

Forecasting a Complete Picture of Future Launches in Terms of Orbital Region and Physical Parameters

Wiebke RETAGNE^{1,a*} and Camilla COLOMBO^{1,b}

¹Politecnico di Milano, Department of Aerospace Science and Technology, Via La Masa, 34, 20156 Milano, Italy

^awiebke.retagne@polimi.it, ^bcamilla.colombo@polimi.it

Keywords: Launch Traffic, Launch Trends, Space Debris Evolution, Machine Learning, Historical Sampling

Abstract This paper provides a step-by-step guide to go from a given future number of launches to a complete picture in terms of orbital region and physical parameters. This paper only describes non-constellation launches. For these non-constellation launches distributions are built using historical data from DISCOS. Here, the orbital parameter distribution is built using Gaussian-Mixture Models. The object class is determined from a simple probability distribution. A special focus is given to the mass of the objects, since this is an important quantity to be considered when talking about the evolution of the space environment. For the mass, an XGBoost algorithm is employed, which uses the other object characteristics as an input. Finally, the area of the object is determined based on the mass.

Introduction

Forecasting the future launch traffic is not only a matter of determining the number of launches per year. To properly assess the effect of these launches on the space debris environment their orbital and physical parameters play a crucial role. Depending on these the orbital lifetime, collision risks and breakup probability differ drastically. This paper aims to develop a complete method to go from a number of launches to building a full picture of each single launch. This includes the distribution in the orbital parameters of the semi-major axis (sma), eccentricity (ecc) and inclination (inc) and the mass and area of each object, as well as their object type (Rocket Body, Payload, Mission-Related-Object (MRO)). The challenge here is that these parameters are all interconnected. Different orbital regions are used by different types of spacecrafts. This influences their area and mass. An earth observation satellite has significantly different characteristics than a GNSS satellite. Hence, the developed method must take this dependency of the orbital region into account. This paper only describes non-constellation launches. For the constellations information about the or-bit and physical characteristics are publicly available. Therefore, here it does not need to be determined statistically, but can be directly taken from the available information. This paper will give a step-by-step overview of the process at arriving at a complete picture of future non-constellation launches, starting from a given number of launches per year. To determine the distributions for each of the mentioned categories, the historical data coming from the ESA database DISCOS [1] is used. The available data is pre-processed by filtering out the debris objects. The data is then used to either create a distribution directly or as training data for models.

Object Class

The object class determines whether an object is a Rocket Body, Payload or Mission Related Object (MRO). This classification already exists in the historical data in DISCOS. Furthermore, depending on this category, parameters such as the explosion probability vary [2]. Therefore,



the division of the launches into these categories will have a definite impact on space debris evolution.

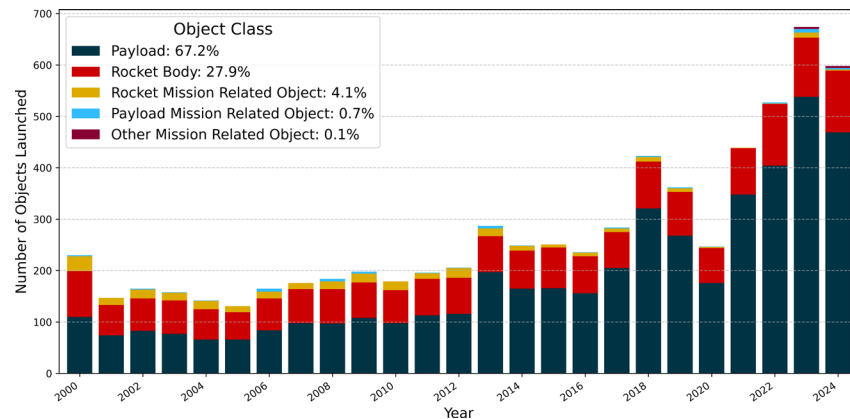


Figure 1: Number of non-constellation launches divided by object class and their total percentages since 2000.

To capture the recent trends in terms of object class distribution the historical data from 2000 to 2024 is considered. As seen in Fig. 1 the majority of launches is determined by Payload objects and Rocket bodies. The total percentages of the object categories in the years between 2000 and 2024 are used to create a categorical distribution. This categorical distribution is then randomly sampled.

Orbital Parameters

The orbital parameters are an important quantity when discussing the evolution of space debris. If the future launches are all concentrated in one orbital region the probability of collisions in that region increase greatly. The orbital parameters of interest are the semi-major axis [km], the eccentricity [-] and the inclination [°]. To capture the recent trend towards the Low Earth Orbit (LEO), the historical data from 2000 to 2024 is used to create a distribution. The distribution is created by using Gaussian Mixture Models (GMMs). A Gaussian Mixture Model is a probabilistic model, which represents data as a combination of several Gaussian distributions. A more detailed explanation can be found in [3]. It is well suited for creating orbital distributions, since the orbital regions naturally cluster around different combinations of sma, ecc and inc. This method for determining future orbital regions was first presented in Wilson et. al. [4]. To ensure that only Earth related missions are considered, large orbital radius with $sma \geq 60,000$ km are filtered out. Before using the historical data to create the GMM clusters a few transformations are applied. These transformations are to make sure that the data samples stay in a realistic range. The eccentricity should always be between 0 and 1. The inclination between 0° and 180°. In the following the data pre-processing steps are described:

- Log transform the semi-major axis. This helps the model handle large numerical ranges.
- Logit transform the eccentricity to avoid infinities when approaching the bounds of 0 or 1.
- Represent the inclination in terms of sinus and cosinus. This ensures that angles that represent the same orientation such as 0° and 360° are treated in the same manner.
- Scale the entire dataset using the standard scaler which standardizes features by re-moving the mean and scaling to unit variance.

These transformations are only done to ensure a stable fit to the historical data. Before assigning each launch to its orbital parameters they are inverse transformed, so that the final launch data is again represented in the correct units. To determine the number of clusters and therefore the number of orbital regions which best fits the distribution in the historical data, the data from 2000 to 2025 is split into a test and train dataset. Different numbers of components are compared using the Bayesian Information Criterion (BIC) [5]. The best number of components was determined to be 8 orbital regions, with the characteristics represented in Tab. 1. The orbital regions span from Low Earth Orbit (LEO), Sun Synchronous Orbit (SSO) to Medium Earth Orbit (MEO), Highly Eccentric Orbit (HEO), Geostationary Orbit (GEO) and the Geostationary Transfer Orbit (GTO).

Table 1: Cluster orbital characteristics by region.

Region	n	sma_mean [km]	ecc_mean [-]	inc_mean [°]
LEO (ISS orbit)	798	6762.8	0.0013	51.65
LEO (Mid inclined)	760	6908.8	0.0064	50.72
LEO (SSO)	2060	6951.2	0.0018	97.71
LEO (Polar)	367	7232.6	0.0133	90.19
MEO (Elliptical)	254	19568.5	0.2405	20.38
MEO/HEO (Inclined)	292	24224.7	0.3685	57.01
GTO	350	25099.1	0.7291	17.34
MEO	268	30777.1	0.1162	0.94
GEO	474	42163.4	0.0003	0.02

LEO is a highly populated region. GMMs tend to split dense clusters into multiple components, which is why the LEO clusters vary only slightly in their characteristics. For the future launches for each year the orbital distribution is sampled from these GMMs.

Distribution of Mass and Area

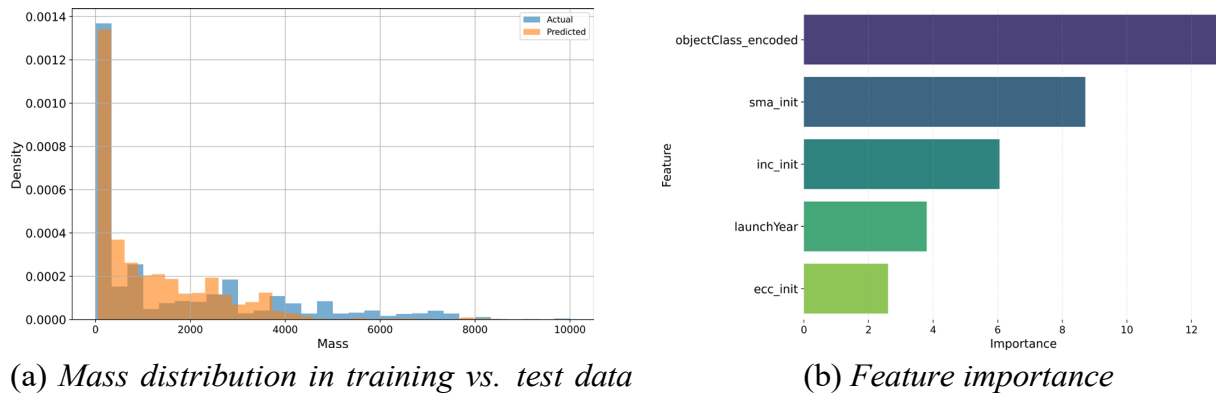
The mass is determined by using an XGBoost algorithm [6]. This is a machine learning algorithm which combines decision trees with gradient boosting. As an input the parameters determined so far are used: the object class, orbital region and type and the launch year. The classification of a satellite based on mass depends on the entity or agency. An overview can be found in A.S. Botelho et. al. [7]. For this work, a very simple classification is adapted. Depending on the mass, the average cross-section area (XSectAvg) also changes. The cross-section area intervals were determined based on the historical data available.

Table 2: Satellite classification by mass and the historical intervals for the average cross-section area.

Classification	Mass [kg]	XSectAvg [m ²]
Nanosatellites	≤10	[0.05, 1.83]
Microsatellites	10-100	[0.59, 13.93]
Minisatellites	100-500	[2.93, 22.10]
Small satellites	500-1000	[9.19, 28.8]
Large satellites	≥1000	[61.40, 1583.48]

When investigating the mass distribution of past launched satellites, it is clear that lower mass satellites are more frequent. In the historical data the biggest recorded mass is from the Discovery (STS-116) space shuttle mission [8]. This mission was launched in 2006 and had a mass of 132931 kg. Just 12 days after launch the shuttle returned to earth. This demonstrates

that the higher mass components in the launch data are more related to human space missions and not satellites that are representative for the space debris evolution. An example of a bigger satellite instead is the Envisat mission [9], launched in 2002, with a mass of 8140 kg. This satellite is characteristic of the mission trends at the time with the launch and operation of large satellites as true space laboratories serving a variety of applications. The XGBoost algorithm is trained for a mass ranging from 0 to 10,000 kg. The results are shown in Fig. 2.



(a) Mass distribution in training vs. test data

(b) Feature importance

Figure 2: The XGBoost model for the mass determination tested against historical data and its feature importance

To train the XGBoost algorithm the full available data from 1957 to 2024 was split into a test and train dataset. The algorithm is validated against the test dataset. This validation can be seen in Fig. 2a. The actual distribution vs. the one predicted by the model is presented. There is good accordance between the two for the lower mass ranges, but the algorithm does not predict many higher mass objects. To correct this difference, in the future a few representative higher mass objects could be added manually. Furthermore, the feature importance can be investigated. Here, as seen in Fig. 2b, the object class drives the assignment of the mass. This is to be expected, since the mass between a Payload object or a Rocket Body might vary significantly. Based on the determined mass the average cross-section area is then assigned by randomly sampling the intervals corresponding to the mass classifications presented in Tab. 2.

Outlook and Conclusions

This paper presented an easy method to go from a given number of future launches to a full picture in terms of object class, orbital region, mass and area. These characteristics are just as important for space debris evolution as the number of launches itself. This method is independent from the model used to predict the future number of launches. The software structure is demonstrated in Fig. 3.

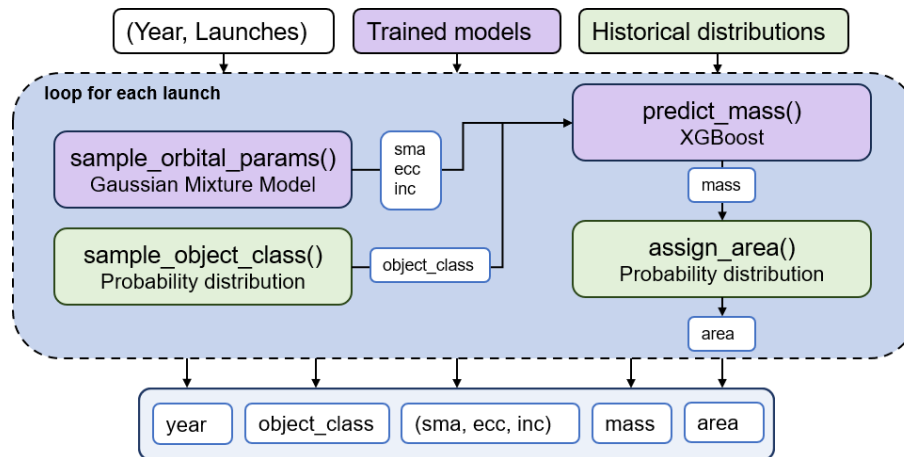


Figure 3: Software architecture to determine a full picture of future launches. The green modules represent historical distributions, and the purple modules describe trained models. The output of each module is marked with a blue border.

The software takes a file representing the future number of launches in the format (Year/Number of launches) and the trained models and distributions based on the historical data as an input. Then, for each launch the different modules are called. Firstly, the orbital parameters are determined using the GMMs, the object class is randomly drawn from the historical categorical distribution. Both of these results are then used as an input into the XGBoost model, to predict a mass based on orbital parameters and object category. The area is then sampled based on the mass classification. The final output file is a file containing the full information for each launch, including the launch year.

This file can then be used as an input into a space debris environment model to further investigate the effects that different launch scenarios can have on the environment as a whole.

Acknowledgments

This research received funding from the European Research Council (ERC) under the European Union's Horizon Europe research and innovation program as part of the GREEN SPECIES project (Grant agreement No - 101089265).

References

- [1] European Space Agency. DISCOS: Database and Information System Characterising Objects in Space. <https://discosweb.esoc.esa.int/>, 2024. Accessed: 2025-07-21.
- [2] Alyssa P. Manis, Mark J. Matney, Phillip D. Anz-Meador, and Andrew B. Vavrin. Time-dependent satellite explosion probabilities for long-term orbital debris environment modeling. *Journal of Spacecraft and Rockets*, 62(3):989-998, 2025. <https://doi.org/10.2514/1.A36118>
- [3] Douglas Reynolds. *Gaussian Mixture Models*, pages 659-663. Springer US, Boston, MA, 2009. https://doi.org/10.1007/978-0-387-73003-5_196
- [4] Callum J. Wilson, Massimiliano Vasile, Jinglang Feng, Keiran McNally, Alfredo Ant'on, and Francesca Letizia. Modelling future launch traffic and its effect on the leo operational environment. In *AIAA SciTech 2024 Forum*, Orlando, Florida, United States, 2024. American Institute of Aeronautics and Astronautics, AIAA. Peer-reviewed, accepted author manuscript. <https://doi.org/10.2514/6.2024-1814>
- [5] Gideon Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461-464, March 1978. <https://doi.org/10.1214/aos/1176344136>

- [6] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. CoRR, abs/1603.02754, 2016. <https://doi.org/10.1145/2939672.2939785>
- [7] A. S. Botelho and A. R. C. Xavier Jr. A unified satellite taxonomy proposal based on mass and size. *Advances in Aerospace Science and Technology*, 4:57-73, 2019. <https://doi.org/10.4236/aast.2019.44005>
- [8] NASA. Sts-116 mission overview. <https://www.nasa.gov/mission/sts-116/>, 2006. Accessed: 2025-11-06.
- [9] European Space Agency (ESA). Envisat mission overview. https://www.esa.int/Applications/Observing_the_Earth/Envisat/, 2002. Accessed: 2025-11-06.