INNOVATIVE METHOD FOR THE COMPUTATION OF SAFETY RE-ENTRY AREA BASED ON THE PROBABILITY OF UNCERTAINTIES IN THE INPUT PARAMETERS

Emilio De Pasquale [(1)], Simone Flavio Rafano Carnà [(2)], Laurent Arzel[(3)], Michèle Lavagna [(4)]

[(1)] European Space Agency, Eurostar Neo Platform Manager, CNES Bat. Auger NEOSAT Project, 18, av. Edouard Belin, 31401 Toulouse Cedex 9, France.
Email: Emilio.De.Pasquale@esa.int
[(2)] Politecnico di Milano, Space Engineering MS student, Aerospace Science and Technology Department, via La Masa 34, 20156, Milan, Italy.
Email: simone.rafano@mail.polimi.it
[(3)] PwC Strategy&, Apollolaan 151, 1077 AR, Amsterdam, The Netherlands
Email: laurent.arzel@strategyand.fr.pwc.com
[(4)] Politecnico di Milano, Associate Professor, Aerospace Science and Technology Department, via La Masa 34, 20156, Milan, Italy
Email: michelle.lavagna@polimi.it

ABSTRACT

The risk reduction measures required for the reentry of a spacecraft at its end of life are regulated in Europe by requirements documented in Space Agencies' instructions and guidelines. According to European Space Agency guidelines [1] the Safety Re-entry Area (SRA) delimits the area where the debris should be enclosed with a probability of 99,999%. The computation and design of SRA is required in the risk assessment of a significant number of space missions like spacecraft in low Earth orbits at its end of life and last stages of launchers that shall be controlled to a destructive reentry.
This paper describes an innovative method to compute the SRA, considering that the input models and its uncertainties are well defined. The method focuses on the statistical distribution of the input parameters uncertainties contrary to classical methods that generates a large number of impact points with Monte Carlo simulation and processes the outputs of this computation.
The paper presents the results of the computation of the SRA for the re-entry in the South Pacific Ocean of the fifth and last European ISS cargo (ATV).
The SRA is also assessed using the classical Monte Carlo approach and the results are compared with this innovative approach highlighting advantages and drawbacks in terms of accuracy, level of conservatism and computational time.

Index Terms— Re-entry, Safety, Statistic, SRA, ATV

1. INTRODUCTION

According to the French Space law, "*the operator responsible of a spacecraft controlled reentry shall identify and compute the impact zones of the spacecraft and its fragments for all controlled reentry on the Earth with a probability respectively of 99% and 99,999% taking into account the uncertainties associated to the parameters of the reentry trajectories*".

Safety boxes are containment contours on the 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 [1]. 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 of interest (10^{-5}) associated to its contour, which makes quite difficult and inaccurate the use of classic statistical techniques. The state of arte methods use Monte Carlo analysis to estimate the SRA, but the number of samples generated determines the accuracy and confidence level achieved while it is proportional to the computational time needed. Therefore the quantile of interest is estimated by extrapolating the density of probability obtained by the Monte Carlo simulation with a number of outputs samples smaller than required [2].

The method described in this paper, hereafter called Inputs' Statistics method, works on the domain of the input uncertainties, being *a priori* statistically modelled, and, consequently, it does not require generating Monte Carlo simulations and processing the statistics of the output. The probability of interest is computed integrating the known multivariate density function of the input parameters and, then, an optimization process is used to find the minimum and maximum within a reduced set of inputs corresponding to an aimed probability.

Three advantages can be recognized:

1) the probability of the impact contour is directly derived by the causes of the impacts (initial conditions and model dispersions) highlighting its physical explanation (the impact is located in a specific point if a set of initial conditions happen);
2) a large amount of computational time is saved since a Monte Carlo simulations are not required;
3) the level of probability can be set arbitrarily small without having any impact on the computational time because it is not sensitive to the probability level to be achieved.

The drawback of the method is related to the simplification that is introduced in the identification of the overall input domain associated to the aimed probability. This simplification may lead to an overestimation of the size of the box but it provides always a conservative solution to the problem. Considering that there is no worst case for a statistical distribution, this assessment can be also retained as a conservative envelope of the optimal solution.

The Input's Statistics method is suitable for many future applications taking advantage of its computational speed and reliability: the destructive controlled re-entry of large structures, including in particular the International Space Station (ISS) and the ISS visiting vehicles at its 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.

2. STATE OF THE ART REVIEW

The theory and applications of extremes and rare events have received an increasing interest. This is primarily due to its practical relevance which is well recognized in different fields such as insurance, finance, engineering, environmental sciences and hydrology [3].

Indeed, extensive work has been devoted in the past to address both the rare events estimation and the extremely low probability computation, which led to the definition of several efficient approaches. A comprehensive but brief overview of the most relevant (for historical or performance reasons) algorithms and methods is provided by Morio et al. in [4]. For a more exhaustive and detailed description, containing the mathematical formulation and several practical applications from the aerospace world, it is suggested the book of the same authors in [5] as well as Falk at al. in [3] and by Kotz and Nadarajah in [6].

It is usually difficult to determine which algorithm is the most appropriate for a specific problem. Indeed, when a very low probability level is sought, often the computational burden is the major concern of the analyst but it is also extremely important to keep always an adequate level of confidence on the final estimated parameter. Thus, in choosing the algorithm for a specific problem, these are the two main aspects that have to be considered and often a compromise between them is required.

Crude Monte Carlo Methods [7,8] allow an easy implementation and do not require any analytical characterization of the transfer function but converge slowly and need a very large number of generated samples for the computation of small probabilities.

In order to minimize the variance and the computational time of the Monte Carlo estimator, Importance sampling techniques [9,10] can be used, emphasizing the sampling of particular values that have more impact in the estimation of a given parameter with respect to other values.

Another approach, efficient to estimate very low probability levels ($P < 10^{-6}$), is adopted by the Adaptive splitting techniques [4,9], which consider supersets of the input region corresponding to the required probability in such a way that the probability of each of these sets can be estimated with a reasonable simulation budget. Then, the probability on the final smallest set is computed by the product of the conditional probabilities.

Some methods have been developed in structural reliability that work on the inputs domain rather than building up an output statistics, similarly to the Inputs' Statistics method presented in this paper. The most simple are the first and second order reliability methods (FORM/SORM) [11,12], which use an optimization process to locate the most probable failure point where they approximate linearly (first order) or quadratically (second order) the contour lines of the transfer function. Accuracy problems can arise when the transfer function has unknown highly non-linear contour surfaces or when the most probable point is not unique [13], but the computational time is extremely short especially if it is compared with sampling-based approaches.

The Extreme value theory [14] provides appropriate distributions to fit extreme events. Generally, there are two related ways of working with extremes in simulated/real data. The first approach, called Block maxima method [15], considers the maximum the variable takes in non-overlapping periods of equal size, for example months or years and fits them using the Generalized Distribution Function over the maxima values inside each block. The second approach, called the Peak over threshold (PoT) method [2,15], considers the distribution of exceedances over a given (high) threshold and fits them using the Generalized Pareto Distribution. As Ferreira and De Haan explain in [16], the PoT method picks up all ''relevant" high observations. The Block maxima method on the one hand

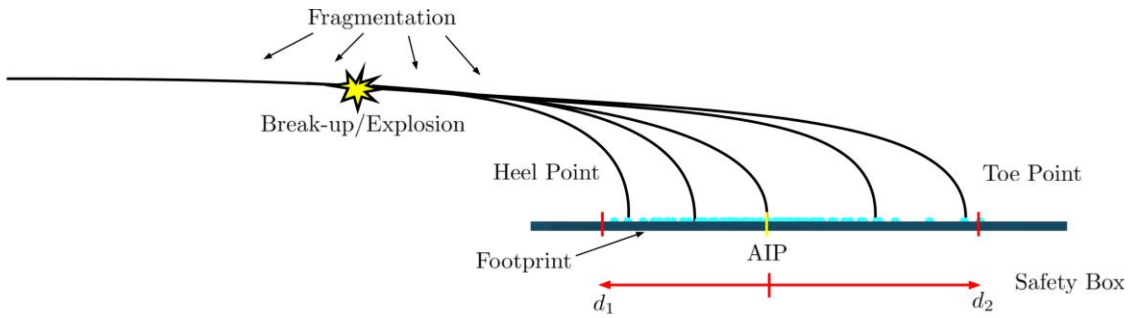

Figure 1: Illustration of atmospheric fragmentation of a space vehicle

misses some of these high observations and might retain some lower observations. Hence the PoT seems to make better use of the available information and it is often the choice in recent applications [15]. The Block maxima method is preferable with respect to the PoT method when the observations are not exactly independent and identically distributed [17] and when only the maxima values are known [18]. As drawback, exactly as for the PoT method, it is difficult to estimate the error on the estimation of the probability. In this work, the PoT method is selected to compare the results provided by the Inputs' Statistics method for a realistic case of atmospheric re-entry of a space vehicle, the Automated Transfer Vehicle. The PoT method has been specifically implemented based on the work of Renaud and Martin in [2] and then recalled by Hourtelle et al. in [19].

3. INPUTS' STATISTICS METHOD

3.1. Definition of the problem

Let group all the uncertain parameters and initial conditions affecting the atmospheric re-entry dynamics of a space vehicles in the vector $\mathbf{X}$. If all the parameters in $\mathbf{X}$ were known, a numerical dynamic propagator could estimate the impact location of the fragment over the Earth's surface. The impact point is identified as the along track signed distance (positive if "in front" and negative if "behind") with respect to a reference point, called Aimed Impact Point (AIP). (see fig.1) This distance is a scalar value Y that is here called ''output''. The deterministic transfer function $f : \mathbb{R}^n \dashrightarrow \mathbb{R}$ describes the transformation from the input $\mathbf{X}$ to the output Y. In addition, let introduce the multivariate probability density function (PDF) associated to the random vector of the inputs $\mathbf{X}$: $p = \mathrm{pdf}_{\mathrm{X}}(\mathbf{X})$. The AIP is computed integrating the re-entry dynamics on the mean value $\boldsymbol{\mu}$ of the input vector $\mathbf{X}$. So accordingly: $f(\boldsymbol{\mu}) = 0$.

The original problem, as cited in the international space law [1], requires the design of the controlled initial conditions (part of the vector of the inputs $\mathbf{X}$) such that the two thresholds d_1 and d_2, with $d_1 < 0$ and $d_2 > 0$ are outside inhabited area. The probability that a fragment impact point falls *outside* the interval $[d_1, d_2]$, across the AIP, shall be less or equal to a given probability level (e.g. $10^{-2}, 10^{-5}$). The two thresholds d_1 and d_2 identify the along track size and location of the safety box and are functions of the chosen probability level. Since the cross-track boundaries of safety boxes are small with respect to the along track ones, they are simply designed as a fixed deviation of +/-100 km (majoring value) from the ground track [2]. Hence, the problem addressed in this paper is the computation of the safety box here identified as the interval $[d_1, d_2]$ and it can be formulated mathematically as the computation of the two thresholds d_1 and d_2, with $d_1 < 0$ and $d_2 > 0$, such that

$$1 - \mathrm{P}(d_1 \leq \mathrm{Y} \leq d_2) \leq \alpha \tag{1}$$

where α is the required probability level. This is the typical formulation used by the state of the art techniques. The main issue of this approach is the computational time. Indeed, since the output statistical distribution of Y is not known a priori, it has to be numerically built using Monte Carlo simulations, which cause a sharp increase of the computational time with the increase of granularity required to estimate the statistical distribution for very low values of the probability density function.

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 modelled using physical considerations and engineering judgment. Once the PDF is defined, the input domain is fully characterized. The probability can be *exactly* computed in the region Ω of the input domain, corresponding to the contour lines of the transfer function f relative to d_1 and d_2, once identified. Accordingly, the problem can be alternatively formulated as the computation of the region Ω such that:

$$1-\int_{\Omega} p(\mathbf{x})\,\mathrm{d}\mathbf{x} \le \alpha \tag{2}$$

with $\Omega = \{\mathbf{X} \in \mathbb{R}^n : d_1 \le f(\mathbf{X}) \le d_2\}$. This is an intrinsically different approach and it is the basis of the Inputs' Statistics method.

Nevertheless, Ω is not unique because two unknowns must be selected: d_1 and d_2 and only one inequality is available from the condition of the required probability level α. Among the family of all the possible choices of Ω, the engineering design of the safety box has as objective to identify that particular Ω which minimizes the distance between the two values d_1 and d_2. This optimal choice of Ω is here called Ω^{Opt}, corresponding to optimal couple of thresholds: d_1^{Opt} and d_2^{Opt}. Since the distance between d_1 and d_2 increases with decreasing required probability level, Ω^{Opt} would be computed with the less restrictive as possible constraint, i.e. equality instead of inequality:

$$1-\int_{\Omega^{\mathrm{Opt}}} p(\mathbf{x})\,\mathrm{d}\mathbf{x} = \alpha \tag{3}$$

Thus, d_1^{Opt} and d_2^{Opt}give the smallest dimensions of the safety box, which satisfies the constraint on the probability level. Since f is a multidimensional dynamic propagator and its contour surfaces are not identified in fast enough computational sequence (e.g. there is not an analytical explicit formulation), the direct computation of the Ω family and especially of Ω^{Opt} is not practically feasible due to computational time limitations.

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

$$\widetilde{\Omega} \cong \Omega^{\mathrm{Opt}} \tag{4}$$

Consequently, the probability P_{IS} of Inputs' Statistics is:

$$\mathrm{P}_{IS} = 1-\int_{\widetilde{\Omega}} p(\mathbf{x})\,\mathrm{d}\mathbf{x} \tag{5}$$

and the corresponding values of the thresholds are $\tilde{d}_1$ and $\tilde{d}_2$.

The idea behind the Inputs' Statistic method is the limitation of the input domain using the n-dimensional contour surfaces of the PDF rather than the computation of the contour surfaces of the transfer function f. Since the mathematical formulation of the PDF is known, its contour surfaces are easily identified. This choice is particularly relevant because it produces the most important features of the new method: generality, speed and conservatism.

So accordingly to the Inputs' Statistics method:

Being $\widetilde{\mathcal{E}}$ the contour surface of the PDF enclosing a probability equal to $1-\alpha$, then $\widetilde{\Omega}$ is the region identified by contour surfaces of the transfer function f corresponding to the thresholds $\tilde{d}_1$ and $\tilde{d}_2$ being the minimum and maximum cases which may occur inside $\widetilde{\mathcal{E}}$. $\tilde{d}_1$ and $\tilde{d}_2$ are the safety box dimensions.

In the next paragraphs, we will firstly explain, under the simplified 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 and then how to compute the safety boxes dimensions. Then, we will generalize the method also to non normal distributed variables.

3.2. Multivariate normal PDF and Ruben series

Let us suppose that all the variables in the input vector **X** are normally distributed random variables. In this simple case, the associated PDF is called Multivariate Normal (MVN):

$$p(\mathbf{x}) = p_{MVN}(\mathbf{x},\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{\sqrt{|\boldsymbol{\Sigma}|\,(2\pi)^n}} e^{-\frac{1}{2}(x-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(x-\boldsymbol{\mu})} \tag{6}$$

where $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}$ are the symmetric positive definite variance matrix and mean values vector of the input vector **X**, respectively. The contour surfaces of the MVN are $n-$dimensional ellipsoids, indicated with $\mathcal{E}$, which can be described as:

$$\mathcal{E}(\mathrm{t}) = \{\mathbf{X} \in \mathbb{R}^n : (\mathbf{X}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu}) \le \mathrm{t}\} \tag{7}$$

where t is the square radius of the ellipsoid and it completely identifies its size.

The Inputs' Statistics method requires the identification of that particular ellipsoid $\widetilde{\mathcal{E}}$, having size $\tilde{\mathrm{t}}$, which corresponds precisely to the required level α, i.e. satisfying the condition:

$$1-\int_{\widetilde{\mathcal{E}}(\tilde{\mathrm{t}})} p_{MVN}(\mathbf{x},\boldsymbol{\mu},\boldsymbol{\Sigma})\,\mathrm{d}\mathbf{x} = \alpha \tag{8}$$

This operation can be done transforming the integral in eq.8 in standard form and approximating it using a particular case of the general Ruben's series in [20] and fully recalled in Genz [21]. It is possible to prove that the probability in eq.8 does not depend on the statistical characteristics of the normal distributed input variables, but only on the size of the ellipsoid identified by the parameter t and on the number of dimensions n. Accordingly, the integral in eq.8 can be expanded in series as:

$$1-\sum_{j=0}^{\infty} c_j F\left(n+2j, \frac{\tilde{\mathrm{t}}}{\beta}\right) = \alpha \tag{9}$$

where $F(l, y)$ is the central χ^2 cumulative distribution function with l degrees of freedom evaluated at y position and β is a parameter. Sheil and O'Muircheartaigh suggest in [22] to take β equal to 29/32 in order to guarantee convergence of the series. The coefficients of the series are given by:

$$c_0 = \beta^{n/2}\,;\ c_j = j^{-1}\sum_{i=0}^{j-1} g_{j-i}c_i \ \text{for } j > 0\,;\ g_j = \frac{n}{2}(1-\beta)^j \tag{10}$$

Thus, eq.9 becomes a non linear algebraic equation to be solved numerically (e.g with the Newton method). Since the evaluation of Ruben's series requires a very low computational time (fractions of second), the computation of the probability over the ellipsoid domain $\widetilde{\mathcal{E}}$ is extremely fast. In tab.1 typical values of $\tilde{\mathrm{t}}$ as function of α and n are reported.

Table 1: Ellipsoid square radius $\tilde{\mathbf{t}}$ as function of probability level $\boldsymbol{\alpha}$ and number of input variables $\boldsymbol{n}$

$\tilde{\mathbf{t}}$	$\boldsymbol{n=1}$	$\boldsymbol{n=2}$	$\boldsymbol{n=5}$	$\boldsymbol{n=10}$	$\boldsymbol{n=15}$
$\boldsymbol{\alpha = 10^{-2}}$	6,635	9,210	15,086	23,209	30,578
$\boldsymbol{\alpha = 10^{-5}}$	19,511	23,026	30,856	41,296	50,493
$\boldsymbol{\alpha = 10^{-7}}$	28,374	32,236	40,863	52,310	62,326
$\boldsymbol{\alpha = 10^{-12}}$	50,844	55,263	65,238	78,471	89,981

3.3. Computation of the safety boxes dimensions

As soon as $\widetilde{\mathcal{E}}$, corresponding to a given α, is identified, the Inputs' Statistics method prescribes to compute the safety boxes dimensions, i.e. $\tilde{d}_1$ and $\tilde{d}_2$, as *the minimum and the maximum values of the transfer function f within the ellipsoid* $\widetilde{\mathcal{E}}$. Using a programming algorithm, $\tilde{d}_1$ and $\tilde{d}_2$ can be numerically estimated:

$$\begin{gathered}\tilde{d}_1 = \min f(\mathbf{X})\\ \text{subjected to } (\mathbf{X}-\boldsymbol{\mu})^{\boldsymbol{T}}\boldsymbol{\Sigma}^{-\mathbf{1}}(\mathbf{X}-\boldsymbol{\mu}) \le \tilde{\mathrm{t}}\\ \mathbf{X} \in \mathbb{R}^n\end{gathered} \tag{11}$$

and

$$\begin{gathered}\tilde{d}_2 = \max f(\mathbf{X})\\ \text{subjected to } (\mathbf{X}-\boldsymbol{\mu})^{\boldsymbol{T}}\boldsymbol{\Sigma}^{-\mathbf{1}}(\mathbf{X}-\boldsymbol{\mu}) \le \tilde{\mathrm{t}}\\ \mathbf{X} \in \mathbb{R}^n\end{gathered} \tag{12}$$

This is a constrained non linear optimization problem that is extensively covered by literature [23] and commercial software. The analysis showed that the particular version of the Barrier method [24] belonging to the "interior point" class embedded in the Matlab Optimization Toolbox and the SNOPT software [25] are efficient tools to solve this problem, which is the final step prescribed by the Inputs' Statistics method.

3.4 Compliance with the safety constraint and characteristics of conservatism

Once the optimization problem in eq.11 and eq.12 is solved, $\tilde{d}_1$ and $\tilde{d}_2$ are the dimensions of the safety box and the final outcome of the analysis. The identification of $\widetilde{\Omega}$ is now straightforward because by definition it is the region in the input space included between the contour surfaces of f corresponding to $\tilde{d}_1$ and $\tilde{d}_2$. Comparing $\widetilde{\Omega}$ with $\widetilde{\mathcal{E}}$ we can verify that the results given by the Inputs' Statistics method *always* satisfy the safety constraint in eq.2. Since, by construction $\widetilde{\Omega}$ includes $\widetilde{\mathcal{E}}$, i.e. $\widetilde{\mathcal{E}}$ is a subset of $\widetilde{\Omega}$:

$$\widetilde{\mathcal{E}} \subseteq \widetilde{\Omega} \tag{13}$$

and since the probability is monotone with respect to two sets which are one the subset of the other, then we can state that:

$$\int_{\widetilde{\mathcal{E}}} p(\mathbf{x})\,\mathrm{d}\mathbf{x} \le \int_{\widetilde{\Omega}} p(\mathbf{x})\,\mathrm{d}\mathbf{x} \tag{14}$$

And, by design of $\widetilde{\mathcal{E}}$ in eq.8, substituting, we prove that $\widetilde{\Omega}$ satisfies the constraint on the probability level:

$$1-\alpha \le \int_{\widetilde{\Omega}} p(\mathbf{x})\,\mathrm{d}\mathbf{x} \tag{15}$$

that is, rearranging the expression, the Inputs' Statistics method satisfies the safety requirement:

$$\mathrm{P}_{IS} = 1 - \int_{\widetilde{\Omega}} p(\mathbf{x})\,\mathrm{d}\mathbf{x} \le \alpha \tag{16}$$

In addition, reminding the definition of Ω^{Opt} in eq.3, it follows that:

$$1 - \int_{\widetilde{\Omega}} p(\mathbf{x})\,\mathrm{d}\mathbf{x} \le 1 - \int_{\Omega^{\mathrm{Opt}}} p(\mathbf{x})\,\mathrm{d}\mathbf{x} \tag{17}$$

and thus we can conclude that:

$$\tilde{d}_1 \le d_1^{Opt} \quad \text{and} \quad \tilde{d}_2 \ge d_2^{Opt} \tag{18}$$

which means that the result in terms of safety boxes dimensions given by the Inputs' Statistics method is always conservative with respect to the optimal solution. In other words, the Inputs' Statistics method designs safety boxes which are larger than or equal to the minimum size.

3.4. Introduction of non-normal distributed variables

The Inputs' Statistics method can be generalized to the case of non normal distributed input variables as long as a transformation $\tau : \mathbb{R}^n \dashrightarrow \mathbb{R}^n$ exists that maps the input vector $\boldsymbol{Z}$, which may have non normal distributed components, into a full normal distributed vector $\mathbf{X}$:

$$\boldsymbol{X} = \tau(\boldsymbol{Z}) \tag{19}$$

If an appropriate transformation τ exists, the PDF of $\boldsymbol{X}$ is still a MVN and the Ruben's series can be applied to integrate the probability over the ellipsoidal contour surfaces of the PDF. Some transformations τ have been proposed depending on the available information on the PDF of $\boldsymbol{Z}$ [5].

In the simple case of having uncorrelated (i.e. diagonal covariance matrix) variables, the transformation in eq. 19 can be applied singularly to each variable. In particular, considering uniform distributed variables, τ is the inverse of the cumulative distribution function (CDF) of the standard normal distribution. This is a particular case of the more general Probability Integral Transform [26]. Indeed, since the CDF of a normal distribution is strictly increasing, being Z_i the i-th uniformly distributed random variable on $[a_i, b_i]$ of $\boldsymbol{Z}$ and defining:

$$X_i = \mathrm{cdf}_{\mathrm{X}}^{-1}\left(\frac{Z_i - a_i}{b_i - a_i}\right) \tag{20}$$

Where $\mathrm{cdf}_{\mathrm{X}}^{-1}$ denotes the inverse of the CDF of the normal distribution:

$$\mathrm{cdf}_{\mathrm{X}}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-\frac{1}{2}\xi^2}\, \mathrm{d}\xi \tag{21}$$

Then, X_i is a normally distributed random variable with zero mean and unitary standard deviation. Therefore, at each iteration of the programming algorithm to compute the safety boxes (eq.11 and 12), the transformation in eq.20 is firstly applied to all the uniform variables Z_i to get the corresponding normal variables X_i and then the X_i variables are constrained to be inside the ellipsoid associated to the probability level α.

Another alternative strategy when dealing with uniform distributions is to consider their full range as input domain to search for the min/max of the transfer function. In this case, the uniform distributed variables are simply let varying freely in the programming algorithm within their assigned interval. The normal distributed variables, instead, are constrained by an ellipsoid computed considering only the number of normal distributed variables. The two approaches are very close each other when an extremely low probability level is sought. They are just two of the infinite choices of the $\widetilde{\Omega}$ domain because both are always compliant to the safety constraint.

3.5. Example of Inputs' Statistics method application

An example of application of the Inputs' Statistics method applied to a simple analytical bi-dimensional function is here presented for clarification purpose. The selected analytical function is:

$$f(x_1, x_2) = (e^{x_1} - 1)(e^{x_2/2} - 1) \tag{22}$$

where x_1 and x_2 are uncorrelated normally distributed random variables with zero mean value and $\sigma_1 = 0.29$, $\sigma_2 = 0.17$. It satisfies the property $f(\boldsymbol{\mu}) = 0$ and its characteristics of non-linearity is similar to the typical functional shape of the atmospheric re-entry transfer function. Considering the SRA probability level, i.e. $\alpha = 10^{-5}$, the corresponding ellipsoid $\widetilde{\mathcal{E}}$ is given by:

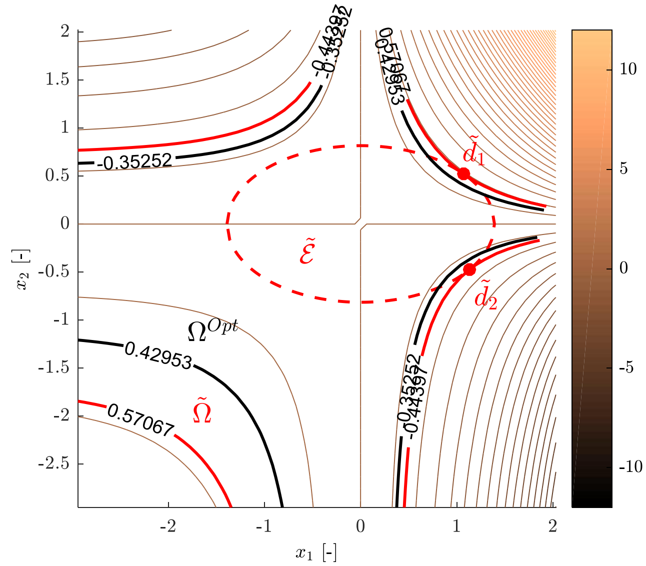

Figure 2: Application of the Inputs' Statistics method to the transfer function f in eq.22.: the area inside the continuous black line is Ω^{Opt}, the area inside the dashed red line is $\widetilde{\mathcal{E}}$ and the area inside the continuous red lines is the region $\widetilde{\Omega}$.

$$\widetilde{\mathcal{E}} \triangleq \frac{x_1^2}{0.0841} + \frac{x_2^2}{0.0289} \leq 23.026 \tag{23}$$

Its boundaries have been plotted in fig.2 in dashed red line as well as the contour lines of the transfer function f.

From the plot, it is shown that the maximum increasing rate is in the upper right quadrant and the maximum decreasing rate is in the bottom right one. Indeed, applying the programming algorithm, we get $\tilde{d}_1 = -0.44397$ and $\tilde{d}_2 = 0.57067$ and the corresponding input points belong to those quadrants. Drawing the contour lines of f corresponding to $\tilde{d}_1$ and $\tilde{d}_2$ (in continuous red line in fig.2), we get the region $\widetilde{\Omega}$. For the generic case of atmospheric transfer function, it is extremely difficult to identify $\widetilde{\Omega}$ because f is a multidimensional numerical propagator. In this simple case, instead, we visualize it and we also integrate the PDF over it. In this way, it is possible to verify that $\widetilde{\mathcal{E}} \subset \widetilde{\Omega}$, satisfying the safety requirement: $\mathrm{P}_{IS} = 1.17e^{-6} \leq \alpha = 10^{-5}$. Furthermore, thanks to the simple analytical expression of f, it is also possible to change iteratively the f contours in order to select those ones that satisfy the constraint with equality and, simultaneously, minimize the distance between d_1 and d_2, i.e. estimate Ω^{Opt}. It is shown in continuous black line in fig.2, corresponding to the optimal values of: $d_1^{Opt} = -0.35252$ and $d_2^{Opt} = 0.42953$. d_1^{Opt} and d_2^{Opt} satisfy eq.18, clarifying how the Inputs' Statistics approximates the optimal results from a conservative direction.

4. CHARACTERISTICS OF THE METHOD

4.1 Direct computation of the probability

The driving idea which led to the development of the Inputs' Statistics method is that the probability that a fragment impacts the ground at certain distance from the Aimed Impact Point is caused by the probability of having certain initial conditions (i.e. input parameters). Therefore, working on the input domain, the probability can be *directly* and *exactly* computed, without the need of building up an output distribution. Most of the State of the Art methods, instead, compute the safety boxes based on the *estimated* probability derived from the simulated impact point distribution obtained by large sampling.

Furthermore, the Inputs' Statistics method has an explicit physical meaning thanks to a direct relationship between the input domain and the output interval. The worst case events are explicitly identified and quantified as well as the physical causes which lead to their occurrence.

4.2 Computational speed

Applying the Inputs' Statistics method, the computation time of the safety boxes is decreased by more than one order of magnitude (typically from hours/days to minutes).

Let us verify this statement with a rough analysis of the orders of magnitude of the computational time, comparing with a classical method based on Monte Carlo simulations. The computation time can be expressed as function of the number of transfer function evaluations. This is dependent on the particular algorithm and software used.

In the tests performed with the Matlab Optimization Toolbox to solve the problem in eq.11 and eq.12 required by the Inputs' Statistics method, it has been observed that:

- the number of optimizer iterations is usually on the order of 10: $O(10)$
- the number of transfer function evaluations for each optimizer iteration is also on the order of 10: $O(10)$

Hence, the total number of transfer function evaluations required by the Inputs' Statistics method is on the order of 100: $O(10^2)$. The time required to evaluate the Ruben's series for the probability computation has been considered negligible, since it is a fraction of second: $O(10^{-1}s)$. Considering the atmospheric dynamics of a re-entry vehicle, the needed time for the single evaluation of the transfer function f is about $1s$. Thus the total time to compute $\tilde{d}_1$ and $\tilde{d}_2$ with the Inputs' Statistics method, here called $T_{Tot_{IS}}$, is on the order of minutes: $T_{Tot_{IS}} = O(\text{minutes})$.

Let us now do a similar rough analysis for a typical state of the art method. The heaviest step in the State of the Art techniques is the computation of the set of output samples for the statistic assessment, i.e. performing the Monte Carlo simulation. Indeed, it is reasonable to consider that the statistical post-process requires a negligible amount of time (usually $O(s)$) with respect to building up the distribution. In addition, the number of samples required for having a significant statistical result and an acceptable confidence level is strongly dependent on the probability level α (the smaller is α the larger is the number of samples required). According to Haya in [27], for the estimation of a probability level $\alpha = 10^{-5}$, it is necessary to perform $O(10^6)$ simulations of the re-entry dynamics transfer function. If $T_{Tot_{SoA}}$ is the total computational time for the State of the Art methods and considering 1s for each sample generation, then $T_{Tot_{SoA}} = O(10^6 s) = O(days)$.

The significant gain in computational time (minutes against days) given by the Inputs' Statistics method with respect to a Monte Carlo based State of the Art method for the probability threshold of 10^{-5} has an improvement factor of about 1000. This gain is even more emphasised for smaller probability levels because the number of samples required for the Monte Carlo simulation rapidly increases with α decrease, whereas the computational time of the Inputs' Statistics method is totally independent with respect to α.

In practise, to reduce this extremely large computational time of the Monte Carlo simulations either the sample time is reduced, e.g. parallelizing the *for* cycle, or less samples are used. Less samples means either using particular statistical techniques (splitting algorithms, importance samples, line sampling etc) to speed up the simulation or just accepting a much lower confidence level.

4.3 Difficult estimation of the conservatism of the results

It has been proven that for construction and for whichever transfer function, the results provided by the Inputs' Statistics method always satisfies the imposed safety constraints. This means that the method does not provide the smallest possible safety boxes, but a slightly larger one, approaching in this way the optimal one from a conservative direction. This consideration cannot be generally stated for the State of the Art methods, which converge to the imposed probability only in the ideal case of generating an infinite number of samples.

Even though the approximation of the inputs' statistics gives conservative results we have to assure that they are not too conservative far from the optimal. From the analysis reported in this paper, the results given by the Inputs' Statistics method are assessed quasi-optimal only for the transfer function associated to safety boxes computation. For particular transfer functions like the propagation of the initial conditions of an atmospheric re-entry, the method gives results quasi-optimal, with a good accuracy compared with the State of the Art methods as shown in the next section.

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

In order to show the effectiveness of the Inputs' Statistics method and its accuracy, a simplified study case of the controlled shallow re-entry of the Automated Transfer Vehicle – George Lamaitre [19] has been assessed with this method. Only four variables affected by uncertainty are here considered: the magnitude of the ΔV of the second de-orbitation manoeuvre (DEO2) ΔV_{man}; the explosion altitude h_{Expl}, the pitch angle of thrust orientation δ, the vehicle overall mass m_0. All the four variables are considered uncorrelated normally distributed and the relative parameters are collected in tab.2.

Table 2: Statistical parameters of the normal distributed variables

	ΔV_{man}	h_{Expl}	δ	m_0
Mean value μ	$47\ m/s$	$77\ km$	$0\ deg$	$15000\ kg$
Standard deviation σ	$0.47\ m/s$	$2\ km$	$0.57 deg$	$106\ kg$

Other parameters, which are required by the dynamic propagator to integrate the atmospheric re-entry trajectory of the Aimed Impact Point, are collected in tab.3.

Table 3: Data required by the dynamic propagator used for the computation of the Aimed Impact Point

Drag coefficient	2.2
Normal surface	$16\ m^2$
Specific Impulse	$312\ s$
Thrust T	$1007\ N$
Atmospheric density	CIRA86-NRLMSISE-00 model with $F10.7 = 149.9$ and $AP = 18.5$

After the explosion, the fragment must be characterized by ballistic coefficient, magnitude and direction of explosion. We distinguish between a long fragment and a short fragment. The long fragment has characteristics that leads it to fall in front of the reference point (positive output) giving the value of $\tilde{d}_1$; the short fragment has characteristics that leads it to fall behind the reference point (negative output) giving the value of $\tilde{d}_2$.

The State of the Art method chosen to compare the results given by the Inputs' Statistics method is the Monte Carlo simulation + Peaks over threshold here applied for the sake of comparison with the simplified set of variable parameters. The probability level considered in this computation is $\alpha = 10^{-5}$ corresponding to the Safety Re-entry Area (SRA). The results have been collected in tables 4 and 5.

Table 4: Results in terms of computational time, number of iterations of the programming algorithm and SRA dimension given by the Inputs' Statistics method.

Results given by Inputs' Statistics method	
Number of iterations of the programming algorithm:	33
Computational time:	850 seconds
SRA dimension	4310 km

Table 5: Results in terms of computational time, number of samples of the statistical distribution and SRA dimension given by the State of Art method: Monte Carlo simulation + Peaks over threshold method

Results given by Monte Carlo Simulation + Peaks over threshold method	
Number of samples:	40000 (half for short frag. and half for the long frag.)
Computational time:	About $22\ hours$
SRA dimension	3900 km

The Inputs' Statistic method converges to a solution with a speed about 100 times faster. The difference in dimension is noticeable but quite small (~10%). The State of Art method gets closer to the optimal solution while the Inputs' Statistic method provides a quick and conservative solution to the problem of the SRA.

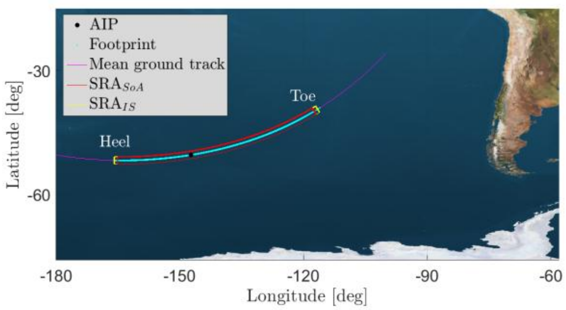

Figure 3: Safety boxes computation using a Monte Carlo based State of the Art method and the Inputs' Statistics method. Comparison of the results.

6. CONCLUSION AND AKNOWLEDGMENTS

A new approach for the estimation of rare events applied to the computation of the SRA has been presented. It has been demonstrated that this approach is always conservative and for the specific application of the SRA it approximates the optimal solution. It has been shown that the time to compute a solution is few orders of magnitude smaller than the time required by the state of art methods: this may allow sensitivity analysis and trade-off of multiple computations of SRA for different initial conditions as it was required in the design of the ATV-GL shallow reentry experiment. The method is also quite not sensitive to the dimension of the input parameters that are subject to statistic distribution of their uncertainties. Limitation of the method has been explained, as the solution is not optimal although it may get quite close to the optimal solution under certain conditions.

A comparison of this method with a state of the art method has been performed in section 5 with the computation of the SRA of ATV-GL shallow reentry confirming the expected good performance.

Potential application of this method to other engineering problems requiring the estimation of rare events caused by multivariable inputs with known statistic distribution of uncertainties will be subject of future works.

The authors acknowledge the support of ESA and CNES Flight Dynamics experts to the study performed during the preparation of ATV-GL shallow reentry mission.

7. REFERENCES

[1] ESA Space Debris Mitigation WG, *ESA Space Debris Mitigation Compliance Verification Guidelines*, 19-February-2015, ESSB-HB-U-002, Issue 1 Revision 0.

[2] Frédéric Renaud, Thierry Martin, *Safety Boxes Sizing for Controlled Re-Entry Application to Automated Transfer Vehicle (ATV)*, Proceedings of the 3rd IAASS Conference: building a safer space together, 21-23 October 2008, Rome, Italy.

[3] Michael Falk, Jürg Hüsler, Rolf-Dieter Reiss, *Laws of Small Numbers: Extremes and Rare Events*, Third edition, Birkhäuser, 2010.

[4] Jérôme Morio, Mathieu Balesdent, Damien Jacquemart, Christelle Vergé. *A survey of rare event simulation methods for static input-output models. Simulation Modelling Practice and Theory*, Elsevier, 2014, 49, pp.287-304

[5] Jérôme Morio, Mathieu Balesdent, *Estimation of Rare Event Probabilities in Complex Aerospace and Other Systems, A Practical Approach*, Woodhead Publishing, Elsevier.

[6] Samuel Kotz, Saralees Nadarajah, *Extreme Value Distributions. Theory and Applications*, Imperial College Press, London, 2000.

[7] Harald Niederreiter, Denis TalayMonte *Carlo and Quasi-Monte Carlo Methods*, Springer, 2004.

[8] Christian P. Robert, George Casella, *Monte Carlo Statistical Methods*, Springer, New York, 2005.

[9] Gerardo Rubino, Bruno Tuffin, *Rare Event Simulation using Monte Carlo Methods*, 3rd chapter: *Splitting techniques*, by Pierre L'Ecuyer, Francois Le Gland, Pascal Lezaud and Bruno Tuffin, 2009 John Wiley & Sons, Ltd.

[10] Dirk P. Kroese, Thomas Taimre, Zdravko I. Botev, *Handbook of Monte Carlo Methods*, 2011 John Wiley & Sons, INC., Publications.

[11] F. Casciati, J. B. Roberts, *Reliability problems: general principles and applications in mechanics of solids and structures*, 3rd chapter: *Methods for structural reliability computations*, by P. Bjerager, International centre for mechanical sciences, Courses and Lectures-No.317, Springer-Verlag Wien GMBH.

[12] Seung-Kyum Choi, Ramana V. Grandhi, Robert A. Canfield, *Reliability-based Structural Design*, Springer-Verlag London Limited 2007.

[13] Bruno Sudret, *Meta-models for structural reliability and uncertainty quantification*, Fifth Asian-Pacific Symposium on Structural Reliability and its Applications, Singapore, 2012.

[14] Jan Beirlant, Yuri Goegebeur, Jozef Teugels, Johan Segers, Daniel De Waal, Chris Ferro *Statistics of Extremes: Theory and Applications* , 2004 John Wiley & Sons, Ltd.Manfred Gilli, Evis Kellezi, *An Application of Extreme Value Theory for Measuring Financial Risk*, Computational Economics 27(1), 2006, 1–23.

[15] Ana Ferreira, Laurens de Haan, *On the block maxima method in extreme value theory: PWM estimators*, The Annals of Statistics 2015, Vol. 43, No. 1, 276.

[16] Richard W. Katz, Marc B. Parlange, Philippe Naveau, *Statistics of extremes in hydrology*, 2001, Advances in Water Resources 25 1287-1304.

[17] Viatcheslav V. Kharin, Francis W. Zwiers, Xuebin Zhang, Gabriele C. Hegerl, *Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations*, 2007, Journal of Climate 20 1419-1444.

[18] Catherine Hourtolle, Sylvain Delattre, Jean-François Goester and Emilio De Pasquale, *Studies about a shallow re-entry for ATV-5*, 25th International Symposium on Space Flight Dynamics, October 19 – 23, 2015, Munich, Germany

[19] Harold Ruben *Probability contents in regions under spherical normal distributions, IV: The distribution of homogeneous and non-homogeneous quadratic functions of normal variables*. 1962, The Annals of Mathematical Statistics, 33:552–570.

[20] Alan Genz, Frank Bretz *Computation of Multivariate Normal and t Probabilities*, Springer Berlin Heidelberg; 1st edition, July 2009.

[21] Sheil WF, O'Muircheartaigh, *The distribution of non-negative quadratic forms in normal variables*. 1977, Journal of the Royal Statistical Society, Series C 26:92–98 Cited on p. 14, 66.

[22] J. Nocedal and S. Wright, *Numerical Optimization*. Springer Series in Operations Research and Financial Engineering, Springer New York, 2006.

[23] R. A. Waltz, J. L. Morales, J. Nocedal, and D. Orban, *An interior algorithm for nonlinear optimization that combines line search and trust region steps*, Mathematical Programming, Vol. 107, No. 3, 2006, pp. 391–408

[24] P. E. Gill, W. Murray, and M. A. Saunders, *SNOPT: An SQP algorithm for large-scale constrained optimization*, SIAM J. Optim., 12 (2002), pp. 979–1006.

[25] George Casella, Roger L. Berger, *Statistical Inference*, 2nd Edition 2001 Duxbury Advanced series

[26] Rodrigo Haya Ramos, Deimos Space S.L.U. (Spain), *Computation of IXV Safety Footprint boundaries*, CCT ORB, Seminar on rare events, extreme values, extreme impacts applied to flight dynamics, Monday, April 27th 2015, ISAE-SUPAERO, Toulouse (FR).