REDUCTION OF THE PROBABILITY BIAS OF THE SINGLE STEP INPUTS' STATISTICS METHOD FOR THE COMPUTATION OF THE SAFETY RE-ENTRY AREA: INTRODUCTION OF A MULTISTEP ALGORITHM

S. F. Rafano Carná[(1)], E. De Pasquale[(2)], L. Arzel[(3)], and M. Lavagna[(4)]

[(1)]*Politecnico di Milano, Via Giuseppe La Masa, 34, 20156 Milano, Italy, Email: simone.rafano@mail.polimi.it*
[(2)]*European Space Agency, 18, avenue Edouard Belin, 31401 Toulouse CEDEX 9, France, Email: emilio.de.pasquale@esa.int*
[(3)]*Toulouse Business School, 20 Boulevard Lascrosses, 31068 Toulouse, France Email: l.arzel@tbs-education.org*
[(4)]*Politecnico di Milano, Via Giuseppe La Masa, 34, 20156 Milano, Italy, Email: michelle.lavagna@polimi.it*

ABSTRACT

The computation of Safety Re-entry Area (SRA) associated to a probability of 99,999% is required for a significant number of space missions that shall be controlled to a destructive re-entry. The dynamics of a spacecraft entering the atmosphere is sensitive to a large number of parameters affected by uncertainty, which might cause risk for the population, air and naval traffic, and ground and sea assets. The extremely low probability of interest requires the identification of events, that occur so seldom and having consequences so catastrophic, that they may be regarded as rare events. This work describes an innovative method, called Inputs' Statistics, efficient in estimating rare events defined above a threshold. It has been originally developed for the computation of the SRA but it may be generalized and extended to solve other rare events problems coming from the aerospace field but not limited to it

Key words: Rare events estimation; Extremely low probabilities; Safety assessment; Destructive re-entry; Safety Re-entry Area; Optimization; Monte Carlo simulation.

1. INTRODUCTION

The interest on extremes and rare events has enormously increased in the last decades because of its practical relevance in many different scientific fields such as insurance, finance, engineering and environmental sciences. But how can a rare event be defined? A rare event is an event that occurs so seldom but having consequences so critical that may result in a catastrophic outcome of the mission. In aerospace engineering, the term refers to events with an occurrence probability typically on the order of 10^{-5}. A representative example is the probability that two aircraft will get dangerously close each other in a given airspace, or the collision probability between operative satellites and space debris.

The rare event of concern in this work is the probability that a fragment resulting from the atmospheric fragmentation of a spacecraft performing a controlled destructive re-entry falls at a so large distance with respect to the aimed impact point that could reach inhabited areas causing thus risk for the population, air and naval traffic, and ground and sea assets. In Europe, the safety compliance verification measures for the re-entry of a spacecraft at its end of life are regulated by restrictive requirements documented in Space Agencies' instructions and guidelines [6]. According to the French Space law: *the operator responsible of a spacecraft controlled re-entry shall identify and compute the impact zones of the spacecraft and its fragments for all controlled re-entry on the Earth with a probability respectively of 99% and 99,999% taking into account the uncertainties associated to the parameters of the re-entry trajectories* [13].

Theses impact zones are usually termed safety boxes, or sometimes safety footprint boundaries. They are containment contours on ground defined such that the probability that a fragment falls outside is below a controlled or known value. The Safety Re-entry Area (SRA) is the safety box associated with the probability 99,999%. It is used to design the re-entry trajectory such that the SRA does not extend over inhabited regions, does not impinge on state territories and territorial waters without the agreement of the relevant authorities. The computation of SRA is required for a significant number of space missions like spacecraft in low Earth orbits at their end of life and last stages of launchers that shall be controlled to a destructive re-entry. The Declared Re-entry Area (DRA) is the safety box associated with the probability 99%; it is used to implement the procedures of warning and alerting the maritime and aeronautic traffic authorities.

The challenge of the SRA design is the extremely low probability associated to its contour (10^{-5}), which makes quite difficult and inaccurate the use of classic statistical techniques. The State of Art methods use Monte Carlo (MC) analysis to estimate the SRA, where the number of generated samples determines the accuracy and confidence level (CL) achieved but is also proportional to the computational time. Since the integration of the at-

Proc. 7th European Conference on Space Debris, Darmstadt, Germany, 18–21 April 2017, published by the ESA Space Debris Office
Ed. T. Flohrer & F. Schmitz, (http://spacedebris2017.sdo.esoc.esa.int, June 2017)

mospheric re-entry transfer function is computationally expensive, usually on the order of one or two seconds in a standard desktop computer, the typical number of simulations is restricted to few thousands. Therefore the quantile of interest is estimated by extrapolating the probability density much beyond the generated data. If the number of outputs samples is smaller than required, the result may be biased or characterized by low confidence [10, 12].
The Inputs' Statistics method has been developed originally with the intent of increasing the efficiency of the computation of the safety boxes [4]. The idea which led to its definition and subsequent implementation relies on the consideration that the probability that a worst case scenario occurs is in some way related to the probability that a particular unlucky combination of initial conditions takes place and consequently leads a fragment fall much further than expected. Trying to follow this intuition led to focus the attention on the input domain of the problem rather than the output. This because the input domain collects all the causes which may lead to foreseeable but also rare catastrophic consequences. Working on the input domain, analyzing it and trying to characterize it with some information coming from the output domain may provide directly some information on the sought probability with an extremely low computational effort. This is the main objective of the Inputs' Statistics method: investigating the causes of the phenomena to get useful information which may speed up the computational intensive overall solution. In addition, the input variables are usually a priori statistically modeled exploiting physical considerations and engineering judgement, assigning a specific probability density to each of them. Once all the input domain is characterized, it is relatively easy to surf through it, integrate directly and exactly the joint probability density over particular subsets and extrapolate local information of the output by evaluating the gradient of the transfer function. Therefore, the input domain is particularly attractive because of the possibility to explore it rapidly as well as the physical meaning which can be derived by its analysis. Nevertheless, it is intrinsically more complex than the output domain because it is always multidimensional. Several uncertain variables have to be considered in a typical realistic problem as the computation of the safety boxes and this may create some difficulties when global rather than local information are required to draw useful conclusions. On the other hand, the output distribution is usually mono-dimensional or, at least, some simplifying hypothesis are introduced to subdivide the problem and study one dimension per time. This is the reason why a possible alternative approach to solve the problem is to built up numerically part of the distribution of the output by performing a MC simulation, but accepting an inevitable increase of the computational time. This strategy is exploited in the Inputs' Statistics method to refine an initial conservative estimation of the result. Indeed, going more in depth of the Inputs' Statistics method, it is possible to distinguish two "versions" of the method:

- the *Single Step* algorithm which provides a preliminary but fast and conservative estimation of the result;
- the *Multistep* algorithm which gets a compromise with the MC simulation to reduce the bias of the Single Step's result.

More specifically, with the Single Step algorithm, the probability of interest is computed integrating the known joint multivariate density function of the input variables and, then, an optimization process is used to find the minimum and maximum (worst) cases within a reduced set of inputs corresponding to a conservative underestimation of the probability. Thus, a large amount of computational time is saved since the MC simulations are not required. The main drawback of the Single Step algorithm is related to the simplification that is introduced in the identification of the input domain associated to the aimed probability. This simplification leads to a bias on the probability estimate but it provides always a conservative solution to the problem. Indeed, when the problem under analysis concerns safety assessment, if a certain error exists, it must be guaranteed that its sign is in accordance with the requirement. Therefore, considering that there is no worst case for the statistical distribution, this assessment can be also retained as a conservative envelope of the optimal solution.
The Multistep algorithm, originally presented in this paper, starts from this preliminary result and corrects it by applying iteratively the Single Step algorithm alternated with "smart" MC simulations over input subsets characterized by low probability density. The number of samples that have to be propagated into the output through the transfer function is thus minimized.
The case study considered in this paper is the computation of the SRA of the Shallow Re-entry of the fifth (and last) Automated Transfer Vehicle (ATV): Georges Lamaître [5]. The results are performed using the classical MC approach and the Inputs' Statistics method. They compared highlighting advantages and drawbacks in terms of accuracy, level of conservatism and computational time. An exhaustive and plenty of examples treatise on the method can be found in [11].

2. PROBLEM STATEMENT

Let $\mathbf{X}$ be a d-dimensional random vector with probability density function (pdf) f and mean value vector $\boldsymbol{\mu}$, and ϕ be a continuous and deterministic scalar function: $\phi : \mathbb{R}^d \to \mathbb{R}$ such that:

$$\mathbf{X} = \{X_1, ...X_d\}^{\mathrm{T}} \to Y = \phi(\mathbf{X}) \tag{1}$$

The vector $\mathbf{X}$ is here called "input" of the problem and the scalar variable Y "output" of the problem. ϕ is the transfer function that transforms the input $\mathbf{X}$ into the output Y. Without loss of generality, the function ϕ satisfies the property:

$$\phi(\boldsymbol{\mu}) = 0 \tag{2}$$

indeed, if a priori this property is not satisfied by a generic transfer function ϕ', then ϕ is simply defined as

$$\phi(\mathbf{X}) = \phi'(\mathbf{X}) - \phi'(\boldsymbol{\mu}) \tag{3}$$

The problem under investigation formally requires to compute two thresholds T_1 and T_2, with $T_1 < 0$ and $T_2 > 0$, such that the probability that the output Y falls outside the interval $[T_1, T_2]$ is lower or equal to a given prefixed value, that is indicated with α:

$$1 - \mathbb{P}(T_1 \leq Y \leq T_2) \leq \alpha \tag{4}$$

where α is a rare event probability, that is $\alpha = 10^{-5}$ or less.

The classic State of the Art (SoA) techniques based on MC methods put a large computational effort in building up numerically the distribution of Y, which is unknown a priori. Calling its density f_Y, the SoA approach can be formulated as looking for $T_1 < 0$ and $T_2 > 0$ such that:

$$1 - \int_{T_1}^{T_2} f_Y(y)\mathrm{d}y \leq \alpha \tag{5}$$

The main issue of this approach is the computational time. Indeed, building numerically the output distribution f_Y causes a sharp increase of the computational time with the decrease of the required probability α in order to keep an acceptable bounded confidence interval.

The problem can be re-formulated in an equivalent way if the attention is focused on the input instead of on the output domain. The input uncertainties are statistically modeled using physical considerations and engineering judgment. Once the pdf f is defined, the input domain is fully characterized. Let us introduce:

Definition 2.1. *Let $\mathbf{X} \in \mathbb{R}^d$ be a continuous random vector and ϕ be a transfer function: $\phi : \mathbb{R}^d \to \mathbb{R}$, the failure domain Ω_f of f relative to the two thresholds $T_1, T_2 \in \mathbb{R}$ with $T_1 < T_2$ is:*

$$\Omega_f = \{\mathbf{x} \in \mathbb{R}^d : \phi(\mathbf{x}) < T_1 \vee \phi(\mathbf{x}) > T_2\} \tag{6}$$

Its complementary is:

$$^c\Omega_f = \{\mathbf{x} \in \mathbb{R}^d : T_1 \leq \phi(\mathbf{x}) \leq T_2\} \tag{7}$$

The probability can be *exactly* computed integrating f over the region Ω_f of the input domain, once identified. Accordingly, the problem can be alternatively formulated as defining the region Ω_f such that it satisfies:

$$\int_{\Omega_f} f(\mathbf{x})\mathrm{d}\mathbf{x} \leq \alpha \tag{8}$$

where $\mathrm{d}\mathbf{x} = \mathrm{d}x_1 ... \mathrm{d}x_d$. This is an intrinsically different approach and it is the basis of the Inputs' Statistics method.

Nevertheless, Ω_f is not unique because two unknowns must be selected: T_1 and T_2 and only one inequality is available from the condition of the required probability level α. There exists an entire family of admissible failure domains. Among all the possible choices of Ω_f, the engineering design has as objective to identify that particular Ω_f which minimizes the distance between the two values T_1 and T_2. Since this distance increases with the decrease of the required probability level, to compute this optimal region it is imposed the less restrictive as possible constraint, i.e. equality instead of inequality:

Definition 2.2. *Let $\mathbf{X} \in \mathbb{R}^d$ be a continuous random vector and ϕ be a transfer function: $\phi : \mathbb{R}^d \to \mathbb{R}$, the optimal failure domain:*

$$\Omega_f^{Opt} = \{\mathbf{x} \in \mathbb{R}^d : \phi(\mathbf{x}) < T_1^{Opt} \vee \phi(\mathbf{x}) > T_2^{Opt}\}$$

with $T_1^{Opt}, T_2^{Opt} \in \mathbb{R}$ and $T_1^{Opt} < T_2^{Opt}$, is defined such that

$$\int_{\Omega_f^{Opt}} f(\mathbf{x})\mathrm{d}\mathbf{x} = \alpha \tag{9}$$

and the difference $T_2^{Opt} - T_1^{Opt}$ is the minimum, that is:

$$T_2^{Opt} - T_1^{Opt} \leq T_2 - T_1 \tag{10}$$

for any possible choice of T_1 and T_2 that satisfy eq.8.

3. SINGLE STEP INPUTS' STATISTICS

Since ϕ is generally a multidimensional dynamic propagator, as for instance in the case of the safety boxes, and its contour surfaces are not identified in fast enough computational sequence (e.g. there is not an analytic explicit formulation), the direct computation of the Ω_f family, and especially of Ω_f^{Opt}, is not practically feasible due to computational time limitations.

Therefore, the Inputs' Statistics method aims at defining a domain $\widetilde{\Omega}_f$ that belongs to the Ω_f family and approximates Ω_f^{Opt}:

$$\widetilde{\Omega}_f \approx \Omega_f^{Opt} \tag{11}$$

Consequently, the probability estimated by the Inputs' Statistics method $\mathbb{P}^{IS}$ is:

$$\mathbb{P}^{IS} = \int_{\widetilde{\Omega}_f} f(\mathbf{x})\mathrm{d}\mathbf{x} \tag{12}$$

and the corresponding values of the thresholds are $\widetilde{T}_1$ and $\widetilde{T}_2$. The idea behind the Inputs' Statistic method is the limitation of the input domain using the d-dimensional contour surfaces of the pdf rather than the computation of the contour surfaces of the transfer function ϕ. Since the mathematical formulation of the pdf is known, its contour surfaces are easily identified.

So accordingly to the Single Step algorithm: *being $\widetilde{\mathcal{E}}$ the d-dimensional contour surface of the pdf enclosing* *a probability equal to $1-\alpha$, then $\widetilde{\Omega}_f$ is the region identified by contour surfaces of the transfer function ϕ corresponding to the thresholds $\widetilde{T}_1$ and $\widetilde{T}_2$ being the minimum and maximum cases which may occur inside $\widetilde{\mathcal{E}}$. $\widetilde{T}_1$ and $\widetilde{T}_2$ are thus the outcome of the method.*

In ref.[4] it is explained, under the hypothesis of having only normal distributed input variables, how to perform the integral of the pdf over the volume enveloped by the pdf contour surfaces. Note indeed that if all the variables in the input vector $\mathbf{X}$ are normally distributed random variables: $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the associated pdf is the Multivariate Normal (MVN) distribution and the contour surfaces of the MVN are d-dimensional ellipsoids, centred on $\boldsymbol{\mu}$ and oriented accordingly to $\boldsymbol{\Sigma}$. It can be extended to the general case as long as a transformation $\tau : \mathbb{R}^d \to \mathbb{R}^d$ exists that maps the random input vector $\mathbf{Z}$, which may have non normal distributed variables, into a full normal distributed vector $\mathbf{X}$. In reference [10], some transformations have been proposed depending on the available information on the pdf of $\mathbf{Z}$.

Three advantages given by the Single Step algorithm can be recognized:

- the probability estimate is directly derived by the causes of the problem (initial conditions and model dispersions) highlighting its physical explanation;
- a large amount of computational time is saved since the MC simulations are not required and the computational time is not sensitive to the probability level to be achieved which can be set arbitrarily small;
- the outcome of the method provides always a conservative solution of the problem.

The main drawback is related to the simplification that is introduced in the identification of the failure domain. This simplification leads to a bias on the probability estimate: the result is always conservative but it is necessary to assure that it is not too far from the optimal one. When this is not the case, often a better accuracy is required. This is why the Multistep algorithm has been introduced, looking for a compromise between the speed of the Inputs' Statistics and the accuracy of the MC approach.

4. MULTISTEP INPUTS' STATISTICS

The idea behind the Multistep Inputs' Statistics is to consider progressively a smaller and smaller d-dimensional contour surface of the pdf (d-dimensional ellipsoid of the MVN) in order to get a smaller $T_2 - T_1$ interval closer to the optimal one $T_2^{Opt} - T_1^{Opt}$. Thus, performing a MC simulation at each step of the algorithm, it is possible to estimate the error that the Inputs' Statistics method gives at the current iteration, which is then subtracted at the following iteration. Since it is required to integrate at each step only the samples that fall outside the current ellipsoid which is a low density region, the algorithm may show good performances especially when compared with the Crude Monte Carlo (CMC) method.

4.1. Algorithm for monotone functions

Let us suppose firstly that the transfer function ϕ is a monotone function with respect to all its variables. This is just a way to illustrate a simplified version of the algorithm which will be generalized in sec.4.5 to generic transfer functions. When the transfer function is monotone, both minimum and maximum belong always to the border of the feasible domain, that is $\widetilde{\mathcal{E}}$. In this way, the contours of ϕ, i.e. $\widetilde{\Omega}_f$, are tangent to $\widetilde{\mathcal{E}}$ and consequently reducing progressively $\widetilde{\mathcal{E}}$, $\widetilde{\Omega}_f$ enlarges until a given tolerance with respect to the probability estimate or the output interval is met.

The increase of $\widetilde{\Omega}_f$ is accomplished iteratively, checking at each step k the error due to the approximation of Ω_f^{Opt} with $\widetilde{\Omega}_f^k$. The error at a generic k-th iteration, indicated with e^k, can be expressed as:

$$e^k = \int\limits_{\widetilde{\Omega}_f^{Opt}} f(\mathbf{x})\mathrm{d}\mathbf{x} - \int\limits_{\widetilde{\Omega}_f^k} f(\mathbf{x})\mathrm{d}\mathbf{x} \tag{13}$$

or equivalently, recalling the definition 2.2: $\int\limits_{\widetilde{\Omega}_f^{Opt}} f(\mathbf{x})\mathrm{d}\mathbf{x} = \alpha$, as:

$$e^k = \alpha - \mathbb{P}^{IS^k} \tag{14}$$

where $\mathbb{P}^{IS^k}$ is the probability estimate given by the Inputs' Statistics at the k-th step.

Then, the ellipsoid $\widetilde{\mathcal{E}}^k$ is reduced at the successive step simply subtracting the error from the previous step:

$$\int\limits_{\widetilde{\mathcal{E}}^{k+1}(\tilde{t}^{k+1})} f(\mathbf{x})\mathrm{d}\mathbf{x} = \int\limits_{\widetilde{\mathcal{E}}^{k}(\tilde{t}^{k})} f(\mathbf{x})\mathrm{d}\mathbf{x} - e^k \tag{15}$$

and thus the updating relation is got:

$$\alpha^{k+1} = \alpha^k + e^k \tag{16}$$

An estimation of the probability $\mathbb{P}^{IS^k}$ is given here through a CMC method. By definition, $\mathbb{P}^{IS^k}$ is the probability that a sample is outside the contours of ϕ associated to $\widetilde{T}_1^k$ and $\widetilde{T}_2^k$. But since $\widetilde{T}_1^k$ and $\widetilde{T}_2^k$ are tangent to the border of $\widetilde{\mathcal{E}}^k$ then $\widetilde{\mathcal{E}}^k \subseteq {}^{\mathrm{c}}\widetilde{\Omega}_f^k$, that is $\widetilde{\mathcal{E}}^k \cap \widetilde{\Omega}_f^k = \varnothing$. Therefore there is not any interest in the distribution inside the current ellipsoid $\widetilde{\mathcal{E}}^k$ and it is necessary to integrate *only* the samples that fall outside it. Since it is a low density region, under some hypothesis, the computational burden keeps limited. Counting how many of these samples have an output which is lower than $\widetilde{T}_1^k$ or larger than $\widetilde{T}_2^k$ and dividing by N, the probability $\mathbb{P}^{IS^k}$ is estimated.

In addition, it is important to note that the samples outside the previous ellipsoid have already been computed at the previous step, so storing them, at the next step it is necessary to characterize only the region included between the previous and the current ellipsoid. To describe mathematically these operations, two sets are introduced, the first containing the indexes of the samples included between two ellipsoids and the second containing the corresponding output values of the transfer function ϕ:

$$\mathcal{I}(t_1,t_2) = \{i \in \mathbb{N} : t_1 \leq (\mathbf{X}_i - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{X}_i - \boldsymbol{\mu}) < t_2, \; i = 1, ..., N\} \quad (17)$$

and

$$\Theta(t_1,t_2) = \{Y_i = \phi(\mathbf{X}_i) \in \mathbb{R} : i \in \mathcal{I}(t_1,t_2)\} \quad (18)$$

where $\mathbf{X}_1, ..., \mathbf{X}_N$ are N iid samples as $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and t_1 and t_2 are the square radius of two generic ellipsoids such that $t_1 < t_2$. Note in particular that using this nomenclature, the indexes of the samples outside an ellipsoid having square radius t can be indicated as:

$$\mathcal{I}(t,+\infty) = \{i \in \mathbb{N} : (\mathbf{X}_i - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{X}_i - \boldsymbol{\mu}) > t, \; i = 1, ..., N\} \quad (19)$$

Thus, $\mathbb{P}^{IS^k}$ can be expressed as

$$\mathbb{P}^{IS^k} = \frac{1}{N} \sum_{i \in \mathcal{I}(\tilde{t}^k,+\infty)} \mathbf{1}_{\phi(\mathbf{X}_i) < \widetilde{T}_1^k \vee \phi(\mathbf{X}_i) > \widetilde{T}_2^k} \quad (20)$$

where $\mathbf{1}_{\phi(\mathbf{X}_i) < \widetilde{T}_1^k \vee \phi(\mathbf{X}_i) > \widetilde{T}_2^k}$ is a function that assumes the value 1 if $\phi(\mathbf{X}_i) < \widetilde{T}_1^k$ or $\phi(\mathbf{X}_i) > \widetilde{T}_2^k$ and 0 otherwise. Algorithm 4.1 gives a schematic and concise description of the Multistep Inputs' Statistics method for monotone functions.

Algorithm 4.1: Multistep Inputs' Statistics for monotone functions

Input: Probability α **Output:** Thresholds T_1^k and T_2^k;
choose
- Maximum number of iterations k_{Max};
- Tolerances ξ_T and ξ_α;
- Number of input samples N through eq.22;

initialize
- Step: $k = 1$;
- Probability outside ellipsoid: $\alpha^1 = \alpha$;
- Ell. $\widetilde{\mathcal{E}}^1$: **find** $\tilde{t}^1$ **such that** $1 - \int_{\widetilde{\mathcal{E}}^1(\tilde{t}^1)} f(\mathbf{x})d\mathbf{x} = \alpha^1$;
- Samples outside $\widetilde{\mathcal{E}}^1$: $\mathcal{I}(\tilde{t}^1,+\infty) = \{i \in \mathbb{N} : (\mathbf{X}_i - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{X}_i - \boldsymbol{\mu}) > \tilde{t}^1, i = 1, ..., N\}$;
- MC simulation outside $\widetilde{\mathcal{E}}^1$: $\Theta(\tilde{t}^1,+\infty) = \{Y_i = \phi(\mathbf{X}_i) \in \mathbb{R} : i \in \mathcal{I}(\tilde{t}^1,+\infty)\}$;
- Set $\widetilde{\mathcal{E}}^0 \equiv \widetilde{\mathcal{E}}^1$, that is $\tilde{t}^0 = \tilde{t}^1$;

repeat for $k \geq 1$
- **compute**
  - $\widetilde{T}_1^k$ and $\widetilde{T}_2^k$ as min/max inside $\widetilde{\mathcal{E}}^k$;
  - Samples between $\widetilde{\mathcal{E}}^k$ and $\widetilde{\mathcal{E}}^{k-1}$: $\mathcal{I}(\tilde{t}^k,\tilde{t}^{k-1}) = \{i \in \mathbb{N} : \tilde{t}_k \leq (\mathbf{X}_i - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{X}_i - \boldsymbol{\mu}) < \tilde{t}^{k-1}, i = 1, ..., N\}$;
  - MC simulation between $\widetilde{\mathcal{E}}^k$ and $\widetilde{\mathcal{E}}^{k-1}$: $\Theta(\tilde{t}^k,\tilde{t}^{k-1}) = \{Y_i = \phi(\mathbf{X}_i) \in \mathbb{R} : i \in \mathcal{I}(\tilde{t}^k,\tilde{t}^{k-1})\}$;
  - **update**
    - Samples outside $\widetilde{\mathcal{E}}^k$: $\mathcal{I}(\tilde{t}^k,+\infty) = \{i \in \mathbb{N} : i \in \mathcal{I}(\tilde{t}^k,\tilde{t}^{k-1}) \vee i \in \mathcal{I}(\tilde{t}^{k-1},+\infty)\}$;
    - Set of outputs outside $\widetilde{\mathcal{E}}^k$ as: $\Theta(\tilde{t}^k,+\infty) = \{Y_i \in \mathbb{R} : i \in \mathcal{I}(\tilde{t}^k,\tilde{t}^{k-1}) \vee i \in \mathcal{I}(\tilde{t}^{k-1},+\infty)\}$;
  - Probability estimate: $\mathbb{P}^{IS^k} = \frac{1}{N} \sum_{i \in \mathcal{I}(\tilde{t}^k,+\infty)} \mathbf{1}_{\phi(\mathbf{X}_i) < \widetilde{T}_1^k \vee \phi(\mathbf{X}_i) > \widetilde{T}_2^k}$;
  - Error: $e^k = \alpha - \mathbb{P}^{IS^k}$;
- **update**
  - Probability outside ellipsoid: $\alpha^{k+1} = \alpha^k + e^k$;
  - Ell. $\widetilde{\mathcal{E}}^{k+1}$: **find** $\tilde{t}^{k+1}$ **such that** $1 - \int_{\widetilde{\mathcal{E}}^{k+1}(\tilde{t}^{k+1})} f(\mathbf{x})d\mathbf{x} = \alpha^{k+1}$;
  - Step: $k = k + 1$

until $k > k_{Max} \vee \frac{\widetilde{T}_2^{k+1} - \widetilde{T}_1^{k+1}}{\widetilde{T}_2^k - \widetilde{T}_1^k} < \xi_T \vee e^k/\alpha < \xi_\alpha$

4.2. Convergence analysis

Let us start with a qualitative discussion of an ideal situation in which two ideal hypotheses hold:

1. the probability $\mathbb{P}^{IS^k}$ at each step is computed exactly, that is we are using an infinite number of samples in the MC simulations;

2. the optimization process used to locate the $\widetilde{T}_1^k$ and $\widetilde{T}_2^k$ is ideal: it converges always to the global minimum/maximum inside the current feasible region $\widetilde{\mathcal{E}}^k$.

Under these hypotheses, it is possible to prove (see [11]) the following three results:

1. positiveness of the error, which is the necessary condition for the convergence of the error;

2. convergence of the error, that is $e^k \rightarrow 0$ for $k \rightarrow +\infty$;

3. convergence of the probability, that is $\mathbb{P}^{IS^k} \rightarrow \alpha$ for $k \rightarrow +\infty$, and from a conservative direction: $\mathbb{P}^{IS^k} \leq \alpha$ for any $k > 0$.

All these convergence theorems are extremely interesting because they allow to justify the consistency of the method: the elimination of the probability bias in the result of the Single Step algorithm. For clarification purposes, we applied the Multistep algorithm to the simple analytical bi-dimensional function, initially introduced in [4]:

$$\phi(\mathbf{X}) = \phi(X_1, X_2) = \left(\mathrm{e}^{X_1} - 1\right)\left(\mathrm{e}^{X_2/2} - 1\right) \tag{21}$$

where X_1 and X_2 are uncorrelated normally distributed random variables: $X_1 \sim \mathcal{N}(0, 0.29^2)$ and $X_2 \sim \mathcal{N}(0, 0.17^2)$, looking for a probability $\alpha = 10^{-5}$. Comparing fig.1 with the Single Step counterpart reported in [4] gives a greater insight into the convergence process. The ellipsoidal domain is decreased step by step

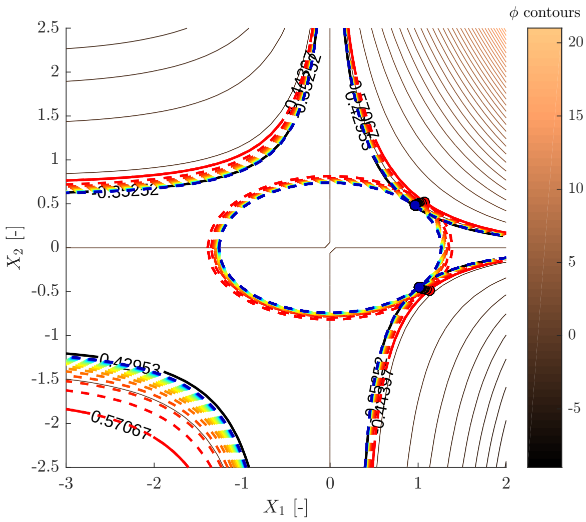
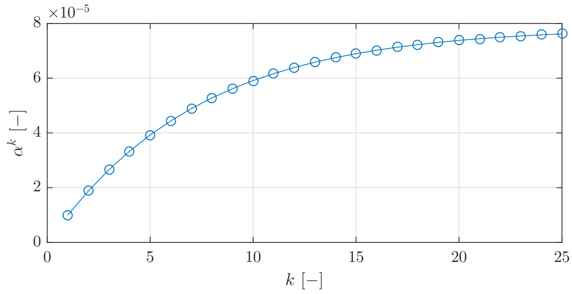


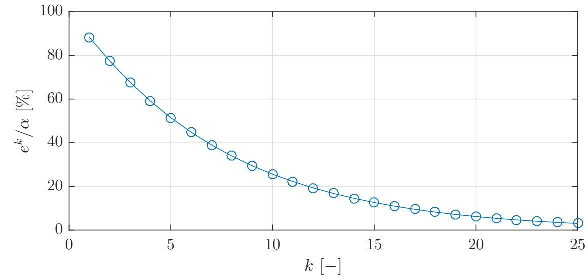
Figure 1: Application of the Multistep algorithm to the exponential analytic function in eq.21.

accordingly to the probability imposed by eq.16, plotted in fig.2a versus the step number. In fig.2b the positiveness and the convergence of the error are reported (theorems 1 and 2) and fig.2c shows the convergence of the probability $\mathbb{P}^{IS^k}$ to α (theorem 3). In fig.2d we see how the output interval $\widetilde{T}_2^k - \widetilde{T}_1^k$ reduces and gets closer to the optimal one $\widetilde{T}_2^{Opt} - \widetilde{T}_1^{Opt}$. However, note that it is not correct to state $\widetilde{T}_1^k$ tends to $\widetilde{T}_1^{Opt}$ and $\widetilde{T}_2^k$ tends to $\widetilde{T}_2^{Opt}$. Indeed a priori it is not guaranteed that the input points giving the optimal interval are equally probable which is instead a constraint imposed by construction by the Inputs' Statistics method.

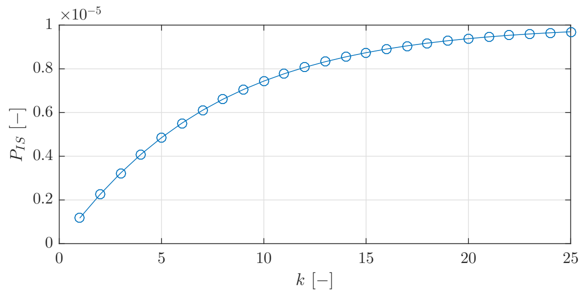
Nevertheless, the convergence cannot be a priori guaranteed in the real case and some problems may arise. First of all, the error may become negative because of a large overestimation of $\mathbb{P}^{IS^k}$. This is typically the case when the used number of samples is too small. To avoid this situation it is suggested to use a number of samples which is big enough to get a significant statistical estimation of a probability on the order of α. The smaller is α, that is, the higher is the rarity to be estimated, the larger is the number of required samples. Glynn, Rubino and Tuffin



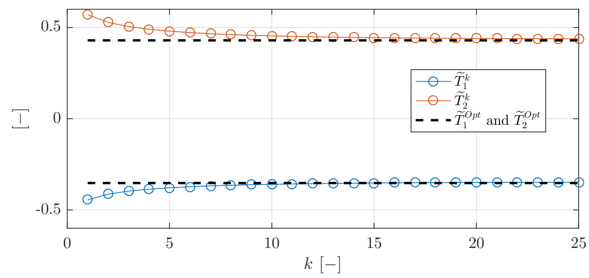
(a) Increase of the probability outside ellipsoid $\widetilde{\mathcal{E}}^k$.



(b) Decrease of the error in percentage w.r.t. α.



(c) Convergence of the probability to α.



(d) Decrease of the threshold interval $\widetilde{T}_2^k - \widetilde{T}_1^k$.

Figure 2: Application of the Multistep algorithm to the exponential analytic function in eq.21. Convergence analysis in ideal condition.

in [7] chapter 4 express the minimum number of required samples as:

$$N > \frac{\left(z_{1-\beta/2}\right)^2}{\mathrm{RE}^2} \frac{1-\alpha}{\alpha} \tag{22}$$

where $z_{1-\beta/2}$ is the $(1-\beta/2)$-quantile of the standard normal distribution $\mathcal{N}(0,1)$: $z_{1-\beta/2} = \Phi_{0,1}^{-1}(1-\beta/2)$, α is the sought probability and RE is the relative error if α is estimated by a CMC method. Rubino and Tuffin in [7] chapter 1 suggest $z = 1.96$, to have 95% CL and a

RE lower than 10% to get a significant reduction of the probability bias.
Another unexpected situation may occur when initially the Single Step algorithm is not so accurate and the first estimation of the output interval is much larger than the optimal one. In this case $\mathbb{P}^{IS^k} \ll \alpha$ and so $\mathbb{P}^{IS^k}$ is highly underestimated using the number of samples suggested in eq.22. Actually, this is not a big issue because it does not affect the progress of the algorithm and the estimation of $\mathbb{P}^{IS^k}$ improves as long as it approaches α. Yet, in this situation the algorithm shows an undesired slowness in the convergence because it proceeds with a step length of the order of α which may require a large number of iterations. It is possible to speed it up through some strategies presented in a sec.4.4.
Finally, a rough convergence may come up when the optimization algorithm is not adapted for the problem under investigation. This produces a non monotone reduction of the output interval along the iterations and consequently an inaccurate estimation of $\mathbb{P}^{IS^k}$, because the contours lines of ϕ are not correctly identified. Unfortunately, there is not a general solution to this issue and a case by case analysis must be carried on. Generally speaking, it is suggested here a predictor corrector strategy, which seems quite efficient specifically for the problem of the safety boxes. As predictor, the *Particle Swarm Optimization* (PSO) [3] is used, which is a swarm algorithm and helps in exploring properly the feasible region. The optimal value given by the PSO is then improved by few iterations of the *Barrier method* [2], a gradient-based method, to converge directly to the solution. In this way, a compromise between exploration of the domain and rapid convergence to a solution is got.

4.3. Computational time discussion

In this section, an analysis of the order of magnitude of the computational time is provided following the same approach used in [4] for the Single Step algorithm expressing it as function of number of transfer function evaluations (t.f.e.). The computational time increases because in the Multistep algorithm the Single Step is applied recursively alternating it with MC simulations. So let's divide the analysis in two parts: MC simulations and optimization processes.
To get a result which is statistically acceptable, the number of input samples N has to be high enough. How much high is prescribed by eq.22. Nevertheless, the Multistep algorithm does not require to propagate all the N samples, but only those that are outside the k-th ellipsoid. In the average, the number of samples that fall outside the first ellipsoid is αN. Going on with the iterations, since α^k increases, more and more points have to be included in the set $\mathcal{I}(\tilde{t}^k, +\infty)$ and so integrated. In the average, after k iterations, it is required the integration of $\alpha^k N$ samples. If α^k is not so far from α, the Multistep Inputs' Statistics shows good performances, but this is not always true. Let's call $N^{MIS_{MC}}$ the number of t.f.e. required by the MC simulations at step k of the Multistep Inputs' Statistics method. Using eq.22, it is possible to express it as

$$N^{MIS_{MC}} > \frac{\alpha^k}{\alpha} \frac{\left(z_{1-\beta/2}\right)^2}{\mathrm{RE}^2} \tag{23}$$

where the term $(1-\alpha)$ has been neglected because for rare event probabilities $\alpha \ll 1$. Now the issue is the estimation of the ratio $\frac{\alpha^k}{\alpha}$ because the other two parameter are usually imposed as: 10% RE and 95% CL, which corresponds to $z_{1-\beta/2} = 1.96$. Quantify the ratio $\frac{\alpha^k}{\alpha}$ is not straightforward and the analysis showed that it depends mainly on three factors:

1. the number of dimensions of the input domain d;
2. the required probability α;
3. the characteristics of the transfer function ϕ.

The three factors are listed in order of importance, that is the ratio increases fast especially when the problem has many uncertain variables. If this ratio is too high, too many samples are required and the Multistep Inputs' Statistics is no more efficient. At this stage of the analysis, it is not available a general analytic formula that is able to estimate a priori the ratio $\frac{\alpha^k}{\alpha}$. So it has been decided to study the problem numerically to provide a parametric estimation of this ratio. Several transfer functions have been evaluated, with different behavior, changing both the required probability and the number of input dimensions. The outcome of this analysis are some look-up tables whose tab.1 is here reported as example. It is suggested to study previously the particular transfer function under investigation, trying to understand its degree of non linearity in the neighbourhood of the optimal condition of the first ellipsoid, and then enter the look-up tables to have an approximate estimation of the ratio $\frac{\alpha^k}{\alpha}$. For example, let us suppose that the transfer function of interest has an exponential behavior with 10 uncertain input variables and that the sought probability is $\alpha = 10^{-5}$. Tab.1 gives that, at convergence, $\frac{\alpha^k}{\alpha}$ should be approximately 710 which implies through eq.23 that $N^{MIS_{MC}}$ is around

Table 1: $\frac{\alpha^k}{\alpha}$ ratio as function of probability level α and number of input variables d for an *exponential* transfer function ϕ

$\frac{\alpha^k}{\alpha}$	$d=2$	$d=4$	$d=6$	$d=8$	$d=10$
$\alpha = 10^{-2}$	1.9	6	13	24	37
$\alpha = 10^{-3}$	2.3	10	28	65	126
$\alpha = 10^{-4}$	2.7	14	51	145	340
$\alpha = 10^{-5}$	2.9	18	80	270	710
$\alpha = 10^{-6}$	3.2	25	121	430	1460

$2.7e^5$. Thus, it is on the order of $\mathcal{O}(10^5)$, which is 100 times faster than the CMC method.
For what concerns the optimization processes, the discussion is not so different with respect to the Single Step algorithm [4]. The number of t.f.e. for a single optimization process can be estimated as $\mathcal{O}(10^2)$. Hence, the total number of t.f.e. required by the optimizer, labeled with $N^{MIS_{Opt}}$, after k iterations of the algorithm is $k\mathcal{O}(10^2)$. Therefore, the total number of t.f.e. required by the Multistep Inputs' Statistics method N^{MIS} may be expresses as sum of the two contributions: $N^{MIS} = N^{MIS_{MC}} + N^{MIS_{Opt}}$, that is:

$$N^{MIS} > \frac{\alpha^k}{\alpha} \frac{\left(z_{1-\beta/2}\right)^2}{\text{RE}^2} + k\mathcal{O}(10^2) \tag{24}$$

Considering the atmospheric dynamics of a re-entry vehicle, the needed time for the single evaluation of the transfer function ϕ is about one second in a standard desktop computer. Let us assume to have for instance 10 uncertain input variables and to be interested in a probability of 10^{-5}. Then, approximately the ratio $\frac{\alpha^k}{\alpha}$ is on the order of 1000 within some tens of iterations. Therefore, $T_{Tot_{MIS}} = \mathcal{O}(10^6 s) = \mathcal{O}(days)$ which is comparable with the other methods that can be found in the literature (see for instance Haya in [8]) and 10 times faster than the simple CMC.

4.4. Speed-up strategy

It should be clear that the MC contribution in eq.24 is totally independent by the number of iterations necessary for the convergence. Indeed, once N is fixed, the number of samples between the first ellipsoid $\widetilde{\mathcal{E}}^1$ and last ellipsoid $\widetilde{\mathcal{E}}^k$ is established and it does not matter if they are integrated all at once or step by step. At this stage of the analysis, it is not available yet a method to speed up this contribution of the computational time. If α^k is much larger than α, the Multistep may not be efficient when compared with other methods in literature. It has been noted that this is the case when the number of dimensions is large. This problem is usually termed *Curse of Dimensionality* (CoD) [1] and can be found in several other methods, especially in the reliability based. Nevertheless, the Multistep Inputs' Statistics is still far from being mature and a future development of the work may introduce a more clever method, instead of the simple CMC, to characterize the domain outside $\widetilde{\mathcal{E}}^k$.
The second contribution to the computational time in eq.24 is given by $N^{MIS_{Opt}}$. Differently with respect to before, it is directly proportional to the number of iterations, because at each step two optimization processes are required. Using eq.16 to compute the successive ellipsoid, the algorithm progresses with maximum step length on the order of α. If the initial approximation is not accurate, that is $\alpha^k >> \alpha$, then too many steps are required to converge to α^k and it may happen that $N^{MIS_{Opt}}$ is even larger than $N^{MIS_{MC}}$ strongly spoiling the performances of the algorithm. The proposed solution to cope with this problem consists in multiply the error in eq.16 by a suitable coefficient, which increases at the beginning to accelerate the algorithm and reduces to one with the error in order to slow down the algorithm when the convergence is reached. For instance the following empirical formulations have been proved to be efficient in several situations:

$$\alpha^{k+1} = \alpha^k + \max\left[\frac{\alpha^k}{\alpha}\left(\frac{e^k}{\alpha}\right)^p, 1\right] e^k \tag{25}$$

or also

$$\alpha^{k+1} = \alpha^k + \max\left[\left(\frac{\alpha^k}{\alpha}\right)^{\left(\frac{e^k}{\alpha}\right)^p}, 1\right] e^k \tag{26}$$

where the parameter p has to be chosen for the specific case, but usually $p = 1$ or $p = 2$ gives good convergence speed. The formulation in eq.25 is suggested for very low probabilities (e.g. 10^{-5} or 10^{-6}) whereas eq.26 is efficient for relatively low probabilities as 10^{-2}. The analysis showed that using this empirical but efficient approach usually no more than 30 iterations of the algorithm are required, even with very unlucky problems having $\alpha^k >> \alpha$.

4.5. Algorithm for generic functions

The objective of this section is to extend the algorithm to general transfer functions ϕ which do not enjoy any more the property of monotonicity. ϕ can be characterized by several local maxima and minima.
Since ϕ is not monotone, it is no more guaranteed that the global maximum and minimum belong to the border of the current ellipsoid $\widetilde{\mathcal{E}}^k$. They may be also inside it. If this happens, it is necessary to decrease $\widetilde{\mathcal{E}}^k$ such that it is tangent to the furthest between the minimum and maximum with respect to the origin. In this way, the problem is transformed as if ϕ was monotone and the algorithm works exactly in the same manner. It reduces again and again $\widetilde{\mathcal{E}}^k$ such to get two thresholds, $\widetilde{T}_1^k$ and $\widetilde{T}_2^k$, that individuate contours of ϕ including a volume corresponding to the sought probability α.

This is true as long as another point inside $\widetilde{\mathcal{E}}^k$ becomes possibly the current global minimum/maximum. This means that $\widetilde{\mathcal{E}}^k$ is still too conservative and has to be decreased to be tangent again to the new global minimum/maximum to re-transform the problem in the monotone one. This is repeated again and again until the contours of ϕ associated to $\widetilde{T}_1^k$ and $\widetilde{T}_2^k$ include the required probability α.
Note that to guarantee the correct convergence of the algorithm, it is important to keep as tangent either the minimum or the maximum as function of their distance with respect to the origin. The furthest between the two must be selected. This because otherwise there is the risk from the other side to exclude from $\widetilde{\mathcal{E}}^k$ a volume including a

probability which could be higher than α and the error becomes thus negative.
The generalization is computationally inexpensive and it is performed as indicated in algorithm 4.2, intended to be added just below line 14 of algorithm 4.1. Note, in particular, that performing the steps in algorithm 4.2 when ϕ is monotone does not have any effect. This is clearly because if the transfer function is monotone, both $\widetilde{T}_1^k$ and $\widetilde{T}_2^k$ belong already to the border of $\widetilde{\mathcal{E}}^k$ and so algorithm 4.2 just returns the same $\widetilde{\mathcal{E}}^k$. Therefore, algorithm 4.1 plus algorithm 4.2 can be applied to a generic transfer function whether or not it is monotone. This is extremely useful because generally the characteristics of ϕ are not known a priori.

Algorithm 4.2: Add-on to algorithm 4.1 for the generalization of the Multistep Inputs' Statistics method to generic transfer functions

get input points corresponding to $\widetilde{T}_1^k$ and $\widetilde{T}_2^k$: $\mathbf{X}_{\widetilde{T}_1^k}$ and $\mathbf{X}_{\widetilde{T}_2^k}$;

compute

 ellipsoids $\widetilde{\mathcal{E}}_{\widetilde{T}_1^k}$ and $\widetilde{\mathcal{E}}_{\widetilde{T}_2^k}$ tangent to $\widetilde{T}_1^k$ and $\widetilde{T}_2^k$, respectively, defined by the square radius $\tilde{t}_{\widetilde{T}_1^k}$ and $\tilde{t}_{\widetilde{T}_2^k}$ as:

$$\tilde{t}_{\widetilde{T}_1^k} = \left(\mathbf{X}_{\widetilde{T}_1^k} - \boldsymbol{\mu}\right)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \left(\mathbf{X}_{\widetilde{T}_1^k} - \boldsymbol{\mu}\right)$$

$$\tilde{t}_{\widetilde{T}_2^k} = \left(\mathbf{X}_{\widetilde{T}_2^k} - \boldsymbol{\mu}\right)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \left(\mathbf{X}_{\widetilde{T}_2^k} - \boldsymbol{\mu}\right)$$

update $\widetilde{\mathcal{E}}^k$ as given by: $\tilde{t}^k = \max\left[\tilde{t}_{\widetilde{T}_1^k}, \tilde{t}_{\widetilde{T}_2^k}\right]$;

5. APPLICATION TO THE ATV-GL SHALLOW RE-ENTRY

The safety boxes geometry is approximated having a rectangular shape and it is usually described in terms of along track and cross track dimensions, as shown in fig.3. The Aimed Impact Point (AIP) is defined as a reference target for a virtual fragment with arbitrary mean characteristics and is used to design the deorbitation manoeuvre plan. The cross track range (C-range) is considered constant for the case of ATV re-entry being much smaller than the along track range. It is taken as a fixed deviation of $\pm 100\ km$ (majoring value) with respect to the ground track, independently from the probability level. The along track range (A-range) is usually divided in up-track range (U-range), if the fragment falls before the AIP, and down-track range (D-range), if the fragment falls after the AIP. They are computed as curvilinear integrations along the ground track from the AIP to the projection of the impact point over the ground track. Therefore, a generic impact point is described by the signed along track distance with respect to the AIP, negative if up-track and positive if down-track: this is the output

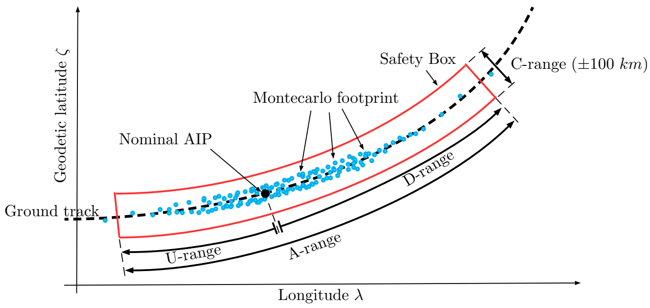


Figure 3: Schematic representation of safety boxes geometry.

Y of the transfer function ϕ. Concerning the problem of the controlled destructive atmospheric re-entry of a spacecraft, the d-dimensional input vector $\mathbf{X}$ usually includes: initial conditions, manoeuvres performances, vehicle characteristics, atmospheric parameters, fragmentation characteristics. All these variables have to be statistically characterized *a priori*. This characterization is not univocal and has to be done case by case, usually using engineering judgement and physical considerations as well as additional studies on the specific variable of the problem. For the specific case of ATV, these information have beed taken from external sources which are the *ESA Space Debris Mitigation Compliance Verification Guidelines* in [6] and the ATV-CC Flight Dynamics work published in [12] for ATV Jules Verne (ATV-JV) and in [9] for ATV Georges Lamaître (ATV-GL). In total, the input variables are 10: 4 normally distributed and 6 uniformly distributed. No correlation among variables is considered. Finally, the suitable input-output formulation to apply a statistical method for the computation of the safety boxes is schematically illustrated in fig.4.

The convergence example of the Inputs' Statistics method here reported concerns the computation of the Safety Re-entry Area (SRA), associated to the probability $\alpha = 10^{-5}$ for the SEDIA ([9]) fragments #3 (short fragment) and #2 (long fragment) for ATV-GL, whose characteristics are collected in tab.2. Since 10 uncertain variables are considered, the Single Step Inputs' Statistics method does not give accurate results and they have to be refined with the Multistep algorithm. The convergence results are reported in the fig.5 and in fig.6.

Table 2: SEDIA fragments used in this work. From ref.[9]

Fragment #	*Mass*	ΔV_{Expl}	*Min* β	*Max* β
2	470	1	3952	5512
3	60	26	2	5

It is really important to use the speed up strategies introduced in sec.4.4 to keep a low number of iterations. For SRA, eq.25 with $p = 1$ shows a good efficiency, giving a convergence in less than 30 iterations. In addition, accordingly to eq.22, for a CL of 95% and a

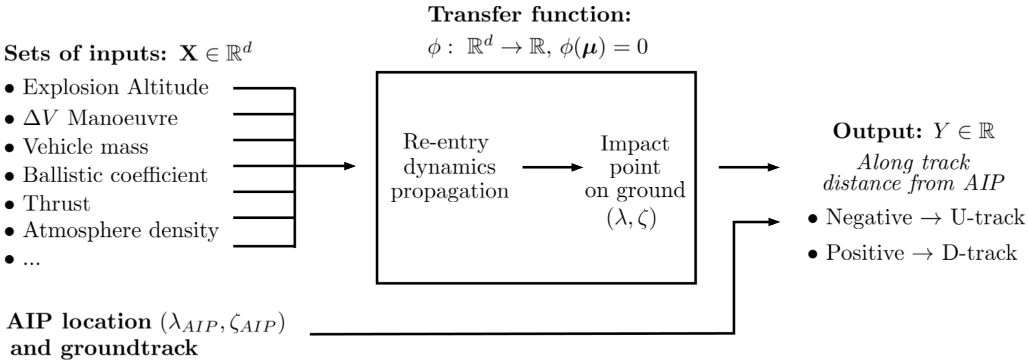


Figure 4: Schematic illustration of the input-output formulation for the safety boxes computation.

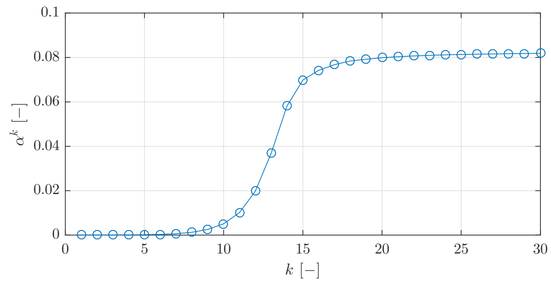
Table 3: Summary of the results given by the Multistep Inputs' Statistics method and compared with CMC and MCS+PoT.

Fragment #	SRA dimension $\widetilde{T}_2^k - \widetilde{T}_1^k$	α	RE	CL	N	$N^{MIS_{MC}}$	$\frac{N^{MIS_{MC}}}{N}\%$	Error w.r.t. MSC+PoT
2	2980	10^{-5}	10%	95%	$3.8e^7$	$3.1e^6$	8%	+1.17%
3	2605	10^{-5}	10%	95%	$3.8e^7$	$2.3e^6$	6%	+2.45%

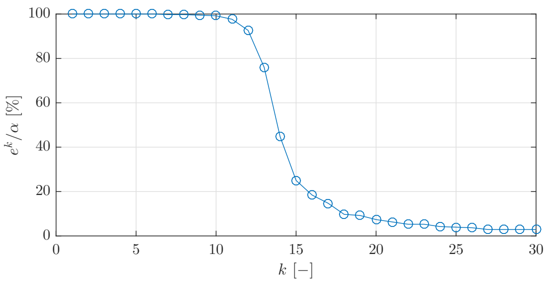
relative error of 10%, the CMC method shall perform the integration of $3.8e^7$ for SRA. For the same confidence level and relative error, the Multistep algorithm requires only 6%-8% of the samples required by CMC method, that is $N^{MIS_{MC}} = 3.1e^6$ for fragment #2 and $N^{MIS_{MC}} = 2.3e^6$ for fragment #3. The results in terms of safety boxes dimensions ($\widetilde{T}_2^k - \widetilde{T}_1^k$) are compared with respect to those given by a MC simulation plus Peaks over Threshold (MCS+PoT) method described in [12]. It is possible to observe a good matching of the results with maximum discrepancy of 2.45% from the conservative direction. All the results are summarized in tab.3. Superimposing the SRA of the single fragments accordingly to the specific AIP, we get the overall safety box of the entire spacecraft. The overall SRA is 4855 km long, shown in fig.7.

6. CONCLUSION AND FUTURE WORK

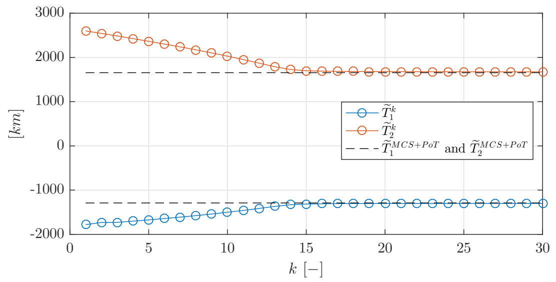
The innovative method described in this work is interesting thanks to its characteristics of conservatism, convergence and speed. It has been named Inputs' Statistics, because it was developed originally with the idea to investigate the inputs domain rather than the output domain of the tranfer function in order to extrapolate precious and fast information about the probability estimate without the need to generate the cloud sampling.
In particular, the Single Step algorithm is always able to provide a conservative result with practically negligible computational time. Introducing some simplifications in the identification of the failure domain, it exploits a deterministic optimization process to find a sub-optimal but immediate solution of the problem. Under some hypothesis, this result may be also within the target accuracy and retained as a good solution of the problem. When it is not the case, it can be considered as a preliminary, sizing and inexpensive solution to be improved in subsequent phase. Indeed, the Single Step is a biased algorithm. This bias introduced in the probability estimate is in some cases too large and the result has to be refined to be acceptable.
The Multistep algorithm achieves this objective finding a compromise between the Single Step algorithm's speed and the sampling techniques' accuracy. More specifically, it alternates optimization processes, used to locate the current most probable failure point inside an ellipsoidal domain, with MC simulations, necessary to characterize the input domain outside the ellipsoid. The error committed at each step is thus identified and subtracted at the successive iteration to get a convergent process that reduces to zero the probability bias.
This approach requires to sample only the least dense region of the input space with the consequent benefit from the computational time point of view. Indeed, the total computational budget can be divided into two contributions. Thanks to simple but effective speed up strategies presented in this work, the contribution associated to the optimization processes is small and generally negligible when compared with the contribution associated to the MC simulations. As a consequence, it is possible to state that the Multistep algorithm is faster then the CMC method with the same confidence level and relative error. This is simply proved by construction and is valid whenever the first contribution is negligible. In addition, the established iterative process keeps the intrinsic characteristic of approaching the optimal result from the conservative direction. This is extremely important when the probability assessment concerns safety issues because an error on the wrong side must be completely avoided. The method gives also the possibility to the analyst to stop the process when the desired tolerance is met both in terms



(a) Increase of the probability outside ellipsoid $\widetilde{\mathcal{E}}^k$.

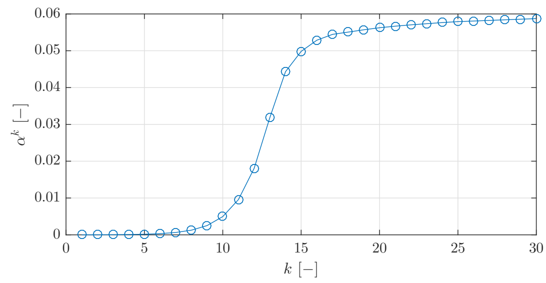


(b) Decrease of the error in percentage w.r.t. α.

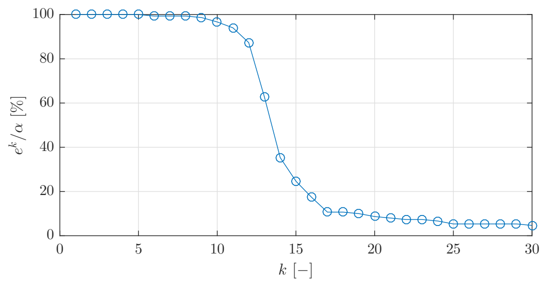


(c) Decrease of the threshold interval $\widetilde{T}_2^k - \widetilde{T}_1^k$.

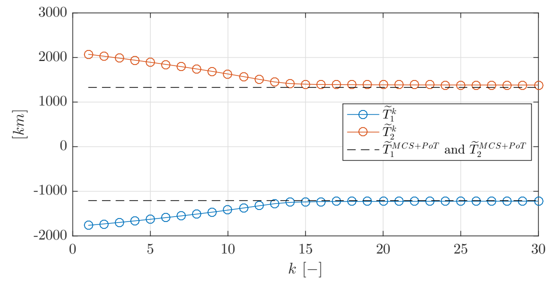
Figure 5: SRA ($\alpha = 10^{-5}$) for fragment #2. Convergence analysis with speed up in eq.25 with $p = 1$. Starting with $N = 3.8e^7$.



(a) Increase of the probability outside ellipsoid $\widetilde{\mathcal{E}}^k$.



(b) Decrease of the error in percentage w.r.t. α.



(c) Decrease of the threshold interval $\widetilde{T}_2^k - \widetilde{T}_1^k$.

Figure 6: SRA ($\alpha = 10^{-5}$) for fragment #3. Convergence analysis with speed up in eq.25 with $p = 1$. Starting with $N = 3.8e^7$.

of probability or output interval size.

The main identified limitation of the Multistep algorithm regards the Curse of Dimensionality (CoD): the convenience of the method decreases with the increase of the number of dimensions of the problem. The CoD is a recurrent issue also in many reliability based algorithms which cannot cope with surrogate models when the problem has more than ten or twelve dimensions. Nevertheless, the Inputs' Statistics method is far from being mature and a large room for improvement still exists. This is due to the fact that is uses the CMC method to characterize the domain outside the ellipsoidal regions. The CMC method is simple and intuitive but many other more sophisticated techniques have been developed in the literature which exploit smart strategies to probe the failure domain or reduce the MC variance. This is particularly the case of Importance sampling and Stratified sampling. Coupling one of these techniques with the Multistep algorithm may strongly enhance its performances because on one hand the domain is sampled in a much more efficient way and on the other the domain to be sampled is still the least dense. This possibility has not been thoroughly investigated in this work and it is remanded to future developments.

In addition, several other open points have been be identified. It is extremely important to characterize the method in terms of robustness and reliability with respect to rarity. An estimator is said to be robust if its quality (i.e. the gap with respect to the true value) is not significantly affected when the required probability tends to zero. A reliable estimator is an estimator for which the confidence interval coverage does not deteriorate when the required probability tends to zero. Those two notions are different: one focuses on the error itself, the other on the quality of the error estimation. It is possible to state that a method is robust if it enjoys the Bounded Relative Error (BRE) property or, in a weaker condition, if it has the Logarithmic Efficiency (LE). Similarly, it is said that an estimator is reliable if it enjoys the Bounded Normal Approximation (BNA) property. Indeed, the variance is often even more sensitive to the rarity than the estimation of the mean itself [7]. The Inputs' Statistics method has been

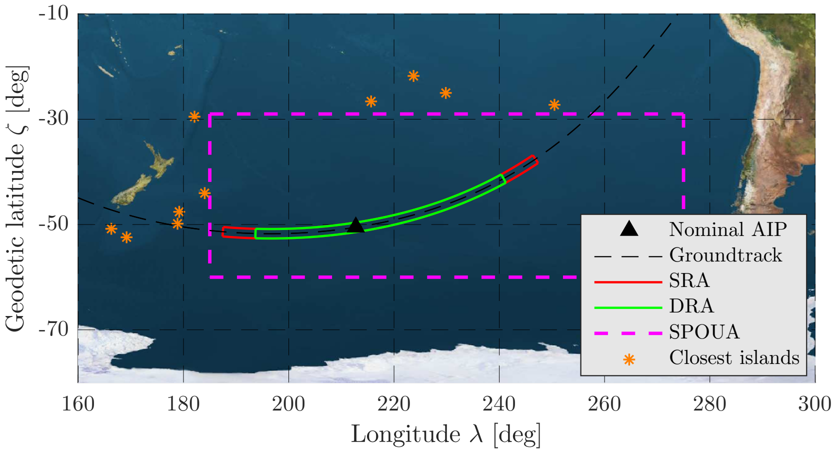


Figure 7: Overall DRA and SRA, as superimposition of the fragments' safety boxes.

characterized in terms of relative error and confidence interval. This was relatively straightforward because those definitions directly come from those corresponding to the CMC method. However, in this work no information on how these characteristics behave when the required probability tends to zero is given. This is an essential point that shall be accurately studied in a future development of the work in order to make the method ready for a real case application.

Another open question concerns the possibility to exploit the information on the location of the input points corresponding to the current local minimum/maximum inside the ellipsoid to identify the existence of privileged input directions. In this respect, what is the impact of such privileged directions on the convergence speed and on the bias correction of the Multistep algorithm? We are confident that answering to this question may give a greater insight into the method and enhance its performances.

Finally, it may worth to discuss briefly about other problems which may be object of application of the Inputs' Statistics method. It is a statistical and numerical scheme aiming at computing two thresholds $T_1 < 0$ and $T_2 > 0$ such that the probability $\mathbb{P}$ that the output is smaller than T_1 or larger than T_2 is smaller or equal to an assigned value. This is the case of the computation of the safety boxes, where the probability is fixed by the international space law to guarantee ground population safety and the two thresholds are up-track and down-track distances defining the safety box. Therefore, it may be used whenever it is required to analyse: the destructive controlled re-entry of large structures, including in particular the International Space Station (ISS) and the ISS visiting vehicles at their End of Life (EoL); the destructive re-entry of large uncooperative satellites orbiting LEO and MEO as conclusive event of the Active Debris Removal (ADR) technology (e.deorbit); the destructive controlled re-entry of last stages of launchers; the uncontrolled atmospheric impact of small asteroids or space objects and debris; the safety compliance verification in case of failure of controlled re-entry spacecraft which are designed to withstand the atmospheric re-entry.

The method can be also specified to the case with a single threshold T and the probability that the output exceeds it is sought. Consider, for instance, the scenario of a space station orbiting in cislunar space, which would undergo numerous rendezvous and docking/undocking activities because the arrival/departure of visiting vehicles, even in unmanned scenarios, from/to Earth, Moon and whichever location in space we can reasonably conceive as a target of interest. The first safety constraint that the mission analyst must assure is that, within a predetermined confidence bound, no collision occurs between the visiting vehicle and the space station in case of aborted manoeuvres and/or off-nominal situations. Initial state vector errors, navigation errors, guidance errors, manoeuvres performances are some of all the dispersions that affect the problem of rendezvous with the station. The output of the transfer function is the scalar minimum distance from the target (i.e. the space station) within a prefixed interval of time (usually 24 hours) of safe coasting trajectory required for any planned burn along the rendezvous trajectory.

The Inputs' Statistics method may be applied also when the problem is reversed: it is required to compute the probability $\mathbb{P}$ such that the output is smaller or larger than a given threshold T. This is in particular the case of the estimation of the probability that a satellite and a space debris collide. The input vector contains the components of the satellite and debris measurements errors on their position and velocity vectors and the output is the minimum distance between the two objects during a given time span. It is required to estimate the probability $\mathbb{P}$ that the distance is smaller than a conflict distance, corresponding to the threshold T of the problem.

In a very different field, we are currently investigating how the method could be use in financial analysis, where Monte Carlo simulations are widely used to assess the risk of future unexpected returns (typically negative return on investment) depending on financial and economic factors which influence the stock markets.

Formalizing the Inputs' Statistics method in the reversed version of the problem and studying its performance and convergence characteristics when applied on all these other real case applications is another main open point which surely deserves to be analysed in the future.

REFERENCES

1. Bellman, R., Adaptive control processes: A guided tour. *Princeton University Press*, 1961.
2. Byrd, R. H., Gilbert, J. C., and Nocedal, J., A trust region method based on interior point techniques for nonlinear programming. *Mathematical Programming*, 89(1):149-185, 2000
3. Clerc, M., Particle swarm optimization, volume 93. *John Wiley & Sons*, 2010
4. De Pasquale, E., Rafano Carná, S. F., Arzel, L., Lavagna, M., Innovative method for the computation of Safety Re-entry Area based on the probability of uncertainties in the input parameters, *6th International Conference on Astrodynamics Tools and Techniques*, 2015 Darmstadt, Germany
5. De Pasquale, E., Steinkopf, M., Francillout, L., Escane, I., Benarroche, P., Delattre, S. and Arzel, L., ATV Shallow Reentry Mission: Challenges and Achievements. In *8thAerothermodynamics Symposium for Space Vehicles*, 2-6 March 2015, Lisbon, Portugal.
6. ESA Space Debris Mitigation Compliance Verification Guidelines, *ESSB-HB-U-002-Issue1*, from Inter-Agency Space Debris Coordination Committee website `http://www.iadc-online.org/`.
7. Glynn, P. W., Rubino, G., Tuffin, B., Robustness properties and confidence interval reliability issues, in collection of Rare event simulation using Monte Carlo methods, *Wiley Online Library*, 2009
8. Haya, R., Computation of IXV Safety Footprint boundaries. *CCT ORB, Seminar on rare events, extreme values, extreme impacts applied to flight dynamics*, ISAE-SUPAERO, Toulouse (FR), 2015.
9. Hourtolle, C., Delattre, S., Goester, JF., De Pasquale, E., Studies about a Shallow Re-entry for ATV-5 *25th International Symposium on Space Flight Dynamics*, Munich, Germany, 2015
10. Morio, J., Balesdent, M., Estimation of Rare Event Probabilities in Complex Aerospace and Other Systems, *Woodhead Publishing*, 2015
11. Rafano Carná, S. F., The Inputs' Statistics method: an innovative method for the computation of the re-entry safety boxes with application to the ATV-GL Shallow Re-entry, *Master Thesis, Politecnico di Milano*, July 2016.
12. Renaud, F., Martin, T., Safety Boxes Sizing for Controlled Re-Entry Application to Automated Transfer Vehicle (ATV). In Proceedings of the *3rd IAASS Conference: building a safer space together*, Rome, Italy, 2008
13. Technical Regulations of the French Space Act. Décrets, arrêtés, circulaires - Textes Généraux, Art.23, Ministère de l'Enseignement Supérieur et de le Rcherche, Journal officiel de la République Français, 31 May 2011, `http://www.iadc-online.org/`.