XLVI CILAMCE
VITÓRIA - ES - BRAZIL - 2025
XLVI IBERO-LATIN AMERICAN CONGRESS ON COMPUTATIONAL METHODS IN ENGENEERING

Monte Carlo-Based Probability Estimation of Asteroid and Comet Impacts on Earth

M. M. D. Domingues[1], M. A. Silva[2]

[1]*Federal University of ABC*
Av. dos Estados, 5001, Bangú, Santo André,09280-560, SP, Brazil
mauro.montezuma@aluno.ufabc.edu.br
[2]*Federal University of ABC*
Av. dos Estados, 5001, Bangú, Santo André,09280-560, SP, Brazil
marcelo.araujo@aluno.ufabc.edu.br

Abstract. Estimates suggest that the probability of a celestial body—such as an asteroid or comet—colliding with Earth's surface increases gradually over time. This study aims to quantify that probability using concepts from Reliability Engineering, specifically applying the Monte Carlo method. The analysis was carried out through the formulation of a failure function and testing of key design variables: the minimum distance of the object from Earth and its relative velocity. The dataset was divided into two categories: real celestial objects mapped by NASA [1], and synthetic objects generated using sampling techniques based on the cumulative distribution function, implemented in both Excel and MATLAB. The results yielded similar failure probability values across the datasets, representing the likelihood of an object impacting Earth's surface: 0.113% for real objects, 0.133% for 15,000 simulated points in Excel, and 0.141% for 1,000,000 points in MATLAB. These outcomes validate the effectiveness of the proposed method, demonstrating that the Monte Carlo approach provides consistent and reliable estimates across different data sizes and platforms. The findings are in line with the assumptions and modeling techniques applied to a wide range of celestial object scenarios.

Keywords: Monte Carlo, probabilistic modeling, Near-Earth objects, impact prediction

1 Introduction

Near-Earth Objects (NEOs), including asteroids and comets, are celestial bodies orbiting within 1.3 AU of the Sun, significant for planetary defense and understanding Solar System formation [2]. Earth faces constant bombardment, from annual small asteroid atmospheric burn-ups, to rare catastrophic impacts capable of widespread damage [3]. Planetary defense initiatives, such as NASA's DART mission, are actively addressing these threats [4].

To quantify impact risks, this study applies the Monte Carlo Method (MCM) from Reliability Engineering [5]. MCM uses extensive random variable simulations, based on key design parameters like relative velocity and minimum distance, to assess failure probabilities through performance functions [6]. This research validates MCM for impact probability modeling by comparing its Excel and MATLAB implementations, also addressing large dataset uncertainties.

2 Impact Probability: Model and Equations

This study quantifies impact probability using real-world NASA data [1] and synthetic datasets generated via the Monte Carlo Method [7] in Excel and MATLAB.

The core of the analysis is a failure function that determines if a celestial body will collide with Earth. This function models the body's motion, considering its relative velocity at minimum distance from Earth. Assuming high altitude, no trajectory deviations, and proximity to Earth, only Earth's gravitational force is considered significant; drag, thrust, and third-body forces are negligible. As gravity is a conservative central force, total energy is conserved, and the force acts normal to the trajectory, equating to centripetal force.

Thus, the performance function for an asteroid or comet near Earth is the difference between the theoretical centripetal force required for a stable orbit and the gravitational pull:

$$g(X) = F_c - F_g = \frac{mv^2}{(d + R_E)} - G\frac{mM_E}{(d + R_E)^2} \tag{1}$$

Where $g(X)$ is the performance/failure function, F_c is centripetal force, F_g is gravitational force, m is the celestial body's mass, v is its relative velocity, d is its minimum distance from Earth, R_E is Earth's radius, G is the gravitational constant, and M_E is Earth's mass.

Simplifying by canceling mass m, as it appears in both force terms and is always positive, the failure function becomes mass-independent:

$$g(X) = \frac{v^2}{(d + R_E)} - G\frac{M_E}{(d + R_E)^2} \tag{2}$$

This expression is used in the Monte Carlo Method to assess failure probability. Here, X_1 (minimum distance to Earth) and X_2 (relative velocity) are random variables, while other terms are constants. Converting units (velocity in $\mathrm{km/s}$, distance in AU) yields:

$$g(X) = \frac{X_2^2 \cdot 10^3}{(X_1 \cdot 149597828,67) + 6378,1} - \frac{39,8736 \cdot 10^7}{((X_1 \cdot 149597828,67) + 6378,1)^2} \tag{3}$$

This equation defines the failure domain: $g(X) < 0$ indicates a collision, and $g(X) \geq 0$ signifies a safe trajectory and is fundamental to the study's probabilistic analysis.

2.1 Monte Carlo Method

The Monte Carlo Method (MCM) is a stochastic simulation approach used for reliability analysis. In this study, MCM tests the failure function $g(X)$ with design variables X_1 and X_2, distinguishing between safety and failure regions to determine the probability of an asteroid or comet impacting Earth.

The MCM steps are:

1. Generate a random number r (0 to 1).
2. Assign the random number to the design variable X using the inverse of its cumulative distribution function (CDF), $X = P^{-1}(r)$.

$$X = P^{-1}(r) \tag{4}$$

3. Evaluate the failure function $g(X)$ with the generated X values.
4. Compute reliability indices: Calculate the reliability index β and failure probability P_f:

$$\beta = \frac{\mu_{g(X)}}{\sigma_{g(X)}} \tag{5}$$

$$P_f = \frac{n(g(X) < 0)}{N} \tag{6}$$

where $\mu_{g(X)}$ and $\sigma_{g(X)}$ are the mean and standard deviation of $g(X)$, N is total simulations, and $n(g(X) < 0)$ is the count of negative $g(X)$ results.

For this study, 15,000 sets of design variables were randomly generated in Excel. To expand the dataset and observe uncertainties, 1,000,000 and 10,000,000 points were generated in MATLAB due to its superior computational capacity. The baseline for random number generation was 14,118 real asteroids and comets mapped by NASA's Jet Propulsion Laboratory, enabling comparison between real and randomly generated data impact probabilities.

2.2 Probability Functions

Probability Density Function (PDF) $p(x)$ and Cumulative Distribution Function (CDF) $P(x)$ are core to describing random variables [8]. The PDF quantifies the likelihood of a variable taking a specific value:

$$\int_{-\infty}^{\infty} p(x)\,dx = 1. \tag{7}$$

The CDF represents the probability that a variable is less than or equal to a given value, obtained by integrating the PDF:

$$P(x) = \int_{-\infty}^{x} p(t)\,dt. \tag{8}$$

For generating random design variables in the Monte Carlo Method, the PDF and CDF of each variable are essential. Based on real NASA data [1], the minimum distance to Earth (X_1) approximates a continuous uniform distribution, and relative velocity (X_2) follows a Gumbel distribution (Fig. 1 and 2).

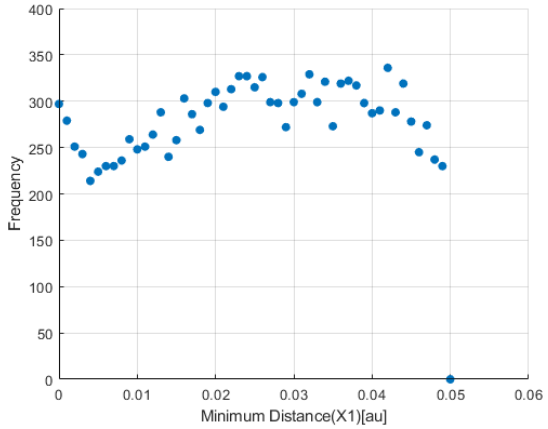


Figure 1. Real minimum distance to Earth distribution

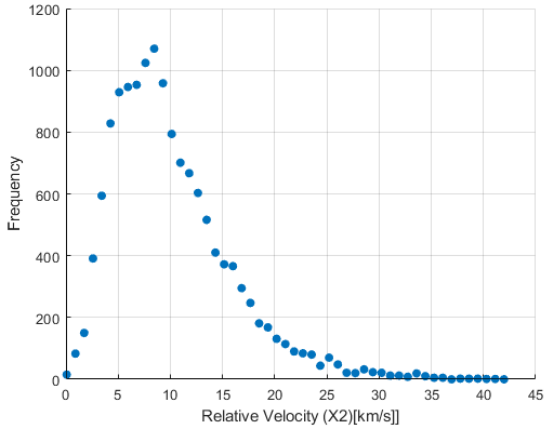


Figure 2. Real relative velocity distribution

To simplify mathematical formulation and better mimic real-world observations, a continuous uniform distribution was chosen for X_1 (Fig. 3) and a triangular distribution for X_2 (Fig. 4).



Figure 3. Continuous uniform distribution.

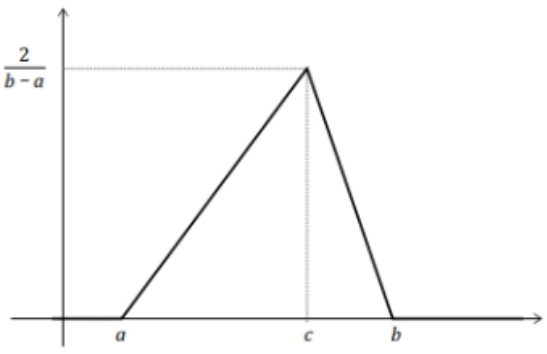


Figure 4. Triangular distribution.

The triangular distribution consists of two linear segments composed of two lines, one increasing and the other decreasing, which form the triangle. Its height h is calculated such that the total area under the PDF is 1:

$$h = \frac{2}{\mu_X + 2,5 \cdot \sigma_X} \tag{9}$$

where μ_X and σ_X are the mean and standard deviation of the design variable. The probability density functions for the two segments, $p_1(x)$ (increasing) and $p_2(x)$ (decreasing), are:

$$p_1(x) = \frac{h}{\mu} x \tag{10}$$

$$p_2(x) = -\frac{h}{2.5 \cdot \sigma} x + h\left(1 + \frac{\mu}{2.5 \cdot \sigma}\right) \tag{11}$$

The cumulative probability functions for these sections, $P_1(x)$ and $P_2(x)$, are:

$$P_1(x) = \frac{h}{2\mu} x^2 \tag{12}$$

$$P_2(x) = \frac{h \cdot \mu}{2} - \frac{h}{5\sigma}(x^2 - \mu^2) + h\left(1 + \frac{\mu}{2.5 \cdot \sigma}\right)(x - \mu) \tag{13}$$

Random values for X_1 (minimum distance) were generated using Excel's 'RAND()' and MATLAB's 'rand()' functions, producing values between 0 and 0.05 AU based on NASA data. For X_2 (relative velocity), values were generated using the inverse CDF of the triangular distribution:

$$\text{If } r \leq P_\mu, \quad X_2 = P_1^{-1}(r),$$
$$\text{If } r > P_\mu, \quad X_2 = P_2^{-1}(r).$$

Here, P_μ is the cumulative probability at the mean relative velocity from real NASA data [1].

3 Results and Discussion

3.1 Monte Carlo Method in Excel

The Monte Carlo process began by generating 15,000 random values for design variables X_1 (minimum distance to Earth) and X_2 (relative velocity). Frequency analysis of these values, shown in Fig. 5 and 6, confirmed the expected uniform distribution for X_1 and triangular distribution for X_2, validating the efficiency of the random number generation.

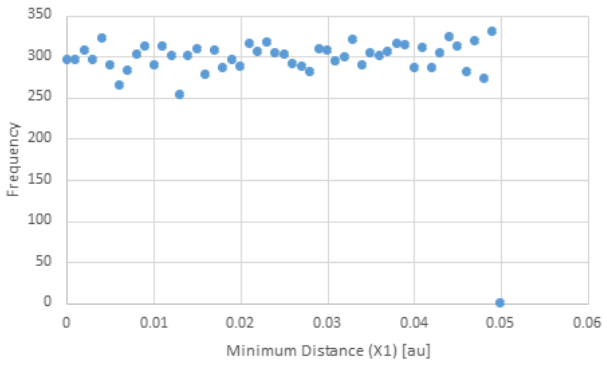


Figure 5. Random X_1 distribution in Excel.

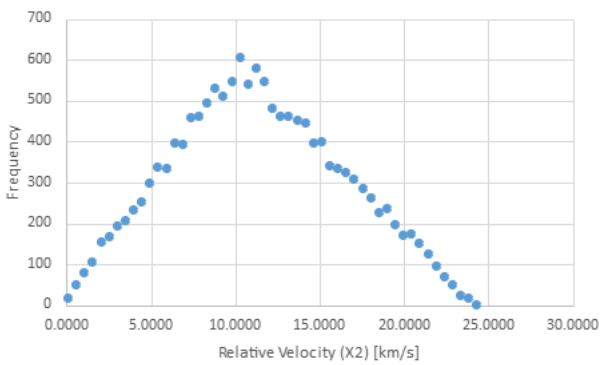


Figure 6. Random X_2 distribution in Excel.

The failure function was tested with all 15,000 generated values. With 20 out of 15,000 trials resulting in negative values (indicating failure), the failure probability (P_f) was calculated as **0.133%** and the reliability index (β) as **22.0%**. These were derived from the mean (μ_g) and standard deviation (σ_g) of the failure function results using equations (2.5) and (2.6).

For comparison, the same analysis was performed on real NASA data [1] within Excel to determine their P_f and β. Table 1 summarizes these results, showing a close correlation between real and randomly generated data.

Table 1. Reliability values for real and random data according to the Monte Carlo Method in Excel.

	Real Data	Random Data
P_f (%)	0.113	0.133
β (%)	18.4	22
P_f Relative Error (%)	-	17.7

The preliminary study shows that the calculated failure probability values are similar for both datasets, with a relative error of 17.7% for P_f, which is considered low given the methodologies used. Further analysis with a larger dataset in MATLAB is planned to ensure method accuracy and reliability for a broader range of celestial bodies.

3.2 Monte Carlo Method in MATLAB

Following the Monte Carlo Method (MCM) approach previously used in Excel, 1,000,000 random points were generated in MATLAB for variables X_1 (minimum distance to Earth) and X_2 (relative velocity). Frequency analyses (Fig. 7 and 8) confirmed the expected continuous uniform and triangular distributions for X_1 and X_2, respectively.

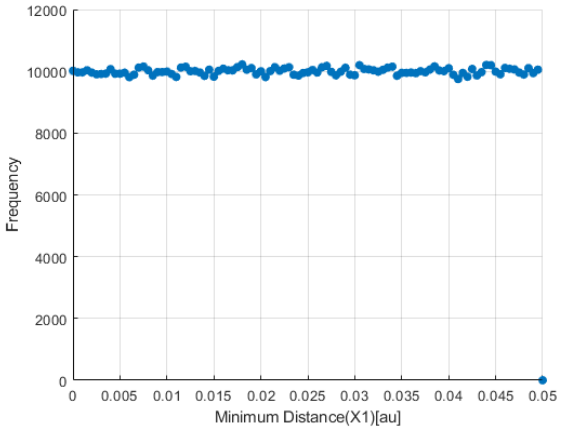


Figure 7. 1,000,000 random X_1 distribution.

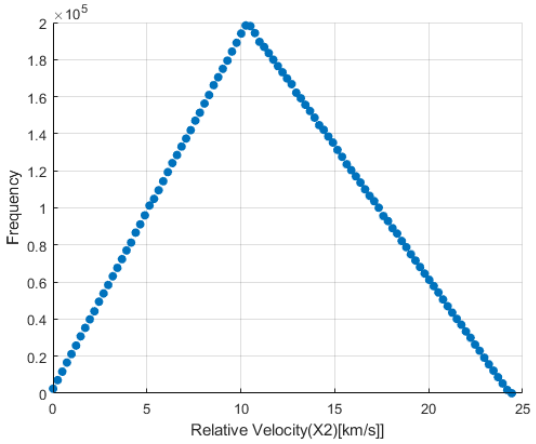


Figure 8. 1,000,000 random X_2 distribution.

A performance function test on these 1,000,000 points yielded a mean failure function value (μ_g) of 0.1431 and a standard deviation (σ_g) of 0.8087. With 1,410 negative function values, the failure probability (P_f) was **0.141%** and the reliability index (β) was **17.7%**:

This process was then repeated with 10,000,000 randomly generated data points for X_1 and X_2. Frequency graphs (Fig. 9 and 10) again verified the consistent distributions, despite the significantly larger dataset.

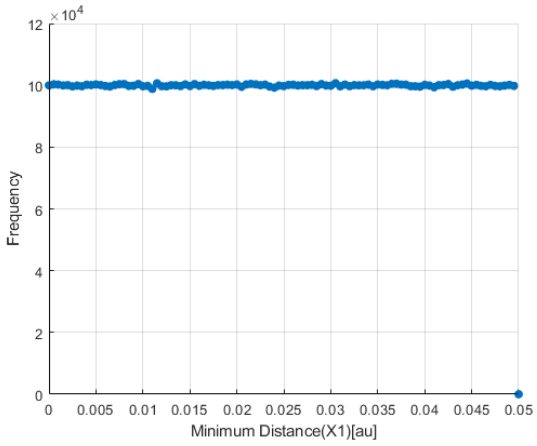


Figure 9. 10,000,000 random X_1 distribution.

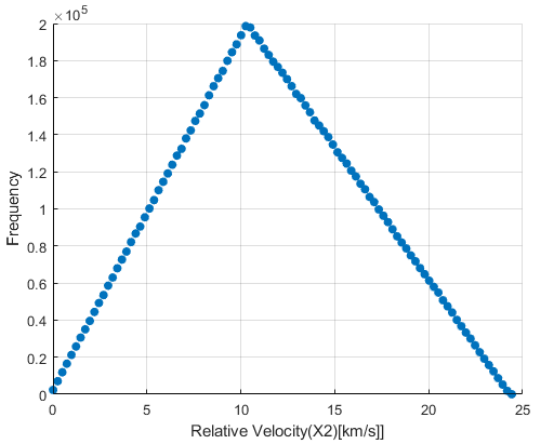


Figure 10. 10,000,000 random X_2 distribution.

The performance function test for 10,000,000 points showed a mean (μ_g) of 0.1438, a standard deviation (σ_g) of 0.8153, and 14,563 negative function values. This resulted in a failure probability (P_f) of **0.146%** and a reliability index (β) of **17.6%**:

Table 2 summarizes the reliability parameters from MATLAB for real data (14,118 points) and the two larger random datasets (1,000,000 and 10,000,000 points).

Table 2. Reliability values for real and random data according to the MCM in Matlab.

	Real Data	1,000,000 Points	10,000,000 Points
P_f (%)	0.113	0.141	0.146
β (%)	18.4	17.7	17.6
P_f Relative Error (%)	-	24.8	29.2

Table 2 shows that the failure probabilities (P_f) for the randomly generated MATLAB data were very close, with only about a 3.5% variation despite a tenfold increase in data points. Compared to NASA's real data, the relative error for P_f in the MCM application was 24.8% for 1,000,000 points and 29.2% for 10,000,000 points. This increased error is attributed to the larger sample space and approximations in sampling techniques, particularly the relationship between triangular and Gumbel distributions.

The calculated reliability indices also remained low, which can be attributed to the high standard deviation of the process variables, reflecting the significant inherent uncertainty in celestial body dynamics. Notably, the reliability index (β) was very similar across real and random data analyses, confirming the method's accuracy in replicating asteroid and comet behavior relative to Earth.

4 Conclusions

This study successfully applied the Monte Carlo Method (MCM) across Excel and MATLAB to determine the probability of celestial body impacts on Earth, verifying its accuracy against real NASA data and across varying sample sizes.

The research established the MCM's theoretical basis and demonstrated its consistent application. Analysis of real NASA data yielded a failure probability of **0.113%. Initial Excel implementation with 15,000 points resulted in a 0.133% failure probability (17.7% relative error to real data), showing good accuracy. Scaling up in MATLAB to 1,000,000 and 10,000,000 points produced impact probabilities of 0.141% and 0.146% respectively, further supporting the method's validity.

The consistency of reliability parameters across platforms and data sizes, as summarized in Table 3, confirms the MCM's robustness. The low relative error between Excel and MATLAB implementations (6.02% for 1M points, 9.77% for 10M points, relative to 15k points) underscores this. The consistently low reliability index

(β) reflects the inherent high uncertainty in modeling celestial body motion, primarily due to the large standard deviation of design variables.

Table 3. Reliability values for random data according to the MCM in Excel and MATLAB.

	15,000 Points (Excel)	1,000,000 Points (MATLAB)	10,000,000 Points (MATLAB)
P_f (%)	0.133	0.141	0.146
β (%)	17.7	17.7	17.6
P_f Relative Error (%)	-	6.02	9.77

Overall, this study successfully determined the probability of asteroid or comet impact on Earth, demonstrating the MCM's effectiveness and its successful replication across different programming environments, with results consistently within acceptable limits compared to mapped celestial bodies.

Acknowledgements. This work is based upon a university work taught by the author M. A. SILVA, which received partial support through a research grant from the Federal University of ABC, Brazil.

Authorship statement. The authors hereby confirm that they are the sole liable persons responsible for the authorship of this work, and that all material that has been herein included as part of the present paper is either the property (and authorship) of the authors, or has the permission of the owners to be included here.

References

[1] N. Content Administrator. Neo earth close approaches, close approach data. *Center for Near Earth Object Studies, Jet Propulsion Laboratory, California Institute of Technology*. Available: `https://cneos.jpl.nasa.gov/ca/`. Accessed: Oct 4, 2023., Mar 2023.
[2] M. Fulchignoni and M. Barucci. Neo impact consequences and hazards. *Comptes Rendus Physique*, vol. 6, pp. 283–289, Apr 2005.
[3] N. Content Administrator. Asteroids fast facts. Technical report, National Aeronautics and Space Administration (NASA). Available: `https://www.nasa.gov/planetarydefense/asteroids`. Accessed: May 9, 2023., Aug 2017.
[4] H. A. B. B. *et al*. A.F. Cheng. Momentum transfer from the dart mission kinetic impact on asteroid dimorphos. *Nature*, vol. 616, pp. 457–460, 2023.
[5] P. Luthra. Reliability: What it means. *Reliability Review*, vol. 8, pp. 10–16, Sept 1988.
[6] da M. A. Silva. Análise de formulações de problemas de confiabilidade estrutural. In *CILAMCE 2019, XL Congresso Ibero-LatinAmerican de Métodos Computacionais em Engenharia*, Natal, RN, Brasil. ABMEC. Nov. 11–14, 2019., 2019.
[7] M. Morales. *Análise de confiabilidade de taludes em condições saturadas-não saturadas via análise limite no espaço cônico quadrático*. Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, RJ, 2013.
[8] S. Ross. Introduction to probability models. *Academic Press*, vol. 11th Edition, 2014.