Noname manuscript No.
(will be inserted by the editor)

Creation of a synthetic population of space debris to reduce discrepancies between simulation and observations

Alexis Petit · Daniel Casanova · Morgane Dumont · Anne Lemaître

Received: date / Accepted: date

Abstract The number of space debris has increased in the orbital environment and consequently, the risk of collision between satellites and space debris or space debris and space debris has become a hot topic in Celestial Mechanics. Unfortunately, just a small fraction of the biggest and brightest objects are visible by means of radar and optical telescopes. In the last years, many efforts have been made to simulate the creation of space debris populations through different models, which use different sources and diverse orbital propagators, to study how they evolve in the near future. Modellizing a fragmentation event is rather complex, furthermore large uncertainties appear in the number of created fragments, the ejection directions and velocities. In this paper, we propose an innovative way to create a synthetic population of space debris from simulated data, which are constrained by observational data, plus an Iterative Proportional Fitting (IPF) method to adjust the simulated population by statistical means. **The final purpose consists in improving a synthetic population of space debris created with a space debris model helped by an additional data set which allows to converge towards a new synthetic population whose global statistical properties are more satisfying.**

Alexis Petit
University of Namur, Namur, Belgium
Tel.: +32 (0)81 72 49 09
E-mail: alexis.petit@obspm.fr

Daniel Casanova
Centro Universitario de la Defensa, Zaragoza, Spain
E-mail: casanov@unizar.es

Morgane Dumont
University of Namur, Namur, Belgium
Tel.: +32 (0)81 72 49 45
E-mail: morgane.dumont@unamur.be

Anne Lemaître
University of Namur, Namur, Belgium
Tel.: +32 (0)81 72 49 08
E-mail: anne.lemaitre@unamur.be

Keywords Space debris · Synthetic population · GEO region · Iterative Proportional Fitting method · Microsimulation

1 Introduction

In 1957 with the launch of Sputnik I the space era began. Since that precise moment, we have been leaving behind all kinds of debris in Space. In particular, the United Nations COmmittee on the Peaceful Uses of Outer Space (UNCOPUOS) defines space debris as all man-made objects, including fragments and elements thereof, in Earth orbit or re-entering the atmosphere, that are non-functional[1]. Such debris include nonfunctional spacecraft, abandoned launch vehicle stages, pieces of debris coming from different missions, explosions (intentional or non-intentional), collisions, satellite-surface degradation due to solar radiation or small impacts, etc.

In the last decades, several authors have dealt with space debris; from modeling the short-term evolution of space debris [Wnuk (1997)] to modeling the long-term evolution of space debris considering the solar radiation pressure [Valk et al. (2009)], the shadowing effects [Hubaux and Lemaître (2013)], or short- and long-term evolution of space debris under different perturbations [Casanova et al. (2015)]. On the other hand, other authors are focused on the global dynamics of space debris [Celletti et al. (2017)]. Furthermore, not only the orbital evolution is significant, but also the study of collisions between space debris and satellites **(for details see [Valsecchi et al. (2002)] and [Rossi and Valsecchi (2006)])**, how to avoid them [Casanova et al. (2014)], or a tool to quantify the catastrophic collision risk and consequences in the coming decades [Rossi et al. (2016)].

However, in this work we will focus on how to give an estimation of the unknown population of space debris. Indeed, the observational means allow us to detect only the biggest and brightest objects in space and consequently, space debris which are smaller than 10 cm in Low Earth Orbit (LEO), and 1 m in Geostationary Earth Orbit (GEO), are difficult to track, and even worse to catalog. **We have to remark that, in LEO region, in situ measurements are available for submillimeter size objects and some statistical information is regularly acquired for objects in the 2-10 cm range (24 h radar staring experiments), and then, it gives strong constraints for the space debris population models. On the other hand, in GEO region, no in situ measurements are available and only sparse data on objects smaller than 1 m are available. Then, using these different sets of data, several models have been developed** to model the space orbital environment and to give an estimation of the unknown population taking into account this data. In particular, Meteoroid and Space Debris Terrestrial Environment Reference (MASTER) [Flegel et al. (2009)] and Orbital Debris Engineering Models (ORDEM) are the most popular space debris models. The evolution of the population generated is then handled by tools like LEO-to-GEO ENvironment Debris model (LEGEND) [Liou et al. (2004)], Debris Analysis and Monitoring Architecture for the

[1] Space Debris Mitigation Guidelines of the Committee on the Peaceful Uses of Outer Space. United Nations. Vienna. 2010

Geosynchronous Environment model (DAMAGE) [Lewis et al. (2001)], and Semi-Deterministic Model (SMD) [Rossi et al. (2009)]. They are semi-deterministic models whose purpose is to model the different sources of the space debris and to propagate the orbit of each individual fragment **or a group of fragments**.

An important issue is the comparison between space debris predictions and optical or radar ground-station observations, or in-situ measures. In particular, in the GEO region, several space debris surveys have been performed since the nineties, and they show the existence of unknown space debris populations [Schildknecht et al. (2004)]. **Modifying the parameters of a space debris model, comparing the obtained results with the existing observations, and iterating this procedure until the simulation and the observable objects fit almost perfectly we can converge toward an accurate model. Then, as a result we have a synthetic population in agreement with the observations. Thus, thanks to different simulations,** in the GEO region the clusters of space debris observed could be explained by just eight (unconfirmed) fragmentations [Jehn et al. (2006)].

The modelling of the orbital environment is a complex task. Even if we know the sources of space debris, many factors are impossible to model with accuracy. Nowadays, the NASA[2] Breakup Model (NBM) is useful for modeling the debris clouds generated by on-orbit explosions and collisions in terms of fragment size and velocity distributions. This model is based on a limited set of controlled terrestrial experiments like hypervelocity impacts, and data coming from the observations of clouds generated by some historical breakups [Johnson et al. (2001)]. The empirical nature of the NBM makes it improvable. A calibration iteration process can be applied to modify the parameters of the space debris models and to obtain a better agreement with the different sets of observational data. However, unknown events, assumptions of the source models, or limited computational ressources, can produce discrepancies. We are also limited by the computational cost of propagating huge populations of space debris since we count approximately 20,000 space debris with a size greater than 10 cm, hundreds of thousands of space debris between 1 cm and 10 cm, and millions of pieces smaller than 1 mm.

The final goal of this work is the creation of a synthetic population of space debris whose global statistical characteristics are similar to the assumed as real, from the synthetic population already created by a space debris model and using additional constraints coming from different sources (for example, observations, simulations, ground-based experiments). This innovative way allows to create new pieces of space debris by using an Iterative Proportional Fitting (IPF) method, which takes into account the statistical properties of additional data such as constraints. Then, we create a new synthetic population. However, we must remark that the IPF technique is independent of the space debris model. The IPF method just modify the initial population by using simple additional constrains, without any influence on the used model.

This paper is organized as follows. Section 2 describes the characteristics of the population in the GEO region according to the Two Line Elements (TLE) catalog

[2] National Aeronautics and Space Administration (NASA)

provided by the United States STRATegic COMmand (USSTRATCOM). In this section, we are able to reproduce through a simulation in a deterministic way the same situation as the one obtained with the TLE catalog. Section 3 introduces the mathematical framework of the IPF method. Then, in Section 4 we apply the IPF method to our simulated population of space debris in two different ways. The first application consists in inferring constraints of the simulated population, and using them to create new pieces of space debris with similar characteristics as the simulated population but saving computational time. The second application consists in using additional data which allow to reduce discrepancies between two sets (for example, simulation and observations). Finally, in Section 5 we conclude this work.

2 Model of the space debris population in the geostationary region

We present a deterministic approach to generate an artificial population of space debris in the geostationary (GEO) region. Then, we ensure the validity of our method by verifying that the small amount of **observed** objects which are known, taken from the well-known Two Line Elements (TLE) catalog, are close to the generated objects in the artificial population. After that, we use our simulated population to create a bigger one by using an Iterative Proportional Fitting (IPF) method, as it will be explained in Section 3.

2.1 The GEO region

The geostationary orbit is defined as the 1:1 gravitational resonance, where the semi-major axis is equal to 42,164 km, i.e., where the orbital period of an object corresponds to one sidereal day (23h 56min 4s). **The GEO protected region is defined as** the ring around the Earth, which is delimited by the geostationary altitude (35,786 km) $\pm$ 200 km, and the latitude $\pm$ 15° around the equatorial plane [IADC (2007)]. This region is very attractive for communication and observation satellites because they have a fixed position in the sky with respect to a ground observer.

Since the launch of the first GEO satellite in 1964 only five fragmentations are known to have occurred **(at the date of 2017/10/22).** The first one took place on June 25, 1978 caused by a malfunction of the battery of satellite Ekran 2 (mass 1,750 kg) and the second fragmentation was caused by a failure of the upper stage Titan 3C Transtage (mass 2,500 kg) on February 21, 1992 [Johnson et al. (2008)]. A second explosion of an upper stage Titan 3C Transtage occurred on June 4, 2014, with the creation of 28 fragments in total [Cowardin et al. (2017)]. On January 16, 2016 an upper-stage **Briz-M** (mass 1,220 kg) breakup took place in a region close to the geostationary ring. It happened at an altitude of 34,866 km (at the subsatellite point 0.17° South, 223° East). Just 10 fragments were observed but several hundreds are expected [Orbital Debris Program Office (2016)a]. A second breakup took place on June 26, 2016 when the satellite BeiDou G2 (mass 1,100 kg) underwent a fragmentation in the GEO region. The resulting orbit presents an apogee of 36,137 km, a perigee of 35,384 km, and an inclination of 4.7°. In particular, at least five fragments were observed but no one was officially cataloged

Table 1 Confirmed breakup events in the GEO region.

Date	1978/06/25	1992/02/21	2014/06/04	2016/01/16	2016/06/29
Type	Spacecraft	Rocket-body	Rocket-body	Rocket-body	Spacecraft
Name	Ekran 2	Titan Trans. 3C-5	Titan Trans. 3C-17	PROTON-M/BRIZ-M	BEIDOU G2
ID SSN	10365	3432	3692	41122	34779
ID COSPAR	1977-092A	1968-081E	1969-013B	**2015-075B**	2009-018A
a [m]	42,163.500	41,826.000	42,926.390	40,979.800	42,138.874
e	$1{,}779{\cdot}10^{-4}$	$8{,}488{\cdot}10^{-3}$	$1.291{\cdot}10^{-2}$	$2.868{\cdot}10^{-2}$	$8.934{\cdot}10^{-3}$
i [deg]	0.100	11.900	8.706	0.174	4.716
Ω [deg]	78.390	21.802	313.195	135.143	61.365
ω [deg]	325.277	76.279	91.579	5.856	195.139
M [deg]	78.390	284.560	269.90	221.106	164.399
Mass [kg]	1,750	2,500	2,500	1,220	1,100

[Orbital Debris Program Office (2016)b]. All the informations about the five fragmentations are summarized in Table 1. **More recently, a second minor breakup occurred for the Titan 3C Transtage (1969-013B) on February 28, 2018 with only one fragment cataloged [Orbital Debris Program Office (2018)], not considered in this work.**

Currently, if we consider the TLE catalog provided by the USSTRATCOM[3], and we filter the objects with a mean motion between 0.9 and 1.1 revolutions per day and whose name contain the tag DEB, we obtain 50 space debris in the GEO region at the date of October 22, 2017. In Figure 1, we plot the repartition of sources. We can observe that almost a half correspond to the fragmentation of the upper-stage Titan 3C and a quarter is related to individual objects. No object is related to BeiDou G2. **Unfortunately, many presumed objects are non detectable for the instruments of the United States Space Surveillance Network (USSSN) and are not cataloged. The Astronomical Institute of the University of Bern (AIUB) and the International Scientific Optical Network (ISON) are both maintaining a catalog of many hundred debris objects in GEO but those catalogs are not publicly accessible.** Nevertheless, in the next subsection, we use the small sample of space debris cataloged to validate qualitatively our simulation.

2.2 Simulation by a deterministic way

NIMASTEP (Numerical Integration of the Motion of Artificial Satellites orbiting a TElluric Planet) is an orbit propagator [Delsate and Compère (2012)], it was written in FORTRAN[4], and it was developed to study the dynamic of space debris in the GEO region taking into account the geopotential, the Moon, the Sun and the solar radiation pressure with shadowing effects. The orbit of a satellite is computed by the integration of the equations of motion, expressed with Cartesian coordinates without averaging process in order to not neglect any resonant effects. An updated version of the orbit propagator includes the atmospheric drag [Petit and Lemaître (2016)]. Moreover, to overcome the computing cost a parallelized hardware architecture was used allowing

[3] `https://www.space-track.org`

[4] FORmula TRANslation (FORTRAN) is a widely used programming language.

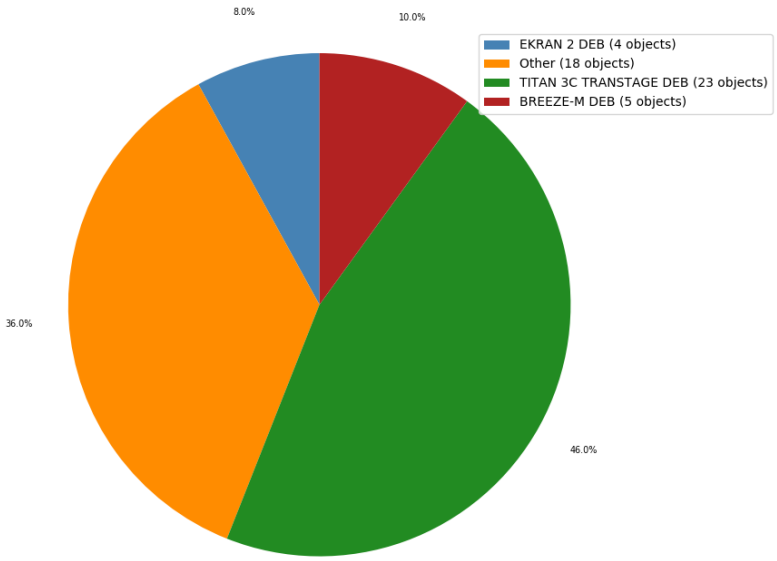

Fig. 1 Percentage of space debris in the geostationary region corresponding to different breakups that took place up to the date 2017/10/22. The objects gathered in the category "Other" are individuals objects not related to one of the major breakups summarized in Table 1.

to propagate several orbits at the same time, hence reducing the computing time. This is particularly suited for studying large populations of space debris. In the case of the GEO region, the atmospheric drag is excluded and we prefer to use a more efficient orbit propagator named Symplec, previously developed to study the stability of a geostationary orbit perturbed by the solar radiation pressure [Hubaux et al. (2012)] [Hubaux and Lemaître (2013)]. Symplec is a symplectic integrator in its principles; however the motions of the Moon and of the Sun are introduced as periodic functions of time, and the energy is then periodically but not linearly, perturbed. This is why we talk about "control" and not about "conservation" of the energy. To avoid a switch on-off in the integration, the passage through the umbra has been smoothed, using hyperbolic functions. Of course, the energy shows small periodic variations, linked to the umbra seasons, but, again, no systematic increase. Symplec code is available to work on several Central Processor Units (CPU) and it is more efficient than NIMASTEP. Thanks to this propagator, just a personal computer is able to propagate several tens of thousands of space debris fragments during several decades in just a couple of hours.

We performed a FORTRAN implementation of the NASA Breakup Model (NBM) following the work of Johnson [Johnson et al. (2001)]. Given a log file with the historical fragmentations the designed implementation is able to generate a cloud of space debris at a given time and a specific location.[5]

The following scheme explains how the implemented code works:

[5] A version of the code is available on the public repository `https://gitlab.obspm.fr/apetit/nasa-breakup-model.git`

Step 1: Read a file containing the initial conditions of a single object or a set of objects.

Step 2: Propagate the orbit of each object by using the software Symplec until the fixed final date or the date of the scheduled fragmentation is reached.

Step 3: Generate the initial conditions of the new fragments by calling the implemented NBM to create a larger population.

Step 4: Continue the propagation of each object until the final date or the date of the next fragmentation event is reached.

Remark that, in our study, we take into account the geopotential until order and degree five, the Sun and the Moon as third bodies, and the solar radiation pressure with shadowing effects. The simulation starts on 1978, and it ends on 2017 and we limit the minimal size of the objects at 10 cm. All scheduled fragmentations are summarized in Table 1.

In Figure 2, we plot the distribution of the fragments generated in the plane of the inclination in function of the right ascension of the ascending node and we also plot the space debris objects contained in the TLE catalog. We observe that the generated clouds of fragments of debris are located around the known location of the TLE objects. This means that our simulation is close enough to the available data. We provide a qualitative evidence of good accordance between the simulated population and the TLE objects. However, a quantitative comparison of the generated population is a difficult task for two main reasons; the small number of objects in the TLE catalog and the huge number of generated fragments. Nevertheless, for the purpose of this work, which is the generation of a cloud of fragments of space debris close to the TLE data, the previous simulation satisfies our expectations.

The simulated population in the GEO region agrees with available data coming from the TLE catalog, but the rest of the population is missing from the catalog. **However, the parameters of the NBM are empirical and they can not fit properly each breakup depending on a particular condition. The nominal parameters which define the distribution laws used by the NBM can not be used for each case. Even if we fit these parameters on the observations as it is the case for MASTER [Horstmann et al. (2017)], the fundamental assumptions (like normal and log-normal laws used, or an isotropic distribution of the increments of velocity of the fragments) can not match the complexity of a real event.**

Nevertheless, as isolated surveys give us additional statistical informations, in the next section we propose a method to reduce the discrepancy between the simulation and a different set of data which come from another simulation performed with different parameters, but in a future work, it could be replaced by real observational data. This method is not a calibration process but it allows to use data of different sources to create a new synthetic population.

3 The Iterative Proportional Fitting process

In this work, we define a synthetic population as the result of a microsimulation technique, which is the process of integrating multiple data to represent a real-world object

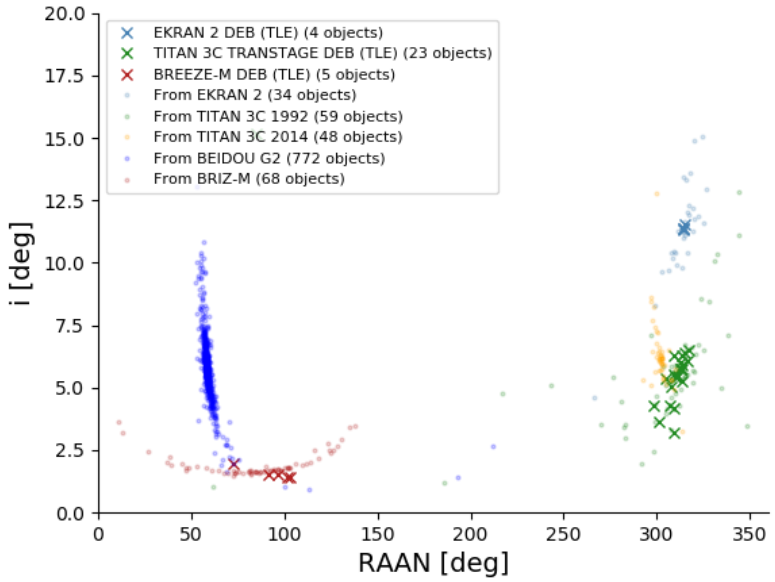

Fig. 2 Inclination versus Right Ascension of the Ascending Node representation of the simulated artificial population, and the 32 objects cataloged by USSTRATCOM and related to the major breakups in the GEO region. We do not consider the space debris not related to the main fragmentations.

into a consistent, accurate, and useful representation. In particular, a complete sample of data is weighted to satisfy controls using a method based on the Iterative Proportional Fitting (IPF) process, which was demonstrated by Deming and Stephan in 1940. For more explanation about the technique, we refer to [Lovelace and Dumont (2016)].

A piece of space debris **can be located by describing its trajectory through six classical orbital elements: the semi-major axis a, the eccentricity e, the inclination i, the right ascension of the ascending node Ω, the argument of perigee ω, and the mean anomaly M, plus the area-to-mass ratio A/m (we neglect other parameters such as the albedo since it produces a low impact on the global dynamics)**. Thus, in our work, those parameters will represent the individuals of the population and they will be used to build the synthetic population of space debris.

In particular, the eccentricity, and argument of perigee are not considered to create the synthetic population since we are in a particular region were the majority of space debris move on quasi-circular orbits, according to the available data coming from the TLE catalog. Furthermore, we are not interested in the precise position of the space debris and consequently, the mean anomaly is neither considered. However, for future works, these two variables will be included to enrich the creation of the synthetic population since for objects with large A/m ratio the eccentricity plays an important role for long time propagations. Thus, in our problem, we deal with four variables, a, Ω, i and A/m, which are discretized to apply the IPF process. Note that, when the different variables are independent, IPF is not necessary since each attribute

of the synthetic population can be generated separately. In our case, to determine the correlation between the variables, we use the Pearson correlation coefficient, which is a measure of the linear association, between -1 and +1. If the variables are uncorrelated, this Pearson correlation approaches 0. In the case of the space debris (coming from EKRAN), Table 2 indicates that for each possible pair of a, Ω and i, this coefficient is higher than 0.8, indicating a high correlation **(and thus dependence)** between these variables. On the other side, A/m seems linearly uncorrelated with each other variable. **Note that by construction, the evolution of A/m depends on the other considered attributes, meaning that there is a dependence even if A/m presents no linear correlation with respect to the three other variables considered one by one.**

Pearson correlation	a	i	Ω	A/m
a	1.000	0.997	0.831	0.022
i		1.000	0.869	0.013
Ω			1.000	-0.049
A/m				1.000

Table 2 Pearson correlation coefficients for **a cloud of space debris coming from the EKRAN breakup and computed with the NBM**.

The way by which the variables are discretized is explained for the semi-major axis, and the other variables will be discretized in a similar way. The semi-major axis will take n possible states (a_1, ..., a_n), each one represents a range of values. For example, and without any loss of generality, we suppose that the semi-major axis take 4 possible states $a \in \{a_1, a_2, a_3, a_4\}$. This means that a piece of space debris is allocated to one region if and only if the semi-major axis has a value between the lower and upper bound (see Table 3) of this region. As an example, if a piece of space debris has a semi-major axis equal to $42,302$ km, this means that the semi-major axis will be allocated to the state a_3.

Possible state (a)	Interval (km)
a_1	[41,000 , 41,500]
a_2	]41,500 , 42,000]
a_3	]42,000 , 42,500]
a_4	]42,500 , 43,000]

Table 3 Semi-major axis take four possible states given through an upper and lower bound

Consequently, each object of the population of space debris can be allocated to a state, as in Table 4, depending on the possible states of each variable of the problem. We limit our analysis to x objects. Each object is defined with its four variables: the first three, the semi-major axis, the inclination and the right ascension of the ascending node have for example four possible states, whereas, the fourth, the area-to-mass ratio has only three possible states.

Space Debris ID	a	i	Ω	$\frac{A}{m}$
1	a_1	i_2	Ω_1	$\frac{A}{m}_1$
2	a_4	i_1	Ω_1	$\frac{A}{m}_1$
3	a_1	i_3	Ω_3	$\frac{A}{m}_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x	a_3	i_2	Ω_4	$\frac{A}{m}_3$

Table 4 Allocation of the individuals into different regions. Semi-major axis, inclination and right ascension of the ascending node are allocated in four different regions, while area-to-mass ratio is classified into three different regions.

Once we have allocated the population of space debris, to apply the IPF process, we have to build a contingency table Π, which is a matrix of dimension $n_a \times n_i \times n_\Omega \times n_{A/m}$, where each cell contains the initial frequencies (usually based on a sample of the population). This table is simply calculated by counting the number of objects in each possible combination of the four variables. In this work, the contingency table is a four dimensional matrix, but for clarity, we show in Table 5 the contingency table Π in a two dimensional case, taking into account only the semi-major axis and the area-to-mass ratio $\frac{A}{m}$. To better understand how to read the contingency table we give two examples; the number of objects whose semi-major axis is equal to a_1 and whose area-to-mass ratio is equal to $\frac{A}{m}_2$ are 4. The number of objects whose area-to-mass ratio is equal to $\frac{A}{m}_3$ is 9.

	a_1	a_2	a_3	a_4	Total
$\frac{A}{m}_1$	1	2	1	1	5
$\frac{A}{m}_2$	4	1	1	0	6
$\frac{A}{m}_3$	1	3	3	2	9
Total	6	6	5	3	20

Table 5 Expression of the contingency table restricted to two dimensions. The last column and the last row give the marginal frequencies. The last cell give the total number of objects.

Before applying the IPF method we have to define constraints. In this aim, we estimate the number of objects to create in each range of values and the total number of space debris in the synthetic population (from additional data). After that, we will describe the IPF process restricted to two dimensions to facilitate the understanding.

Let Π be the contingency table and each cell be denoted by $\Pi_{i,j}$. The marginal controls for the i-th row and j-th column are noted m_i and m_j respectively. The IPF process is an iterative method, which will weight the frequencies to fit the marginal controls one after the other. If we write Π^t the contingency table at the t-iteration, the row-fitting is implemented as,

$$\Pi^t_{i,j} = \Pi^{t-1}_{i,j} \frac{m_i}{\sum_k \Pi^{t-1}_{i,k}} \quad \forall i, j, \tag{1}$$

and the column-fitting is implemented as,

$$\Pi_{i,j}^{t} = \Pi_{i,j}^{t-1} \frac{m_j}{\sum_k \Pi_{k,j}^{t-1}} \quad \forall i,j. \tag{2}$$

For example, to give the intuition behind this formula, we illustrate the process with Tables 6, 7 and 8, already limiting us to a contingency table with two dimensions. In Table 6, we add the theoretical marginal controls inferred from additional data coming from other sources, i.e. isolated observations, new simulations, statistical data, etc. In the last row and last column, we have the total number of space debris in the new population.

	a_1	a_2	a_3	a_4	Total	Theoretical total
$\frac{A}{m}1$	1	2	1	1	5	10
$\frac{A}{m}2$	4	1	1	0	6	12
$\frac{A}{m}3$	1	3	3	2	9	11
Total	6	6	5	3	20	
Theoretical total	12	9	8	4		33

Table 6 Expression of the contingency table restricted to two dimensions. The current and theoretical (constrained) totals are in the last two rows and columns of the matrix.

Fitting the rows consists in reweighting the cells to perfectly fit to the theoretical marginals of the area-to-mass ratio. For example, in Table 7, the cell $(1,1)$ is multiplied by 10 and divided by 5.

	a_1	a_2	a_3	a_4	Total	Theoretical total
$\frac{A}{m}1$	$1 \times \frac{10}{5}$	$2 \times \frac{10}{5}$	$1 \times \frac{10}{5}$	$1 \times \frac{10}{5}$	5	10
$\frac{A}{m}2$	$4 \times \frac{12}{6}$	$1 \times \frac{12}{6}$	$1 \times \frac{12}{6}$	$0 \times \frac{12}{6}$	6	12
$\frac{A}{m}3$	$1 \times \frac{11}{9}$	$3 \times \frac{11}{9}$	$3 \times \frac{11}{9}$	$2 \times \frac{11}{9}$	9	11
Total	6	6	5	3	20	
Theoretical total	12	9	8	4		33

Table 7 Expression of the contingency table restricted to two dimensions. First iteration of IPF for the row fitting.

The first step of the first iteration still needs to update the cells and the current totals. Table 8 indicates that after the fitting for the area-to-mass ratio, we perfectly follow the constraint for this variable, but not for the other one. This is the reason why each iteration performs this process for each constraint.

Since we are dealing with an iterative procedure we need a stopping condition. Thus, we consider that the convergence of the method is reached when the difference between two consecutive contingency tables is close to zero (less than the machine epsilon (10^{-16}). Thus, we compute the distance by using the following equation :

$$D(\Pi_{i,j}^{t}, \Pi_{i,j}^{t-1}) = \sum_{i,j} |\Pi_{i,j}^{t} - \Pi_{i,j}^{t-1}|. \tag{3}$$

	a_1	a_2	a_3	a_4	Total	Theoretical total
$\frac{A}{m}1$	2	4	2	2	10	10
$\frac{A}{m}2$	8	2	2	0	12	12
$\frac{A}{m}3$	1.22	3.67	3.67	2.44	11	11
Total	11.22	9.67	7.67	4.44	33	
Theoretical total	12	9	8	4		33

Table 8 Expression of the contingency table restricted to two dimensions: table after the fit of the first constraint at the first iteration.

In particular, we fix this stopping condition to 10^{-13} when the chosen discretization produces a contingency table of 960 cells.

In Figure 3, we show in a flowchart the followed process to apply the IPF method. We see that two sets of data are used: the data coming from the simulations (at the left), and the additional data used as constraints (at the right). The first one gives the cross-table with the frequencies associated, and the second one gives the constraints of the new population. Then, we apply the IPF process to create the new cross-table, which will be converted in a new population of objects. Nevertheless, the created cross-table by the IPF process contains real values, but integers are required. We apply the Truncate, Replicate and Sample (TRS) method [Lovelace and Ballas (2013)]. First, we truncate each cell of the contingency table to keep only the integer part. Then, the total population is smaller than the target population. We have to add fragments in the contingency table to complete the population. For this purpose, the decimal part is kept in a weight table, and normalized to obtain the sum of the weights equal to one.Then, the probability to add a fragment is given by this weight table. The missing fragments are chosen by following these probabilities.

The final part consists of a conversion of the new contingency table in the form of Table 3, i.e in a list of fragments whose variables are defined by a class. To obtain real values, we compute a random value in the bounds of the class, following the distribution law used to determine the constraint.

4 Applications

Once we have simulated a population of space debris in the GEO region by a deterministic way considering all the breakup events that took place in the last decades, and once we have explained how the Iterative Proportional Fitting (IPF) method works, we can apply both tools together. The first application consists of applying the IPF method to create a bigger population of space debris by using as constraints the inferred data from the simulated population in **Section 2**. Then, the second application consists of analyzing how constraints will influence in the creation of synthetic populations.

4.1 First application: generation of new space debris

Without loss of generality, we propose to focus the study on the cloud created by the breakup of Ekran 2. The IPF method will modify the population simulated according

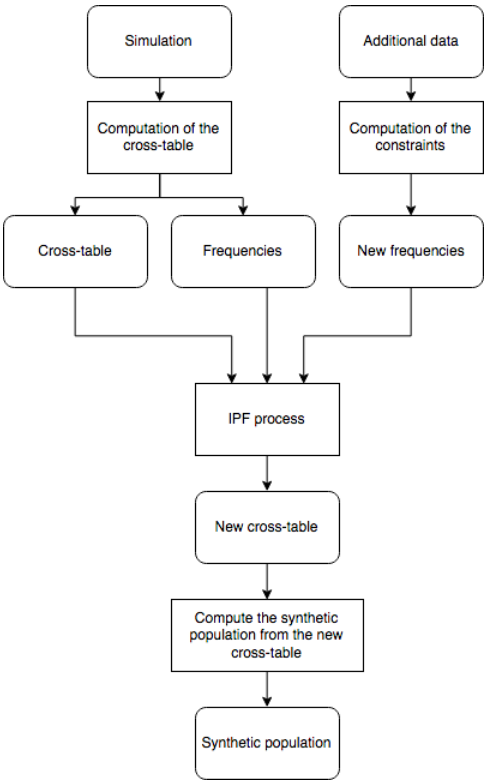

Fig. 3 Flowchart of the IPF process.

to the inferred constraints of the considered population. The final purpose is to validate the IPF method, but also to use this methodology to create a reliable population from a small sample, and save computational time.

In Figure 4, we plot the distribution of each variable obtained by the selection of the 460 fragments (with a minimum size of 1 cm) of the cloud created by the breakup of Ekran 2, and propagated until the date of October 16, 2016.

Taking into account the fact that the orbits of the fragments differ from the parent body due to the isotropic velocity increment, we can assume the distributions of the variables a, i and ω follow **Cauchy** laws centered around the mean values **(a Kolmogorov-Smirnov confirms this assumption with p-values always greater than 0.4)**. We keep in mind that this is not true for all cases and it depends on the dynamics. Then, a more complex distribution law **(or the empirical distribution)** could fit in a better way, however, we consider that the **Cauchy** law is well suited and useful for showing the proposed methodology. For the A/m ratio we use a lognormal law since the NBM uses this kind of distribution. Furthermore, in Fig. 4 we observe that the selected distributions fit perfectly with the available data. These distributions are used to compute the new frequencies of the synthetic population by using a Monte-

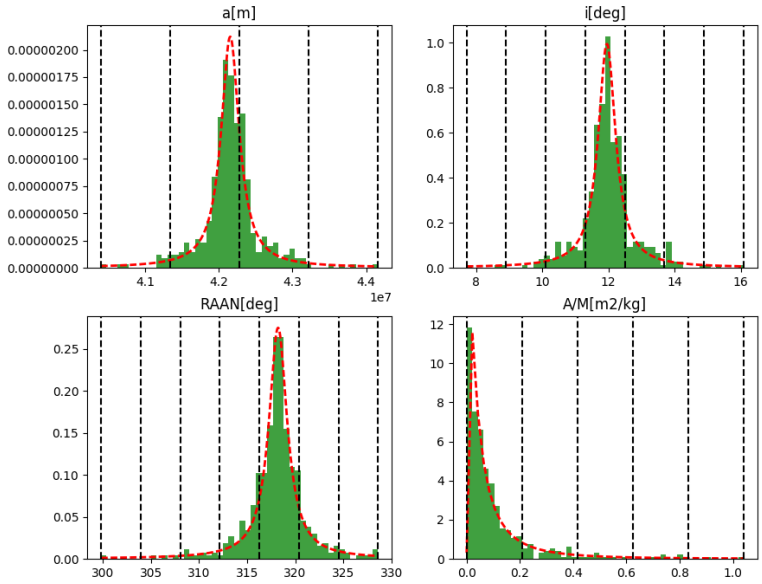

Fig. 4 The distribution of the variables a, i, Ω, $\frac{A}{m}$ of the simulation and the fit by a **Cauchy** law for the first three ones and by a lognormal law for the last one.

Carlo method and fixing the total number of objects of this bigger population. In this example, we double the number of fragments of the initial population.

In order to illustrate the convergence of the IPF method, we plot in Figure 5 the distance computed with equation (3) at each iteration. We observe that the IPF method converges in less than 100 iterations and it stops when a minimal distance equal to 10^{-13} is reached. Note that the computation is fast and takes just several seconds.

Finally, in Figure 6, we show a comparison between two families of objects; the ones corresponding to the simulated population (460 objects), corresponding to the ones illustrated in Figure 2, and the ones corresponding to the synthetic population, created thanks to the IPF method (920 objects). Recall that we do not compare the entire population, we focus on a particular region, i.e. objects whose right ascension of the ascending node is in a range 300-330°, and whose inclination is in the range 8-16°.

Figure 6 indicates that the synthetic population seems to be located in the same region as the population in terms of inclination and RAAN. The Pearson correlation coefficient between RAAN and inclination is **0.75**, meaning that the IPF has conserved a high positive correlation between these variables. However, the simulation showed a slightly higher correlation (0.869 as seen in Table 2). **A hypothesis test of Kolmogorov-Smirnov gives a p-value greater than 0.4 when testing if inclina-**

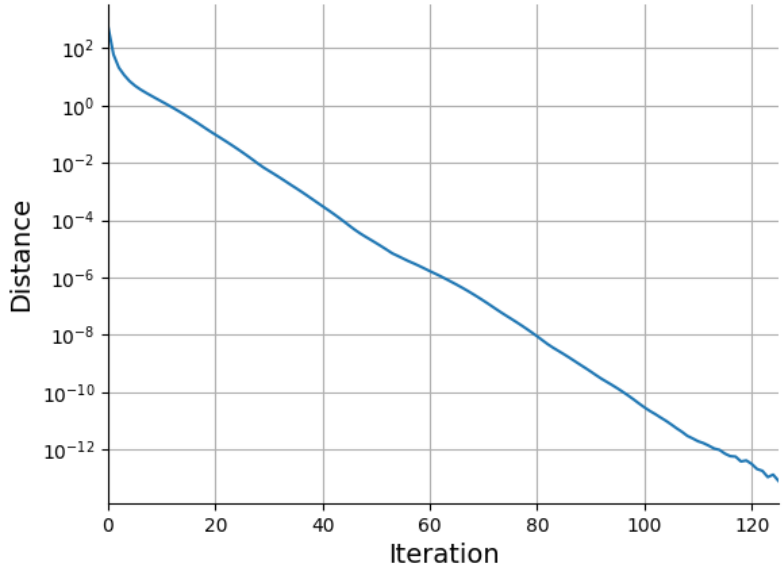

Fig. 5 Evolution of the distance computed with the equation (3) at each iteration.

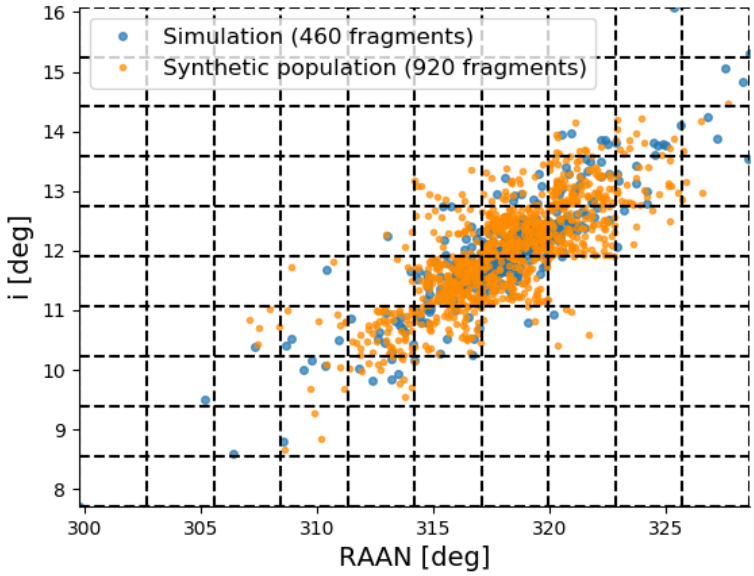

Fig. 6 Comparison between the simulation and the synthetic population.

tion and RAAN follows the same Cauchy distribution as initially. That means, as Figure 7 shows, the synthetic population follows the assumed distribution.

In conclusion, for this application, the method succeeded in creating a larger population of space debris by keeping high correlations and following the distributions used as constraints.

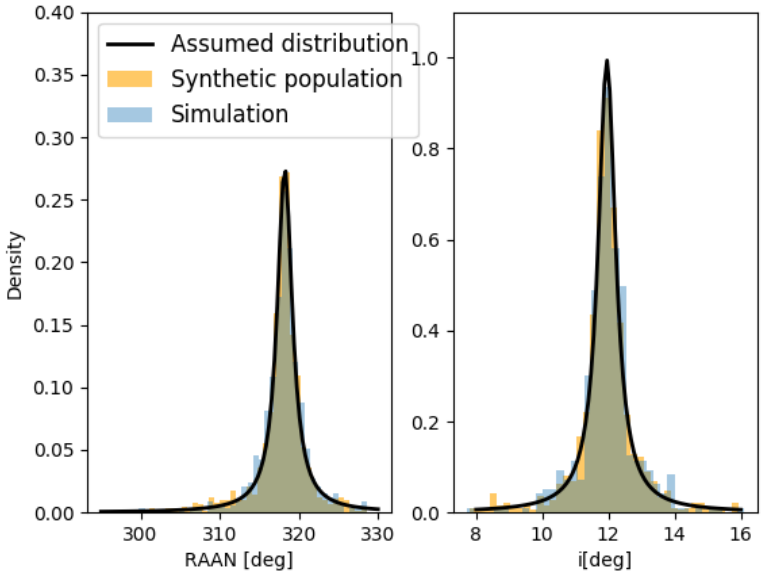

Fig. 7 Distribution of RAAN and inclination for the simulation and the synthetic population.

4.2 Second application: reducing discrepancies between IPF model and the simulated population

In this second application we consider again the case of Ekran 2 population, but this time constraints differ from the population simulated **in Section 2**. Indeed, we alter the NASA Breakup Model (NBM) and we obtain a different cloud of space debris in the simulation. The fragments of this cloud have different distributions and consequently, we infer different constraints, with which the IPF method will produce a different synthetic populations. Thus, we can observe how different constraints in the initial simulation can influence the creation of the synthetic population.

The NBM gives an increment of velocity ΔV^{NBM} for each fragment, but in the reality, the ejection velocity of fragments will depend on the energy of the event and then, when the cause is unknown, it is impossible to estimate the ejection velocity. We will consider two different cases; one of them with high ejection velocities of

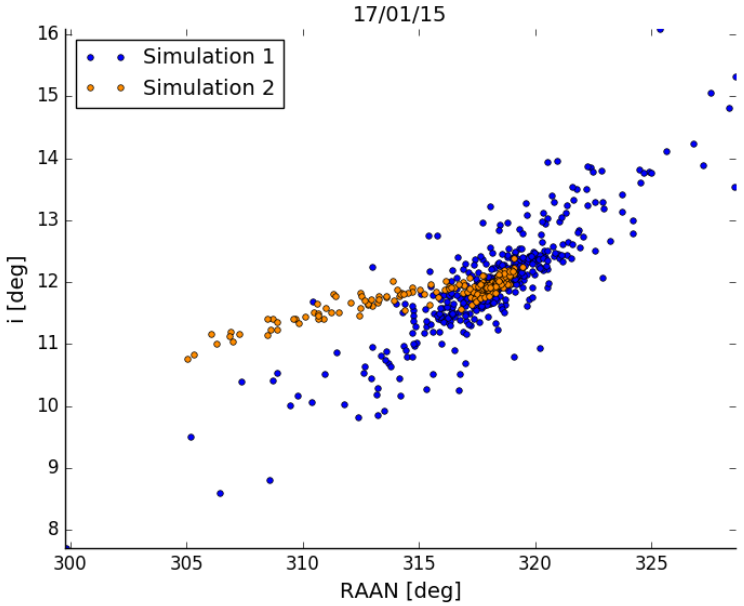

Fig. 8 The first simulation of the Ekran 2 cloud is performed with the nominal increments of velocity. The second simulation is performed dividing by ten the increments of velocity.

the fragments, and a second case with low ejection velocities. For this purpose, we introduce a factor β to obtain the increment of modified velocity $\Delta V^{modif} = \beta\ \Delta V^{NBM}$.

We assume that the ejection velocities of the fragments produced by the explosion of the satellite Ekran 2 are ten times smaller than the ejection velocities considered in the simulation presented in Section 2, i.e., we use $\beta = \frac{1}{10}$. Moreover, we take only the set of debris with a characteristic size above 1 cm. In Figure 8, we compare the distribution of the cloud produced with the nominal increments of velocity (simulation presented in Section 2) and the cloud produced with the modified ones. We observe that the second cloud (simulation 2) appears less expanded in the considered plane.

We keep the first simulation produced with the nominal values of the NASA breakup model as our initial population. We apply the IPF method using new constraints computed with the second simulation, where the increments of velocity were divided by ten. Note that this is an extreme test for the method. Indeed, the second simulation does not follow the same correlation as the first one. The synthetic population created is compared with the population of the first simulation in Figure 9. We can observe that, as desired, the shape of the scatter plot has been changed by the procedure. However, even if this change is in the direction suggested, the synthetic population does not really follow the tendency of simulation 2 used as constraint (shown in Figure 8). As seen on Figure 8, a high linear correlation is present for both simulations (0.869 for simulation 1 and 0.910 for simulation 2), but not following the same relation. This confuses the IPF process and the synthetic population has a correlation of only **0.529**. Thus, the positive correlation is still present but less evident.

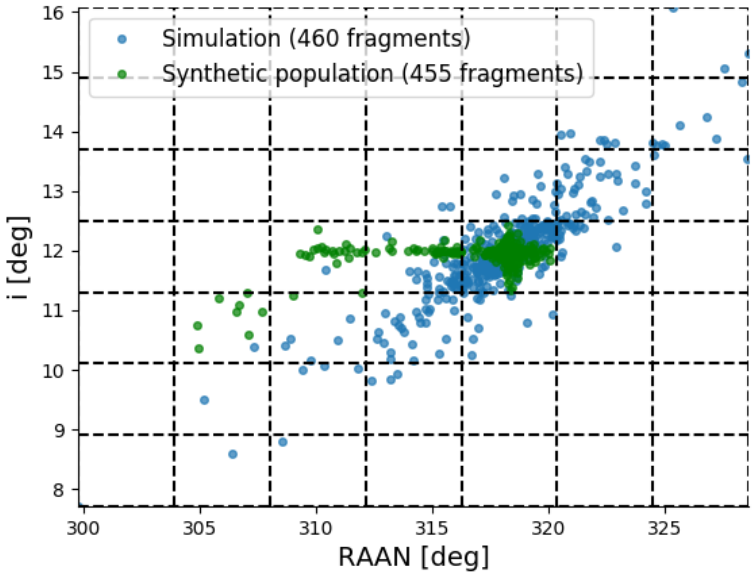

Fig. 9 Comparison between the nominal simulation and the synthetic population created with the modified simulation.

This is caused by the dense "square" of dots around a RAAN value of 318 degrees that we can observe on Figure 8.

Figure 10 contains the two simulations, the synthetic population (created from the constrains of the second simulation) and the linear regressions associated with each set of space debris. The regression lines of the simulation 2 and the synthetic population stand close together in comparison to simulation 1, indicating a good (but not perfect) improvement of the simulation with the IPF.

Try to guess a simulation from another one thanks to microsimulation is a challenging task. It is possible to summarize the followed procedure:

Step 1: Construct discretized constraints after a continuous fit of the distribution for simulation 2.

Step 2: Run an IPF initialized to simulation 1 with the constraints defined before. This step gives the number of object per discretized zone (the square on the graph).

Step 3: Create a population of space debris by determining for each object a specific attribute for each variables (thanks to the known cell that gives the range, and thanks to the continuous distribution of each variable).

This section shows that this method is quite satisfactory and we could improve by adapting **the first and third steps** to the application. More precisely, we have several propositions for a future work.

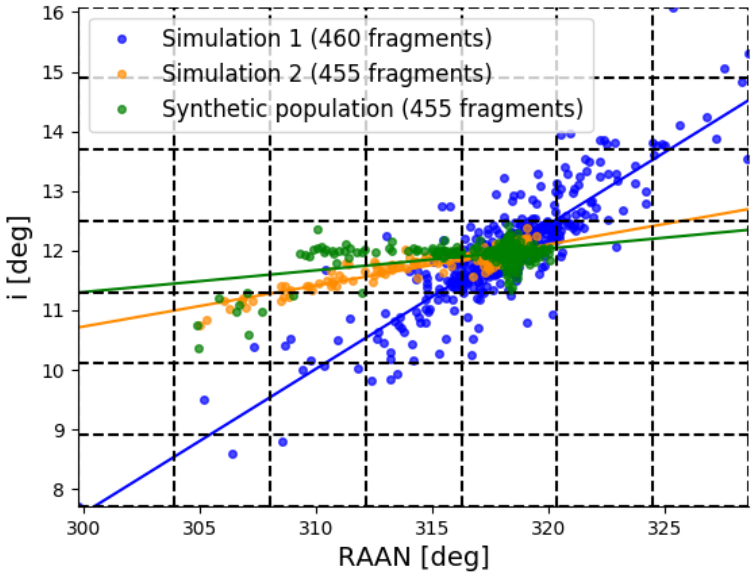

Fig. 10 Comparison between the two simulations and the synthetic population created with the modified simulation.

- Try to use as constraint a density function adapted to the simulation (in this work we just took a log-normal distribution for A/m and a **Cauchy** distribution for all other variables).
- Discretize the environment, i.e. a grid of cells of same dimension is used in this work, but to use a grid based on the quantiles could give better results.
- Tackle the zero cell problem of the IPF method, i.e. when the target simulation needs individuals in a cell not present in the simulation, they will never be in the synthetic population. This zero cell problem can be avoid as explained in [Suesse et al. (2017)].

Note that step three of the explained procedure is a stochastic process, meaning that running the code several times could give slightly different results.

5 Conclusion

In this paper we **first present a deterministic approach to generate an artificial population of space debris in the geostationary (GEO) region in agreement with the population provided by the USSTRATCOM catalog. Then, we use that generated population to create a new one by using a microsimulation method (IPF technique). This method** is based on a process of integrating multiple data to represent a real-word object into a consistent, accurate, and useful representation including

in the model additional constraints. The purpose is to create a population of space debris whose global characteristics **are closer to a population assumed as real since it merges from different sets of observational data used as constraints. Indeed, the empirical nature of the simulations could give large discrepancies between the simulated space debris population (with a space debris model) and the observed space debris population. Furthermore, the limitations of our knowledge about the events occurred in orbit, the assumptions made in the source models, the limited computational resources, limit the improvements of the calibration iterative process of the model parameters in order to obtain a space debris model in accordance with observations.**

In this work we show how to create a synthetic population of space debris by using an IPF technique. We provide two applications. The first one consists of creating a synthetic population with the same statistical properties than the assumed as real, but it includes more objects; in this application the new pieces are infered from a similar set of data. The second application provides a synthetic population different from the initial one. In this case, the final goal is to show the influence of the statistical properties used as contraints in the creation of a synthetic population of space debris.

This model is a first idea of producing a synthetic population of space debris, which explains the main idea of the process and shows the relevance of this method for the space debris models. In a future work it will be interesting to use real data as constraints and applied to a population generated by a space debris model. Then, a work will have to be done to improve the discretization process in order to suit in a better way the natural distribution of a cloud of space debris. Moreover, we will have to investigate how the zero cells in the contingency table, a well-known problem in the microsimulation community, represent a difficulty for the weighting process. We limited our application on objects with a size greater than 1 cm mainly produced by breakups to show with the easiest case how the process work but we can generalize it to other populations generated by Ejecta, NaK droplets or residuals from solid rocket motor firings.

This research proposes to join two fields: microsimulation and space debris dynamics by applying the IPF method in order to create synthetic populations of space debris. This innovative approach is a first attempt which could be complemented by testing and comparing other methods of microsimulation based on heuristic methods (simulated annealing, genetic algorithms).

Acknowledgements The work of A. Petit is supported by a F.R.I.A Ph.D grant. The work of D. Casanova **was supported by the Spanish Ministry of Economy and Competitiveness, Project no. ESP2017–87113–R (AEI/FEDER, UE); and by the Aragon Government and European Social Fund (group E24_17R).** This research used resources of the “Plateforme Technologique de Calcul Intensif (PTCI)” (http://www.ptci.unamur.be) located at the University of Namur, Belgium, which is supported by the F.R.S.-FNRS under the convention No. 2.4520.11. The PTCI is member of the ”Consortium des quipements de Calcul Intensif (CCI)” (http://www.ceci-hpc.be).

References

[Casanova et al. (2014)] Casanova, D., Tardioli, C., Lemaître A.: Space debris collision avoidance using a three-filter sequence. Mon. Not. R. Astron. Soc. 442 (4), 32353242, (2014)

[Casanova et al. (2015)] Casanova, D., Petit, A., Lemaître, A.: Long-term evolution of space debris under the J2 effect, the solar radiation pressure and the solar and lunar perturbations. Celest. Mech. Dyn. Astr. 123, 223-228, (2015)

[Celletti et al. (2017)] Celletti, A., Efthymiopoulos, C., Gachet, F., Gale, Pucacco, G.: Dynamical models and the onset of chaos in space debris. Int. J. Nonlin. Mech. 90, 147-163, (2017)

[Cowardin et al. (2017)] Cowardin, H., Anz-Meador, P. D., Reyes, J. A.: Characterizing GEO Titan IIIC Transtage Fragmentations Using Ground-based and Telescopic Measurements. Advanced Maui Optical and Space Surveillance (AMOS) Technologies Conference, 36, (2017)

[Delsate and Compère (2012)] Delsate, N., Compère, A.: NIMASTEP: a software to modelize, study, and analyze the dynamics of various small objects orbiting specific bodies. Astron. Astrophys. 540 (A120), (2012)

[Flegel et al. (2009)] Flegel, S., Gelhaus, J., Wiedemann, C., et al.: The MASTER-2009 Space Debris Environment Model. Proceedings of the Fifth European Conference on Space Debris. 672, 15 (2009)

[Horstmann et al. (2017)] Horstmann, A., Stoll, E., Krag, H.: A Validation Method of ESA's MASTER 1 cm Population in Low Earth Orbit. Advanced Maui Optical and Space Surveillance (AMOS) Technologies Conference, id.87 (2017)

[Hubaux et al. (2012)] Hubaux, Ch., Lemaître, A., Delsate, N., Carletti, T.: Symplectic integration of space debris motion considering several Earths shadowing models. Adv. Space Res. 49 (10), 1472-1486, (2012)

[Hubaux and Lemaître (2013)] Hubaux, C., Lemaître, A.: The impact of Earths shadow on the long-term evolution of space debris. Celest. Mech. Dyn. Astr. 116, 79-95, (2013)

[IADC (2007)] Inter-Agency Space Debris Coordination Committee (IADC). IADC-02-01. Space Debris Mitigation Guidelines. Revision 1, (2007)

[Jehn et al. (2006)] Jehn, R., Ariafar, S., Schildknecht, T., Musci, R., Oswald, M. Estimating the number of debris in the geostationary ring. Acta Astronaut. 59 (15), 84-90, (2006)

[Johnson et al. (2001)] Johnson, N. L., Krisko, P. H., Liou, J.-C., Anz-Meador, P. D.: NASA's new breakup model of evolve 4.0. Adv. Space Res. 28 (9), 1377-1384, (2001)

[Johnson et al. (2008)] Johnson, N. L., Stansbery, E., Whitlock, D. O., Abercromby K. J., Shoots D.: History of on-orbit satellite fragmentations, 14th edition. Orbital Debris Program Office, (2008)

[Lewis et al. (2001)] Lewis, H. G., Swinerd, G., Williams, N., Gittins, G.: DAMAGE: a dedicated GEO debris model framework. Proceedings of the Third European Conference on Space Debris. 1, 373-378 (2001)

[Liou et al. (2004)] Liou, J.-C., Hall, D. T., Krisko, P. H., Opiela, J. N.: LEGEND a three-dimensional LEO-to-GEO debris evolutionary model, Adv. Space Res. 34 (5), 981-986 (2004)

[Lovelace and Ballas (2013)] Lovelace, R., Ballas, D.: Truncate, replicate, sample: A method for creating integer weights for spatial microsimulation. Comput. Environ. Urban Syst., 41 (1), 1-11, (2013)

[Lovelace and Dumont (2016)] Lovelace, R., Dumont, M.: Spatial microsimulation with R. CRC Press (2016)

[Orbital Debris Program Office (2016)a] Orbital Debris Program Office. Orbital Debris Quarterly News. 20 (1-2), 3-4, (2016)

[Orbital Debris Program Office (2016)b] Orbital Debris Program Office. Orbital Debris Quarterly News. 20 (4), 2-3, (2016)

[Orbital Debris Program Office (2018)] Orbital Debris Program Office. Orbital Debris Quarterly News. 22 (2), 4, (2018)

[Petit and Lemaître (2016)] Petit, A., Lemaître, A.: The impact of the atmospheric model and of the space weather data on the dynamics of clouds of space debris. Adv. Space Res. 57 (11), 2245-2258, (2016)

[Rossi and Valsecchi (2006)] Rossi, A., Valsecchi, G.B.: Collision risk against space debris in Earth orbits. Celest. Mech. Dyn. Astr. 95 (14), 345356, (2006)

[Rossi et al. (2009)] Rossi, A., Anselmo, L., Pardini, C., Jehn, R., Valsecchi, G. B.: The New Space Debris Mitigation (SDM 4.0) Long Term Evolution Code. Proceedings of the Fifth European Conference on Space Debris. 672 (90), (2009)

[Rossi et al. (2016)] Rossi, A., Lewis, H., White, A., Anselmo, L., Pardini, C., Krag, H., Bastida Virgili B.: Analysis of the consequences of fragmentations in low and geostationary orbit. Adv. Space Res. 57 (8), 1652-1663, (2016)

[Schildknecht et al. (2004)] Schildknecht, T., Musci, R., Ploner, M., Beutler, G., Flury, W., Kuusela, J., De Leon Cruz, J., De Fatima Dominguez Palmero, L. Adv. Space Res. 34 (5), 901-911, (2004)

[Suesse et al. (2017)] Suesse, T., Namazi-Rad, M., Mokhtarian P., Barthélemy J.: Estimating Cross-Classified Population Counts of Multidimensional Tables: An Application to Regional Australia to Obtain Pseudo-Census Counts. Journal of Official Statistics, 33 (4), 1021-1050, (2017)

[Valk et al. (2009)] Valk, S., Delsate, N., Lemaître, A., Carletti, T.: Global dynamics of high area-to-mass ratios GEO space debris by means of the MEGNO indicator. Adv. Space Res. 43 (10), 1509-1526, (2009)

[Valsecchi et al. (2002)] Valsecchi G.B., Rossi A.: Analysis of the Space Debris Impacts Risk on the International Space Station. In: Celletti A., Ferraz-Mello S., Henrard J. (eds) Modern Celestial Mechanics: From Theory to Applications. Springer, Dordrecht. (2002)

[Wnuk (1997)] Wnuk, E.: Space debris - The short term orbital evolution in the earth gravity field. Celest. Mech. Dyn. Astr. 66, 71-78, (1997)