



POLITECNICO DI MILANO

SCHOOL OF INDUSTRIAL AND INFORMATION ENGINEERING

DEPARTMENT OF AEROSPACE SCIENCE AND TECHNOLOGY

MASTER OF SCIENCE IN SPACE ENGINEERING

Outlier identification in Two-Line Element sets of space debris

Supervisor:

Prof. Mauro Massari

Candidate:

Vasile Stratan

Matricola: 873957

Academic Year 2019-2020

“There must be something in books, things we can’t imagine, to make a woman stay in a burning house; there must be something there. You don’t stay for nothing”

-Ray Bradbury - Fahrenheit 451-

To my family, for giving me the possibility to follow this path.

To professor Mauro Massari for his support and guidance.

TABLE OF CONTENTS

Nomenclature	8
List of Figures	10
List of tables	11
Abstract	12
1 Introduction	14
2 Space Debris problem	21
2.1 Debris Sources	21
2.2 Hazards	24
2.3 Detection and Tracking.....	26
2.3.1 United States of America	26
2.3.2 European Space Agency	29
2.3.3 Russian Federation	30
3 Two-Line Elements.....	32
3.1 Two-Line Elements description	32
3.2 Two Line Elements sources.....	39
4 Two Line Elements Filtering	42
4.1 Filtering Algorithm	42
4.2 Two Line Element update-correction filter	43
4.3 Two Line elements - epoch gaps detection	45
4.3.1 Mean plus standard deviation method	48
4.3.2 Median plus median absolute deviation method.....	48
4.4 Orbital elements filters	50
4.4.1 Outlier detection - First method.....	50
4.4.2 Outlier detection - Second method.....	54
4.4.3 Deleting negative Bstar.....	56

4.5 Optimization of the orbital element filters.....	56
4.5.1 Optimization of the first method filter.....	56
4.5.2 Optimization of the second method filter	65
5 Results	67
5.1 The performance of the first method.....	67
5.2 The performance of the second method	69
6 Conclusion and further development.....	71
Bibliography.....	72
Appendices.....	74

NOMENCLATURE

ESA - European space agency

NASA - National Aeronautics and Space Administration

JAXA - Japan Aerospace exploration Agency

Roscosmos - Roscosmos State Corporation for Space Activities

ASI – Agenzia Spaziale Italiana (Italian Space agency)

CNES - Centre national d'études spatiales (National Centre for Space Studies)

DLR - *Deutsches Zentrum für Luft- und Raumfahrt* (German Aerospace Center)

KARI - Korea Aerospace Research Institute

ISRO - Indian Space Research Organization

NSAU - National Space Agency of Ukraine

UKSA – UK Space Agency

PL - Payload

PF - Payload Fragmentation Debris

PD - Payload Debris

PM - Payload Mission Related Object

RB - Rocket Body

RF - Rocket Fragmentation Debris

RD - Rocket Debris

RM - Rocket Mission Related Object

UI - Unidentified

GEO - Geostationary Orbit

IGO - Inclined Geosynchronous Orbit

EGO - Extended Geostationary Orbit
NSO - Navigation Satellites Orbit
GTO - GEO Transfer Orbit
MEO - Medium Earth Orbit
LEO - Low Earth Orbit
HAO - High Altitude Earth Orbit
MGO - MEO-GEO Crossing Orbits
HEO - Highly Eccentric Earth Orbit
UFO - Undefined Orbit
ESO - Escape Orbits
TLE – Two-line elements
EVA -Extra vehicular activity
EOL - End of life
ODPO -The Orbital Debris Program Office
EISCAT - European Incoherent Scatter Scientific Association
JSPOC - Joint Space Operations Centre
SD – Standard Deviation
MAD – Median Absolute Deviation
IQR – Interquartile Range

LIST OF FIGURES

<i>Figure 1 Graphic representation of the debris amount in different orbits around Earth. [1].....</i>	<i>15</i>
<i>Figure 2 Number evolution for different class of objects from 1960 to 2019 [2].....</i>	<i>16</i>
<i>Figure 3 Mass evolution for different object class from 1960 to 2019, (ESA Space Debris Office 2019) [2].....</i>	<i>16</i>
<i>Figure 4 Evolution of the number of object for different orbits, (ESA Space Debris Office 2019) [2] ...</i>	<i>17</i>
<i>Figure 5 Evolution by mass of different class of orbits, (ESA Space Debris Office 2019) [2].....</i>	<i>17</i>
<i>Figure 6 Distribution of objects in Low Earth Orbit (ESA Space Debris Office 2019) [2].....</i>	<i>18</i>
<i>Figure 7 The point of collision of the Iridium 33 with Kosmos 2251 over northern Siberia [8].....</i>	<i>22</i>
<i>Figure 8 Simulated debris cloud formed after 20 minutes following the impact [8].....</i>	<i>23</i>
<i>Figure 9 Simulated debris fields after 50 minutes following the impact [8].....</i>	<i>23</i>
<i>Figure 10 The resulted damage of Sentinel 1 solar array after it was struck by debris (Credit: ESA).</i>	<i>25</i>
<i>Figure 11 The position of SSN systems [10]</i>	<i>28</i>
<i>Figure 12 Measurement data used by NASA ODPO to describe debris population [11].....</i>	<i>29</i>
<i>Figure 13 TIRA space observation radar [12].....</i>	<i>30</i>
<i>Figure 14 SPACE-TRACK.ORG user interface for TLE download.....</i>	<i>40</i>
<i>Figure 15 The Space Track TLE Retriever software interface.....</i>	<i>41</i>
<i>Figure 16 The workflow breakdown of the filtering process</i>	<i>42</i>
<i>Figure 17 The mean motion representation for the TLE set of the object 25543</i>	<i>44</i>
<i>Figure 18 Cumulative distribution of the update frequency for the object “25543”</i>	<i>46</i>
<i>Figure 19 Selection frequency of outlier detection methods [18].....</i>	<i>47</i>
<i>Figure 20 Effect of outliers on threshold based techniques [18].....</i>	<i>47</i>
<i>Figure 21 Visualisation of the data binning process and how the tail discontinuity is found.....</i>	<i>49</i>
<i>Figure 22 The result of running the gap detection algorithm on the set of the object “13025”</i>	<i>50</i>
<i>Figure 23 The sliding window used for the linear regression and residual computation.</i>	<i>51</i>
<i>Figure 24 The representation of focal point shifting in the sliding window</i>	<i>55</i>
<i>Figure 25 Smoothed and resampled mean motion series for the object “33542”</i>	<i>58</i>
<i>Figure 26 Inserted artificial outliers in the smoothed and resampled mean motion series of the object “33542”</i>	<i>59</i>
<i>Figure 27 The result of running the mean motion filter on the set enriched with outliers</i>	<i>60</i>
<i>Figure 28 The result in terms of false positive/false negatives for multiple iteration process</i>	<i>61</i>

LIST OF TABLES

<i>Table 1 A sample of TLEs date and time format derived from the epoch</i>	<i>44</i>
<i>Table 2 The set of orbital debris objects for different orbits used for the training of the filters</i>	<i>57</i>
<i>Table 3 The quantity of inserted single, double, and triple consecutive outliers in mean motion series</i>	<i>62</i>
<i>Table 4 Initial settings of the first method to run the training algorithm for the mean motion filter ..</i>	<i>62</i>
<i>Table 5 Initial settings of the first method to run the training algorithm for the eccentricity</i>	<i>63</i>
<i>Table 6 Initial settings of the first method to run the training algorithm for the inclination</i>	<i>64</i>
<i>Table 7 Initial settings of the second method used for the training algorithm of the mean motion.....</i>	<i>65</i>
<i>Table 8 Initial settings of the second method used for the training algorithm of the eccentricity</i>	<i>66</i>
<i>Table 9 Initial settings of the second method to run the training algorithm for the inclination</i>	<i>66</i>
<i>Table 10 Means and Standard deviations for the obtained false negatives/positives of each filter using first method.....</i>	<i>70</i>
<i>Table 11 Means and Standard deviations for the obtained false negatives/positives of each filter using second method.....</i>	<i>70</i>

ABSTRACT

The well-known problem of the space field, orbital debris, has been around for many decades already. It became clear that the humankind can pollute not only the air and water on Earth, but also the space environment is close of being considered “uninhabitable”. Letting aside the solutions to reduce the number of current and future debris, how do we maintain operational the global space program? The threat imposed by space debris can be catastrophic for the future space exploration and for people on the Earth’s surface. The available solution is to track the debris during atmospheric re-entry and predict the location of the impact, or in the case of operational satellites, to execute anti-collision maneuvers. As the information about many thousands of objects is limited, the only source of orbital information used for orbit prediction is given by Two Line Elements. The aim of this thesis is to build an algorithm that detects the outlying data in the TLE set of each object of interest. Two different methods were developed. To improve the efficiency of each method, objects belonging to different types of orbits were used during the optimization. The last part of the thesis presents and compares the performance of the two methods. For the assessment of the efficiency of each method, 100 TLE sets were used, and the results are presented in terms of missed outliers.

ABSTRACT

Il noto problema dell'ambito spaziale, i detriti spaziali, esiste già da molti decenni. È ormai chiaro che l'umanità è riuscita a inquinare non solo l'aria e l'acqua sulla Terra, ma anche l'ambiente spaziale sta per essere considerato "inabitabile". Tralasciando le soluzioni per diminuire il numero di detriti attuali e futuri, come manteniamo operativa l'attività nell'ambito spaziale? La minaccia rappresentata dai detriti spaziali può essere catastrofica per la futura esplorazione spaziale e per le persone sulla superficie della Terra.

La soluzione disponibile è tracciare i detriti durante il rientro atmosferico e anticipare la posizione dell'impatto, e nel caso di satelliti operativi, eseguire manovre di anticollisione. Siccome i dati su molte migliaia di oggetti sono limitati, l'unica fonte di informazioni orbitali utilizzate per la previsione dell'orbita è la Two Line Elements. L'obiettivo della tesi è di costruire un algoritmo che rilevasse i dati anomali contenuti nei set TLE di ogni oggetto analizzato. Sono stati sviluppati due metodi diversi. Per migliorare l'efficienza di ciascun metodo, sono stati utilizzati oggetti appartenenti a diversi tipi di orbite durante il processo di ottimizzazione. L'ultima parte della tesi presenta e confronta le prestazioni dei due metodi. Per la valutazione dell'efficienza di ciascun metodo, sono stati utilizzati 100 set di TLE, e i risultati sono presentati in termini di valori anomali mancati.

1 INTRODUCTION

Sputnik 1 (“Простейший Спутник 1”) was the first artificial Earth’s satellite launched in space by the Soviet Union on 4th October 1957. It was the day that changed the pace of modern industry, and the start of the Space Race between United States and Soviet Union.

Even though the resulting technological advancement favoured only military application in the beginning, after some years it became clear that the wide range of opportunities which space offers is a good promising also for civil application. Having such potential, the space began to be heavily exploited. After so many years of sending rockets and payloads into orbits, space environment became polluted.

The man-made objects floating in space at different altitudes are referred as *space debris* and includes, but not limited to, segments from the launchers (e.g. fairings, upper stages, etc.), connection/separation elements (e.g. exploding bolt), fragments which did result as consequence of satellite collisions, decommissioned satellites, etc.



Figure 1 Graphic representation of the debris amount in different orbits around Earth. [1]

Ignored in the beginning; due to its potential threat to spacefaring nations, it became a phenomenon worth paying attention to.

The problem was first taken in account by the United States and it reached International community by mid-1970. The biggest concern comes after the postulate of Donald Kessler in 1978, stating that the continuous increase in space debris and the lack of actions to mitigate it will lead to a chain reaction making the space hazardous for space flight. Accordingly, on the initiative of NASA, a global bilateral knowledge transfer was initiated leading to the establishment of Inter-Agency Space Debris Coordination Committee (IADC) in 1993 by ESA (Europe), NASA (USA), JAXA (Japan) and ROSCOSMOS (Russian Federation) and later joined by other space agencies, such as ASI (Italy), CNES (France), CNSA (China), CSA (Canada), DLR (Germany), KARI (South Korea), ISRO (India), NSAU (Ukraine), and UKSA (United Kingdom).

The evolution of space debris in time is characterized in detail in ESA's Space Environment Report [2]. The report describes the space environment across history including object classifications and definitions. Moreover, the author makes reference to space debris mitigation guidelines and metrics.

It could be observed in Figure 2 to Figure 5 the exponential growth in the number and mass of space debris. Different class of objects and orbits have an exponential grow, meaning an increasing awareness.

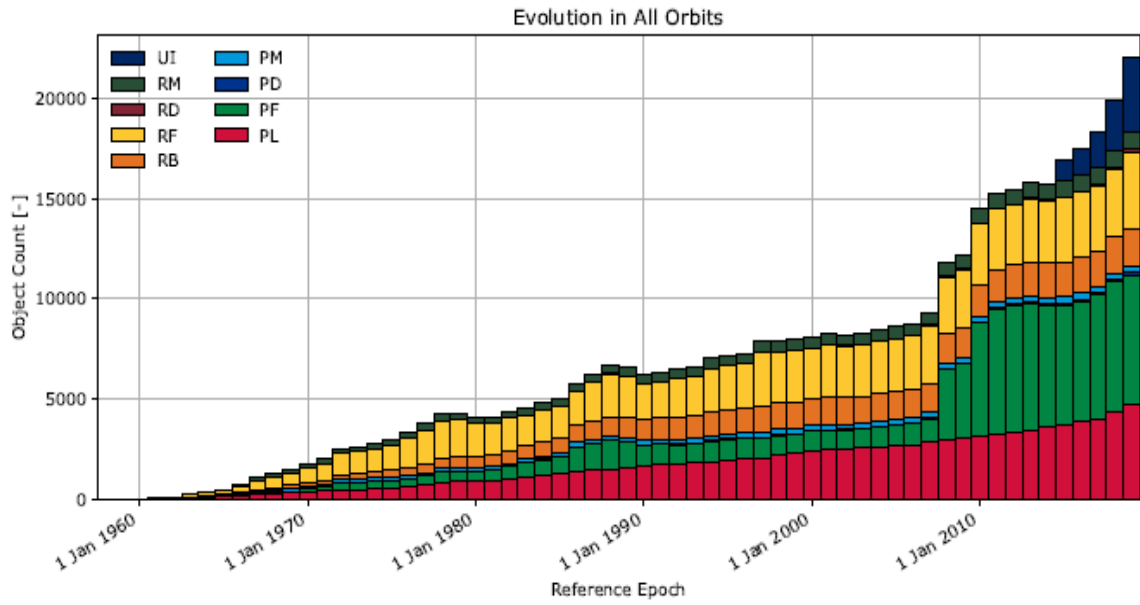


Figure 2 Number evolution for different class of objects from 1960 to 2019 [2]

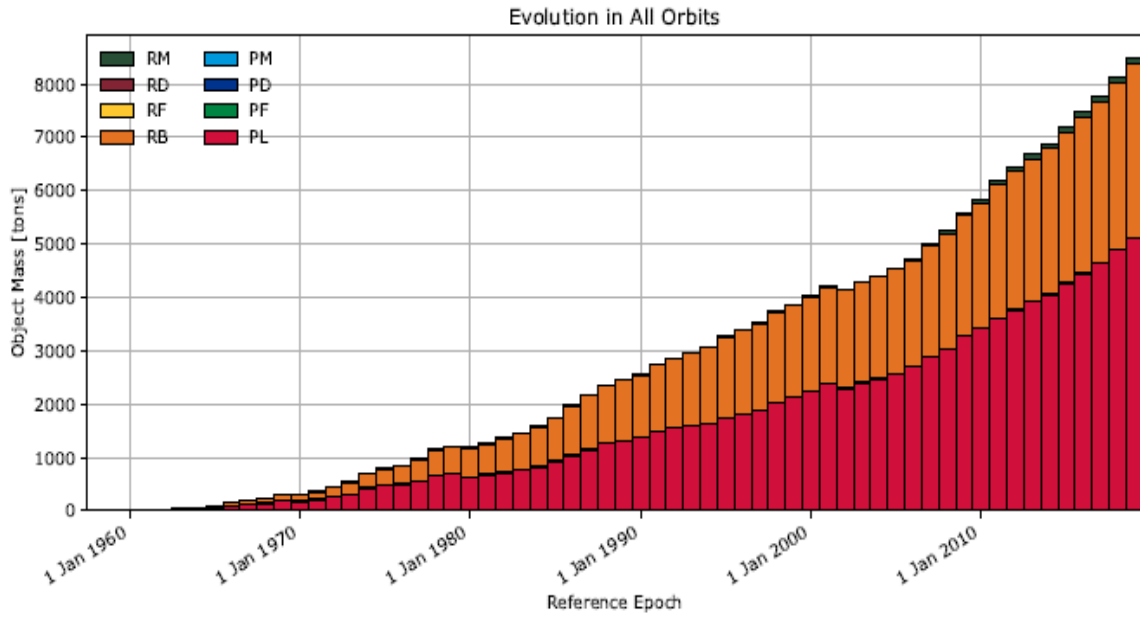


Figure 3 Mass evolution for different object class from 1960 to 2019, (ESA Space Debris Office 2019) [2]

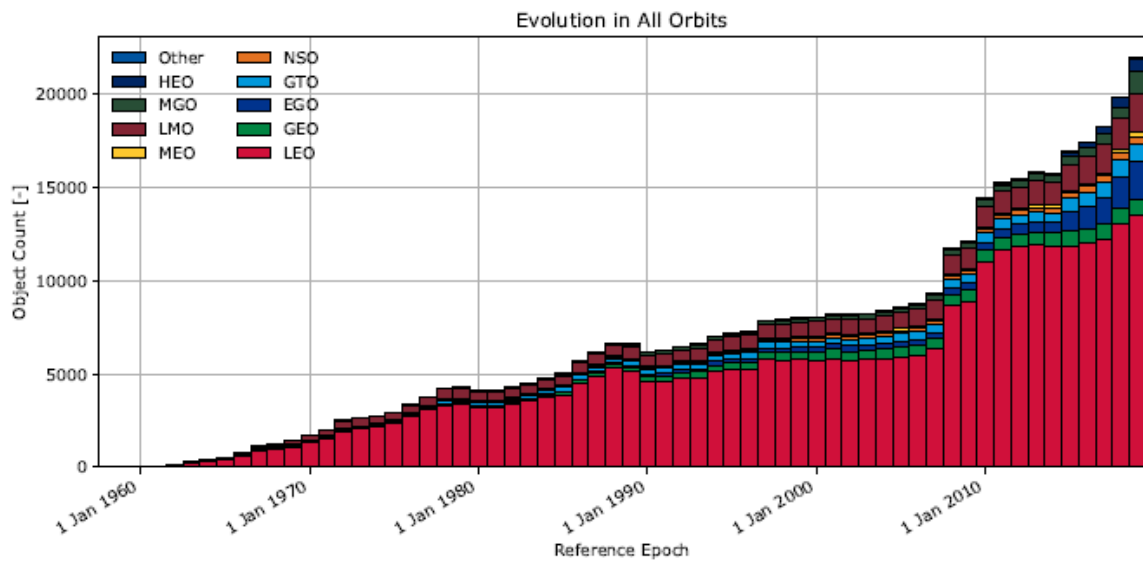


Figure 4 Evolution of the number of object for different orbits, (ESA Space Debris Office 2019) [2]

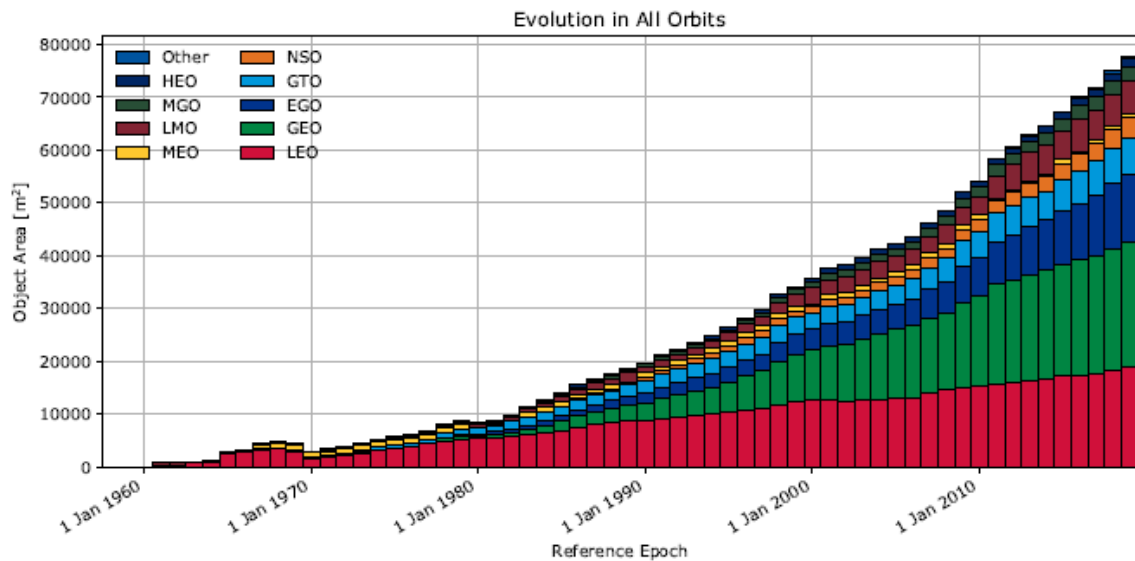


Figure 5 Evolution by mass of different class of orbits, (ESA Space Debris Office 2019) [2]

Each orbit is affected by space debris in the same way, but the effects might be more drastic for each of them. Considering GEO, the inserted satellites are on average more

expensive and the impact with a space debris would lead to an economic impact. Meanwhile, LEO is the busiest orbit and for now, is the only one that supports crewed application like International Space Station. Therefore, space debris collisions could have an economic impact but also endanger the life of the crew.

With the commercialization of space, the human activity in LEO would grow, starting with Axiom Space, the first commercial Space Station, and followed by new rising companies that are interested in using LEO for private research or even tourism.

How safe it is to continue bringing humans in space? In Figure 6 the density of object both residing and crossing LEO is represented. One can observe that the lower inclinations and lower altitudes, that are of interest for human applications, are crossed by a high number of debris. Thus, the safety of astronauts is under question.

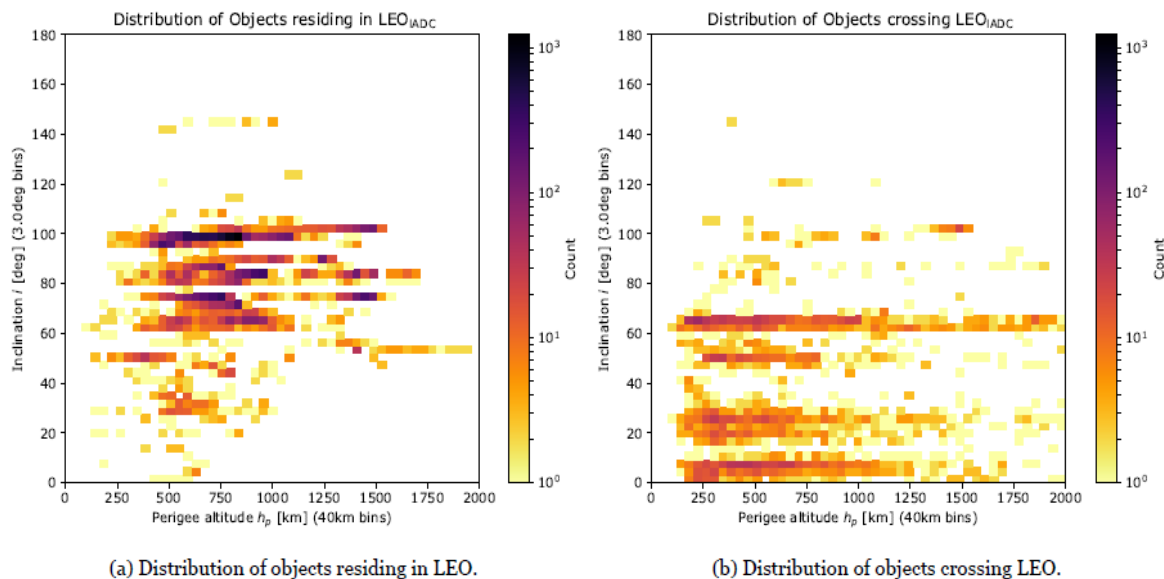


Figure 6 Distribution of objects in Low Earth Orbit (ESA Space Debris Office 2019) [2]

To reduce the risk of space collisions with orbital debris, collision avoidance manoeuvres are performed, therefore moving from the path of incoming threat. Nowadays, one of the many tasks of space agencies is to keep tracking as many debris as possible. Both United States and Russian Federation own networks of radars to detect, track and catalogue space objects. Each detected object is included in catalogues as Two-Line Elements (TLE) sets which are a standardized way to include all orbital elements of the object and other useful information. The deterministic data will go in these catalogues including statistical information on number of the detected objects.

The available and published TLE can be used to propagate the orbit of desired object in time to predict its behaviour like the re-entry time and location but also to predict a possible impact with an in-service satellites and spacecrafts. However, the Two-Line Elements sets does not include the accuracy of the orbit and cannot express the accuracy of conjunction analysis or re-entry prediction. Therefore, simply propagating the TLEs using SGP4 means limited accuracy.

The missing covariance matrix can be determined from the TLE set using different methods, and then, it can be used for orbit propagation. For this reason, it is important to process more qualitative observations. Unfortunately, each set of two-line elements can be affected by outlying data. One can find released TLE with corrupted data, with measuring errors in orbital elements, or even with data belonging to other objects. To increase the accuracy of using TLE in future analysis, the aim of this work is to find an efficient way of filtering out data from the TLE sets that is hindering the orbit determination process.

Literature Review

The space debris research field has become a primary topic for main players of the space exploration agencies. Therefore, there are several works treating the use of TLE elements of space debris and satellites for orbit determination.

The work done by Song et al [3] uses Two-Line Elements to detect orbit anomalies occurred due to collisions between two satellites, breakup of satellites, explosions or due to environmental orbit anomaly. The method called SACM, is presented by using the relation between the change of orbit parameters and velocity increments

$$\begin{cases} \Delta v_U = \frac{n\sqrt{1-e^2}}{2\sqrt{(1+2e\cos f+e^2)}} \cdot \Delta a \\ \Delta v_N = \frac{n}{\sin E} \left(\frac{\cos f + e}{\sqrt{(1+2e\cos f+e^2)}} \cdot \Delta a - \frac{a\sqrt{(1+2e\cos f+e^2)}}{1-e^2} \cdot \Delta e \right) \\ \Delta v_W = \frac{na^2\sqrt{1-e^2}}{r \cos u} \cdot \Delta i \end{cases} \quad (1)$$

Where $\Delta v_U, \Delta v_N, \Delta v_W$ are velocity increments of the along-track, in plane and out plane. The author discards from the start 20% of the data with larger values, considering them as outlying or anomalous. For the remaining data, the anomaly detection process consists in using the data of the last three months and considering the orbital data at any time as basis. The semi-major axis changes (Δa) of the other data with respect to the basis is then computed, obtaining a series of time intervals Δt and Δa . The absolute value of Δa is taken as a variable, and for a sample of a time interval the mean and standard deviation is computed and used to define a threshold for the semi-major axis change anomaly detection.

Another work discussing the use of two-line elements for space object tracking and re-entry predictions was performed by Vallado and Cefola [4]. In their paper, it is mentioned the challenges related to the accuracy of TLE due to the absence of a covariance but also due to unknown manoeuvres. Large manoeuvres could lead to a satellite to become lost in the formation of a TLE, resulting to incorrectly assigned observation.

Lemmens and Krag [5], describes two developed methods for detecting space events occurred to space objects using their two-line elements. The first method, TLE Consistency Check (TCC), was designed to be insensitive to natural events and detects events by checking the consistency between arbitrary two-line element sets of the same object. The approach is to compare the propagated state and the following released state and to analyse the difference. The second method, called TLE Time Series Analysis (TTSA) makes use of outlier statistics on the entire series.

Similar to Lemmens and Krag, another paper on space event detection using TLE was published by Patera [6]. In his work, he makes use of the energy derived from the parameters of the two-line elements. Events like manoeuvres, collisions and interaction with atmosphere would be noticed in the change of energy. The energy was computed as kinetic plus potential based on spherical Earth from mean motion.

$$E = -0.5 \left(\frac{\pi n G M}{12 \cdot 60^2} \right)^{\frac{2}{3}} \quad (2)$$

A moving window is used for the detection of space events to fit a curve between a sample of points. The residual between the fitted curve and actual data or the derivative of the fitted polynomial are determined. The events are detected by analysing the residuals or the derivatives, and based on statistics of these data, large variation are detected.

2 SPACE DEBRIS PROBLEM

In the introduction chapter it was already presented what space debris represents and how they affect the future of space. It is important to specify that space debris refers to man-made objects that became inactive in orbit or parts of them. By knowing the exact condition that favours the increase of orbital debris, the next missions could be planned and designed to minimize the new introduced debris into orbit.

As of October 2019, The US Space Surveillance Network reported approximately 20,000 objects on Earth's orbits and 2,218 operational satellites. Anyway, this is only the amount of large enough objects to be detected. It is estimated that about 128 million pieces smaller than 1 cm, 900,000 pieces of 1-10 cm, and 34,000 pieces larger than 10 cm are orbiting around Earth. Despite their small dimensions, the aftermath of a collision with a functional satellite could be critical to its mission and the number of debris would further be increased.

2.1 Debris Sources

Due to the high cost of placing a satellite in an orbit, its returning back to Earth was not considered worth the expenses. Therefore, starting from the beginning of space era, different processes have influenced the multiplication of debris population. Multiple sources of debris formation exist and each of them should be treated with attention:

- Upper Stages - During the launch of a payload in space, to inject it in the orbit the upper stage needs to have the velocity required by the payload on its first orbit. After deployment, the upper stage stays in the same orbit.
- Lost Equipment - The maintenance of the International Space Station requires EVA, where astronauts are carrying instruments and components outside the station. During all the years of ISS activity, but also previous space stations like MIR and Skylab, mistakes were made, and the astronauts managed to lose

cameras, wrenches, pliers, tool bags and screwdrivers.

- Decommissioned satellites - At the EOL of a satellite, the contact and control of its systems are lost, leaving it in the orbit for very long time. If the propulsion unit tanks are not completely empty, after the degradation, the leaks could lead to an explosion, breaking the satellite or the spacecraft in multiple pieces that are inserted in different orbits.
- Solid rocket motor small debris - After the firing of SRM upper stages or kick-motors, dust particles and slag are created and placed in the orbit.
- Collisions – Perhaps the biggest source of space debris. It occurs when satellites lack or underperform collision avoidance manoeuvres. On 10th February 2009, Iridium 33, a communication satellite owned by USA collided with Cosmos 2251, an inactive Russian communication satellite, about 760 km above the Taymyr Peninsula in Siberia. In the Figure 7 to Figure 9 it is shown the follow up of this collision event.

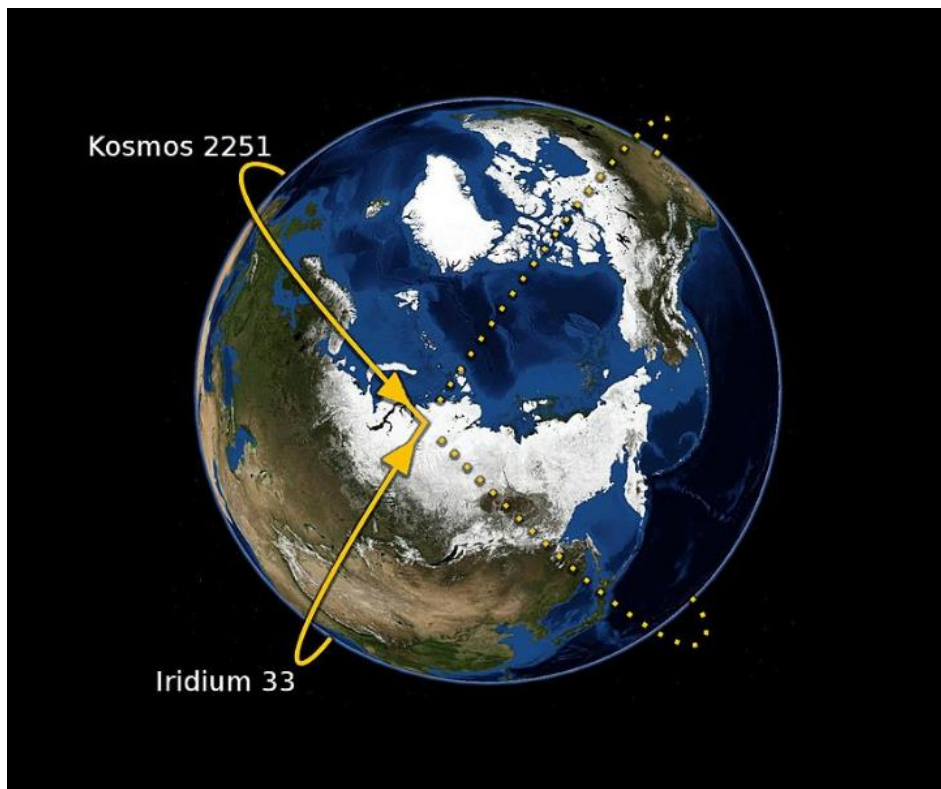


Figure 7 The point of collision of the Iridium 33 with Kosmos 2251 over northern Siberia [8]



Figure 8 Simulated debris cloud formed after 20 minutes following the impact [8]

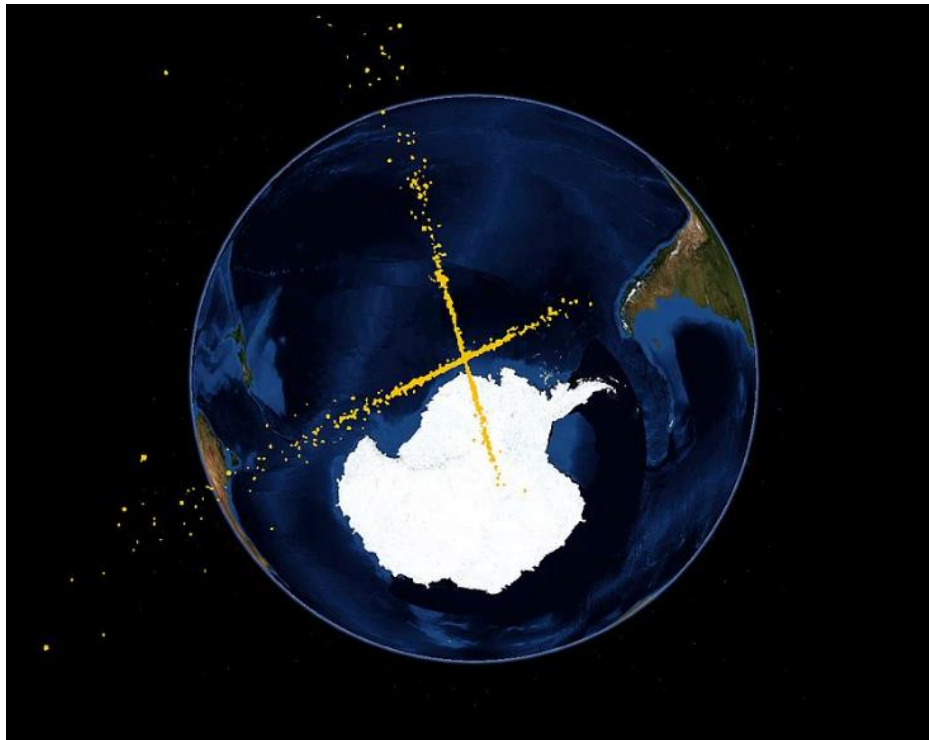


Figure 9 Simulated debris fields after 50 minutes following the impact [8]

NASA estimated ten day after the collision that the event created at least 1,000 debris pieces larger than 4 cm. US Space Surveillance Network had catalogued by 2011 over 2,000 large debris fragments.

The worst event of debris formation was caused by Chinese military after a test of anti-satellite system, destroying Fengyun-1C weather satellite, leading to an addition of more than 3,000 new debris fragments which make more than 20% of all space debris and possibly being the cause of the failure of Russian satellite BLITS.

2.2 Hazards

Space debris poses a risk for operating spacecrafts and satellites. As the amount of debris is increasing, so does the risks. If necessary measures are not considered, the LEO might end up being unusable for orbiting systems.

Different missions might be affected in a different way by the debris of a wide range dimensions.

For the unmanned spacecrafts, even though being designed with a certain safety factor, the solar panels are weaker and collision with small debris might destroy cells, as shown in Figure 10 or the entire panel, reducing the output power they can provide, thus compromising the functionality of the spacecraft. Moreover, low mass impacts cause the wear of the solar panels by changing the optical properties of the external layer or induce plasma which could cause an electrical short, again making the cell useless.

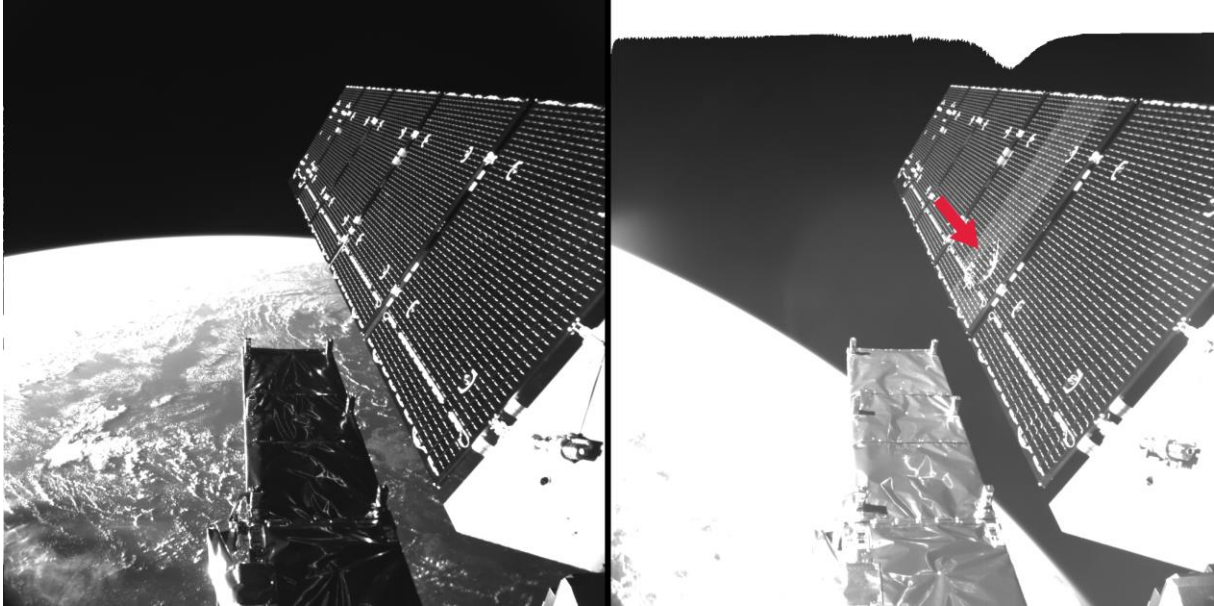


Figure 10 The resulted damage of Sentinel 1 solar array after it was struck by debris (Credit: ESA)

In the case of crewed mission, beside the above-mentioned risks, the life of astronauts might be at risk due to collision with space debris. Avoidance manoeuvres are executed when the probability of a collision is higher than the threshold. In the Space Shuttle program, multiple occasions for avoidance manoeuvring have taken place when the probability for a collision was likelier than 1 in 200.

In the case of the International Space Station, if there is a chance of collision higher than 1 in 10,000, the avoidance manoeuvre is committed. However, to do the manoeuvre in time, several days of warning are necessary to make sure the ISS completes the avoidance manoeuvre procedures.

What about the danger space debris poses for the human life on ground? It is well known that most of the debris will burn in the atmosphere during the re-entry, but large object might reach the ground. According to NASA, in the last year, every day a catalogued debris falls on Earth. Even though they rarely impacted on properties, there have been some cases when people were injured. In 1969 five Japanese sailors were injured by space debris and in 1997 a woman from Oklahoma was hit on the shoulder by a 10×13 cm part of the propellant tank of Delta II. In 2001 a Star 48 Payload assist module rocket upper stage crashed in the Saudi Arabian desert.

Thus, considering the impact of debris on people's safety and the success of space application, it is of high importance to control the space debris we are creating in future mission. Moreover, more attention must be paid for detecting and tracking a larger amount of debris with a higher accuracy.

2.3 Detection and Tracking

Information about the space debris state is helpful in the determination of the future hazards. Besides the large pieces, which are easier to detect and track, the small ones requires a bigger effort from Space agencies.

The detection of near-Earth orbital debris is achieved by using both space and ground observations, with the help of ground- based radars and optical telescopes, space-based sensors, and ground-based laboratory experiments.

A small amount of space objects in Earth's orbit are being tracked and catalogued. For each detected object, a set of orbital parameters and parameters related to the origin of the object creates a data entry called Two-line element.

A new set of data for each trackable element is generated on an as-needed basis. The frequency of the updates depends on many factors like orbit type, manoeuvring or non-manoevring, or even the phase of the object's life. Therefore, a new set might be released with a frequency of less than a period up to few days.

The process of keeping an eye on the detected and catalogued objects is expensive and currently it is affordable only for the USA's Space Surveillance Network (SSN) and Russia's military Space Surveillance System (SSS) and Europe Space Agency.

2.3.1 United States of America

After the start of space activities, North American Aerospace Defence Command (NORAD) initiated a database with all the objects that reached orbit as a way of possible threat detection. The US Space Surveillance Network (SSN) has 29 radars and optical sensors spread across the world, as shown in Figure 11. Despite the network's military purpose, it is used also for debris detection.

The radars and sensors are divided in three categories:

- Dedicated sensors
 - Ground-based Electro-Optical Deep Space Surveillance (GEODSS) sites
 - Space Surveillance Telescope (SST)
 - MOSS - an Electro-Optical (E-O) surveillance system located at the Moron Air Base, Spain

- GLOBUS II radar
- AN/FPS-85 Space Track Radar
- AN/FPS-133 Air Force Space Surveillance System, also known as the Space Fence and its replacement Space Fence
- Midcourse Space Experiment (MSX) / Space Based Visible (SBV) satellites
- Collateral Sensors
 - Maui Space Surveillance System (MSSS) and Advanced Electro-Optical System (AEOS) telescope, co-located with a GEODSS station in Maui, Hawaii
 - Haystack Ultrawideband Satellite Imaging Radar (HUSIR), Haystack Auxiliary Radar (HAX) and Millstone Hill Radar
 - ALTAIR and ALCOR radars at the Ronald Reagan Ballistic Missile Defence Test Site, Kwajalein Atoll
 - Ascension Range Radar, locate at the Eastern Space lift Range
 - Ground-Based Radar Prototype (GBR-P), located Ronald Reagan Ballistic Missile Defence Test Site, Kwajalein Atoll
- Auxiliary sensors
 - Solid State Phased Array Radar System (SSPARS) / AN/FPS-132 Upgraded Early Warning Radar (UEWR) system of system, deployed at multiple sites
 - AN/FPS-108 Cobra Dane
 - AN/FPQ-16 Perimeter Acquisition Radar Characterization System (PARCS)

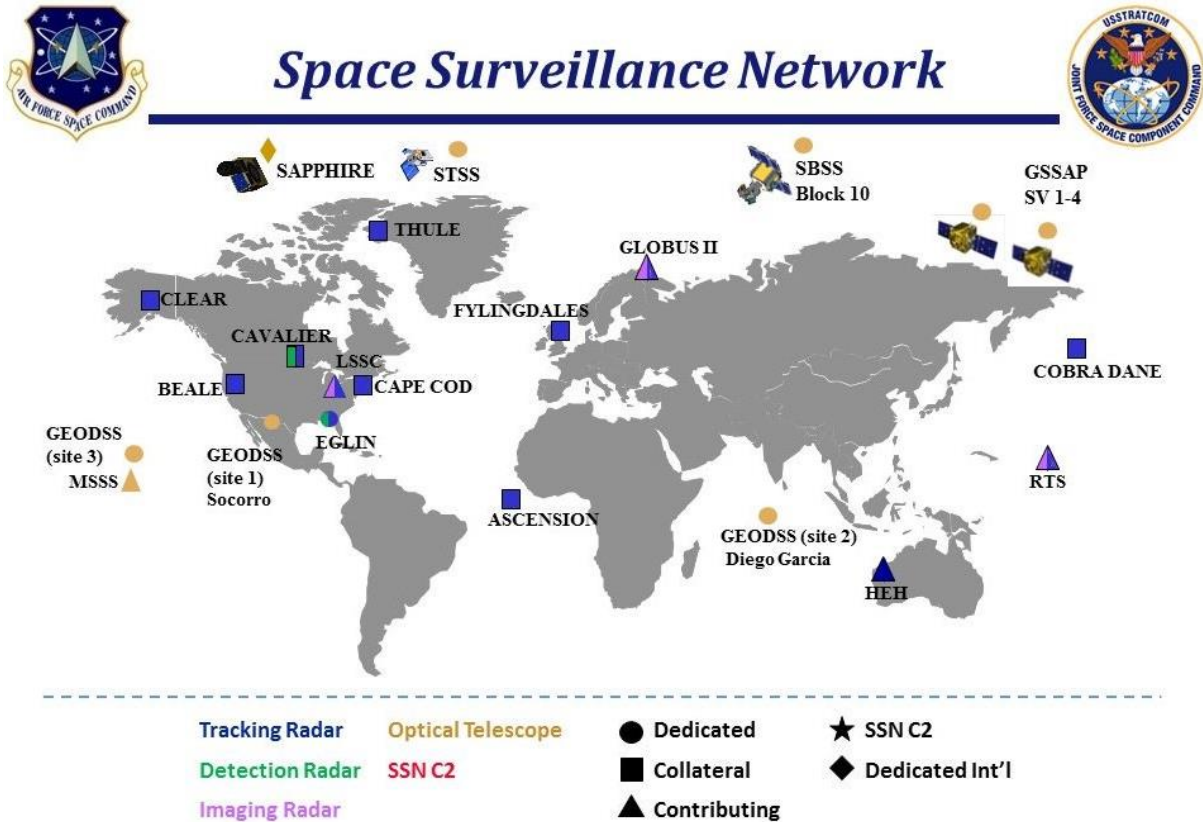


Figure 11 The position of SSN systems [10]

As SSN is dealing with larger object in Earth’s orbit, the task of keeping an eye on the small, but still threatening objects, is performed by ODPO. It uses ground-based radars to characterize the debris environment in lower orbits of the Earth.

By using the Haystack Ultrawideband Satellite Imaging Radar (HUSIR) operated by MIT’s Lincoln Laboratory [11], they are able to detect debris in the range of 5mm to 30 cm. Characterizing the debris population size by size, altitude and inclination it was concluded that there are about 500,000 debris fragments in the order of 1 cm. The NASA ODPO makes use also of the Haystack Auxiliary Radar (HAX) which is less sensitive but has a wider field-of-view. The Goldstone Orbital Debris Radar can collect data for NASA of orbital debris of about 2 mm in Low Earth Orbit.

Combining radar measurements with optical into the ORDEM model allows NASA to obtain information on debris at all altitudes. The optical measurements are done by the Eugene Stansbery Meter-Class Autonomous Telescope (ES-MCAT) and the Optical Measurement Centre (OMC). In Figure 12 it is represented the combined SSN and NASA data base.

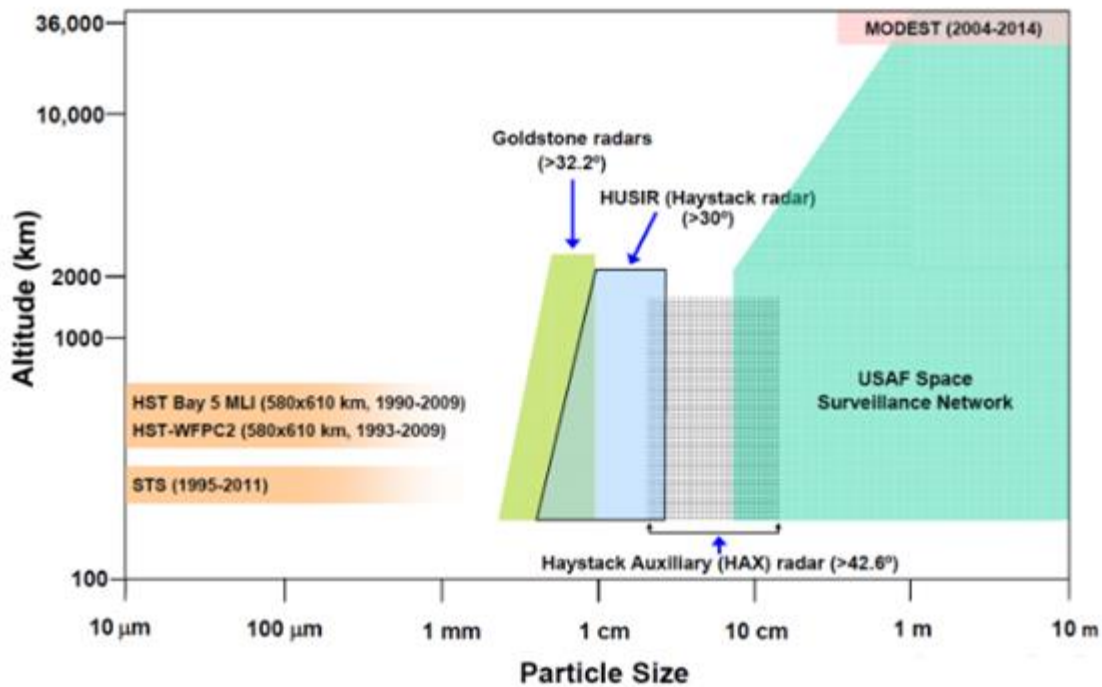


Figure 12 Measurement data used by NASA ODPO to describe debris population [11]

2.3.2 European Space Agency

ESA's Space Situational Awareness Programme (SSA), is managing the Space Surveillance and Tracking system (SST) Agency aiming to predict the potential collisions in space or on ground (for re-entry objects) by maintaining its space catalogue

ESA uses the services of the Tracking and Imaging Radar (TIRA), presented in Figure 13, located at the Fraunhofer Institute for High Frequency Physics and Radar Techniques (FHR) near Bonn for:

- searching and tracking space objects
- characterisation of the space debris environment
- tracking re-entering objects
- imaging space objects
- validation of space debris modelling



Figure 13 TIRA space observation radar [12]

The EISCAT radar based in Svalbard whose mission is to perform ionospheric measurements, is equipped with a dedicated space debris computer making it capable to detect debris in LEO down to some centimetres [12]. EISCAT was used to monitor the debris cloud of China's FengYun-1C.

The Optical Ground Station at the Teide Observatory on Tenerife, Spain can observe objects near GEO down to 10-15 cm, being top ranked worldwide. Around 75% of its detections are new objects that are not being contained in the US Space Surveillance Catalogue.

2.3.3 Russian Federation

The Russian Space Surveillance System (SSS) detects, track and catalogues orbital debris by using the old radar threat-warning network including Krona system at Zelenchukskaya in the North Caucasus and Nakhodka. The network includes 10 radars in the UHF, VHF and C-band to detect the objects in low orbits and another 12 optical instruments for the objects in higher orbits.

The network, though it suffers from worldwide coverage, leading to breakups in observations, transmits around 50000 measurements daily to the Space Surveillance Centre, to be processed and to update the space catalogue.

By 2027, a new space surveillance network called Milky Way is planned to be developed. It will include 65 ground-based optical telescopes, one LEO surveillance satellite and one optical sensor installed on the ISS.

SSS also plans to apply Artificial Intelligence to its milky way network, thus increasing the detection and prediction capabilities.

3 TWO-LINE ELEMENTS

In 1960, Max Lane developed a mathematical model to predict the location of satellites based on a minimum set of parameters. After publishing his first paper on this topic in 1965, introducing the Analytical Drag Theory, he was joined by K. Cranford. Together they upgraded the model by introducing various harmonic effects due to Moon-Sun interactions to the first spherical-symmetric non-rotating atmosphere model. As the improved version became the standard model for NORAD in the 70s, it led to the creation of the TLE.

The development of the Simplified General Perturbation (SGP) model started by Lane reached an apogee with the SGP4. The first released SGP4 propagator code was released in the Spacetrack Report Number 3 [13]. A detailed paper regarding the SGP4 code was published by Vallado and Crawford [14].

3.1 Two-Line Elements description

A Two-Line Element (TLE) set is a data format that encodes the orbital parameters together with origin of a man-made object in Earth orbit at a specific epoch. TLE are general perturbation mean elements produced using the least squares estimation from space observations. It can be also found in the format of Three Line Element, which is a derived form containing an additional line that specifies the name of the object. The name of the object might be uncalled for as the format also contains a specific identification code. The Joint JSPOC is responsible of producing new sets of TLE. The new arrived information is processed into a new TLE which becomes available to the users approximatively every 8 hours [4]. Therefore, the users have delayed access to new sets, which restricts the range of TLE applicability.

The format was original proposed for punch cards, encoding a set of elements on two 80-columns cards, and later was replaced by text files with two 69 -column ASCII lines.

The format of a TLE is the following:

```
0 AAAAAAAAAA
1 NNNNNU NNNNNAAA NNNNN.NNNNNNNN +.NNNNNNNN +NNNNN-N +NNNNN-N N NNNNN
2 NNNNN NNN.NNNN NNN.NNNN NNNNNNN NNN.NNNN NNN.NNNN NN.NNNNNNNNNNNNNNN
123456789112345678921234567893123456789412345678951234567896123456789
```

The acceptable characters are:

Numbers 0-9, capital alphabetical numbers A-Z, the period (.), the space (), the plus (+) and the minus (-).

Rules:

- Columns with period or space cannot have other character.
- Columns with an N can have only numbers from 0-9, but sometimes a space.
- Columns with an A can have any character A-Z, any number 0-9, or space.
- The column with C can have only a character describing the classification of the respective object:
S- secret
C – classified
U - unclassified.
- Columns with a plus (+) can have either plus (+) minus (-), or space ()
- Columns with a minus (-) can have either a plus (+) or minus (-)

To facilitate the interpretation of a Two-Line Element, the example of International Space Station is described.

```
ISS (ZARYA)
1 25544U 98067A 04236.56031392 .00020137 00000-0 16538-3 0 9993
2 25544 51.6335 344.7760 0007976 126.2523 325.9359 15.70406856328906
-----
1234567890123456789012345678901234567890123456789012345678901234567890
```

Line 0

- Field 0 - Columns 01-24 – Common name
Example: “ISS (ZARYA)”

This field indicates the name of the identified object according to NORAD's catalogue. In the case of a rocket body, the name is followed by 'R/B' while in the case of a debris the name of the satellite it belonged to followed by 'DEB'.

Line 1

- Field 1 - Column 01 - Line Number

It is important the #1 is present at the beginning of the line. In the case of its absence, the TLE is considered invalid.

- Field 2 - Columns 03-07 – Satellite Catalogue Number

Example: "25544"

This is the Object Catalogue Number assigned by NORAD to the specific object. Each identified debris or satellite has a distinct number. Objects with an identifier smaller than 10,000 are aligned to the right and padded with zeros or spaces to the left.

- Field 3 - Column 08 – Classification

Example: "U"

This character specifies the classification of the current object. Unclassified "U" TLEs are the only one available to public use while the Classified "C" and Secret "S" are not available.

- Field 4 – Columns 10-11 - International Designator (last two digits of the launch year)

Example: "98"

- Field 5 – Columns 12-14 – International Designator (Launch number of the year)

Example: "067"

Since the Designator gives only the last two digits of the launch year, one should add 1000 if it is equal or greater than 57 or add 2000 if it is below 57.

- Field 6 – Columns 15-17 – International Designator (Sequential ID of a piece in a launch)

Example: "A"

“A” indicates that the object is a primary payload. Subsequent lettering indicates secondary payloads and rockets that were directly involved in the launch process. Any debris detected from the original object is catalogued as a subsequent letter (B, C, D, etc.) in the order of when they were first detected. Some satellites have broken up into hundreds and even thousands of pieces, each having its own International ID number. Debris can exceed the 26-letter alphabet, and so two (or more) letters can be used (AA, AB, AC, etc).

- Field 7 – Columns 19-20 – Epoch year (last two digits of the year)

Example: “04”

The same rule applies. Since the only two digits of the year are indicated, one should add 1000 if it is equal or greater than 57 or add 2000 if it is below 57.

- Field 8 – Columns 21-32 -Epoch (Day of the year and fractional portion of the day)

Example: ”236.56031392”

The day of the year is indicated using Coordinated Universal Time (UTC).

- Field 9 – Columns 34-43 – First derivative of mean motion

Example: “.00020137”

The "Mean Motion Dot" is defined as half of the first time derivative of the Mean Motion, measured in $\text{orbits}/\text{day}^2$. It basically defines how the Mean Motion changes from day to day so that some orbit propagation software can continue to predict the location of the satellite accurately over longer periods of time from the TLE Epoch. This value can be negative or positive. Not all orbit propagators read or use this value.

- Field 10 – Columns 45-52 – Second derivative of mean motion

Example: “ 00000-0”

The Mean Motion Double Dot is defined as one sixth the second time derivative of the mean motion measured in $\text{orbits}/\text{day}^3$. It basically defines the rate of change of the Mean motion dot from day to day so that some orbit propagation software can continue to predict the location of the satellite accurately. The

mean motion double dot is normally zero unless the satellite is being maneuvered or is undergoing orbit decay.

The format of the Mean Double Dot is:

Sign: (nil or -): The sign of the digits succeeding it. No sign indicates positive;

First five digits: A number with a decimal point assumed to precede the first digit;

Sign (nil or -): The sign of the exponent. No sign indicates positive; and

Last digit: The exponent (base 10).

In the case of "-12345-6" within a TLE, it would mean $-0.12345 \cdot 10^{-6}$ orbits/day³ or in scientific notation: $-1.2345 \cdot 10^{-7}$ orbits/day³. Not all orbit propagators read or use this value.

- Field 11 – Columns 54-61 – B^* Drag term

Example: "16538-3"

The B-Star Drag Term is used for the SGP4 type propagator and estimates the effects of atmospheric drag on the satellite's motion. It is defined by the following equation:

$$B^* = \frac{C_D \rho_0 A}{2m}$$

Where: B^* = The drag term

C_D = The Coefficient of Drag

ρ_0 = The Atmospheric Density

A = The Cross-sectional Area of the Satellite

m = The Mass of the Satellite

The units are $Earth_{Radii}^{-1}$

The format of the B^* Drag Term in a TLE is as follows:

Sign: (nil or -): The sign of the B-Star Drag Term. No sign indicates positive

First five digits: A number with a decimal point assumed to precede the first digit.

Sign (nil or -): The sign of the exponent. No sign indicates positive

Last digit: The exponent (base 10)

The ISS B-Star Drag Term "16538-3" translates to a B-Star Drag Term of $+0.16538 \cdot 10^{-4} \text{ Earth}_{\text{radii}}^{-1}$

- Field 12 – Column 63 – Ephemeris type

Example: "0"

Air Force Space Command uses this value to indicate the orbit model used to generate a TLE. In practice, this value is set to 0 for all external viewing. No external orbit propagators read or use this value.

- Field 13 – Columns 65-68 – Element set number

Example: "999"

The Element Set Number is used to distinguish a specific satellite TLE from its predecessors and successors. Whenever a new TLE is generated for a particular satellite, the Element Set Number for that satellite is normally incremented.

"999" = the 999th TLE generated for the ISS from NORAD since its launch.

- Field 14 – Column 69 – Checksum

Example: "3"

The Line 1 Checksum is determined by adding all the previous numbers in Line 1 and taking the last digit in the final sum. All letters, periods and plus signs are taken as "0". Negative signs are taken as "1".

This is mainly used to verify the first line's authenticity and/or its integrity upon receipt.

Line 2

- Field 1 - Column 01 - Line Number

It is important the #2 is present at the beginning of the line. In the case of its absence, the TLE is considered invalid.

- Field 2 – Columns 03-07 - Satellite Catalogue number

Example: "25544"

This is the Object Catalogue Number assigned by NORAD to the specific object. Each identified debris or satellite has a distinct number. Objects with an identifier smaller than 10,000 are aligned to the right and padded with zeros or spaces to the left. This field should be the same as in the first line, otherwise it is considered invalid.

- Field 3 – Columns 09-16 - Inclination

Example: "_51.6335"

The inclination indicates the angle formed between the object's orbital plane with the equatorial plane. It can range anywhere from 0 to 180 degrees. An orbit with an inclination between 0 and 90 degrees is called *prograde* orbit. An orbit with an inclination between 90 and 180 degrees is called *retrograde* orbit.

- Field 4 – Columns 18-25 – Right Ascension of the Ascending Node

Example: "344.7760"

The RAAN indicates the geocentric Right Ascension of an orbital object as it intersects the Earth's equatorial plane traveling northward. Its value can range anywhere from 0 to 360 degrees.

- Field 5 – Columns 27-33 – Eccentricity

Example: "0007976"

The eccentricity is a unitless value. Within a TLE set it is found without a decimal point that must be assumed. For the current example, the true eccentricity should be taken as 0.0007976.

- Field 6 – Columns 35-42 – Argument of Perigee

Example: "126.2523"

The Argument of Perigee is defined as the angle within the satellite orbit plane measured from the Ascending Node to perigee point along the direction of travel. The value can range anywhere from 0 to 360 degree.

- Field 7 – Columns 44-51 – Mean Anomaly

Example: "325.9359"

The Mean Anomaly indicates where the satellite was located within its orbit at the specified epoch. It can range from 0 to 360 degrees.

- Field 8 – Columns 53-63 – Mean Motion

Example: "15.70407856"

The Mean Motion is defined as the number of revolutions per day around the Earth that the satellite completes in exactly 24 hours. Theoretically, the mean motion can range between 0 and 17 orbits per solar day.

- Field 9 – Columns 64-68 – Orbit Number

Example: “32890”

The Orbit Number indicates how many orbits has the satellite completed from the launch to the specified epoch.

- Field 10 – Column 9 - Checksum

Example: ”6”

The Line 1 Checksum is determined by adding all the previous numbers in Line 1 and taking the last digit in the final sum. All letters, periods and plus signs are taken as "0". Negative signs are taken as "1".

This is mainly used to verify the first line's authenticity and/or its integrity upon receipt.

3.2 Two Line Elements sources

The Two-Line Element sets for unclassified objects are available for public use via “SPACE-TRACK.ORG”. The website allows the user to search for specific objects based on their Catalogue Number and can generate two lines format as well as three-line formats for the last released set or for every set released in a period specified by the user. The searching interface of the source is represented in Figure 14.

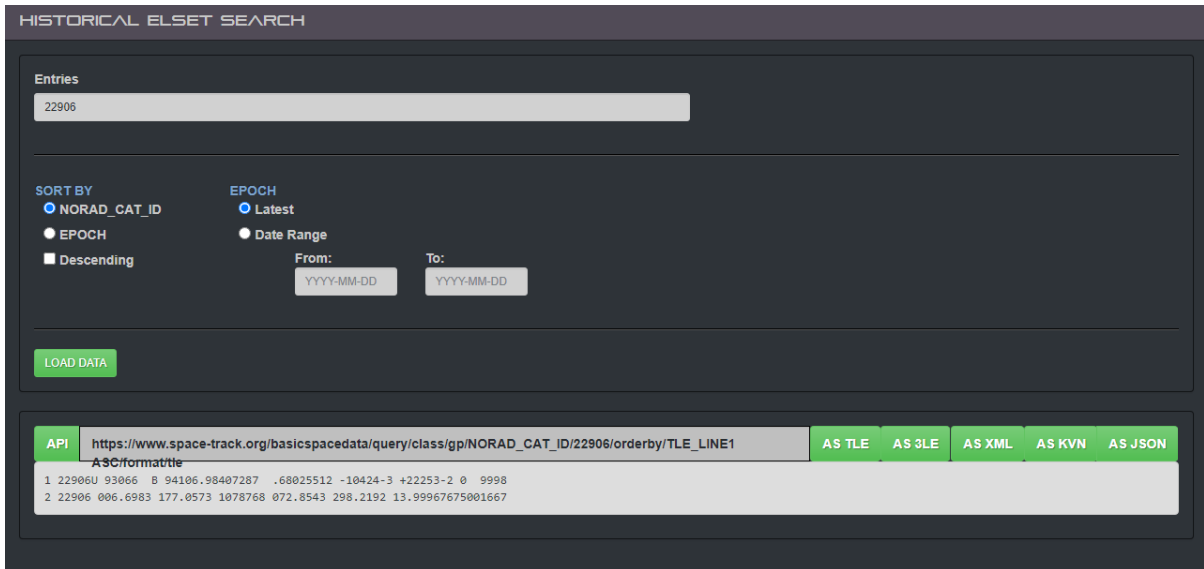


Figure 14 SPACE-TRACK.ORG user interface for TLE download

Another useful source of TLE sets is “CelesTrack” platform by Dr. T. S. Kelso. Many important objects can be found and downloaded from CelesTrack. It includes the satellite constellations of multiple operators, and what is more important for our work, the debris clouds produced by collisions are included. One should be interested in a specific set of objects that require constant update with the last released TLEs, CelesTrack offers the *SpaceTrack TLE Retriever* software. It allows to registered users to easily download and keep track on released TLE.

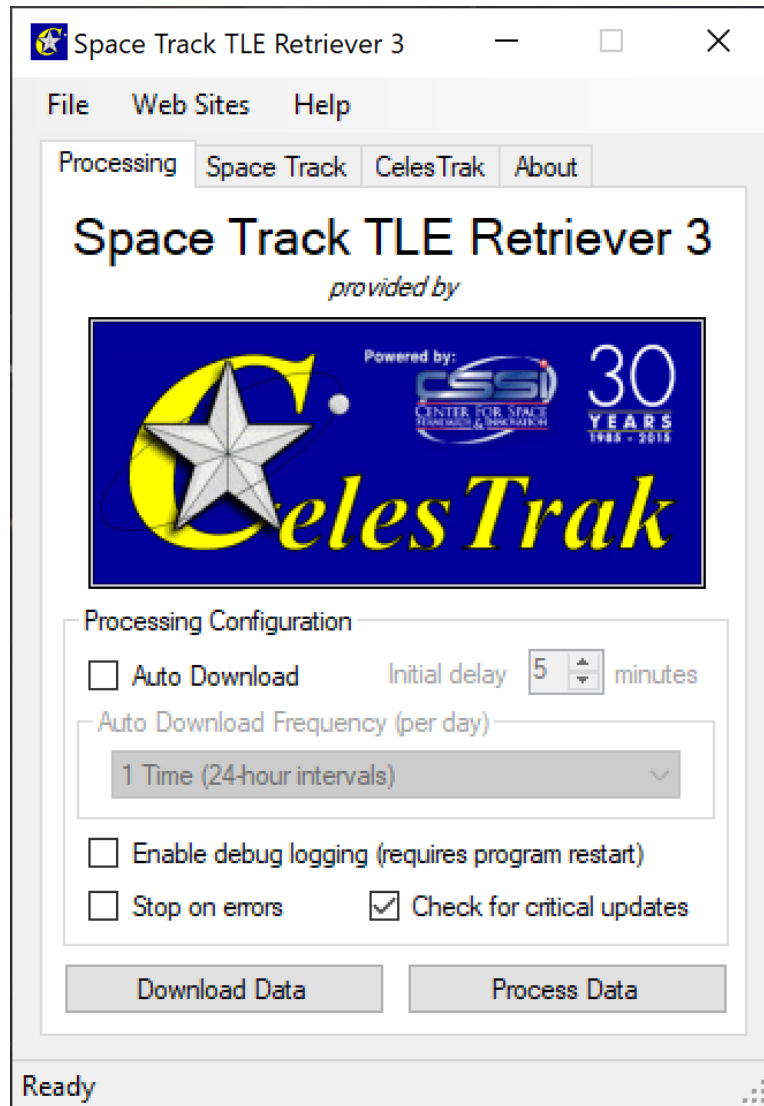


Figure 15 The Space Track TLE Retriever software interface

4 TWO LINE ELEMENTS FILTERING

Every satellite operator has the need to assess the conjunction and perform collision avoidance manoeuvres. Unfortunately, this process is not very adequate as result of a lack in information about other space objects. Therefore, the only available information about the debris and other satellites is contained in TLE. However, the predictions based on TLE sets are not sufficiently accurate. Moreover, there are some issues regarding TLE sets that could reduce the performance of prediction even more. It is not uncommon for a TLE set to contain entries of low quality, or even worse, entries of a different object. Consequently, if we want to increase the accuracy of the propagator, all the aberrant or incoherent TLEs from an object's set must be detected and removed.

4.1 Filtering Algorithm

The filtering of a TLE set is not a straightforward process. A staged approach will be used in this thesis because besides the incoherence in the TLE data, other aspects might influence the filtering performance. The entire workflow of the filtering process can be followed in Figure 16.



Figure 16 The workflow breakdown of the filtering process

First, upon reading every TLE set, the developed code will check the checksum indicator to assess the validity of each line. If any of the two lines of each TLE would result as corrupted, both lines would be deleted from the collection.

Afterwards, the sequence of operations represented in Figure 16, will detect all the irregularities in the data.

4.2 Two Line Element update-correction filter

One of the elements of a Two-Line Element set of a certain space object that can be derived is the Epoch. As every other parameter of the set is strictly related to the epoch, one can have a look on the time passed between the upload of consecutive TLEs. One can observe that the update interval is not constant. Depending on the nature of the orbital object, update frequency can significantly differ. Satellites in LEO could be updated several times a day because of the unpredictable results of the atmospheric drag [15]. Meanwhile, satellites in low drag orbits are more predictable, therefore an update in one or two weeks is considered sufficient.

Besides the update frequency differences between categories of space objects, a closer look inside a certain TLE set reveals a lack of evenness in the updates of that set. One can find a released TLE at a very short time interval from the previous one. The cause for such a short update frequency could be the need for a correction to be made. Given the limited information on how the TLE are produced and released, how one could be certain if the newly released TLE is a correction or it is just another TLE released for unknown to us reason? As a matter of facts, it is not possible to determine the true reason. Past work by Lemmens and Krag [16] considers a TLE as a correction if it was uploaded earlier than half an orbital period, while Kelecy et al. [17] filters the TLE collection by keeping only one TLE in a 24-hour window.

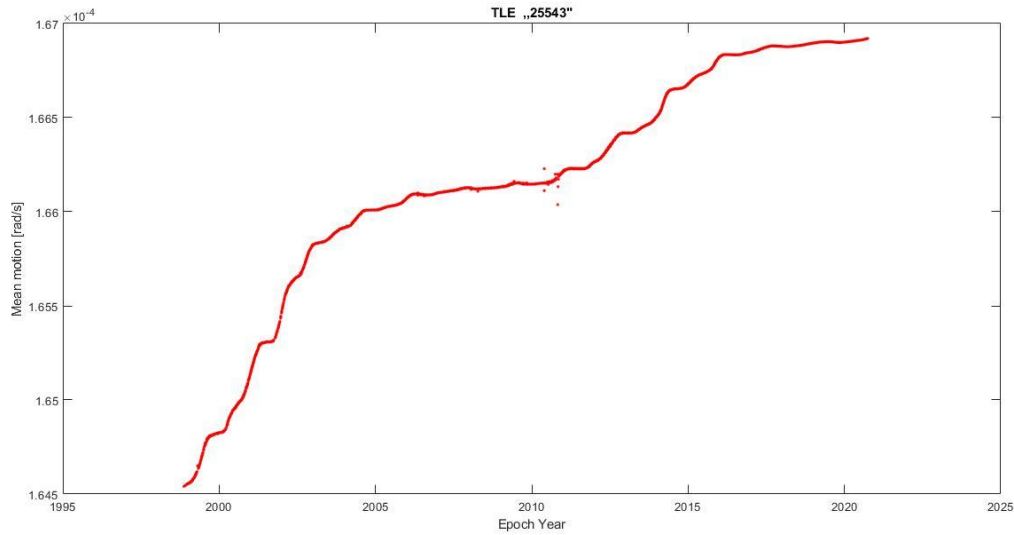


Figure 17 The mean motion representation for the TLE set of the object 25543

Having read and transformed all the parameters of the TLE collection in orbital elements, each epoch is analysed and the time difference between each consecutive entry is computed. To have a better understanding of how different the update frequency between consecutive TLE are, a random sample of epochs transformed in date-time format Table 1, was extracted from the “25543” collection. It can be easily observed that the time between TLEs can range from few hours up to weeks in the case of debris objects.

Year	Month	Day	Hour	Minute	Second
1999	2	23	17	11	56.916672
1999	2	25	0	53	42.018432
1999	2	25	22	1	18.288192
1999	2	26	8	35	14.263008
1999	2	28	23	58	37.741152
1999	3	6	6	42	57.274848
1999	3	7	3	50	34.621152
1999	3	8	0	58	6.249504
1999	3	11	13	27	42.3864

Table 1 A sample of TLEs date and time format derived from the epoch

In the case of a detected correction, the current TLE element will be kept while the preceding one will be discarded. The TLEs released later than half an orbital period are considered as early releases. Before the application of the correction filter, there were 6866 TLEs in the “25543” set. A total of 650 points, accounting for approximately 9.47% were considered as “corrected” by the algorithm and discarded. Had we used the method of Kelecy, keeping only one TLE in a 24-hour window, the amount of discarded points would be 4138 which accounts for 60.3%. Discarding more than half of the elements from the data set does not seem a good solution as it could greatly influence the orbit determination or re-entry prediction. For the sake of keeping the TLE set as “unharmful” as possible, in this thesis a correction will be considered if the upload happens earlier than half an orbital period.

4.3 Two Line elements - epoch gaps detection

The next step in the pre-processing of TLE set consists in the detection of any substantial gap in the update frequency. If large gaps are allowed, the state will not be propagated with a good certainty from the beginning of a gap up to the end of it. Therefore, the analysis in the further steps will be erroneous from the beginning. Depending on the orbit type, TLE can suffer of errors in numerical propagation of the state, or not be affected at all.

As the time interval between consecutive TLEs can vary from less than an orbital period up to tens of revolutions, thus defining a solid threshold to detect the large gaps is unreasonable. A robust procedure should be adopted to make sure that for different objects belonging to different types of orbit, the process will be autonomous and based on update distribution. By looking at the cumulative distribution of the update frequency in Figure 18, it is noticeable that a large difference in the update intervals of the current TLE exists based on the discontinuity starting around 11 days. However, one cannot be sure, the gap starting near 10 days might be also big enough to start the discontinuity

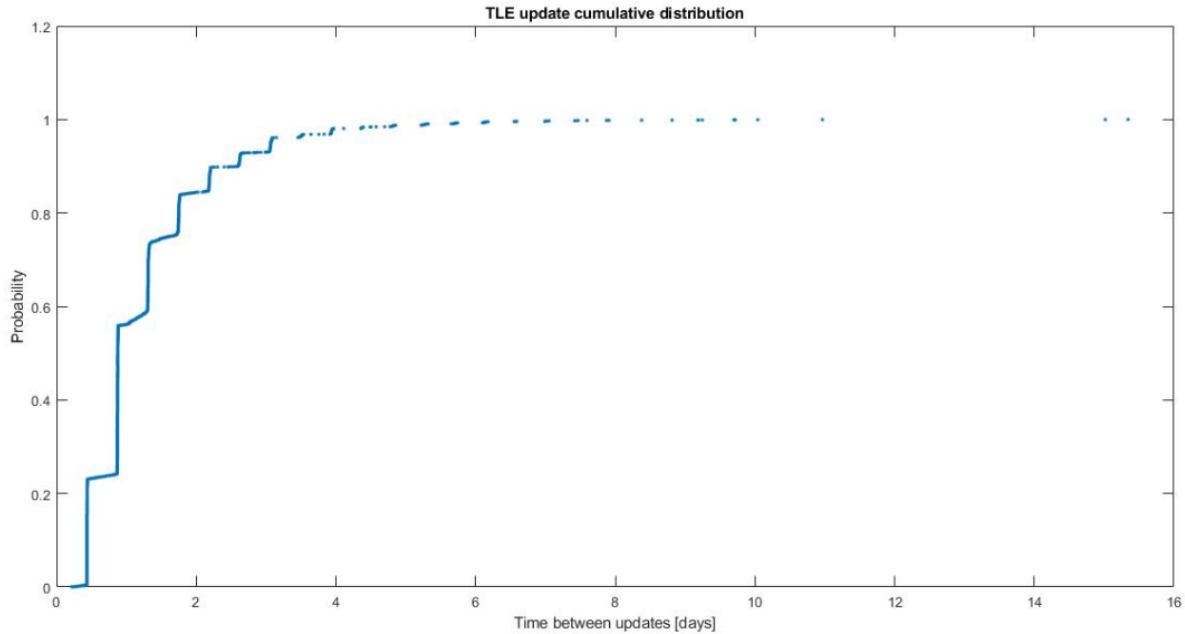


Figure 18 Cumulative distribution of the update frequency for the object “25543”

The implemented algorithm for the detection of the discontinuity in the update cumulative distribution is based on binning the data in windows. After computing the time interval between each consecutive TLE, the obtained set of data is analysed in order to detect the outlying time separations. The decision of considering a certain point as an outlier is usually backed up by an outlier detection method rather than a visual inspection by the user. Choosing the right method is an important step for any statistical analysis, as the outliers can affect the drawing of conclusions. Many methods are available nowadays, but which one fits the model the best? An answer to this question could serve the survey performed by Yang et al [18]. By doing a survey on a large set of research papers focused on outlier detections, they released the outlier detection methods preferred by researchers and the frequency they are using them, presented in Figure 19.

All these techniques are affected by outliers to a certain degree. The presence of large value outliers tends to “pull” the behaviour of all Standard Deviation, Median Absolute Deviation, Interquartile Range, thus, overestimating the threshold and failing to detect all the outliers, as it can be seen in Figure 20. Anyway, out of the automated techniques, the SD and MAD methods are allowing more control over the threshold definition. Consequently, only these two will be described and weighted for the selection of the final method to be used.

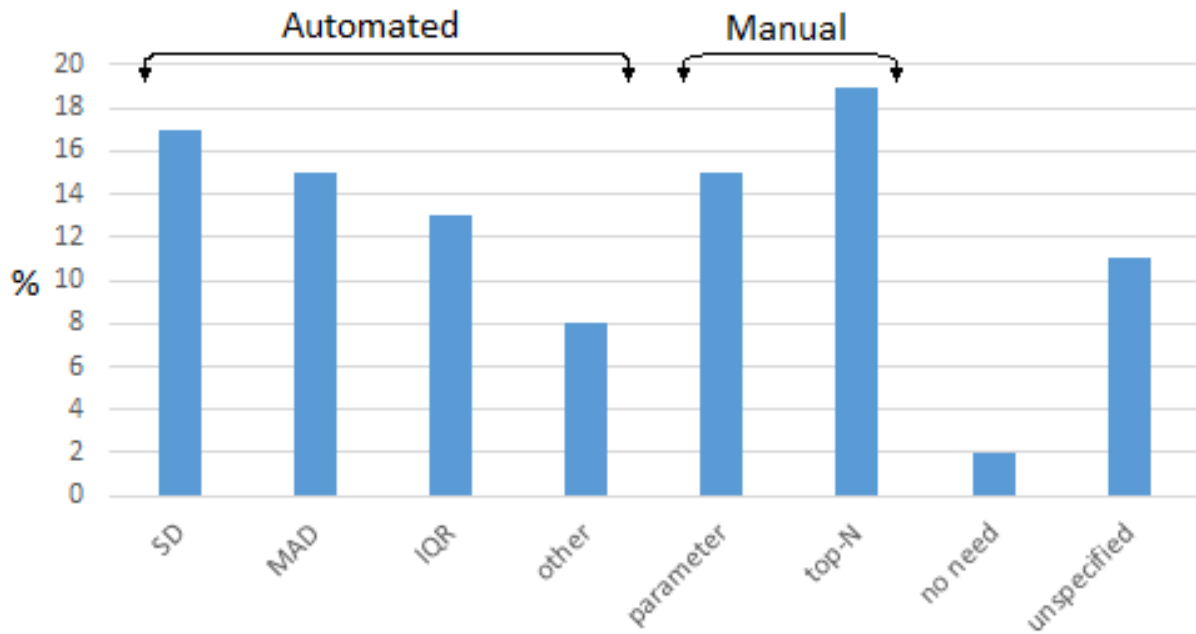


Figure 19 Selection frequency of outlier detection methods [18]

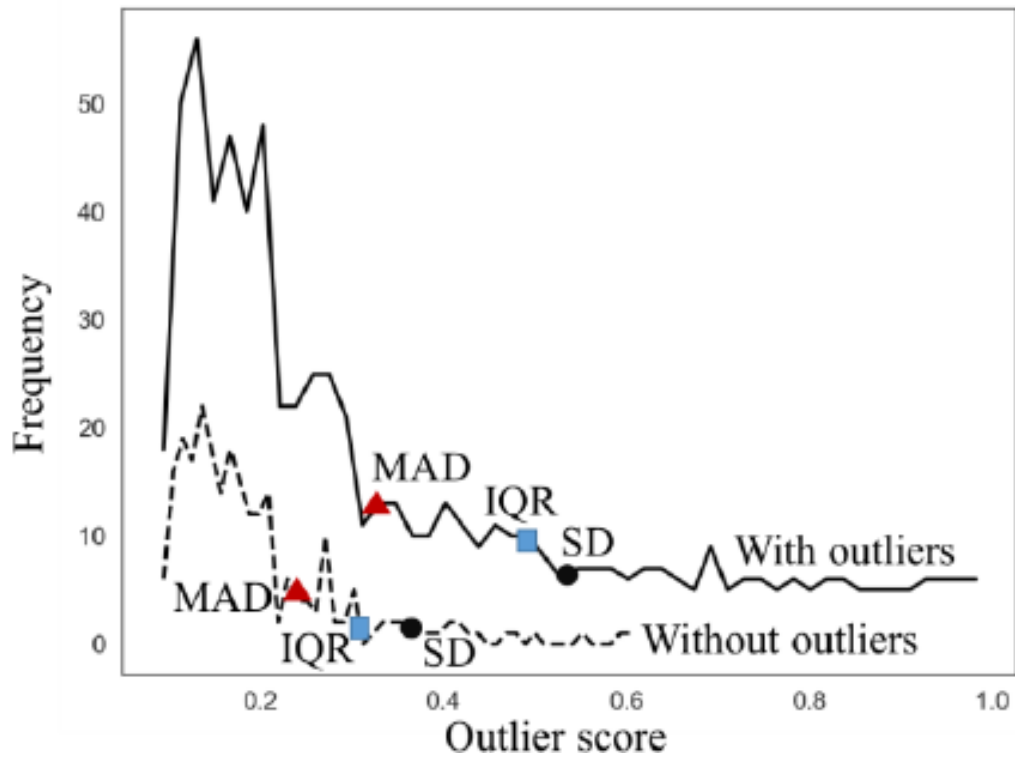


Figure 20 Effect of outliers on threshold based techniques [18]

4.3.1 Mean plus standard deviation method

The first method, mean plus standard deviation, relies on taking the width of the window as the mean of all the time separations between the TLEs plus a number of Standard Deviations of the same set.

$$width = Mean \pm k \cdot SD \quad (3)$$

The constant k defines the conservative level desired by the researcher. The common practice to take $k = 3$ is based on the characteristics of a normal distribution for which 99.87% of the data are within range Leys C. et al. [19]. Meanwhile, taking $k = 2$ or $k = 2.5$ is also acceptable if we are less demanding.

The issue related to this method is the sensitivity of the mean and of the standard deviation to outliers. Large time difference between consecutive outliers could highly alter the mean leading to missing out possible large gaps.

4.3.2 Median plus median absolute deviation method

The second method used for outlier detection; median plus median absolute deviation has the form:

$$Median - k \cdot MAD < x_i < Median + k \cdot MAD \quad (4)$$

For our problem, since the width cannot be defined by negative numbers, it will be limited to the upper bound.

$$width = Median + k \cdot MAD \quad (5)$$

It was started by Carl F. Gauss and then re-introduced by Hampel in 1974. Like Mean, the Median also is a measure of central tendency while being less sensitive to outliers. Another aspect that makes this method appealing is the independence of MAD to the sample size.

$$MAD = c \cdot Median_i(|x_i - Median_j(x_j)|) \quad (6)$$

Where $c = 1.4826$, is a constant that assumes normality of the data, disregarding the abnormality induced by the outliers.

The value of k must be taken according to the needs of the user. A $k=3$ is used for a very conservative approach, while a value of 2.5 and 2 leads to moderately conservative and poorly conservative, respectively.

Accordingly, after weighting the stated methods, it was decided to use the second approach. Now, in order to avoid the influence of very large gaps, only the values in the 95th percentile were used to compute the median and the median absolute deviation. Using equation (5) the width of the window was computed, and the data was grouped in consecutive bins. The discontinuity starts when the first empty bin is detected. In Figure 21, the result of applying the above described algorithm is represented. The first window that did not contained any data indicates the start of the tail disjoint. Therefore, every time separation between two consecutive TLEs that is higher than the one circled, represents large gaps in the TLE set.

Should large gaps be identified in a TLE, it is important to divide the series in sequences delimited by the identified gaps. For each set of data, the indexes of the starting and ending of every gap are computed and used in the further steps of the filtering process. Figure 22 shows a total of 6 large gaps detected. It is important to specify that the code will discard a TLE that is both preceded and succeeded by a large gap, as it cannot be propagated accurately.

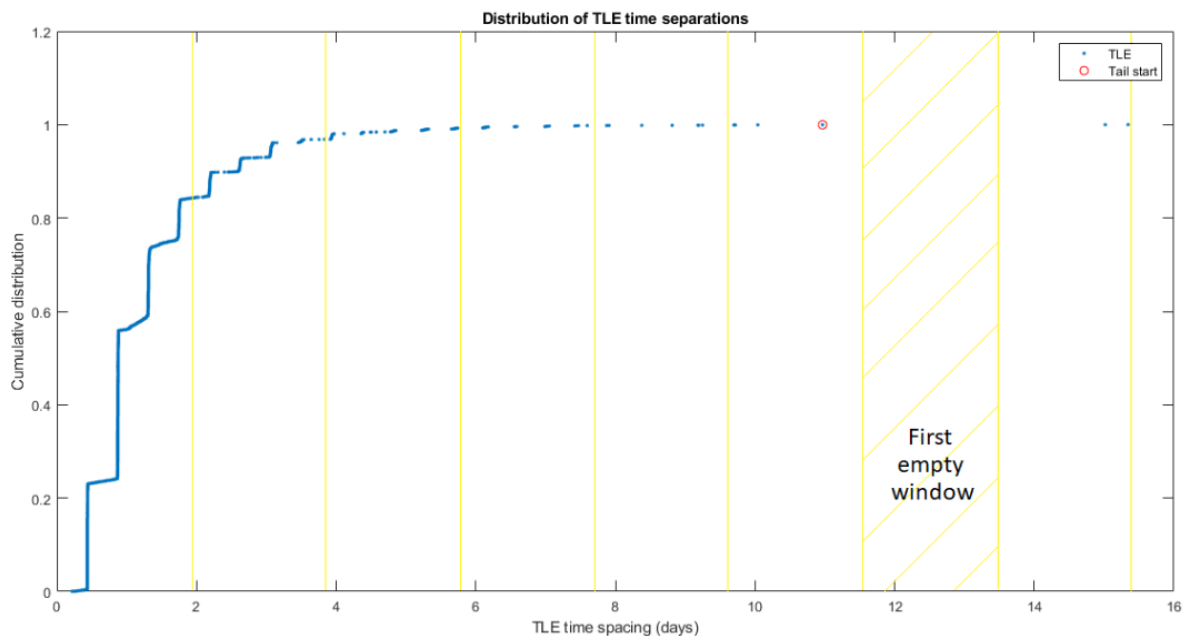


Figure 21 Visualisation of the data binning process and how the tail discontinuity is found.

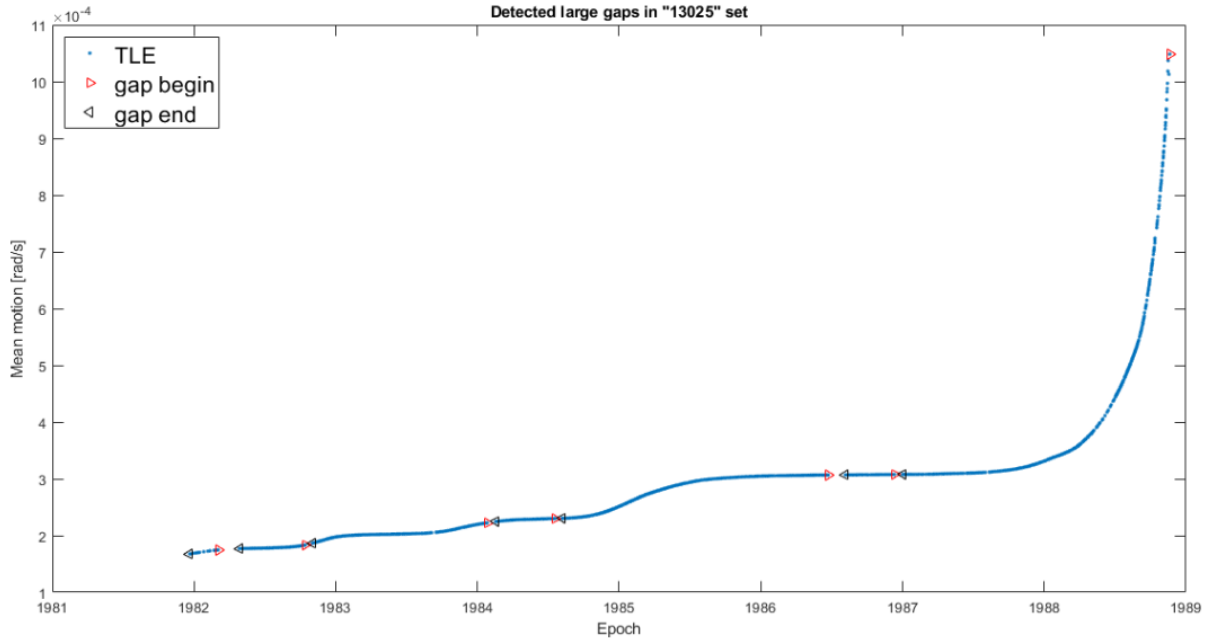


Figure 22 The result of running the gap detection algorithm on the set of the object “13025”

4.4 Orbital elements filters

Once a TLE series is analysed for large gaps in the epoch, if there are any detected, the series should be split in sequences from a gap to another. It is very important that the further processing is performed on each sequence separately, as the gaps in the epoch decreases the evenness of the series and it cannot be determined if any event happened. The simple outlier detection methods used in the previous section might not be the best choice for this step. Since the orbit is evolving, for objects suffering of drag, the rate of change in the orbital elements will differ from the rate of change in the epoch. Therefore, two different outlier detection methods for the orbital elements are built in this thesis. The orbital elements needed to be filtered are the mean motion, eccentricity, inclination and B^* . However, to ease the optimisation, avoiding the scaling issues, instead of using the eccentricity the perigee radius will be used.

4.4.1 Outlier detection - First method

The first method used for outlier detection in orbital elements is similar to the method described by Lidtke et al. [7], and is based on a sliding window approach to define the threshold for outlier detection.

A sliding window method is an efficient method to analyse nonlinear behaviour of statistic data. It is advantageous when it comes to linear fitting. As TLEs contain a large number of points, using a smaller sample of data points makes it more accurate in the determination of data's behaviour.

Algorithm Description

The reader should recall that the algorithm is implemented for each TLE sequence found between two large gaps. By sliding a window containing N_{TLE} from the beginning to the end of the each sequence, a polynomial of specific degree is regressed through these points of the window and is propagated to the next point following the window, as shown in Figure 23.

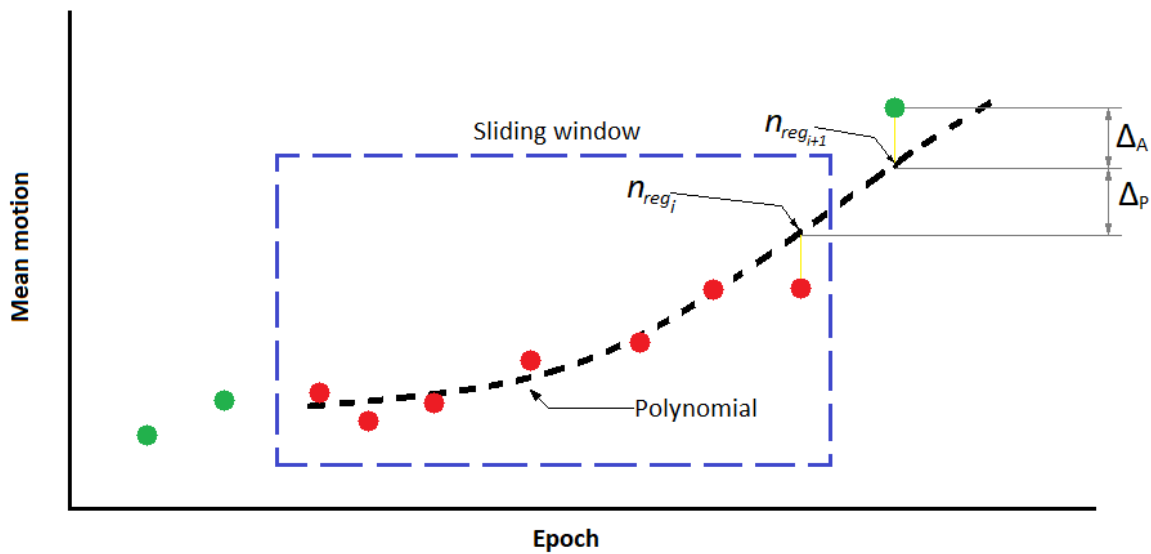


Figure 23 The sliding window used for the linear regression and residual computation.

For each iteration, the window slides by one step acquiring the succeeding point while discarding the first contained point, maintaining its length. Two quantities are computed: the residual Δ_A is defined as the difference between the mean motion in first point following the window and the regressed value and the residual Δ_P is defined as the difference between the regressed value of the point following the window and the regressed value of the last point inside the window.

$$\Delta_A = n_{i+1} - n_{Regressed_{i+1}} \quad (7)$$

$$\Delta_P = n_{Regressed_{i+1}} - n_{Regressed_i} \quad (8)$$

Based on previous determined quantities, the relative tolerance T_r is computed as:

$$T_R = \frac{\Delta_A}{\Delta_P} \quad (9)$$

If the variation of the orbital element in the sliding window will be very small, Δ_P will go to zero. Consequently, the relative tolerance will grow above unity, marking the current TLE as an outlier even in the case the point is consistent with the neighbouring ones. To make sure no TLE is falsely marked as an outlier, an absolute tolerance quantity T_A is defined as the ratio between the residual Δ_A and the value of the regressed orbital element in the last point of the sliding window.

$$T_A = \frac{\Delta_A}{n_{regressed_i}} \quad (10)$$

For a TLE to be considered as an outlier, now it is required for both tolerance quantities to fall above a defined threshold.

Regression methods

For the sliding window method to work, it is necessary to find a good polynomial fitting method that is compatible with the fitting needs.

The first regressor considered is the *Theil-Sen robust linear regression*. It is a linear regression method that fits data in a robust way, thus, it is not affected by the outliers or noise in the same way a least square method would be, [20].

The working principle of the regressor is simple. For a set of N pairs (x, y) of data, it computes the slope of each line pairing two points and then chooses the median as the estimate of the regression slope. Using that slope, a line is passed through all the pair (x, y) to obtain N intercepts. The median of the intercepts is taken as the estimate of the regression intercept.

$$a_1 = \text{median} \left\{ \frac{y_i - y_j}{x_i - x_j} \right\} \text{ for all } i \neq j \quad (11)$$

$$a_0 = \text{median}(y - a_1x) \quad (12)$$

The second regression method, based on polynomial fitting algorithm, is the so-called LOESS/LOWESS algorithm [21].

The major advantage that the method offers is the possibility to locally weight the regression function.

Taking the points contained in the sliding window, the first step would be to define the weighting function based on the distance from focal point to remaining ones.

$$w(x_k) = \left(1 - \left| \frac{x_i - x_k}{d_i} \right|^3 \right)^3 \quad \text{for } k = 1, \dots, N \quad (13)$$

Where, x_k is the focal point, and d_i is the distance between the focal point and the furthest ones. Therefore, the closest points weight more while the furthest point will have zero gain.

Once all the gains in the window are computed, a line is regressed between the points. LOWESS method will fit a 1st order polynomial while LOESS fits a 2nd order polynomial.

$$\text{LOWESS:} \quad \hat{y}_k = a + bx_k \quad (14)$$

$$\text{LOESS:} \quad \hat{y}_k = a + bx_x + cx_k^2 \quad (15)$$

After the weighted fitting based on the distance between the points is finished, a second set of robust weights is computed based on the distance of each regressed point, to the original set. The regressed points that are further from the real points will be given a lower gain than the closer one, thus, reducing the influence of the outlying data. The robust weights are computed as:

$$G(x_k) = \begin{cases} \left(1 - \left(\frac{|y_i - \hat{y}_i|}{6 \cdot \text{median}(|y_i - \hat{y}_i|)}\right)^2\right)^2 & \left|\frac{|y_i - \hat{y}_i|}{6 \cdot \text{median}(|y_i - \hat{y}_i|)}\right| < 1 \\ 0 & \left|\frac{|y_i - \hat{y}_i|}{6 \cdot \text{median}(|y_i - \hat{y}_i|)}\right| \geq 1 \end{cases} \quad (16)$$

The robust weighting becomes zero if a residual is greater than 6 medians. Therefore, the effect of outliers is cancelled.

The robust weightings, multiplied with the proximity weightings, are used for re-estimating a linear regression function within the individual ranges:

$$\text{LOWESS:} \quad \sum_k w(x_k)G(x_k)(y_k - a - bx_k)^2 \quad (17)$$

$$\text{LOESS} \quad \sum_k w(x_k)G(x_k)(y_k - a - bx_k - cx_k^2)^2 \quad (18)$$

After analysing both methods, considering that the sliding window takes a minimal set of points, a 1st order polynomial fitted through a small sample is better than a 2nd order. Deciding to go with a 1st degree regressor, Theil-Sen is very efficient in terms of RME for small samples and is faster than the LOWESS, thus it is selected as fitting method.

4.4.2 Outlier detection - Second method

Another adopted method to detect outlying data is based on defining a threshold using a dispersion quantity. As in the case of the previous method, a better fitting is obtained by using a sliding window of N_{TLE} for nonlinear data. The window slides from the beginning of a TLE sequence up to its last point. For each iteration, the median of the points inside the window is computed and is then subtracted from the orbital element in the focal point of that window. By doing this for every point in the TLE series a time series of differences is created. The focal point of the window should be in the middle; therefore, the size of the window is better always to be an odd value. Moreover, considering the number of points contained in the window, at the beginning and at the end of the series, the focal point cannot be placed in the middle of it. Therefore, to also cover the tails of the sequences, the code will freeze the window at the beginning while

the focal point will change its position. After the focal point reaches the central position, from that point on the window will slide, keeping the focal point in the middle. In Figure 24 is represented the way a sliding window containing 9 elements moves through the data set of a certain orbital element. One can observe that the window is in the same position until it reaches the 5th iteration while the focal point is moving. After reaching the 5th iteration, the focal point stays in the middle while the window is the one moving.

Having obtained the series of differences for each sequence of the TLE set, the same process of sliding window is repeated on the new set. This time, the computed quantity is the mean absolute deviation. Having obtained the median and the mean absolute deviation in every point of the sequence, a threshold is defined. The upper and lower bounds of the threshold are defined as (19) and (20).

$$T_{U_i} = Median_i + k \cdot MAD_i \quad (19)$$

$$T_{L_i} = Median_i - k \cdot MAD_i \quad (20)$$

If any point in the analysed orbital element set is bigger than the upper threshold or smaller than the lower threshold, it will be marked as an outlier. The value of k will be determined in the sub-paragraph related to the training of the filters.

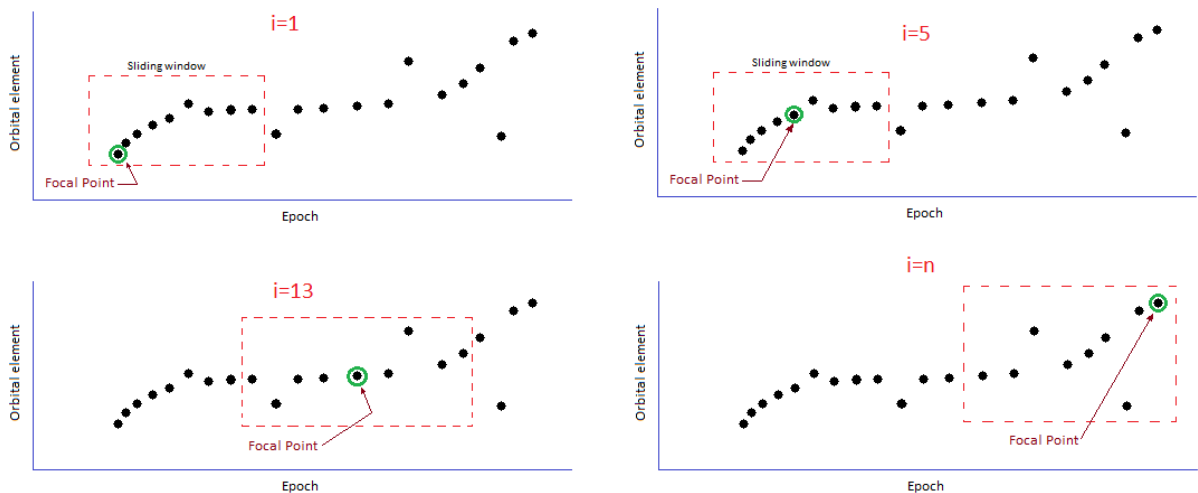


Figure 24 The representation of focal point shifting in the sliding window

4.4.3 Deleting negative B_{star}

The last possible action that could improve the accuracy of orbit determination using TLEs is to discard from the set all the TLE that have negative B_{star} . Since this term is used by NORAD to estimate the drag effect of the atmosphere on the satellite or debris, a negative value would mean that the orbital energy is increasing. The increase of orbital energy is possible only due to solar radiation pressure, but for LEO the effect of it is insignificant. Therefore, by deleting all negative terms from all the sets improves the accuracy in using Two Line Elements.

4.5 Optimization of the orbital element filters

The above described methods for outlier detections are based on defining a threshold quantity. A user defined set of thresholds are not recommended as the results might be unreliable, since a correct decision is dependent on operator's clarity of thinking but also human mind tends to choose a solution that approves its theories and expected results rather than reflecting the reality. Therefore, both methods require an optimization that would make them suitable for use in the filtering of TLE.

4.5.1 Optimization of the first method filter

As it was previously discussed, the method is using a sliding window to fit a line in the contained points and to compare the fitted point outside the window with the last one inside of it. The computed quantities relative and absolute tolerances will mark a TLE as an outlier if for an analysed orbital element both will exceed a defined threshold. How do we set the best threshold, considering that every TLE set is different than the other?

In order to make sure the filter performs well; it will be trained beforehand to determine the parameters for which the detected outliers are in fact data points whose behaviour worsens the data set.

The filtering algorithm will be trained for the following orbital elements: mean motion, perigee radius, and inclination. Each of the orbital parameters will have some distinction in their tuning settings. To have a detector that is efficient for all available

TLEs, the training was performed on a set of multiple objects of different types of orbits. A set of 15 TLE sets, of each type of orbit (*GTO*, *GEO*, *HEO*, *MEO*, *LEO*), was used as there are important distinction between the shape and parameters of each of them. In Table 2 the used TLE sets for each orbit type are shown.

Type of orbit	Debris catalogue number
<i>GEO</i>	<i>11436, 23438, 32288</i>
<i>GTO</i>	<i>13025, 25543, 33542</i>
<i>HEO</i>	<i>11909, 40297, 44553</i>
<i>LEO</i>	<i>33758, 36012, 38216</i>
<i>MEO</i>	<i>15702, 25502, 43248</i>

Table 2 The set of orbital debris objects for different orbits used for the training of the filters

Mean motion filter

In the current algorithm, each of the 15 inserted TLEs undergoes a transformation. As the inserted TLEs serve as dummies for the training of the filtering code, firstly they are smoothed to reduce the influence of existing outliers and yet preserving the shape and the characteristics of each TLE. It is important to use a real TLE set because only this way one can have the certainty of the best detection capability. The smoothing is achieved using the same Theil-Sen linear regressor with a defined size of the sliding window. The resulting sets are outlier-free imitations of the 15 TLEs used. Since each used TLE for the training are different it is useful to introduce a similarity between them. Accordingly, to focus on finding the best settings for the filter, regardless on the update time passed between consecutive TLE, the smoothed set is being resampled on a 24-hour basis. The resampling is performed using linear interpolation.

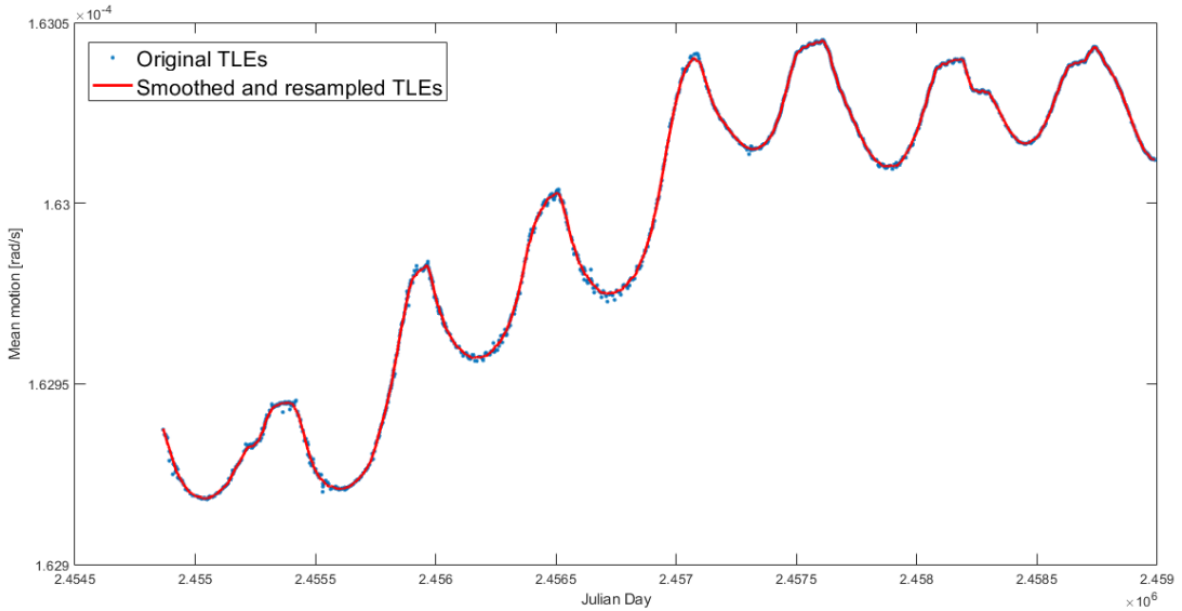


Figure 25 Smoothed and resampled mean motion series for the object “33542”

The smoothed and resampled set of mean motions now resembles a set of virtually outlier free data. To simulate a real data set containing outlying points, singular as well as sets of double and triple consecutive outliers were introduced. Another way of making it more difficult for the filter to detect outliers is to insert them at random locations.

The insertion of an outlier in a certain location of an orbital element is performed by scaling the value of the respective point. Therefore, an outlying point at a random location will be inserted according to the following rule.

$$n_{outlied} = n + n \cdot random(magnitude_n) \quad (21)$$

The range of magnitudes for the mean motion was selected to cover values several times smaller than the respective value as well as several times larger.

Thus,

$$magnitude_n = \pm[0.001 \quad 0.0025 \quad 0.005 \quad 0.01 \quad 0.05 \quad 0.1 \quad 0.2] \quad (22)$$

Depending on the number of elements a TLE contains, the number of outliers will be different. The inserted outliers of random magnitude and place at a random location will look like in Figure 26

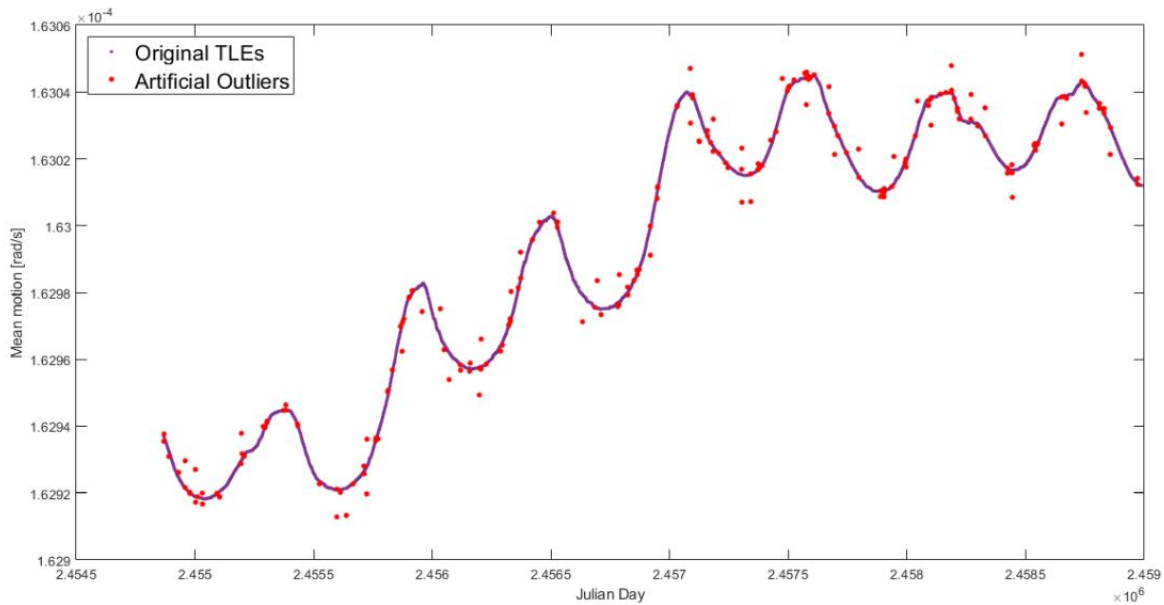


Figure 26 Inserted artificial outliers in the smoothed and resampled mean motion series of the object “33542”

Once all the 15 TLE series had been inserted with artificial outliers, all of them are prepared to be analysed by the MATLAB script, implementing the first method of outlier detection described in section 4.4.1. The code returns a vector of indexes of the detected outliers. By comparing the detected outliers with the artificially inserted ones, two quantities are defined, *false negatives* and *false positives*. The false negatives are outliers that have been artificially inserted in the preparing phase but were not detected by the filter. On the contrary, the false positives are outliers detected by the algorithm that were not artificially inserted Figure 27.

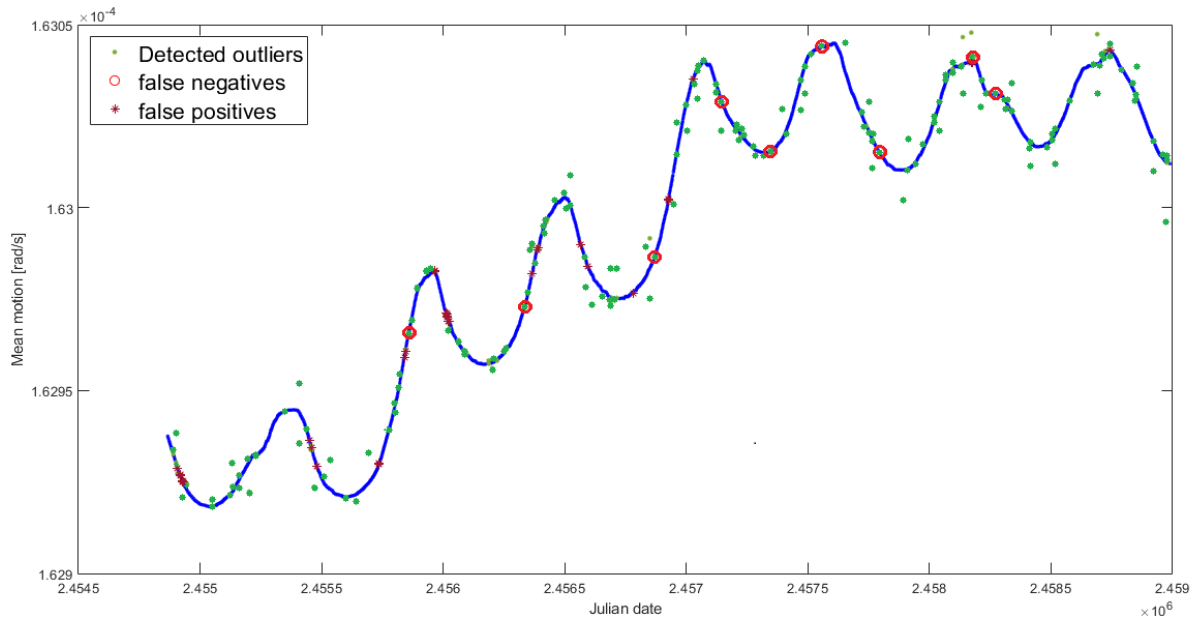


Figure 27 The result of running the mean motion filter on the set enriched with outliers

The purpose of the training phase is to make sure that the resulted false positives and false negatives are minimal. Finding a set of data to reduce both the false negatives and positives to zero is unlikely, but the focus of this thesis is to minimize their number. By using larger thresholds for relative and absolute tolerance the amount of resulting false positives is decreased, but the number of false negatives is increased. On the contrary, a smaller value for relative and absolute tolerance will reduce the number of false negatives but will increase the number of false positives. Focusing on minimizing just one of the quantities is not the best practice since minimising the false negatives without constrains could lead to a high increase in the number of false positives. Discarding so many points means decreasing the accuracy of the orbital determination process.

The number of the inserted outliers is much lower than the total points in one set. Therefore, for each iteration of the optimization algorithm, the resulted false negatives were normalized by the number of inserted outliers and the resulted false positives were normalized by the total number points in the TLE series. For a balance between false positives and false negatives, the solution closer to the origin is selected, Figure 28, where each point represents the solution of an iteration.

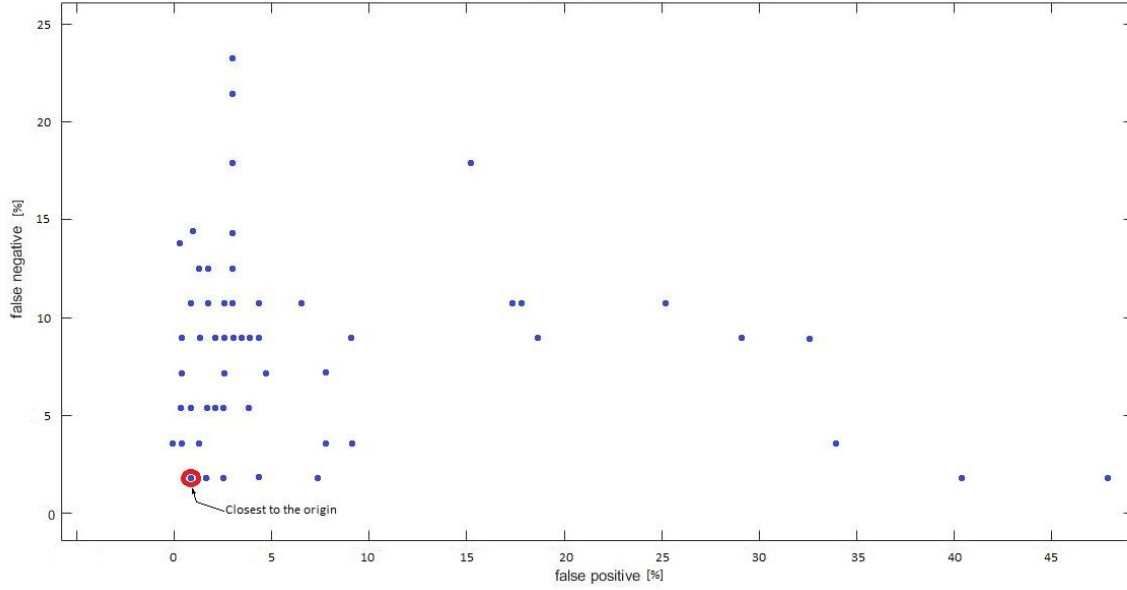


Figure 28 The result in terms of false positive/false negatives for multiple iteration process

In order to find the suitable parameters leading to the optima; the MATLAB Genetic Algorithm of the Global Optimization Toolbox, was used. The best solution of the above described model is found by using a fitness function.

$$f_{fitness_n} = \left\{ \begin{array}{c} T_{T_{rn}} \\ T_{T_{an}} \\ N_{smoothing_n} \\ N_{filtering_n} \end{array} \right\} \quad (23)$$

Where, $T_{T_{rn}}$ is the threshold for the relative tolerance, $T_{T_{an}}$ is the threshold for absolute tolerance, $N_{smoothing_n}$ is the size of the moving window used in the smoothing of the TLE prior inserting the artificial outliers and $N_{filtering_n}$ is the size of the moving window used in the filtering function. For each of the 15 sets, a percentage of outliers was introduced since the range of elements in the used sets are between several hundred up to 10 thousand. The ratio $N_{outliers}/N_{real\ points}$ should be the same for each TLE set. The amount of artificially inserted outliers in the mean motion is shown in Table 3.

Type of outlier	Amount [%]	Amount
Singular	2	1294
Double	0.5	323
Triple	0.05	33
Total outliers		2039

Table 3 The quantity of inserted single, double, and triple consecutive outliers in mean motion series

As the fitness function used in the genetic algorithm contains four variables, to converge to the best solution, a set of lower and upper boundaries Table 4 for the searching of the solution was introduced in the code.

	Lower boundary	Upper boundary
T_{Tr_n}	10^{-2}	4
T_{Ta_n}	$5 \cdot 10^{-10}$	$5 \cdot 10^{-2}$
$N_{smoothing_n}$	3	40
$N_{filtering_n}$	3	20

Table 4 Initial settings of the first method to run the training algorithm for the mean motion filter

The reader should be aware that the lower boundary on both smoothing and filtering steps for the number of points inside the sliding window should be ≥ 3 since the code is built to work with a window containing at least 3 points.

Perigee radius filter

As in the case of the mean motion filter. Proper settings and training are required for the perigee filter before it can be used. The 15 objects TLE set is smoothed using the same Theil-Sen linear regression for the same reason stated in the paragraph describing mean motion filter optimization. After performing the resampling on a 24-hour basis. A set of artificial outliers are introduced in each TLE set following the same approach. The perigee radius is computed using relation (24) and

$$r_p = a(1 - e) \quad (24)$$

$$r_{p_outlied} = r_p + r_p \cdot \text{random}(\text{magnitude}_{r_p}) \quad (25)$$

The magnitude range used for simulating outliers in perigee radius is not the same since the order of magnitude of perigee radiuses are different than the one of the mean-motions. Therefore, a new set that would improve the detection behaviour was used.

$$\text{magnitude}_{r_p} = \pm[0.001 \quad 0.005 \quad 0.01 \quad 0.05 \quad 0.1 \quad 0.2] \quad (26)$$

The fitness function for the perigee radius filter optimization is:

$$f_{\text{fitness}_{r_p}} = \left\{ \begin{array}{l} T_{T_{r_p}} \\ T_{T_{a_r_p}} \\ N_{\text{smoothing}_{r_p}} \\ N_{\text{filtering}_{r_p}} \end{array} \right\} \quad (27)$$

where, $T_{T_{r_p}}$ is the threshold for the relative tolerance, $T_{T_{a_r_p}}$ is the threshold for absolute tolerance, $N_{\text{smoothing}_{r_p}}$ is the size of the moving window used in the smoothing of the perigee radius set prior inserting the artificial outliers and $N_{\text{filtering}_{r_p}}$ is the size of the moving window used in the filtering function of the perigee radius set.

The amount of artificially inserted outliers in the eccentricity is the same used as for mean motion, Table 3.

The boundaries for the genetic algorithm to start the optimization are:

	Lower boundary	Upper boundary
$T_{T_{r_p}}$	10^{-2}	5
$T_{T_{a_r_p}}$	$5 \cdot 10^{-9}$	0.5
$N_{\text{smoothing}_{r_p}}$	3	40
$N_{\text{filtering}_{r_p}}$	3	30

Table 5 Initial settings of the first method to run the training algorithm for the eccentricity

Inclination filter

The last orbital element filtered, the inclination, as well as the previous orbital parameters, follows the same procedure.

$$i_{outlied} = i + i \cdot \text{random}(\text{magnitude}_i) \quad (28)$$

$$\text{magnitude}_i = \pm[0.001 \quad 0.005 \quad 0.01 \quad 0.05 \quad 0.1 \quad 0.2] \quad (29)$$

The fitness function for the eccentricity filter optimization is:

$$f_{fitness_i} = \left\{ \begin{array}{c} T_{T_{r_i}} \\ T_{T_{a_i}} \\ N_{smoothing_i} \\ N_{filtering_i} \end{array} \right\} \quad (30)$$

Where, $T_{T_{r_i}}$ is the threshold for the relative tolerance, $T_{T_{a_i}}$ is the threshold for absolute tolerance, $N_{smoothing_i}$ is the size of the moving window used in the smoothing of the TLE prior inserting the artificial outliers and $N_{filtering_i}$ is the size of the moving window used in the filtering algorithm of the inclination series. The quantity of the inserted outliers is the same.

And the set of boundaries used for the genetic algorithm is:

	Lower boundary	Upper boundary
$T_{T_{r_i}}$	10^{-1}	3
$T_{T_{a_i}}$	$5 \cdot 10^{-9}$	$5 \cdot 10^{-1}$
$N_{smoothing_i}$	3	40
$N_{filtering_i}$	3	40

Table 6 Initial settings of the first method to run the training algorithm for the inclination

4.5.2 Optimization of the second method filter

The training algorithm for the second method algorithm for outlier filtration is very similar to the training of the 1st method. The same 15 TLE sets are used. Each orbital element to be filtered is being smoothed using the Theil-Sen robust estimator to obtain virtual outlier free TLE sets. The obtained sets are enriched with the artificial outliers in random location all over the epoch characterizing the orbital element. The quantities of the inserted artificial outliers and the ranges of the magnitudes used are the same used for the elements in the first method Table 3. The only difference consists in the structure of the genetic algorithm model and the set of initial boundaries that are used to find for the best solution for the problem.

Mean motion

For the mean motion, the new fitness function is defined as

$$f_{fitness_n} = \left\{ \begin{array}{c} k_n \\ N_{smoothing_n} \\ N_{filtering_n} \end{array} \right\} \quad (31)$$

Where, the quantity k_n represents how many mean absolute deviations above the median should be taken to correctly detect the outliers, $N_{smoothing_n}$ is the number of TLEs inside the moving window used in the smoothing of the orbital element series prior inserting the artificial outliers and $N_{filtering_n}$ is the number of TLEs inside the moving window used for the outlier detection process.

The initial settings for the optimization process are:

	Lower boundary	Upper boundary
k_n	1	20
$N_{smoothing_n}$	3	50
$N_{filtering_n}$	3	40

Table 7 Initial settings of the second method used for the training algorithm of the mean motion

Perigee radius filter

The optimization of the second method for the outlier detection in the perigee radius set used the fitness function similar to the one for the mean motion:

$$f_{fitness_{r_p}} = \begin{pmatrix} k_{r_p} \\ N_{smoothing_{r_p}} \\ N_{filtering_{r_p}} \end{pmatrix} \quad (32)$$

For this case, the initial settings are presented below:

	Lower boundary	Upper boundary
k_{r_p}	1	20
$N_{smoothing_{r_p}}$	2	40
$N_{filtering_{r_p}}$	3	30

Table 8 Initial settings of the second method used for the training algorithm of the eccentricity

Inclination filter

Finally, for the last orbital element filter, the fitness function is:

$$f_{fitness_i} = \begin{pmatrix} k_i \\ N_{smoothing_i} \\ N_{filtering_i} \end{pmatrix} \quad (33)$$

And the initial boundaries:

	Lower boundary	Upper boundary
k_i	1	20
$N_{smoothing_i}$	3	40
$N_{filtering_i}$	3	30

Table 9 Initial settings of the second method to run the training algorithm for the inclination

5 RESULTS

As the focus of this thesis is to improve the quality of TLE sets by filtering out the irregularities and the outliers, the analysis was performed on a collection of objects belonging to debris in different orbits. After discarding the corrections and splitting the sets in sequences separated by large time, two methods for outlier detection were developed. Therefore, we need to analyse the outcome of both methods and compare the performances achieved by each of them.

5.1 The performance of the first method

Using the algorithm described in section 4.4.1, for each filtered orbital elements a set of parameters are obtained that should be used as predefined inputs for the filter before it is used in the filtering process.

Mean motion

In the case of the mean motion, after running the genetic algorithm on the 15 TLE set, for the inserted number of outliers found in Table 3, it was found that only 13 were missed by the filter, accounting for approximately 0.64% of the total outliers. Additionally, after proceeding with the training phase, the resulting parameters of the filter that ensure the optimal filtering behavior are:

$$T_{T_{r_n}} = 0.0845$$

$$T_{T_{a_n}} = 5.65 \cdot 10^{-4}$$

$$N_{filtering_n} = 7$$

$$N_{smoothing_n} = 21$$

Perigee Radius

As it was stated in section 4.5.1, instead of using the eccentricity, the perigee radius set was used. From all the inserted outliers, after running the filter, 22 outliers were missed, which would account for approximately 1.08% of the total outliers. The optimized settings for the filtering are presented below:

$$T_{T_{rrp}} = 3.7$$

$$T_{T_{arp}} = 6.82 \cdot 10^{-4}$$

$$N_{filtering_{rp}} = 8$$

$$N_{smoothing_{rp}} = 23$$

Inclination

For the last orbital element filtered, running the filter resulted in 17 missed outliers, accounting for approximately 0.83 % of the total inserted outliers. The optimized parameters for the filtering of the inclination set:

$$T_{T_i} = 1.68$$

$$T_{T_{a_i}} = 5.49 \cdot 10^{-4}$$

$$N_{filtering_i} = 7$$

$$N_{smoothing_i} = 32$$

5.2 The performance of the second method

For the second method, the obtained results are presented in the same format.

Mean motion

By running the second detection algorithm for the same number of artificially inserted outliers, 74 outliers were missed, accounting for 3.03 %. The optimized parameters of the filter are:

$$k_n = 5$$

$$N_{filtering_n} = 6$$

$$N_{smoothing_n} = 16$$

Perigee radius

In the case of perigee radius set, by using the perigee radius with the same number of outliers introduced as in the first method, 39 outliers were missed, accounting for 1.9% of the total outliers. The optimized parameters for the filter are:

$$k_{r_p} = 3$$

$$N_{filtering_{r_p}} = 7$$

$$N_{smoothing_{r_p}} = 37$$

Inclination

For the last filtered element, a total of 76 missed outliers were counted, accounting for approximately 3.7% of the total outliers. Finally, the optimized parameters for the inclination filter are:

$$k_i = 2$$

$$N_{filtering_i} = 5$$

$$N_{smoothing_i} = 32$$

Having determined the optimized settings used for filter, it is important to assess the efficiency and compare both methods to determine which one is more performant. Accordingly, a set of 100 TLE sets belonging to different type of orbits was used. For each set, the quantity of false negatives/positives were determined using the same number of artificial outliers as in the training phase. The results of both methods can be found in Annex 1 and Annex 2. For both methods it can be observed that while for some objects the filters work perfectly, for other ones, there are few missed outliers as well as wrong identified outliers. The mean and the standard deviations of the resulted false negatives and positives displayed in Table 10 and Table 11 could be used as an indicator to assess and compare the efficiency of both methods.

	False negatives		False positives	
	Mean	SD	Mean	SD
Mean motion	1.05	2.88	0.44	0.94
Perigee radius	2.07	4.07	0.12	0.13
Inclination	3.04	6.12	0.44	0.42

Table 10 Means and Standard deviations for the obtained false negatives/positives of each filter using first method

	False negatives		False positives	
	Mean	SD	Mean	SD
Mean motion	1.84	3.85	1.46	2.25
Perigee radius	2.40	4.09	1.43	1.06
Inclination	6.81	10.55	3.2	2.19

Table 11 Means and Standard deviations for the obtained false negatives/positives of each filter using second method

By comparing the obtained standard deviations of the two methods it can be observed that the second method is less effective than the first one for every orbital element. Moreover, for the same method, the filter performances are reduced for some orbital elements with respect to the other. The main contributor to the different performance lies in the distribution of data for each orbital element. Since the determination of the eccentricity and inclination of the orbital debris is highly prone to measurement noise, it is expected that the filter would be less efficient in detecting their outliers.

6 CONCLUSION AND FURTHER DEVELOPMENT

The obtained results presented in the previous paragraph are the basis of drawing conclusion regarding the use of filters for outlier detection in Two-Line Elements. The first method, based on thresholds defined on the difference between the real data and the propagated value using a linear regression, works well. Hence, using the optimal parameters obtained in section 5.1, a large collection of TLE of interest can be pre-processed before using it for orbit determination.

The filters were optimized based on a set of 15 TLEs belonging to 5 different orbit types. Therefore, one can expect that instead of giving the best performance for one orbit type, a trade-off between efficiency and universality occurred. For future developments, instead of focusing on a universal filter, it is expected that an optimization for each type of orbit should be done. Moreover, the optimization should be performed on a set of training sets higher than 15 elements, as was performed in this thesis. Thus, obtaining different settings (threshold and window width), a specific filter adapted just to one type of orbit is expected to increase the efficiency of the filter.

BIBLIOGRAPHY

1. ESA. (2020). Retrieved from https://www.esa.int/Safety_Security/Space_Debris
2. ESA Space Debris Office. (2019). ESA's Annual Space Environment Report. Darmstadt: GEN-DB-LOG-00271-OPS-SD.
3. W.D. Song, R.L. Wang, J. Wang. A simple and valid analysis method for orbit anomaly detection, *Advances in Space Research*, Volume 49, Issue 2, 15 January 2012, Pages 386-391.
4. David A. Vallado, Paul J. Cefola. Two-Line Element Sets – Practice and Use, 63rd International Astronautical Congress, Naples, Italy. 2012.
5. Stijn Lemmens, Holger Krag, Two-Line Elements based maneuver detection methods for satellites in Low Earth Orbit, *Journal of Guidance Control and Dynamics*, Vol 37, No. 3, May-June 2014.
6. Russell P. Patera, Space event detection method, *Journal of Spacecraft and Rockets*, Vol. 45, No. 3, May-June 2008.
- 7 Aleksander A. Lidtke, David J. Gondelach, Roberto Armellin, Camilla Colombo, Hugh G. Lewis, Quirin Funke, Tim Flohrer, Processing Two Line Elements sets to facilitate re-entry prediction of spent rocket bodies from the geostationary transfer orbit, The 6th International Conference on Astrodynamics Tools and Techniques, March 2016.
- 8 https://en.wikipedia.org/wiki/2009_satellite_collision
- 9 https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-1/Copernicus_Sentinel-1A_satellite_hit_by_space_particle
- 10 <http://www.satnews.com/story.php?number=1567042258>
- 11 <https://www.orbitaldebris.jsc.nasa.gov/measurements/radar.html>
- 12 https://www.esa.int/Safety_Security/Space_Debris/Scanning_and_observing2
- 13 Felix R. Hoots, Ronald L. Roehrich, 1980, Spacetrack report no.3, Models for propagation of NORAD Elements sets, Package Compiled by TS Kelso, 1988
- 14 David A. Vallado, Paul Crawford, SGP4 Orbit Determination, AIAA/AAS Astrodynamics Specialist Conference and Exhibit, 2008

- 15 Dr. T. S. Kelso, Frequently Asked Questions: Two-Line Element Set Format, *Satellite Times*, Volume 4, Number 3
- 16 S. Lemmens and H. Krag, “Two-line-elements-based manoeuvre detection methods for satellites in low earth orbit,” *Journal of Guidance, Control, and Dynamics*, vol. 37, pp. 861–868, 2014
- 17 T. Kelecy, D. Hall, K. Hamada, and D. Stocker, “Satellite manoeuvre detection using two-line element data,” *Advanced Maui Optical and Space Surveillance Technologies Conference*, Wailea, Maui, HI, USA, 2007
- 18 Jiawei Yang, Susanto Rahardja, Pasi Fränti, Outlier detection: How to threshold outlier scores? *The International Conference*, 2019.
- 19 Christophe Leys, Christophe Ley, Olivier Klein, Philippe Bernard, Laurent Licata, Detecting outliers: Do not use standard deviation around the mean, use absolute <http://extremelearning.com.au/the-siegel-and-theil-sen-non-parametric-estimators-for-linear-regression/> , by *Dr Martin Roberts*
- 21 Cleveland, W.S. (1979). Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association*, Vol. 74, pp. 829-836

APPENDICES

Annex 1 – False negatives and false positives outliers detected for 100 tested elements using the first method

TLE object	mean motion		perigee radius		inclination	
	false negatives [%]	false positives [%]	false negatives [%]	false positives [%]	false negatives [%]	false positives [%]
23883	0.00	0.26	0.00	0.09	4.73	0.60
41548	2.27	0.31	4.55	0.15	25.00	0.23
41839	0.00	0.40	0.00	0.00	0.00	0.32
41840	0.00	0.08	0.00	0.16	11.63	0.48
43282	0.00	0.00	4.17	0.00	0.00	0.38
38699	9.80	0.26	5.88	0.00	0.00	0.39
43284	0.00	0.51	0.00	0.00	0.00	0.13
43285	4.35	0.14	26.09	0.27	4.35	0.14
43447	0.00	0.68	0.00	0.17	11.11	0.17
43449	0.00	0.74	10.00	0.00	5.00	0.00
43453	4.54	0.29	0.00	0.72	9.09	0.14
43454	0.00	0.56	8.70	0.14	0.00	0.14
25000	0.00	0.37	0.00	0.01	0.00	0.32
43456	0.00	0.71	4.35	0.42	4.35	0.28
43514	21.42	0.21	7.14	0.00	14.29	0.00
43515	11.11	0.17	11.11	0.00	11.11	0.34
43516	0.00	0.29	4.55	0.00	0.00	0.15
43517	11.76	0.00	17.65	0.00	17.65	0.20
44458	0.00	0.00	0.00	0.00	37.50	1.08
44459	0.00	0.00	0.00	0.00	0.00	0.00
10089	0.82	0.31	0.62	0.09	0.62	0.35
19857	1.59	0.39	1.27	0.13	2.23	1.03
22850	0.00	0.39	0.00	0.06	4.50	1.96
24675	1.11	0.30	0.37	0.08	0.37	0.43
26090	0.00	0.49	0.00	0.01	0.00	0.28
27039	0.00	0.32	0.00	0.08	0.47	0.96
27602	0.75	0.22	0.75	0.07	3.01	1.12
27832	0.00	0.28	0.00	0.00	0.51	0.48
28998	0.61	0.28	0.61	0.06	1.22	0.32
32405	0.00	0.23	0.00	0.05	4.23	0.46
35557	0.80	0.24	0.00	0.03	1.60	0.29
35816	0.00	1.14	2.25	0.21	2.25	0.14

36034	0.00	0.00	16.67	0.00	0.00	0.00
37266	0.00	0.24	0.00	0.05	2.48	0.32
37817	0.93	0.39	0.00	0.18	5.56	0.58
38750	1.00	0.39	0.99	0.10	24.75	1.84
41837	0.00	0.35	0.00	0.00	22.83	3.34
41946	0.00	0.07	0.00	0.14	2.17	0.35
44068	0.00	0.60	5.13	0.00	12.82	0.69
10485	0.00	9.57	9.09	0.00	0.00	0.87
11550	0.21	0.34	0.82	0.06	0.00	0.34
11888	1.31	0.31	0.00	0.06	0.66	0.42
12680	0.45	0.38	2.02	0.26	0.67	0.42
12834	0.00	0.32	0.46	0.12	0.46	0.40
13897	2.24	0.32	2.24	0.32	2.74	0.53
14191	0.24	0.38	1.88	0.11	1.41	0.38
15206	0.00	0.43	0.48	0.16	0.72	0.34
18949	0.26	0.37	0.26	0.11	0.26	0.41
19611	0.00	0.33	0.59	0.03	0.00	0.36
23645	0.28	0.41	0.83	0.09	0.00	0.38
26629	0.00	0.33	0.35	0.10	0.00	0.32
26964	0.00	0.33	0.54	0.12	1.62	0.33
27883	1.50	0.37	2.50	0.70	3.00	0.42
27884	2.17	0.35	3.80	0.09	1.09	0.35
27888	2.14	0.40	1.07	0.52	1.07	0.38
27897	0.53	0.33	2.13	0.07	2.66	0.37
35504	1.69	0.35	1.69	0.24	2.26	0.39
35505	0.00	0.50	5.19	0.29	1.30	0.33
42720	0.00	0.37	0.00	0.16	0.80	0.45
18725	0.00	0.18	2.63	0.09	0.00	0.62
18728	0.27	0.36	0.54	0.11	0.00	0.33
18729	0.00	0.30	0.00	0.09	0.53	0.30
18730	0.26	0.43	0.27	0.11	0.27	0.43
18737	0.00	0.29	0.27	0.04	0.80	0.29
18742	0.29	0.32	0.30	0.06	0.30	0.22
18746	0.00	0.32	0.27	0.10	0.27	0.30
18747	0.00	0.36	0.27	0.11	0.00	0.34
18753	0.27	0.43	0.00	0.11	0.27	0.29
18754	0.27	0.30	0.00	0.02	0.54	0.30
18755	0.00	0.40	0.53	0.10	0.00	0.31
18756	0.79	0.49	0.00	0.12	0.27	0.33
18756	0.00	0.34	0.00	0.04	0.00	0.38
18758	0.54	0.42	0.27	0.07	0.27	0.43
18801	0.27	0.31	0.53	0.04	0.00	0.37
18803	0.00	0.44	0.27	0.15	0.27	0.37

18804	0.54	0.31	0.27	0.07	1.34	0.37
19005	0.00	0.34	0.00	0.04	0.27	0.38
26111	0.27	0.38	0.00	0.10	0.00	0.32
26601	0.85	0.36	1.28	0.12	0.00	0.40
19753	0.00	0.36	0.44	0.11	0.44	0.33
28944	0.56	0.40	0.56	0.06	0.28	0.41
32025	1.22	0.16	0.00	0.50	0.00	0.28
32026	0.00	0.54	0.67	0.04	1.33	0.28
32033	0.00	0.26	0.00	0.02	0.00	0.19
32034	0.00	0.48	0.00	0.04	0.00	0.41
32035	0.67	0.34	0.00	0.04	0.00	0.41
32043	0.00	0.41	0.00	0.22	0.00	0.22
32047	0.00	0.22	0.00	0.04	0.67	0.28
32493	0.00	0.28	3.36	0.26	0.67	0.33
33473	0.00	0.36	1.47	0.17	3.68	0.38
34239	0.00	0.38	0.00	0.05	1.53	0.30
34241	0.00	0.33	6.93	0.20	3.96	0.46
34254	0.93	0.24	1.85	0.12	1.85	0.21
34257	2.82	0.23	5.63	0.23	1.41	0.36
34257	1.47	0.43	5.88	0.05	0.00	0.24
34258	2.35	0.50	0.00	0.04	2.35	0.27
35833	0.00	0.37	0.86	0.25	4.31	1.26
36958	0.00	0.67	2.50	0.00	2.50	0.34
38563	4.35	0.41	0.00	0.14	4.35	0.68
40060	0.00	0.24	0.00	0.14	1.47	0.58

Annex 2– False negative and false positive outliers detected for 100 tested elements using the second method

TLE object	mean motion		perigee radius		inclination	
	false negatives [%]	false positives [%]	false negatives [%]	false positives [%]	false negatives [%]	false positives [%]
23883	0.00	2.38	0.00	2.23	15.27	0.53
41548	2.27	2.86	0.00	1.31	29.55	1.70
41839	12.20	0.08	0.00	0.48	0.00	1.59
41840	0.00	13.13	2.33	2.39	37.21	0.16
43282	0.00	7.89	0.00	0.89	8.33	0.76
38699	7.84	0.78	0.00	2.60	9.80	2.93
43284	0.00	2.66	0.00	2.41	29.17	0.38
43285	0.00	4.48	8.70	1.09	13.04	1.49
43447	16.67	0.68	5.56	1.86	33.33	2.03

43449	0.00	3.54	0.00	1.03	35.00	1.77
43453	0.00	1.15	13.64	0.86	22.73	1.00
43454	0.00	2.79	0.00	1.26	26.09	1.54
25000	0.38	0.99	0.00	2.92	5.34	0.62
43456	0.00	1.27	8.70	1.41	13.04	1.13
43514	0.00	1.86	7.14	2.06	21.43	3.30
43515	5.56	5.16	5.56	1.03	5.56	2.24
43516	0.00	3.21	0.00	0.15	22.73	0.58
43517	5.88	5.12	5.88	1.57	41.18	2.56
44458	0.00	13.72	0.00	0.36	37.50	0.72
44459	14.29	0.00	28.57	0.38	0.00	1.14
10089	0.21	0.66	3.29	0.35	1.23	3.60
19857	0.96	0.62	0.64	1.66	6.37	1.56
22850	0.00	0.49	0.64	1.24	15.76	2.04
24675	1.86	0.80	0.74	1.07	0.74	1.86
26090	0.00	5.42	0.43	0.47	0.43	1.09
27039	0.47	1.06	0.00	1.16	6.16	1.99
27602	0.75	0.24	4.51	1.07	10.53	1.60
27832	0.00	2.23	0.00	1.02	11.79	1.02
28998	0.00	2.45	0.00	0.58	3.05	0.42
32405	0.00	1.69	0.00	0.85	7.75	1.44
35557	0.00	1.49	0.00	1.18	6.40	1.39
35816	11.24	0.14	0.00	1.00	3.37	1.46
36034	0.00	3.47	0.00	0.50	0.00	0.99
37266	8.26	0.00	0.00	1.54	6.61	1.32
37817	0.00	0.00	0.00	1.67	17.59	0.88
38750	0.99	1.25	0.00	0.46	35.64	1.25
41837	0.00	0.00	0.00	0.70	38.04	0.95
41946	2.17	0.00	0.00	0.84	2.17	2.31
44068	2.56	0.00	2.56	0.26	7.69	0.60
10485	18.18	0.29	0.00	2.61	9.09	3.77
11550	0.00	2.26	6.60	0.23	1.03	4.11
11888	0.00	1.12	2.62	0.62	2.62	3.75
12680	0.67	0.86	3.81	0.94	1.57	4.14
12834	0.23	0.99	3.23	0.87	1.15	3.04
13897	1.75	0.27	3.49	1.02	1.25	3.62
14191	0.71	2.57	7.29	0.34	1.41	4.43
15206	0.48	1.16	3.35	0.46	0.96	3.17
18949	1.32	1.55	5.53	0.88	4.47	2.02
19611	0.00	4.23	5.88	0.18	1.47	4.27
23645	0.28	2.40	4.16	0.86	0.55	3.66
26629	0.35	2.46	7.80	0.24	1.06	3.66
26964	0.00	3.29	1.08	0.11	3.78	0.56

27883	1.50	1.32	4.50	1.19	2.00	3.55
27884	1.09	1.11	4.35	0.37	1.09	2.84
27888	2.14	0.79	8.02	0.47	0.00	4.09
27897	0.00	0.94	6.91	0.40	1.06	2.82
35504	2.26	1.26	3.39	0.76	0.56	3.81
35505	0.00	0.21	11.69	0.50	0.00	1.54
42720	0.00	0.58	4.80	0.39	1.60	1.78
18725	0.00	0.00	0.00	0.44	2.63	3.19
18728	1.08	0.10	0.54	3.05	1.34	4.89
18729	0.27	0.13	0.00	2.84	1.06	5.85
18730	0.53	0.20	0.00	2.22	1.06	5.70
18737	0.80	1.23	0.00	3.43	0.00	8.58
18742	0.00	0.53	0.90	3.34	1.50	5.62
18746	1.33	0.10	0.00	1.72	1.06	9.48
18747	1.07	0.03	0.00	1.40	0.53	11.71
18753	0.53	1.09	0.00	3.62	1.59	5.40
18754	0.00	0.13	0.80	3.05	0.27	5.75
18755	0.27	0.11	0.27	2.38	1.06	5.63
18756	0.00	0.21	0.53	3.21	1.06	7.50
18756	0.53	0.01	0.00	2.04	1.33	5.58
18758	0.54	0.02	2.41	0.82	1.07	7.67
18801	0.53	0.45	1.06	3.35	2.39	5.05
18803	0.00	0.10	0.00	4.90	1.88	5.53
18804	0.54	0.28	0.00	2.90	1.88	5.01
19005	0.00	1.18	0.00	3.77	2.14	6.17
26111	0.00	0.11	0.00	3.34	0.27	5.12
26601	0.00	0.00	1.28	2.68	2.13	6.72
19753	0.00	1.14	0.44	4.21	0.88	6.47
28944	0.28	2.02	0.85	2.02	0.85	2.24
32025	1.83	3.34	0.00	2.91	0.00	5.55
32026	0.00	0.13	0.00	0.47	2.67	1.79
32033	4.67	0.04	6.67	0.19	1.33	3.32
32034	0.67	0.07	0.00	2.36	1.34	5.40
32035	2.00	0.90	0.00	0.95	0.00	4.01
32043	2.00	0.22	2.00	0.60	0.00	3.69
32047	0.00	1.95	0.00	1.02	1.34	4.78
32493	2.01	0.20	8.05	1.29	0.00	3.49
33473	0.00	0.00	0.00	1.87	0.00	4.36
34239	0.00	0.40	0.00	0.50	2.29	3.55
34241	1.98	0.30	10.89	1.05	1.98	5.93
34254	0.00	0.24	0.00	1.75	0.93	5.11
34257	7.04	0.45	1.41	0.99	2.82	4.01
34257	0.00	0.62	0.00	1.14	4.41	2.47

34258	1.18	0.50	0.00	1.37	0.00	4.43
35833	18.10	0.03	0.00	1.32	7.76	1.21
36958	7.50	0.34	0.00	1.68	0.00	4.97
38563	0.00	0.55	4.35	0.96	0.00	2.73
40060	1.47	0.91	0.00	0.29	1.47	2.21