 1980's, the Department of Defense's (DoD) Orbital Debris Program has also conducted terrestrial tests, including warhead static arena testing, to model and characterize satellite fragmentation [11]. Similarly, a previous paper [12] has been presented using terrestrial data produced from a static arena method but including more modern camera systems to produce stereoscopic imaging.

Using this static arena test data from the U.S. Naval Air Warfare Center's Weapons Division (NAWCWD) at China Lake, and associated simulation data used to expand the training dataset, also provided by NAWCWD, multivariate gaussian mixture models (GMMs) and expectation maximization (EM) were fit to intersection points on a polar-azimuth plane at various radii of intersection. Each GMM represented the distribution for a given characteristic and made up the training dataset. Then, K-nearest neighbors regression (K-NN) was used to estimate the spatial, velocity, and mass distributions of the fragments.

This presented paper aims to expand these proposed ideas, performed on terrestrial structures, by applying them to more realistic satellite debris data.

1.2 Objectives

Previously, the higher accuracy in terrestrial fragmentation data, in addition to the uniform shape and mass of each object used in the static arena test, made it desirable to test the proposed method.

Now, this paper presents a similar technique on a system with more realistic orbital object behavior. Real debris data, in this case Iridium 33 debris information, following the 2009 collision with Cosmos 2251, is used to test this method and produce spatial distributions from the collision over time, rather than over various intersection radii from an initial distance.

2. Material and methods

The first TLE set data recorded after the collision for the Iridium 33 satellite debris was retrieved from CelesTrak [8]. Only data first recorded within a year after the event was considered to maintain accuracy. Before using this as training data, periodic variations, i.e., oblateness perturbation, drag, and gravitational effects, are reconstructed using the SGP4 propagator [8,9,10]. The data was propagated from the time of impact, estimated to be approximately February 10, 2009, 16:56 UTC, for 5 minutes at 1 second timesteps.

Then, the propagated data is transformed from the Earth-Centered, Earth-Fixed (ECEF) reference frame to a coordinate system centered around Iridium 33's expected ECEF coordinate at each timestep if the event had not occurred. This provides the user with the distance the debris data has travelled from its expected orbit. These cartesian coordinates are then transformed into longitude and latitude coordinates before fitting a GMM with 8 components to represent the debris distribution or spread at each timestep. The successful GMM fits are used as the training data.

The general machine learning procedure used to estimate the number of fragments within a certain latitude-longitude section follows the previous work referenced in the introduction, which can be referenced in the schematic (see Fig. 1) [12]. However, the inputs are no longer the terminal state and radius, but instead the Julian date.

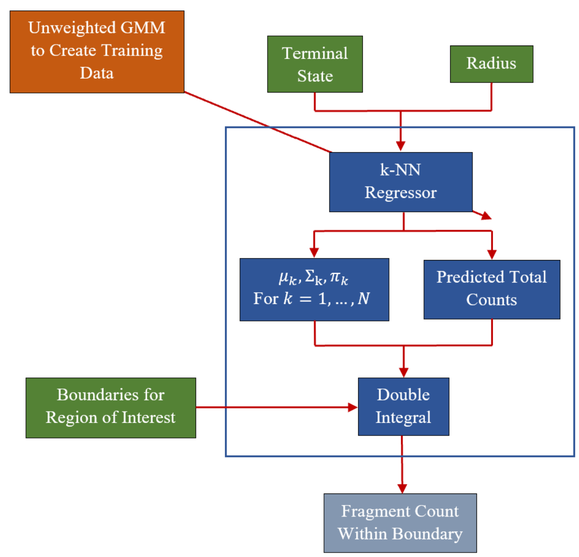

Fig. 1. Schematic of the methodology used to predict the number of debris objects within a given area.

MATLAB was used for the training data preparation, while python and the Scikit-learn toolbox were used for the regression model [13].

3. Theory and calculation

The following subsections discuss the theory behind the methodology used in this work.

3.1 Gaussian Mixture Models

After collecting longitude-latitude information at each timestep, probability distributions could be used to create training data, representing the spatial distribution of the debris fragments. Gaussian mixture models (GMMs) are the distribution method of choice as they can model more complex information than single gaussian distributions alone, as they are weighted linear combinations of individual gaussian distributions. GMMs have three defining parameters, which can be adjusted to tune the model: the mean, μ, represents the location of each mode of the distribution, or center, the covariance, Σ , represents the spread of the data distribution, and mixing coefficients, π_i . In the multivariate case the GMM is the probability distribution defined as:

$$p(\vec{X}) = \sum_{i=1}^{N} \pi_i \mathcal{N}(\vec{X}|\vec{\mu}_i, \Sigma_i) \quad (1)$$

where N is the number of components, $\sum_{i=1}^{K} \pi_i = 1$ where π_i are the mixing coefficients, and $\mathcal{N}(\vec{x}|\vec{\mu}_i, \Sigma_i)$ is multivariate distribution defined as:

$$\mathcal{N}(\vec{X}|\vec{\mu}, \Sigma) = \frac{1}{\sqrt{(2\pi)^{dim}|\Sigma|}} exp\left(-\frac{1}{2}(\vec{X}-\vec{\mu})^T \Sigma^{-1}(\vec{X}-\vec{\mu})\right) \quad (2)$$

where $\vec{X}$ is the random variable vector, Σ is a symmetric covariance matrix, μ is a vector containing the means, and dim is the dimension of the dataset [14,15]. In this work a full covariance with 8 components is utilized.

3.2 K-Nearest Neighbors Regression

Regression methods are used to create a function to describe an input-output relationship. K-nearest neighbors regression (K-NN) uses an average of the closest observations in a local neighborhood with a specified number of neighbors to produce a model. K-NN computes the mean as:

$$f_{knn}(x_q) = \begin{cases} \dfrac{\sum_{i=1}^{k} w_i f(x_i))}{\sum_{i=1}^{k} w_i}, & D(x_q, x_i) \neq 0 \\ f(x_1), & D(x_q, x_i) = 0 \end{cases} \quad (3)$$

where x_q is the query point, k is the size of the neighborhood, x_i is a close point, or a neighbor, and w is the weight defined as:

$$w_i = \frac{1}{D(x_q, x_i)} \quad (4)$$

[16]. In the case that the minimum distance between the pair of objects is zero, $D(x_q, x_i) = 0$, the weight is undefined. Hence, the corresponding output training data of the single closest point is used, instead of taking an average. The distance metric used for this method is Euclidean distance, measured as the physical length of a segment between points, and the size of the neighborhood, k , is a chosen integer, and must be adjusted to avoid overfitting and underfitting.

3.3 Boundaries of Interest

After estimating the spatial distribution, the number of debris fragments at each timestep within a certain area of the plane can be calculated using double integrals defined:

$$Number\ of\ Debris = N_{total} \iint_S p_N(x)\, dS \quad (5)$$

where N_{total} is total estimated number of debris at the timestep and dS is described as

$$dS = d\phi d\theta \quad (6)$$

where ϕ is the latitude angle and θ is the longitude angle [12].

4. Results and Discussion

To visually assess how well the GMM fit, the probability density function was plotted over the data points. The resulting GMM parameters were then used to produce 493 randomized points, representing the number of debris fragments considered. This was compared to the actual debris locations (Fig. 2). The qualitative result shows that, for this timestep, the GMM fits well to the data, even though the contours do not visually include all the points.

When selecting the regression model, K-nearest neighbors regression with $k = 2$ yields the best results. Other values of k, 2 to 10, were considered before coming to this conclusion, with $k = 2$ yielding the only silhouette score above 0, indicating that all other values of k, or number of clusters, poorly fit the data. This is most likely a result of the limited amount of data points. Trials with random forest regression were also performed.

To validate the model, as done in [12], Monte Carlo simulations were used. For 15 timesteps outside of the training dataset, 20 different randomized regions of the plane were generated. The resulting mean squared error (MSE) and standard deviation (SD) of the model estimated number of debris compared to the actual model values are shown in Table 1.

Table 1. Mean squared error and standard deviation of the fragment count estimations.

MSE, Debris Fragments	SD, Debris Fragments
1.651	2.537

For 4 different timesteps, of the 15 tested, the resulting GMM fit and predicted GMM fit for one of the randomized areas of interest, shown as the black rectangle, are displayed in Fig. 3-Fig. 5.

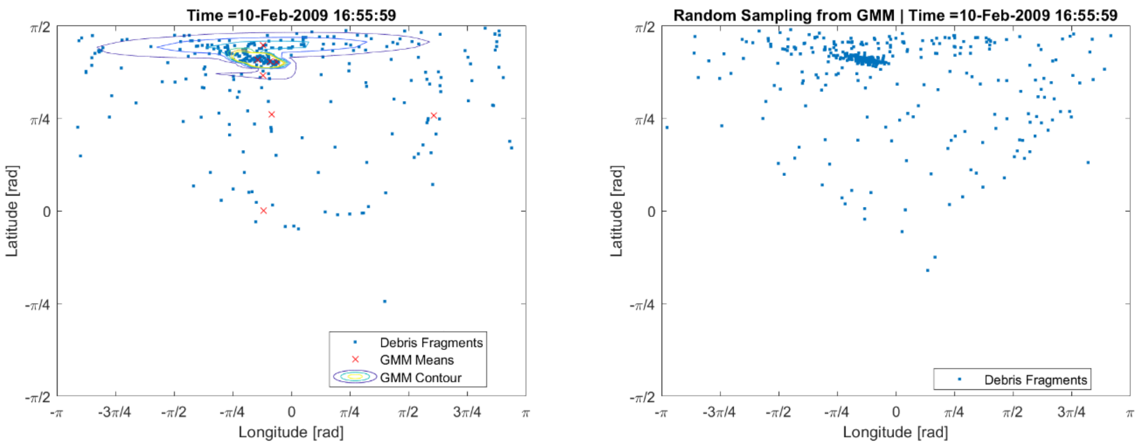

Fig. 2. The GMM fit to the first timestep (left) and the resulting random sampling produced by the GMM parameters.

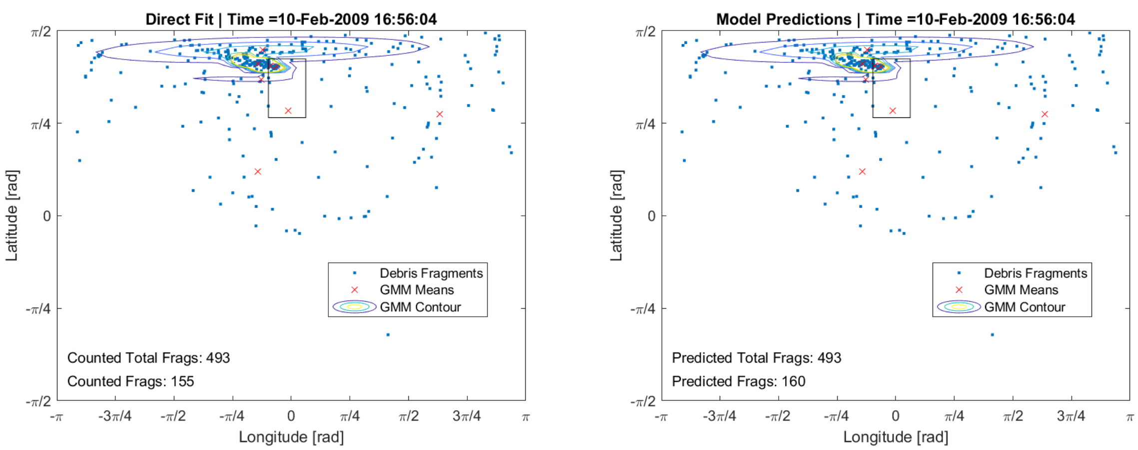

Fig. 3. Example of the Monte Carlo simulation for a testing timestep.

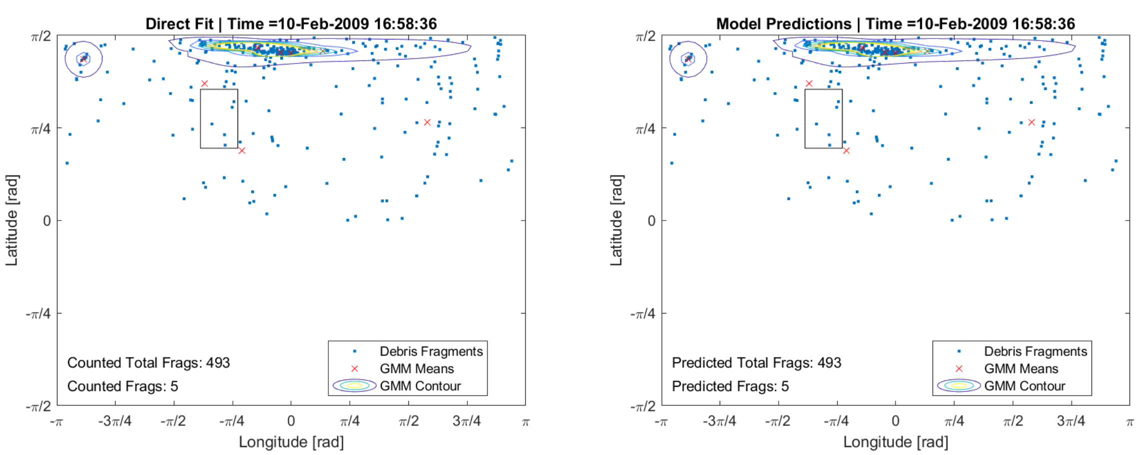

Fig. 4. Example of the Monte Carlo simulation for a testing timestep.

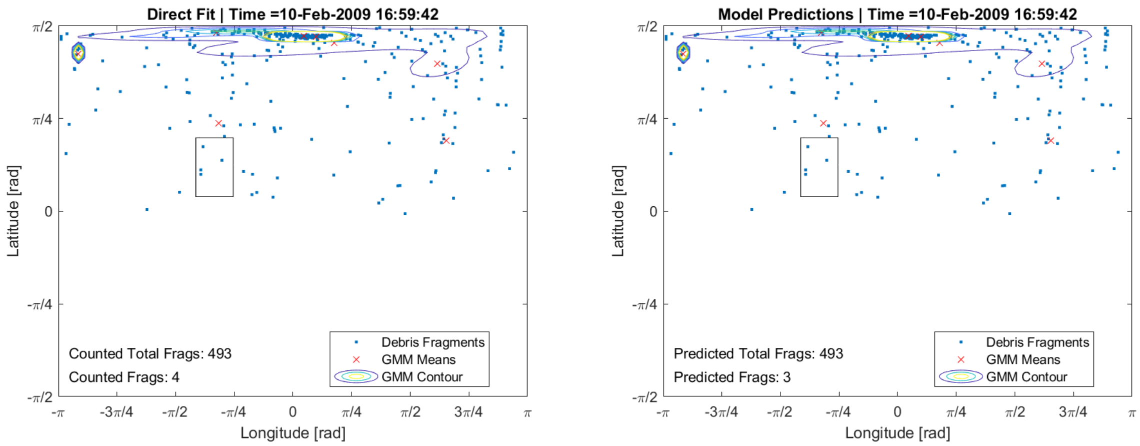

Fig. 5. Example of the Monte Carlo simulation for a testing timestep.

Visually, the predicted GMM seems to fit close to the directly fit GMM. However, if more debris objects can be considered for the GMM in the future, it may fit better to the data. Additionally, though the mean squared error is small, when sections with smaller counts are considered, slight differences in the GMM parameters are more significant.

5. Conclusions

These results prove that this method can be applicable to space debris estimations. After propagating TLE sets for the Iridium 33 debris, training data was generated using GMMs. Training K-NN with $k = 2$ was found to produce the most accurate spatial distribution results. Using these distributions, the total values within a chosen boundary of interest were calculated. The resulting distributions were reviewed using Monte Carlo simulations.

While the model still requires improvements in data accuracy and amount, it proves to make successful predictions, displaying the potential in using this methodology to predict debris characteristics. In the future, this work should be applied to predict the velocity and mass distributions of satellite debris.

However, data accuracy will need to be assessed and further work should be performed with TLE's to make the data more reliable. Additionally, to continue work with space debris, more data will need to be considered for proper predictions. With limited information collected at the time of the event, as well as a limited number of debris systems, other sources, such as simulations, should be considered to expand the training dataset.

This data is also only useful for this collision event specifically. Break-up characteristics, such as orientation, size, material, and method of break-up, will need to be considered before this method can be generalized. With future work, this method will become more generalizable and particles currently too small to track may be characterized.

Acknowledgements

Previous data was provided by the U.S. Naval Air Warfare Center Weapons Division at China Lake and high-fidelity simulations were also performed by the U.S. Naval Air Warfare Center Weapons Division. Current data is retrieved from Celestrak [8]. Investigations are supported by the U.S. Air Force Office of Scientific Research (award number FA9550-20-1-0200). K. E. Larsen also thanks the SMART Scholarship Program and the Intuitive Machine and Columbia Sportswear Advancing Women in Technology Fellowship for financial support in the pursuit of her degree.

References

[1] J.C. Liou, Orbital Debris Modeling, NASA Orbital Debris Program Office, (2012).
[2] N.L. Johnson, P.H. Krisko, J.-C. Liou, P.D. Anz-Meador, NASA's New Breakup Model of Evolve 4.0, Advances in Space Research, (2001) 1377–1384.
[3] P.H. Krisko, J.-C. Liou, NASA Long-Term Orbital Debris Modeling Comparison: Legend and Evolve, IAC-03-IAA.5.2.03, 54th International Astronautical Congress, Bremen, Germany, 2003, 29 September to 3 October.
[4] J.-C. Liou, N.L. Johnson, Instability of the Present LEO Satellite Populations. Advances in Space Research. 7, (2008) 1046–1053.
[5] DebriSat. Astromaterials Research & Exploration Science: NASA Orbital Debris Program Office, NASA, https://www.orbitaldebris.jsc.nasa.gov/measurements/debrisat.html.
[6] M. Rivero, B. Shiotani, M. Carrasquilla, N. Fitz-Coy, J.-C. Liou, M. Sorge, T. Huynh, J. Opiela, P. Krisko, H. Cowardin, DebriSat Fragment Characterization System and Processing Status, IAC-16-A6.2.8, 67th International Astronautical Congress, Guadalajara, Mexico, 2016, 26-30 September.
[7] Space-Track. https://www.space-track.org.

[8] T.S. Kelso, CelesTrak, https://celestrak.org.
[9] F.R. Hoots, R.L. Roehrich, SPACETRACK REPORT NO. 3 Models for Propagation of NORAD Element Sets, 1988.
[10] D.A. Vallado, P. Crawford, R. Hujsak, T.S. Kelso, Revisiting Spacetrack Report #3, AIAA/AAS Astrodynamics Specialist Conference, Keystone, Colorado, 2006, 21-24 August.
[11] J.C. Connell, W. Tedeschi, D. Jones, Examples of Technology Transfer from the SDIO Kinetic Energy Weapon Lethality Program to Orbital Debris Modeling, (1991).
[12] K.E. Larsen, R. Bevilacqua, O.S. Mulekar, E.L. Jerome, T.J. Hatch-Aguilar, Predicting Dynamic Fragmentation Characteristics from High-Impact Energy Events Utilizing Terrestrial Static Arena Test Data and Machine Learning, Acta Astronautica, (2023) 67-81.
[13] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, et al., Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research. 12 (2011) 2825–2830.
[14] P. Eslambolchilar, A. Komninos, M. Dunlop, Machine Learning Basics, in: Intelligent Computing for Interactive System Design: Statistics, Digital Signal Processing, and Machine Learning in Practice, Association for computing machinery, New York, 2021, pp. 143–193.
[15] J. McGonagle, G. Pilling, A. Dobre, V. Tembo, A. Kurmukov, A. Chumbley, et al., Gaussian Mixture Model, Brilliant Math & Science Wiki, https://brilliant.org/wiki/gaussian-mixture-model/.
[16] O. Kramer, Unsupervised K-Nearest Neighbor Regression, (2011).