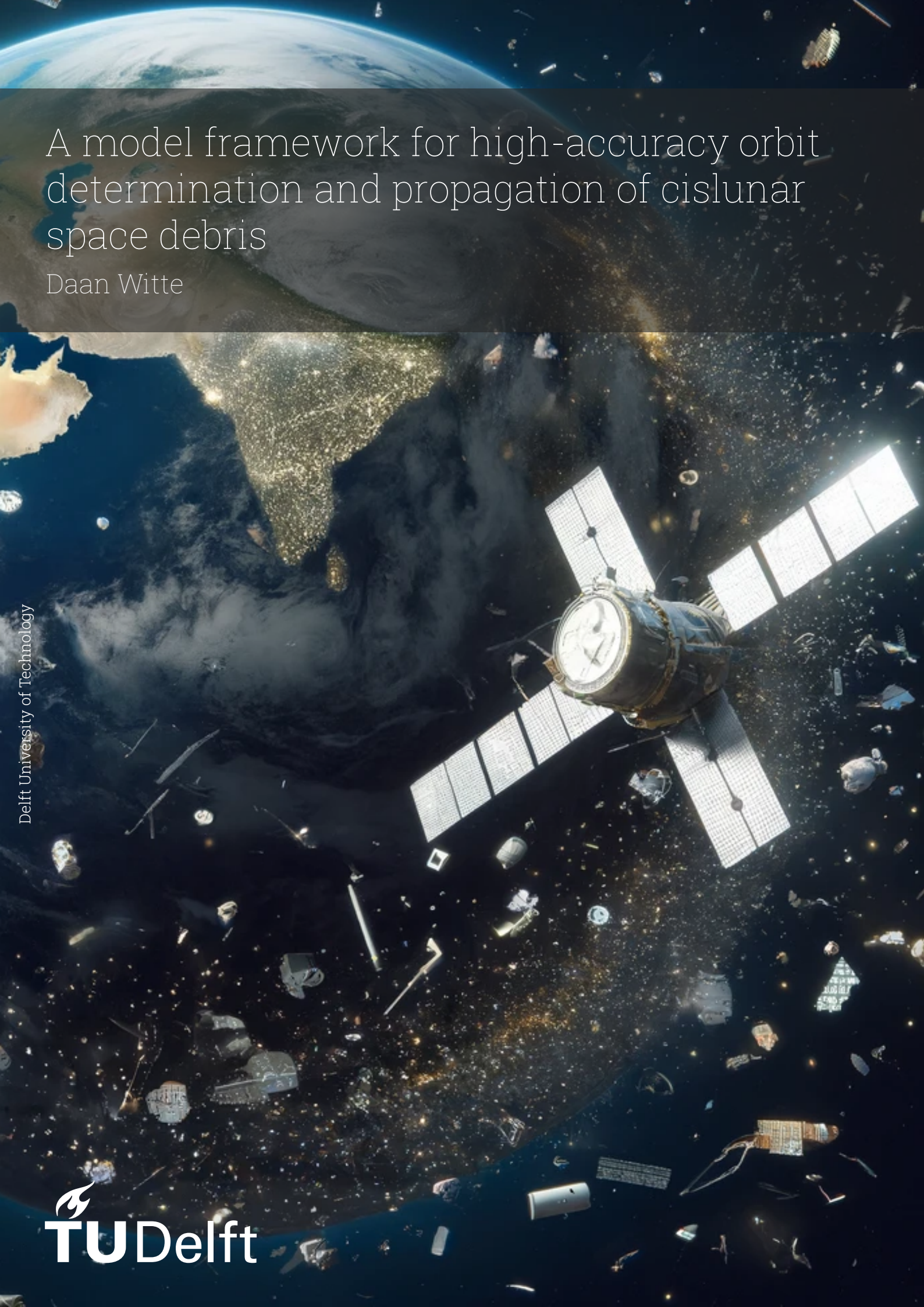
A model framework for high-accuracy orbit determination and propagation of cislunar space debris

Daan Witte



Delft University of Technology

TU Delft

A model framework for high-accuracy orbit determination and propagation of cislunar space debris

by

Daan Witte

to obtain the degree of Master of Science in
Aerospace Engineering at Delft University of Technology

Student number:	4670302	
Project duration:	November 2023 - June 2024	
Thesis committee:	Ir. K.J. Cowan MBA	Committee chair
	Dr. M. Langbroek	Thesis supervisor
	Dr. S. Gehly	External examinator
	Dr. ir. D. Dirkx	Additional examinator

Cover: Space debris visualization generated using Dall-E 3.

TU Delft

Preface

This M.Sc. thesis marks the end of my studies at the faculty of Aerospace Engineering. For the last year, I have worked on developing a model framework for cislunar space debris and it has been quite a ride. The countless model framework iterations, debugging hours and over 200 GitHub commits, have taught me a lot.

I would like to thank my supervisors, Marco and Dominic, for their continuous and proactive support in this process. Your guidance helped me overcome challenging hurdles, gave me important insights and overall made the thesis very enjoyable. Furthermore, I would like to thank Bill for the many hours he spent compiling high-quality datasets, without which this thesis would not have been possible. Finally, a huge thanks to my family and friends for always having my back throughout all my endeavours.

I am glad to share that this paper has been selected for an oral presentation at the International Astronautical Congress 2024 in Milan. Hopefully, we will be able to inspire others to explore the field of cislunar space debris.

Daan Witte
Rotterdam, June 2024

Summary

In recent years, interest in cislunar space has grown exponentially. Countries like the US, China and Israel are all competing in this second Lunar space race. At the same time, our space debris problem is escalating and the risk of the Kessler syndrome occurring is constantly growing. Space Situational Awareness (SSA) aims to locate all objects in various regions, preventing this catastrophe from ever happening. Unfortunately, the massive amount of space debris and our physical limitations make continuous tracking infeasible. This has created a need for high-accuracy orbit determination and propagation model frameworks. Research on this topic has been done extensively for near-Earth space debris, but there is little experience with cislunar space. The chaotic nature of cislunar space makes modelling space debris orbits especially challenging.

This research has focused on the development of a model framework that is able to perform orbit determination and propagation for cislunar space debris with high accuracy, whilst also quantifying the effect of uncertainty realistically over time. We apply our framework to the Chang'e 2 and 3 upper stages. Both objects have ~10 years of optical observations stored in the Minor Planet Center (MPC) format. 13 different estimation windows are chosen amongst the two objects. This allows for analysis on a diverse range of orbit and observation characteristics, like stability and estimation window length. Model performance is analyzed using Root-Mean-Square Error (RMSE), Period of Sufficient Accuracy (PSA) and computation time. PSA is a metric, developed for this framework, that indicates how long the propagated orbit will be in field of view of an average observer (being 0.5 degrees). Both in-sample and out-of-sample RMSE are analyzed during development to prevent overfitting. Orbit determination was performed using Weighted Least-Squares. Development of the other components of the model framework was a three step process. First, preliminary design choices have been made for the model framework based on literature. Afterwards, a generic model framework has been developed that achieves high accuracy with reasonable computational load for all use cases. The main design choices to optimize for during generic model development, are the dynamical model and integrator configurations. Finally, tailoring of the framework on each individual use case (estimation window) results in optimal performance. Design choices during tailoring are estimation of radiation pressure coefficient and/or observation bias, adding an a priori covariance matrix and applying realistic observation weights.

Using this approach, many interesting findings were made. Firstly, a generic model framework has been found, estimating only initial state with a sufficiently accurate dynamical model and integrator configuration. This framework is able to efficiently achieve good propagation accuracy for all use cases and reaches PSAs from 50 days up to 2 years, depending mainly on orbit stability of the use case. Tailoring has shown significant improvements in model framework performance for 8 out of 13 use cases, decreasing out-of-sample RMSE between 20-95% and increasing PSA up to 250 days compared to the generic framework. Estimation of the radiation pressure coefficient was especially important during tailoring, and a priori covariance matrices were often required to prevent overfitting of the estimation. Comparing the various estimation windows, it has been found that estimating on 7-10 months of observations seems most robust. This window length consistently achieves the best orbit determination quality and propagation accuracy. Estimating on less than 4 months of data is sometimes sufficient, but can result in a poor fit. The tailored model framework performs well on windows that are unstable in-sample, it is able to balance out large orbital elements fluctuations during orbit determination. It is also able to propagate out-of-sample orbits accurately in non-linear environments. Though, it does still struggle with propagating orbits accurately over highly non-linear close Moon approaches out-of-sample. Finally, the effect of uncertainty over time, analyzed using Monte Carlo simulation, is highest for short windows (<4 months) and is highly dependent on orbit linearity. Uncertainty is still relatively high in the current model framework and is expected to decrease by improving radiation pressure modelling and by applying empirical accelerations.

To summarize, the model framework, PSA metric and uncertainty propagation results can guide observers in what cislunar space debris is at risk of being lost. Contributing to cislunar SSA, this work can be implemented as a tool to prevent collisions and advice observers on which cislunar space debris objects require new observations.

Contents

List of Figures

List of Tables

Nomenclature

Abbreviations

Abbreviation	Definition
AB(M)	Adam-Bashforth(-Moulton)
BS	Bulirsch-Stoer
CCA	Consider Covariance Analysis
CCD	Charge-Coupled Device
CH2	Chang'e 2 upper stage
CH3	Chang'e 3 upper stage
Decl.	Declination
DOPRI	Dormand Prince
EOM	Equation of Motion
FOV	Field Of View
GEO	Geostationary orbit
GMM	Gaussian Mixture Model
GS	Ground Station
H0	Null Hypothesis
KS	Kolmogorov-Smirnov test
IAU	International Astronomical Union
IS	In-Sample
LEO	Low Earth Orbit
MC	Monte Carlo
MEE	Modified Equinoctial Elements
MPC	Minor Planet Center
NASA	National Aeronautics and Space Administration
NEA	Near-Earth Asteroid
OOS	Out-Of-Sample
PCE	Polynomial Chaos Expansion
PDF	Probability Density Function
PM	Point-Mass
PSA	Period of Sufficient Accuracy
RA	Right Ascension
RAAN	Right Ascension of the Ascending Node
RK	Runge-Kutta integrator
RKF	Runge-Kutta-Fehlberg
RMSE	Root-Mean-Square Error
SC	Spacecraft
SH	Spherical Harmonics
(S)RP	(Solar) Radiation Pressure
SSA	Space Situational Awareness
TDB	Temps Dynamique Barycentrique
TLE	Two-Line Element
Tudat	TU Delft Astrodynamics Toolbox
UT	Unscented Transform
UTC	Universal Time Coordinates
USM	Unified State Model
V&V	Verification & Validation

Symbols

Symbol	Definition	Unit
A	Object area	$[m^2]$
C_{nm}	Spherical harmonics coeff. of order n and degree m	[-]
C_r	Radiation pressure coefficient	[-]
d	Total residual or angular separation	[arcsec]
f	Focal length	[m]
$\boldsymbol{f}(x,t)$	First order differential equation of EOM	[-]
$\boldsymbol{h}$	Observables matrix	[rad]
$\boldsymbol{H}$	Jacobian	[-]
J	Least-Squares Loss	[-]
k	true-to-formal error ratio	[-]
m	Degree of spherical harmonics	[-]
m	Mass of object	kg
n	Order of spherical harmonics	[-]
N	Sample size	[-]
P	Power of radiating body	[W]
$\boldsymbol{P}$	Covariance matrix of parameters	[-]
$\boldsymbol{P}_0^{\text{apr}}$	A priori covariance matrix	[-]
P_{nm}	Legendre polynomial of order n and degree m	[-]
r	Distance between 2 objects	[m]
$\boldsymbol{r}$	Position vector	[m]
$\dot{\boldsymbol{r}}$	Velocity vector	[m/s]
$\ddot{\boldsymbol{r}}$	Acceleration vector	$[m/s^2]$
R	Body radius	[m]
S	Sensor width	[m]
S_{nm}	Spherical harmonics coeff. of order n and degree m	[-]
t	Time	[s]
t_0	Initial time of estimation window	[s]
t_E	Final time of estimation window	[s]
$\boldsymbol{W}$	Weight matrix	[-]
$\boldsymbol{x}$	State vector	[-]
$\dot{\boldsymbol{x}}$	Derivative of state vector	[-]
$\boldsymbol{x}(t)$	State vector over time	[-]
$\dot{\boldsymbol{x}}(t)$	Derivative of state vector over time	[-]
$\boldsymbol{x}_0$	Initial state and parameter vector	[-]
$\boldsymbol{x}_0^{\text{apr}}$	Initial parameter guess	[-]
x, y, z	Cartesian position	[m]
$\dot{x}, \dot{y}, \dot{z}$	Cartesian velocity	[m/s]
$\boldsymbol{z}$	Observation vector	[-]
α	Right ascension	[arcsec]
δ	Declination	[arcsec]
λ	Longitude	[rad]
μ	Gravitational parameter	$[m^3/s^2]$
ϵ_{obs}	Observation bias	[arcsec]
ϵ_α	Right ascension bias	[arcsec]
ϵ_δ	Declination bias	[arcsec]
ϵ	Error (true or formal)	[arcsec]
σ	Standard deviation	[-]
σ_i	Mean measurement error	[arcsec]
σ_μ	Standard error of the mean	[-]
τ	Light time	[s]
ϕ	Latitude	[rad]

1 Introduction

The space industry is constantly improving its ability to launch vehicles into space. On the other hand, it is equally important to think about the long-term sustainability of these activities, especially in terms of the space debris left behind in our orbits. Space debris refers to all artificial objects orbiting the Earth (or other bodies) that do not serve a certain function. These objects usually originate from decommissioned spacecrafts, collisions and explosions [Klinkrad, 2006]. Back in 2021, NASA was already able to track at least 27000 pieces of space debris larger than a softball and half a million pieces larger than 1 cm [Garcia, 2021]. Unlike most active objects, the orbits of space debris can not be simply controlled. This is why orbit maneuvers are sometimes required to prevent collisions with space debris. Collisions have already happened, an example being the collision of the satellites Iridum 33 and Cosmos 2251 which created a massive cloud of debris [Braun et al., 2017]. Collisions with space debris would not only damage our assets, but could trigger the Kessler syndrome [Kessler et al., 2010]. A phenomenon in which a collision causes a chain reaction of collisions, mainly due the high object density, resulting in massive amounts of debris. The Space Situational Awareness (SSA) program, started by ESA, aims to locate all objects in different space regions and contributes significantly to collision prevention.

At the same time, the space industry has set their eyes on cislunar space [Laurini et al., 2018]. Which has been triggered by a second space race of reaching, colonizing and potentially industrializing the Moon. The increase in resources directed towards this region is a worldwide phenomenon. Well known examples of this trend are the Artemis program of the United States which aims to return humans to the Moon [Smith et al., 2020], Israel's lunar surface mission [Temming, 2019] and the Chinese Lunar Exploration Program (also known as the Chang'e project) with the Chang'e-4 mission exploring the far side of the Moon [Li et al., 2021]. Cislunar space has different names (e.g. XGEO) and definitions. In this research it implies the entire region of space between the Earth and Moon that is outside of near-Earth orbits.

Most space debris is currently located in near-Earth orbits [Klinkrad, 2006]. The number of artificial objects and space debris in cislunar space is still relatively small [Frueh, Howell, K. DeMars, et al., 2021]. But its tracking will become increasingly more important, considering the growing number of cislunar space missions. Furthermore, cislunar space has not been exempt of accidents. An example being the explosion of the Apollo 13 oxygen tank, which was followed by the famous words: 'Houston, we've had a problem'. This explosion created a large cloud of space debris around the spacecraft which could be seen from Earth [Administrator-NASA, 2022]. Cislunar SSA will not only prevent collisions with cislunar space debris for future lunar missions, but can also assist in object identification for optical surveys. This is specifically interesting for cislunar space debris, as these objects are often incorrectly characterized as a near-Earth asteroid, a prominent example being the WT1190F object which impacted Earth in November 2015 [Micheli et al., 2018].

Although observers can track objects in a wide variety of orbits [Blake, 2021], continuous tracking of space debris is limited to a certain extent. We only have so many instruments that can achieve this, while also being constrained by physical limits like loss of signal due to blocking of the Earth or Moon. Furthermore, for cislunar space specifically, the region of space to cover is much larger. The increased distance will also decrease the signal strength of the telescope. This is why numerical modelling of space debris orbits is becoming highly relevant. If we can track objects for a certain period and then

predict their orbit moving forward (based on numerical orbit determination and propagation), we can direct our instruments elsewhere. This solves the limitations posed by tracking and allows for optimal awareness of space debris objects in our orbits. A lot of work has been done on space debris orbit determination and propagation for near-Earth orbits [Celletti et al., 2017][Hofsteenge, 2013][Lemaître, 2019], but this field of expertise is still limited for the cislunar region.

It should be noted, some work has been done on propagating space debris in stable or periodic cislunar orbits [N. Boone et al., 2021][Wilmer et al., 2021][Wilmer et al., 2022]. These are orbits positioned in one of the Lagrange points. This is relevant as objects in these orbits tend to stay there for a long time, which is why many spacecrafts also occupy these orbits [Holzinger et al., 2021]. But these studies do not cover orbit determination and propagation of space debris in unstable orbits, while most cislunar orbits are unstable [Holzinger et al., 2021]. Furthermore, the researchers simplified the orbit determination and propagation problem by using the Bi-Circular Restricted Four-Body Problem with a highly simplified dynamical model. This model framework is unlikely to reach high accuracy for unstable cislunar orbits due to its non-linear orbital behaviour and is thus not generally applicable to cislunar space debris as a whole. Numerical orbit determination and propagation including all significant dynamics is more computationally expensive, but is likely also more accurate for cislunar space. The final paper written on the topic describes the need to investigate orbits outside of the stable cislunar regime, as this is deemed important for the growing field of Lunar missions [N. R. Boone et al., 2023]. Lastly, some numerical orbit estimation of unstable cislunar space debris orbits has already been performed prior to this research by the independent astronomer B. Gray using the Find_Orb software[1]. Although no research on its performance and suitability for the overall cislunar space debris problem has yet been published. The corresponding Find_Orb ephemeris are used as input in the orbit determination algorithm of this work.

The majority of the orbits in cislunar space are inherently chaotic [Holzinger et al., 2021]. No-longer is only the Earth's gravity a prominent acceleration on the object, but also the Moon's gravity is now a driving acceleration. This creates a system that, except in the Lagrange points, is unstable and easily affected by perturbations [Celletti et al., 2017]. It is this chaotic nature, that calls for an orbit determination and propagation model framework that is able to estimate and propagate a wide variety cislunar space debris orbits over time with high accuracy. Moreover, it should be quantified how long the model framework stays sufficiently accurate and what the uncertainty is over time. The model framework is developed using optical observations of the Chang'e 2 and Chang'e 3 upper stages, which are remnants of the Chang'e missions [Li et al., 2021]. These objects are positioned in chaotic cislunar Earth-centered orbits of which the apogee lies beyond the Moon. 13 observation periods are chosen for the two objects on which orbit determination is performed, described in the paper as estimation windows. The model framework is designed to accurately propagate orbits and corresponding uncertainty over time for cislunar space debris. Which can be used as a tool to not only prevent collisions, but to also guide observers in which cislunar objects are at high risk of being lost and therefore require new observations.

1.1 Research questions

The objective of this research is to develop and apply a model framework that can be used for high accuracy orbit determination and propagation of various cislunar space debris orbits, at the same time also quantifying the effect that uncertainty has over time. Therefore, the main research question is formulated as:

How can long-term cislunar space debris orbits be estimated and propagated with sufficient accuracy, and how does the uncertainty of the resulting orbits evolve over time?

This breaks down into the following sub-questions:

1. What model framework configuration achieves the best propagation accuracy with reasonable computational load, for any cislunar space debris orbit?
2. What is the theoretical accuracy that the model framework should achieve to be effective for cislunar Space Situational Awareness and what is the maximum propagation duration for which the model framework can maintain this accuracy?

[1] https://www.projectpluto.com/find_orb.htm

3. How is the orbit estimation quality and propagation accuracy affected by various cislunar orbit and estimation window characteristics?
4. What is the effect of estimating the radiation pressure coefficient and observation bias on orbit estimation quality and propagation accuracy of cislunar space debris?
5. What is the effect of model and observation uncertainty over time for cislunar space debris orbits?

1.2 Report structure

The main part of this research is written in a paper format and can be found in chapter 2. Afterwards, the previously mentioned research questions will be answered in chapter 3, including recommendations for future research. Four appendices provide supplementary information and results. In Appendix A, the selection process and visualizations of the cislunar space debris use cases are presented. Afterwards, Appendix B discusses the development process, intermediate and final results of generic model development and tailoring on individual use cases. Then, in Appendix C the uncertainty propagation process and results are discussed in greater detail. Finally, Appendix D presents the preliminary, and final verification and validation tests performed on this model framework.

2

Paper

The paper format can be found on the next page.

A model framework for high-accuracy orbit determination and propagation of cislunar space debris

D. Witte
Faculty of Aerospace Engineering, Delft University of Technology
GitHub repository: `https://github.com/daanwitte99/Thesis/tree/dev`

ABSTRACT

Cislunar space is getting increasingly important. Countries like the US and China are directing their attention towards the Moon with the Artemis and Chang'e missions. Simultaneously, space debris pollution of our orbits is escalating, increasing the risk of the Kessler syndrome occurring. Space Situational Awareness (SSA) aims to prevent this. The sheer amount of space debris makes continuous tracking unfeasible. This creates the need for accurate orbit determination and propagation between observations. Model frameworks have been developed extensively for space debris in near-Earth orbits, but there is little experience with cislunar space. This region is more challenging because of its unstable nature, increasing the risk of losing track of objects.

In this research, a model framework is developed, using the open-source Python package Tudat, that can estimate and propagate long-term cislunar space debris orbits accurately from optical data, whilst quantifying the uncertainties realistically over time using Monte Carlo simulation. Orbit determination has been performed using Weighted Least-Squares. The algorithm can estimate (amongst other parameters) initial states of the objects, model parameters like radiation pressure properties, and observation bias. We apply our framework to the Chang'e 2 and 3 upper stages. A selection of 13 estimation windows has allowed for analysis on a diverse range of orbital and observation characteristics. Moreover, the effect of close Moon approaches on orbit determination quality and propagation accuracy has been investigated.

A generic model framework has been found that achieves sufficient accuracy with reasonable computational load for all 13 use cases and can be used as a foundation for other cislunar space debris studies. The generic framework only estimates initial state, and uses a relatively simple dynamical model and integrator configuration. This allows sufficiently accurate propagation up to 2 years out-of-sample, depending strongly on the stability of the orbit. Afterwards, the generic model framework is tailored on individual use cases to improve its performance. Parameters like the radiation pressure coefficient and observation bias are estimated in this process. The tailored model framework improved performance for 8 out of 13 use cases (compared to the generic model framework), decreasing out-of-sample RMSE between 20-95% and increasing period of sufficient accuracy with up to 250 days. Estimating on 7-10 months of observations results in the best orbit determination quality and propagation accuracy for the cislunar use cases. The tailored model framework performs robustly for various non-linear orbits, except for out-of-sample close Moon approaches. Several solutions are proposed to solve this issue. Finally, it is found that the effect of uncertainty for cislunar space debris orbits over time is significant in the current framework. Uncertainty over time is especially large when estimating on short estimation windows (<4 months) and for orbits experiencing non-linear behaviour.

1. Introduction

Conventionally, most of the focus in the space industry has been directed towards near-Earth orbits [NASA, 2020]. After all, this is where most of our assets are located. However, this focus has extended in recent years towards cislunar space. Cislunar space officially means the entire region of space between the Earth and Moon [Holzinger et al., 2021], but generally only involves the region beyond geostationary orbits. Prominent examples of this world-wide trend are NASA's Artemis [M. Smith et al., 2020] and Lunar Gateway program [Johnson et al., 2021], ESA's Moonlight program [Ventura-Traveset et al., 2021] and the Chinese Lunar Exploration Program (also known as the Chang'e project) [Li et al., 2021]. This space race will start growing the number of objects in cislunar space.

The increasing amount of space activity both in near-Earth and cislunar regions, simultaneously grows the threat that space debris poses to space flight. In 2021, NASA was able to track at least 27000 pieces of space debris larger than a softball and half a million pieces larger than 1 cm [Garcia, 2021]. This number has likely grown since then, increasing the collision risk between our assets and pieces of space debris. Collisions would not only damage our assets, but could also trigger the Kessler syndrome [Kessler et al., 2010]. The Kessler syndrome is a phenomenon in which the object density in space is so high, that one collision can cause a cloud of space debris which cascades into other collisions, eventually making space unusable. The threat that the growing space population causes, is the reason for the Space Situational Awareness (SSA) program. The program aims to locate all objects in different space regions, reducing the risk of collisions ever occurring [Kennewell et al., 2013].

Although cislunar space is significantly less populated compared to near-Earth orbits [Frueh, Howell, K. DeMars, et al., 2021], SSA and collision prevention will become increasingly more important as interest in the region grows. Furthermore, cislunar SSA assists in object identification of

cislunar space debris, as these objects are often incorrectly identified as asteroids [Micheli et al., 2018]. Currently, many objects are being tracked by optical (and other) instruments [Blake, 2021], but continuous tracking is limited. This is solved by orbit determination and propagation. If we can track objects for a certain period and estimate their orbits from these observations, we can achieve optimal knowledge of space debris locations over time.

Extensive research has been done on space debris orbit determination and propagation for near-Earth orbits [Hofsteenge, 2013][Celletti et al., 2017][Lemaître, 2019]. But the increased influence of the Moon's gravity in cislunar space often results in unstable orbits with non-linear orbital behaviour, which is different from near-Earth regions [Holzinger et al., 2021]. Furthermore, some work has been done on propagating space debris in cislunar space, but this only covers the stable Lagrange orbits [Boone et al., 2021][Wilmer et al., 2021][Wilmer et al., 2022]. The researchers apply a simplified Bi-Circular Restricted Four-Body Problem approach, which does not take into account perturbations like solar radiation pressure, to propagate the stable space debris orbits for 50 days following a collision in cislunar space. Unstable orbits are known to be strongly affected by perturbations [Celletti et al., 2017]. For these reasons, the current near-Earth and cislunar orbit determination and propagation model frameworks are not suitable for unstable cislunar orbits. In other words, there is a large opportunity for a model framework that is able to perform long-term numerical orbit determination and propagation with high accuracy not only for stable, but also for unstable cislunar orbits. This is the main goal of this research.

As mentioned, the space debris population in cislunar space is still relatively small. There are several remnants of the Chinese Lunar Exploration Program which have been or still are in cislunar orbit. The model framework will be designed for two space debris objects left behind by this program, being the Chang'e 2 and 3 upper stages [Xu et al., 2014]. With plenty of observations spanning over 10 years, these objects are found to be suitable use cases. The optical observations are collected by the Minor Planet Center (MPC) and stored by Project Pluto. The model framework is developed using the open-source Python/C++ package, TU Delft Astrodynamics Toolbox (Tudat) [Dominic Dirkx, Mooij, et al., 2019][Dominic Dirkx, Marie Fayolle, et al., 2022]. Many different design choices are considered and tested in the model framework. Examples are dynamical models, integrator settings, propagation scheme and orbit determination algorithm [Montenbruck et al., 2000]. Orbit determination is performed on various observation periods in this 10 year span, also called estimation windows. First, a generic model framework is developed that is generally applicable for all estimation windows. Afterwards, tailoring of the model framework will be performed for each individual window. A selection of performance metrics has been defined including Root-Mean-Square Error (RMSE) [Chai et al., 2014] and Period of Sufficient Accuracy (PSA), which are used to compare model performance for various configurations. This enables development of a model framework applicable for a diverse set of cislunar space debris orbits, and also allows for analysis on the effect of orbit and observation characteristics on model performance. A deep-dive is included on the effect of close Moon approaches on orbit determination quality and propagation accuracy. Finally, orbit determination and propagation of cislunar space debris involves many uncertainties. Monte Carlo simulation is used to quantify the effect of uncertainty over time [Luo et al., 2017].

The paper is structured as follows. First, the cislunar space debris use cases (Chang'e 2 and 3 upper stages) used for model framework development, are discussed in section 2. Afterwards, in section 3 the methodology behind the model framework and the corresponding development process is presented. Then, section 4 discusses the results of the developed model framework. Finally, conclusions are drawn and recommendations are made in section 5.

2. Use cases

Before diving into the methodology behind development of a model framework that can accurately estimate and propagate various cislunar space debris orbits (see section 3), it is vital to define a suitable set of cislunar space debris use cases to develop on.

Two objects are deemed suitable for model framework development, the Chang'e 2 and 3 upper stages. Moreover, a set of observation periods (use cases) have been defined for the two objects on which orbit determination will be performed, henceforth called 'estimation windows'. This will be elaborated upon in the following sections.

2.1. Objects

Two objects are chosen from the population of cislunar space debris, the Chang'e 2 and Chang'e 3 upper stages (identifiers: 2010-050B and 2013-070B). Two pieces of space debris which have orbited or are still orbiting Earth in cislunar space, and have good observation coverage over time (observations spanning over ~10 years). The upper stages are remnants of the Chinese Lunar Exploration Program [Xu et al., 2014]. A complete set of observations of these objects can be retrieved from Project Pluto databases[1], originally gathered by the Minor Planet Center (MPC). The Chang'e 2 and 3 upper stages are boosters of a Long March 3C and 3B rocket respectively[2][3] [Cen, 2011]. Each booster has an estimated dry mass of 5000 kg. An initial radiation pressure coefficient C_r of 1.2 and maximum surface area of 37.14 m^2 is found for the cannonball radiation pressure model (elaborated in subsection 3.1). C_r will be estimated more accurately in the orbit determination process.

Chang'e 2 booster observations start in the launch year of 2010 and span to 2021, when it entered a heliocentric orbit and was lost to observers. Chang'e 3 booster observations start in the launch year of 2013, it is still visible now. Orbit characteristics of both objects over time can be analyzed from the Keplerian elements derived from Two-

[1] https://projectpluto.com/pluto/mpecs/pseudo.htm

[2] https://www.nasaspaceflight.com/2010/10/live-long-march-3c-launch-change-2-probe/

[3] https://nssdc.gsfc.nasa.gov/nmc/spacecraft/display.action?id=2013-070A

Line Elements (TLEs). The TLEs have been created by an independent researcher using the Find_Orb software package[4][5]. The Keplerian elements over time including times of observation are shown in Figure A.1 and Figure A.2 of the appendix. The Keplerian elements show that the orbit of the Chang'e 2 booster is inherently more unstable than the orbit of the Chang'e 3 booster. For simplicity, the abbreviations Chang'e 2 (or CH2) and Chang'e 3 (or CH3) will be used in figures and tables, indicating the respective boosters. The terms upper stage and booster are interchangeable, both are used throughout the paper.

2.2. Estimation window selection

One of the main research goals is analyzing the effect that different cislunar orbit and estimation window characteristics have on model performance (accuracy, computational load etc.). To analyze these effects, a wide variety of estimation windows is required on which the model framework will be tested. In this research the terms in-sample (IS) and out-of-sample (OOS) are used when referring to model performance on an estimation window. In-sample means inside the estimation window and includes observations on which orbit determination is performed, out-of-sample is the period after the estimation window and includes unseen observations w.r.t. which the propagated orbit is tested. This is further elaborated upon in subsection 3.6.

An overview of the chosen estimation windows and their characteristics can be found in Table 1. A visualization of the estimation windows on the semi-major axis, derived from Find_Orb TLEs, is shown in Figure 1. 3D visualizations of the orbits, based on estimations of the tailored model framework described in subsection 4.2, can be found in Figure A.3 and Figure A.4 of the appendix.

In total, 13 different estimation windows have been selected for the Chang'e 2 and 3 upper stages. To draw conclusions on the effect of in-sample and out-of-sample orbit stability on model performance, a wide variety of stability configurations have been picked. Visual analysis of the Keplerian elements over time (using the previously mentioned TLEs) is used as a measure for orbit stability. Furthermore, to find insights on the effect of estimation window length or number of observations on model performance, varying window lengths have been chosen. A short window means the estimation is performed on less than 4 months of observations, a medium window means 4-6 months and a long window means over 6 months of data. As a figure of merit, one orbit around the Earth typically takes around one month. Windows with the same stability configuration for a specific object, are designed to end on the exact same date. This way, out-of-sample performance analyses and findings on effect of window length are more straightforward.

The unstable out-of-sample windows of the Chang'e 2 booster (windows 5-6) end right before a close approach to the Moon on the 12th of June 2019, which can be seen in Figure A.1. The object gets so close, that it results in a sharp increase in semi-major axis and a 60 degree drop in inclination. The unstable in-sample window of the Chang'e 2 booster (window 7) includes this close approach in-sample. Unstable windows of the Chang'e 3 booster (window 5-6), as seen in Figure A.2, include a similar but much later out-of-sample encounter with the Moon on the 17th of September 2016. These encounters can be used to identify the effect that close Moon approaches have on model performance in-sample and out-of-sample.

Finally, each window has also received a code which identifies its length, in-sample stability and out-of-sample stability (used for reference in the results). For example, 7-MUS stands for Window 7, (M)edium length, (U)nstable in-sample, (S)table out-of-sample.

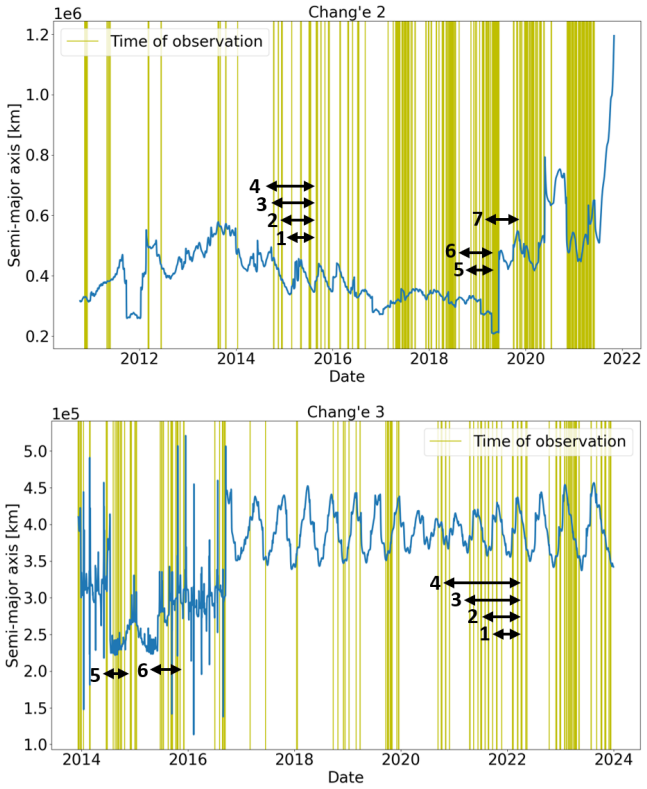


Figure 1: Estimation window visualization on semi-major axis of the Chang'e 2 and 3 upper stages

3. Methodology

This research involves modelling space debris orbits in cislunar space. Cislunar orbits are different from near-Earth orbits, as they violate the Keplerian motion assumptions more strongly due to the large third body perturbation of the Moon [Holzinger et al., 2021]. This creates an inherently unstable or chaotic system, which is easily affected by perturbations and can be challenging to propagate accurately for longer periods [Celletti et al., 2017]. A framework of models will be designed for orbit determination and propagation of a wide variety of cislunar space debris orbits, henceforth called the model framework.

A simplified overview of the model framework can be seen in Figure 2. Major design choices in the model framework are dynamical models, orbit integrators & propagators, orbit determination algorithms and much more. All elements in this framework will be elaborated upon in

[4] https://www.projectpluto.com/find_orb.htm
[5] https://github.com/Bill-Gray/tles

Table 1: Overview of estimation windows, their characteristics and identification codes

Object	Nr.	Window period	Nr. of obs.	Wind. length	IS stability	OOS stability	Code
Chang'e 2	1	2015-5-6 to 2015-7-14	42	Short	Stable	Stable	1-SSS
	2	2015-2-24 to 2015-7-14	46	Medium	Stable	Stable	2-MSS
	3	2014-11-15 to 2015-7-14	65	Long	Stable	Stable	3-LSS
	4	2014-10-11 to 2015-7-14	75	Long	Stable	Stable	4-LSS
	5	2019-2-7 to 2019-4-24	121	Short	Stable	Unstable	5-SSU
	6	2018-11-11 to 2019-4-24	138	Medium	Stable	Unstable	6-MSU
	7	2019-4-21 to 2019-10-31	181	Medium	Unstable	Stable	7-MUS
Chang'e 3	1	2021-10-19 to 2022-2-9	18	Short	Stable	Stable	1-SSS
	2	2021-7-31 to 2022-2-9	41	Medium	Stable	Stable	2-MSS
	3	2021-4-17 to 2022-2-9	68	Long	Stable	Stable	3-LSS
	4	2020-11-4 to 2022-2-9	79	Long	Stable	Stable	4-LSS
	5	2014-8-5 to 2014-10-23	34	Short	Stable	Unstable	5-SSU
	6	2015-7-1 to 2015-10-18	70	Short	Unstable	Unstable	6-SUU

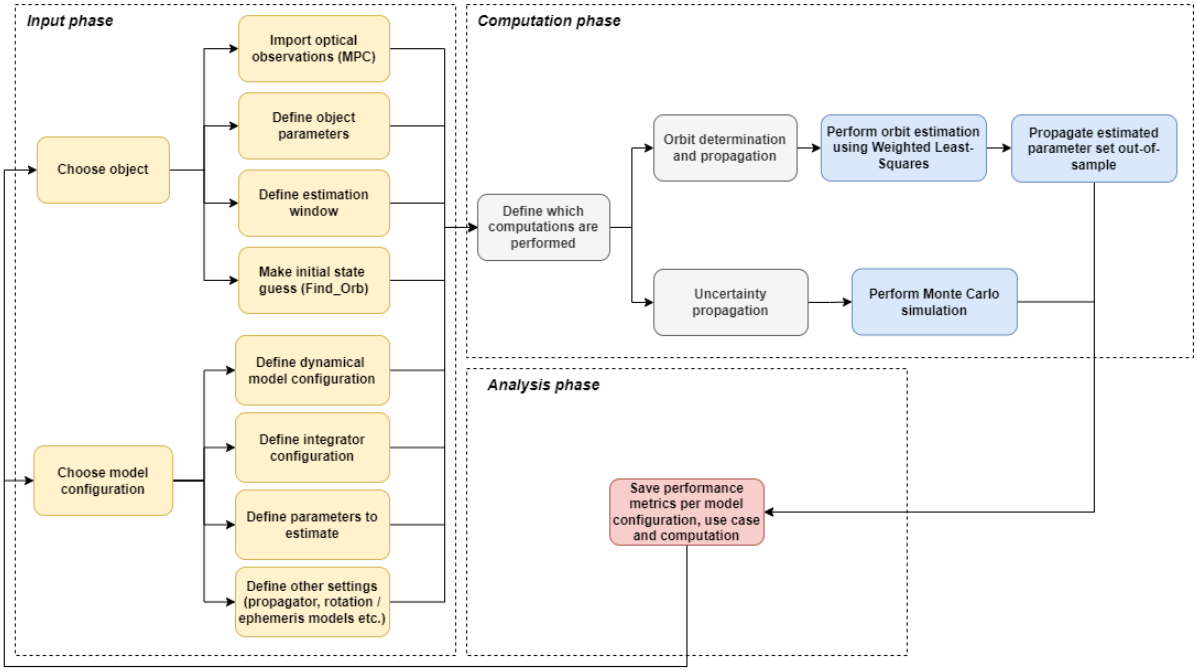


Figure 2: Overview of model framework

subsection 3.1 to subsection 3.6. The development process of the model framework will be described in subsection 3.7.

3.1. Dynamical models

Small accelerations can have a large impact on cislunar orbits over time [Holzinger et al., 2021][Celletti et al., 2017], due the unstable nature of this region. The general equation of motion (EOM) of an object with n accelerations acting on it, can be described as Equation 2.1 [Wakker, 2015]. A wide variety of accelerations act on the space debris. An overview of accelerations and respective dynamical models is shown in Table 2, including whether the models are tested in model development or not. Afterwards, the dynamical models and reasoning behind exclusion is discussed.

$$\ddot{\boldsymbol{r}} = \sum_{i=0}^{n} \ddot{\boldsymbol{r}}_i \tag{2.1}$$

Table 2: Overview of accelerations and dynamical models

Dynamical models	Tested?
Gravity - Point-mass	Yes
Gravity - Spherical harmonics (up to [20, 20])	Yes
Gravity - Schwarzschild relativistic effect	Yes
Gravity - Ocean tides	No
Gravity - Solid Earth tides	No
RP - Solar radiation pressure	Yes
RP - Earth radiation pressure	Yes
RP - Moon radiation pressure	No
RP - Cannonball model	Yes
RP - Panelled model	No
RP - Shadow effects	Yes
RP - Yarkovsky-Schach effect	No
RP - Poynting-Robertson drag	No
Other - Atmospheric drag	Yes
Other - Empirical accelerations	No

Gravitational accelerations of planetary bodies are usually implemented as a point-mass or spherical harmonics model. Point-mass models are generally applied for sufficiently small third-body perturbations [ISO, 2016]. Spherical harmonic models take the mass distribution inside a body into account and are usually used for the main gravitational acceleration and large third-body perturbations (e.g. the Moon). Higher spherical harmonics degree and order leads to better model accuracy and worse computational time. However, this effect of increasing model accuracy for higher degree and order (n, m), is less important the farther an object is from the gravitational source. This can be deduced from the distance r in the acceleration equation for spherical harmonics, shown below [Montenbruck et al., 2000]:

$$\ddot{\boldsymbol{r}}_{SH} = \nabla \left(\frac{\mu}{r} \sum_{n=0}^{\infty} \sum_{m=0}^{n} \frac{R^n}{r^n} P_{nm} \sin(\phi) (C_{nm} \cos(m\lambda) + S_{nm} \sin(m\lambda)) \right) \tag{2.2}$$

Finally, several gravitational model corrections can be made to improve accuracy of Earth's gravitational model. The Schwarzschild relativistic correction accounts for the non-Newtonian dynamics of the cislunar problem [McCarthy et al., 2004]. Ocean and solid Earth tide corrections can account for the time varying mass distribution of Earth [Montenbruck et al., 2000], but the improvement in gravity model accuracy is deemed insignificant compared to the effort their implementation requires.

Another important acceleration in cislunar space is radiation pressure. The main source of radiation pressure is direct solar radiation pressure. The Earth and Moon are significantly smaller sources of radiation pressure. Especially modelling of solar radiation pressure has shown to have a big impact on the long-term motion of space debris [Casanova et al., 2015][Hubaux et al., 2013].

Two popular methods of modelling the different radiation pressure accelerations on the object are the cannonball and panelled radiation pressure model. The cannonball model simplifies the radiation pressure problem. It assumes the object to be a sphere of a common material with a constant radiation pressure coefficient and is thus easy to apply [Scheeres et al., 2011][Celletti et al., 2017][Carrico, Policastri, and Lutz, 2018]. For clarity, the cannonball radiation pressure model is described below [Montenbruck et al., 2000]:

$$\ddot{\boldsymbol{r}}_{cb} = C_r \frac{P}{4\pi r^2} \frac{A}{m} \boldsymbol{r} \tag{2.3}$$

The acceleration depends on power of the radiating body P, distance from the body r, object radiation pressure coefficient C_r, area A and mass m. The object specific parameters for the Chang'e 2 and 3 booster are quantified in subsection 2.1. Instead of simplifying the radiation pressure model, one could also choose the panelled approach. This is used for high-accuracy orbit determination problems like the TOPEX/Poseidon mission [Marshall et al., 1992]. In the panelled radiation pressure model, the object is divided into a set of surfaces representing the shape of the object, on which the radiation pressure is applied. This can become challenging when performing orbit determination on space debris of which the shape, material and especially the rotation is unknown. Corrections can be made to account for shadow effects of large bodies like Earth. A conical shadow model is often used for the cannonball model [Montenbruck et al., 2000]. Other radiation related perturbations like the Yarkovsky-Schach effect and Poynting-Robertson drag can theoretically be considered, but are expected to be relatively small with respect to the main accelerations [Murawiecka et al., 2018][Lhotka et al., 2016] and add unnecessary complexity.

Finally, aerodynamic drag is a perturbation that is often taken into account for space debris in Low Earth Orbits (LEO), which are orbits with an altitude below 2000 km. But for space debris in higher orbits, the atmospheric density becomes so small that aerodynamic drag is often neglected [Celletti et al., 2017]. The unstable nature of cislunar orbits does not exclude space debris from entering LEO, thus aerodynamic drag should at least be considered in the model framework. The NRLMSISE-00 atmospheric model is used to model the atmospheric conditions at a specific altitude [Picone et al., 2002].

As can be seen, the dynamics of cislunar space debris can be complex to model perfectly. A way to prevent having to model all dynamics precisely, which could lead to unreasonable computational loads, is to include so-called empirical accelerations in the system [Montenbruck et al., 2000]. These empirical accelerations are estimated to account for any unmodeled accelerations. Due to time constraints, empirical accelerations have not yet been considered in this model framework.

3.2. Orbit propagation

The EOMs found by combining Equation 2.1 with different dynamical model configurations from Table 2, are ordinary differential equations. These equations need to be solved numerically to know the state of the object over time. The differential equations are solved for a specified set of initial conditions $(\boldsymbol{x}(t_0))$, which will be elaborated upon in subsection 3.3.

The EOM is a 6-dimensional first-order differential equation, which can be described as shown in Equation 2.4 [Montenbruck et al., 2000] and solved using numerical integration to approximate the state vector $\boldsymbol{x}(t)$.

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}(t, \boldsymbol{x}) = \begin{pmatrix} \dot{\boldsymbol{r}} \\ \ddot{\boldsymbol{r}} \end{pmatrix} \rightarrow \boldsymbol{x}(t) = \int_{t_0}^{t_1} \boldsymbol{f}(t, \boldsymbol{x}) dt \tag{2.4}$$

To solve this, a wide variety of numerical integrators are applied in the field of orbit propagation and have been considered in this research. In the end, a variable step-size RKF7(8) integrator with an error tolerance of 10^{-10} is applied [Fehlberg, 1968]. It has been found to be most effective from various integrator tests performed during the development process, described in subsection 3.7. This integrator performs single step integration with step size

control, which makes it suitable for rapidly changing non-linear dynamics and thus suitable for cislunar space.

After choosing a numerical integrator to solve the equations of motion, it is also important to define which propagation scheme is used. This inherently means, choosing what coordinate system to represent the equation of motion in for numerical integration. The Cowell propagator, or method of Cowell, is deemed to be most suitable for the cislunar space debris problem. Advantages of the Cowell propagator are that the EOM is straightforward and the scheme does not depend on any behaviour assumptions, which makes it robust and easy to use. On the other hand, computational load can be relatively large [Montenbruck et al., 2000]. Since robustness is important for the potential non-linearity of cislunar orbits and computational load is not a leading requirement, Cowell is applied.

Using the Cowell propagator, the Cartesian state is integrated and propagated over time. The equation of motion, state and state derivative are defined as follows [Wakker, 2015]:

$$\ddot{\boldsymbol{r}} = \sum_{i=0}^{n} \ddot{\boldsymbol{r}}_i \quad | \quad \boldsymbol{x} = \begin{pmatrix} \boldsymbol{r} \\ \dot{\boldsymbol{r}} \end{pmatrix} \quad \dot{\boldsymbol{x}} = \begin{pmatrix} \dot{\boldsymbol{r}} \\ \ddot{\boldsymbol{r}} \end{pmatrix} \tag{2.5}$$

3.3. Orbit determination

Orbit determination is the process of estimating the initial state and model parameters used for solving the EOM with numerical integration, described in subsection 3.2. The parameters are estimated from a set of orbit observations, which in this case are optical observations. Weighted Least-Squares [Montenbruck et al., 2000] is utilized for orbit determination because of its robust and ease of use properties.

Weighted Least-Squares estimation is a form of Least-Squares in which observations can receive different weights based on the mean measurement error of each observation. Least-Squares is an orbit determination technique that tries to find the set of orbit and model parameters that minimizes the sum of the squared residuals between the modeled observations and the actual observations. The parameter set is usually a combination of the initial state ($\boldsymbol{x}(t_0)$) and several force parameters like the cannonball radiation pressure coefficient C_r. But one can also estimate parameters like observation bias ϵ_{obs}, which are measurement uncertainties and will be elaborated upon in subsection 3.5. The parameters are assumed to be (sufficiently) constant.

One important note, the vector $\boldsymbol{x}_0$ includes both the initial state vector and other estimated parameters for the following formulations. Thus, $\boldsymbol{x}_0$ could for example be defined as:

$$\boldsymbol{x}_0 = [x(t_0), y(t_0), z(t_0), \dot{x}(t_0), \dot{y}(t_0), \dot{z}(t_0), C_r, \epsilon_{obs}]^T \tag{2.6}$$

In this research, single-arc orbit determination is performed for simplicity. This means one set of parameters $\boldsymbol{x}_0$ is estimated for the entire set of observations. An initial parameter guess $\boldsymbol{x}_0^{\text{apr}}$ is required for initialization of the Weighted Least-Squares algorithm. The general solution of the Weighted Least-Squares problem is defined as:

$$\Delta\boldsymbol{x}_0^{\text{lsq}} = (\boldsymbol{H}^T\boldsymbol{W}\boldsymbol{H})^{-1}(\boldsymbol{H}^T\boldsymbol{W}\Delta\boldsymbol{z}) = \boldsymbol{P}(\boldsymbol{H}^T\boldsymbol{W}\Delta\boldsymbol{z}) \tag{2.7}$$

$\boldsymbol{H}$ is the Jacobian and contains the partial derivatives of the modeled observations w.r.t. estimated parameters, $\Delta\boldsymbol{x}_0^{\text{lsq}}$ is the difference between the parameter set of the Weighted Least-Squares solution and a reference set, and $\Delta\boldsymbol{z}$ contains the difference between the actual and predicted observations from the simulated trajectory. Furthermore, $(\boldsymbol{H}^T\boldsymbol{W}\boldsymbol{H})^{-1}$ is the covariance matrix $\boldsymbol{P}$ of the parameters and $\boldsymbol{W}$ is the weight matrix or covariance matrix of the observations, being [Montenbruck et al., 2000]:

$$\boldsymbol{W} = \text{diag}\left(\sigma_1^{-2}, \ldots, \sigma_n^{-2}\right) = \begin{pmatrix} \sigma_1^{-2} & & 0 \\ & \ddots & \\ 0 & & \sigma_n^{-2} \end{pmatrix} \tag{2.8}$$

Here, σ_i is the mean measurement error in the observations due to random noise and systematic errors (described in subsection 3.4). It is assumed that the measurement errors are uncorrelated and have a Gaussian distribution. If either assumption is violated, the weighting matrix will become non-diagonal [Montenbruck et al., 2000]. The solution $\Delta\boldsymbol{x}_0^{\text{lsq}}$ is found iteratively to minimize the loss function:

$$\boldsymbol{J}(\Delta\boldsymbol{x}_0) = (\Delta\boldsymbol{z} - \boldsymbol{H}\Delta\boldsymbol{x}_0)^T\boldsymbol{W}(\Delta\boldsymbol{z} - \boldsymbol{H}\Delta\boldsymbol{x}_0) = \boldsymbol{\rho}^T\boldsymbol{W}\boldsymbol{\rho} \tag{2.9}$$

Every time $\Delta\boldsymbol{x}_0^{\text{lsq}}$ is found, it is used to compute a new set of reference parameters $\boldsymbol{x}_0^{\text{new}} = \boldsymbol{x}_0^{\text{ref}} + \Delta\boldsymbol{x}_0^{\text{lsq}}$, which is then used to propagate a new reference orbit (see subsection 3.2). This is done until the relative change of the loss function shown in Equation 2.9, achieves a certain threshold. The residuals of an ideal fit would be perfectly Gaussian, since only random observation noise will be left in the residuals. This will be further elaborated upon in subsection 3.4. Thus, the in-sample residual distribution needs to be analyzed to ensure a good fit of the estimation on the data.

It can be useful to include information on accuracy of the initial parameter guess. This is achieved using an a priori covariance matrix $\boldsymbol{P}_0^{\text{apr}}$, which penalizes large deviations from the initial guess $\boldsymbol{x}_0^{\text{apr}}$ [Montenbruck et al., 2000]. Applying $\boldsymbol{P}_0^{\text{apr}}$ can be used to prevent overfitting. Overfitting is overconfidence of the estimation on the observation set, which results in poor performance when comparing the resulting orbit propagation to unseen data. The modified solution and loss function are shown in Equation 2.10 and Equation 2.11, where $\Delta\boldsymbol{x}_0^{\text{apr}} = \boldsymbol{x}_0^{\text{apr}} - \boldsymbol{x}_0$.

$$\Delta\boldsymbol{x}_0^{\text{lsq}} = \left((\boldsymbol{P}_0^{\text{apr}})^{-1} + \boldsymbol{H}^T\boldsymbol{W}\boldsymbol{H}\right)^{-1} \cdot \left((\boldsymbol{P}_0^{\text{apr}})^{-1}\Delta\boldsymbol{x}_0^{\text{apr}} + \boldsymbol{H}^T\boldsymbol{W}\Delta\boldsymbol{z}\right) \tag{2.10}$$

$$J = \left(\Delta \boldsymbol{x}_0 - \Delta \boldsymbol{x}_0^{\text{apr}}\right)^T (\boldsymbol{P}_0^{\text{apr}})^{-1} \left(\Delta \boldsymbol{x}_0 - \Delta \boldsymbol{x}_0^{\text{apr}}\right) + \boldsymbol{\rho}^T \boldsymbol{W} \boldsymbol{\rho} \tag{2.11}$$

In this problem, the optical observations that are used as input in the Weighted Least-Squares algorithm, are topocentric right ascension (α) and declination (δ) observations in the J2000 frame, elaborated upon in subsection 3.4. Thus, the observations are defined w.r.t. the observer. A general orbit determination process looks as follows.

First, the observer delivers the topocentric measurements of the spacecraft position in the form of right ascension and declination. Then, a guess is made for the initial state and potential other parameters. The initial state guess for the use cases described in section 2, has been retrieved from orbit estimations made by an independent researcher using Find_Orb[6]. Find_Orb is an individual orbit determination software package developed by Project Pluto. Initial guesses for other parameters are found from literature. Afterwards, the software (being Tudat), propagates the geocentric spacecraft state (and the variational equations) over time. The geocentric Cartesian observer state can be retrieved based on known ground station locations. This is then used to calculate a Cartesian topocentric state $\bar{\boldsymbol{r}}$ over time, as shown in Equation 2.12 [Montenbruck et al., 2000] and Figure 3. This equation does not yet include an aberration correction, elaborated upon in subsection 3.4.

$$\bar{\boldsymbol{r}} = \bar{\boldsymbol{r}}_{sc} - \bar{\boldsymbol{r}}_{obs} \tag{2.12}$$

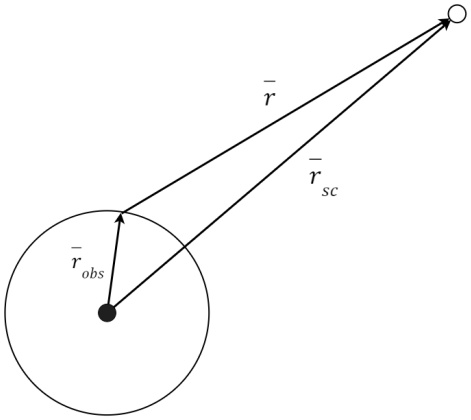


Figure 3: Difference between topocentric and geocentric state

The Cartesian topocentric state $\bar{r}$ is then transformed to polar coordinates using [Montenbruck et al., 2000]:

$$\bar{\alpha} = \arctan\left(\frac{\bar{y}}{\bar{x}}\right) \quad \bar{\delta} = \arctan\left(\frac{\bar{z}}{\sqrt{\bar{x}^2 + \bar{y}^2}}\right) \qquad \bar{r} = \sqrt{\bar{x}^2 + \bar{y}^2 + \bar{z}^2} \tag{2.13}$$

Here, $\bar{x}, \bar{y}, \bar{z}$ is the propagated Cartesian position and $\bar{\alpha}, \bar{\delta}, \bar{r}$ is the propagated polar position (right ascension, declination and distance). It can thus be seen that the modelled observables ($\bar{\alpha}$ and $\bar{\delta}$) only depend on the propagated Cartesian position. If propagated Cartesian position is inaccurate due to for example poorly defined dynamics, this will result in errors in the modelled observables.

Afterwards, the modelled angular position over time ($\bar{\alpha}$ and $\bar{\delta}$) is compared to the measured one and a residual is calculated. The residual is then used to calculate the Least-Squares loss and a differential correction $\Delta \boldsymbol{x}_0$, as shown in Equation 2.9 and Equation 2.7 respectively. The differential correction is applied to the previous parameter set and another iteration is performed (if necessary).

Finally, the Jacobian H, required for calculating the differential correction and loss, contains the partial derivatives of the modelled observables (α, δ) w.r.t. estimated parameters $\boldsymbol{x}_0$ and is defined as follows [Montenbruck et al., 2000]:

$$H = \frac{\partial \boldsymbol{h}(\boldsymbol{x_0})}{\partial \boldsymbol{x_0}} = \begin{bmatrix} \frac{\partial \alpha(\boldsymbol{x_0})}{\partial \boldsymbol{x_0}} & \frac{\partial \delta(\boldsymbol{x_0})}{\partial \boldsymbol{x_0}} \end{bmatrix} \tag{2.14}$$

In other words, H can be found by taking the derivative of $\bar{\alpha}$ and $\bar{\delta}$, shown in Equation 2.13, w.r.t. the parameter vector $\boldsymbol{x}_0$.

3.4. Optical tracking

As mentioned in subsection 3.3, orbit determination is performed on optical observations in this research. Space debris orbits are often captured using optical instrumentation at several ground stations, which utilize so-called Charge-Coupled Devices, or CCD cameras [Klinkrad, 2006]. The observations are made in sidereally driven staring mode [Klinkrad, 2006], meaning the object appears as a streak of light on a fixed background of stars. A representation is shown in Figure 4.

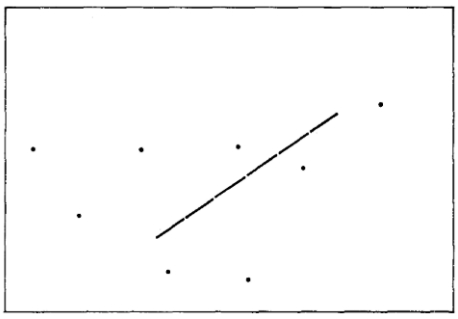


Figure 4: Representation of an image made with sidereally driven staring mode [Veis, 1963]

The utilized space debris observations have originally been collected by the Minor Planet Center[7] (MPC), but can still be retrieved from Project Pluto[8]. MPC is an organization, mandated by the International Astronomical Union (IAU), responsible for designating and tracking minor bodies in our solar system. MPC also collects data on artificial satellites, which are often found as a byproduct in near-Earth asteroid surveys [Micheli et al., 2018]. MPC cooperates with many ground stations and observers

[6] https://www.projectpluto.com/find_orb.htm
[7] https://www.minorplanetcenter.net/iau/mpc.html
[8] https://projectpluto.com/pluto/mpecs/pseudo.htm

around the world to collect this data. Optical observations stored by MPC are required to have a precision of at least 1 arcsecond[9]. It is constrained by the MPC that at least 2 observations are made of each object per night. The total observation should take at least 30 minutes, but a general recommendation is to have 1 hour of coverage for 2-3 observations. As mentioned in subsection 3.3, the observation errors are assumed to be uncorrelated, which is further justified by the spacing regulations of the MPC.

Optical observations in the MPC format are in terms of right ascension (α) and declination (δ) in a topocentric J2000 frame. Each observation contains (amongst other information) the time of observation, object identifier and observatory code[10]. Though, the observations stored by the MPC are required to have a high accuracy, they are not free from imperfections.

An important correction to apply, is the aberration or light-time correction [Moyer, 2005]. Simply put, the observed object position at time t is actually the position of the object at time $t - \tau$, where τ is the time it took the light to reach the observer. This introduces a small error in the measured position, which essentially means that Equation 2.12 should be defined as:

$$\bar{r}(t) = \bar{r}_{sc}(t - \tau) - \bar{r}_{obs}(t) \tag{2.15}$$

Iteratively estimating the light time τ, as shown by [Montenbruck et al., 2000], allows to correct the object state for the aberration effect. Because of the nature of the aberration effect, the correction becomes more significant for objects travelling at higher speeds and greater distances from the observer. Since the effect is already about 0.6 arcseconds for LEO objects [Montenbruck et al., 2000] and the cislunar space debris problem is a highly elliptical one with greater distance from the Earth, it will most likely have a significant impact and thus has to be applied. The aberration correction is applied automatically by Tudat.

Unfortunately, there is also some level of uncertainty present in the optical observations, which can be attributed to observation noise. These are errors created by the instrumentation itself. A big challenge is the inconsistency in instrumentation and how observations are made at different ground stations [Klinkrad, 2006]. This induces varying noise levels in measurements.

In general there are different noise sources present in the data, resulting in either observation bias, time bias or random error. The different types are described below [Kjeldsen et al., 1992]:

- **Observation bias**: A bias introduced in the dataset by the instrument that made the observation. This can be caused by resolution constraints, tracking errors, inaccurate ground station coordinates etc.
- **Time bias**: Time bias corresponds to errors present in the time of observation. This is different from observation bias as its magnitude is time dependent. For example, positional error due to time bias will be larger in perigee, caused by increased angular velocity relative to the observer [Langbroek, 2023].
- **Random error**: Some random error is always present in measurements and is typically assumed to follow a Gaussian distribution.

As described earlier, the total bias of the measurements can be no larger than 1 arcsecond, according to the MPC guidelines. But it has been shown in previous studies that some ground stations do not meet this constraint [Vereš et al., 2017]. Furthermore, since cislunar space debris can travel at much higher speeds in perigee than the objects the surveys are meant to observe, the time bias alone might even be up to 3 arcseconds [Langbroek, 2023].

An overall measurement bias (ϵ_{obs}) can be estimated using orbit determination for observations of each ground station, as explained in subsection 3.3. The overall bias will likely estimate both observation and time bias in its magnitude. Since time bias is variational, one should either estimate a separate bias for each streak of observations or accept that the overall bias is a rough approximation. Moreover, one must be careful with estimating other unmodeled effects into the bias and drawing conclusions. Propagating the effect of observation (and model) uncertainty over time will be elaborated upon in subsection 3.5.

3.5. Uncertainty

Realistic uncertainty propagation and analysis is important for Space Situational Awareness (SSA) [Frueh, Howell, K. DeMars, et al., 2021][Cano et al., 2022], as it is a measure of the reliability of your propagated orbit over time. In general, uncertainty sources can be divided into two different types. Epistemic and aleatory uncertainty [Luo et al., 2017]. The main difference between the uncertainty types, is the way the errors are quantified. An overview of the most significant uncertainty sources in cislunar space is shown in Table 3.

Table 3: Overview of uncertainty sources

Uncertainty source	Type
Earth's gravity uncertainty	Epistemic
Moon's gravity uncertainty	Epistemic
Third body gravity uncertainty	Epistemic
Radiation pressure uncertainty	Aleatory
Observation bias uncertainty	Aleatory
Time bias uncertainty	Aleatory
Random noise uncertainty	Aleatory

Epistemic uncertainty, also called model simplification or type B uncertainty [Luo et al., 2017][Hoffman et al., 1994], refers to the reducible error usually caused by the decision to use a simpler model instead of the state-of-the-art option. This choice is often made to improve the computational efficiency of the model and is driven by the model requirements. The nature of epistemic uncertainties, means that the errors can be calculated deterministically. For example, can quantify the uncertainty induced by a lower degree/order spherical harmonics of Earth, by

[9] https://minorplanetcenter.net/iau/info/Astrometry.html
[10] https://www.minorplanetcenter.net/iau/info/OpticalObs.html

propagating the same orbit with a high degree/order model and comparing model performance. Various epistemic uncertainties will be introduced and quantified during generic model framework development. The generic model framework is developed in such a way that the epistemic uncertainties are negligible, as elaborated upon in subsection 3.7.

Aleatory uncertainty, also called imperfect knowledge or type A uncertainty [Luo et al., 2017][Hoffman et al., 1994], is caused by the fact that it is practically impossible to model an environment perfectly, and that there is always noise present in observations. This causes an uncertainty in the system that is difficult and sometimes impossible to reduce. Aleatory uncertainties usually drive the requirements set for a problem. Prominent examples are uncertainties in the estimated state caused by observation bias or imperfections in the cannonball radiation pressure model. Aleatory uncertainties can not be quantified using simple deterministic analysis. The assumed probabilistic nature of these errors means that statistical analysis of probability distributions is required to quantify its effect over time [Luo et al., 2017]. The goal of statistical analysis is to predict the state and uncertainty at a later point in time, given an initial state and uncertainty. Covariance is used to represent the initial uncertainty, which assumes a Gaussian uncertainty distribution. It can be unrealistic for non-Gaussian uncertainty caused by highly non-linear behaviour, this will be investigated. Different linear and non-linear uncertainty propagators have been considered to predict uncertainty over time. In the end, Monte Carlo simulation is found to be most suitable.

Monte Carlo (MC) simulation is a widely applied method of uncertainty propagation for various objects and orbits [Luo et al., 2017][Virgili et al., 2021]. In this research, the MC approach approximates the Probability Density Function (PDF) of the state at time t, by randomly simulating N parameter sets from the mean and covariance of the parameter estimation at an initial time, and propagating the samples to a later time t [Luo et al., 2017]. The distribution of the propagated samples approximates the uncertainty over time. The larger the sample size, the better the approximation of the future PDF becomes.

Advantages of the MC method are that it can reach a very high accuracy and is easy to use. But a big drawback is that it can be inefficient [Luo et al., 2017]. The many repeated sample calculations, required for Monte Carlo, cause it to quickly become computationally expensive compared to other methods. The high accuracy that the MC method can achieve, makes it a very robust but brute-force option to use for uncertainty propagation. The MC method is already applied for the application of cislunar SSA, where the uncertainty of many different cislunar orbits has been propagated successfully [Frueh, Howell, K. J. DeMars, et al., 2021]. In this research, as will be explained in section 2, only two objects and in total 13 estimation windows are analyzed. The high accuracy constraints that cislunar space poses, and the low computationally load that the limited use cases bring, make Monte Carlo suitable to the problem at hand. One important note, the accuracy of MC simulation depends heavily on whether the initial uncertainty distribution is correct or not. The initial uncertainty distribution is found from the estimated covariance at that time, retrieved from orbit estimation and described in subsection 3.3. This process will be elaborated upon below.

As mentioned, Monte Carlo simulation requires a mean and covariance at a specific time for initialization. In this work, MC is initialized at the end of the observation period (or estimation window) on which the orbit is estimated, defined by t_E. The mean is the vector containing the state at this time and other estimated parameters (e.g. C_r and ϵ_{obs}), both retrieved from orbit estimation. The covariance matrix is retrieved from the estimator at the start of the estimation window t_0 and propagated to t_E using the variational equations [M. Fayolle et al., 2022]. This approach assumes linear uncertainty propagation in-sample.

It is important to note, that the estimated covariance matrix is often an underestimation of the true covariance at a specific time. This is because the formal error of the estimation, and the true error (or residual) are not equal. The formal error is equal to the square root of the diagonals of the covariance matrix $\boldsymbol{P}$, shown in Equation 2.7, and is thus a measure of the assumed errors in the state. The true error tends to be much higher than the formal error. This difference is generally caused by imperfections in the dynamical model and any noise left in the observations [M. Fayolle et al., 2022][D. Dirkx et al., 2017], causing overconfidence of the orbit determination algorithm on the formal errors of the fit.

To compensate for this, an approximation method is proposed. One can calculate an average ratio k between the formal and true errors, computed using the average of the median of the out-of-sample k ratios in α and δ residuals (shown in Equation 2.16). Since k ratio tends to fluctuate between α and δ as well as over time, taking the median should negate the effect that outliers may have on the average k ratio.

$$k_{avg} = \frac{\text{median}\left(\frac{\epsilon_{\alpha,true}}{\epsilon_{\alpha,formal}}\right) + \text{median}\left(\frac{\epsilon_{\delta,true}}{\epsilon_{\delta,formal}}\right)}{2} \tag{2.16}$$

This k ratio method assumes linear uncertainty propagation over time, thus careful consideration has to be put in the out-of-sample window that is used to approximate it and on the resulting uncertainty propagation. Afterwards, the true covariance matrix is calculated using:

$$\boldsymbol{P}_{true} = \boldsymbol{P}_{formal} \cdot k_{avg}^2 \tag{2.17}$$

Applying a constant k ratio to approximate true covariance is used in reference studies for covariance analysis on the JUICE mission [M. Fayolle et al., 2022][D. Dirkx et al., 2017]. Using the mean and true covariance, one can utilize a multivariate Gaussian Random Number generator[11] to find N different parameter sets. These parameter sets are then applied for N orbit propagations, resulting in propagation

[11] https://numpy.org/doc/stable/reference/random/generated/numpy.random.normal.html

of (aleatory) uncertainty for different use cases and model configurations over time.

As described in previous paragraphs, the current Monte Carlo approach makes several assumptions. These assumptions need to be tested, to verify and validate whether the initial generated uncertainty distribution is sufficiently correct. Firstly, covariance propagation from t_0 to t_E assumes that in-sample uncertainty can be linearly propagated. This is easily verified by performing MC simulation from t_0 and from t_E, and ensuring out-of-sample MC distributions are sufficiently similar to each other. If this is not the case, then the propagated covariance at t_E is likely a poor representation of the actual uncertainty distribution, resulting in errors in the subsequent propagated uncertainty distributions. This can for example be caused by highly non-linear orbital behaviour in-sample. Moreover, the N random samples used as propagation input at t_E, should represent a Gaussian distribution for the initial uncertainty distribution to be valid. Furthermore, linear uncertainty propagation is assumed out-of-sample for the average k ratio approach to find a realistic initial uncertainty distribution. As long as uncertainty propagation is linear, the state distribution of the MC samples should also stay Gaussian. This information can be used to analyze for what period the k ratio approximation holds out-of-sample. Whether the initial state and propagated state distributions are still Gaussian, can be tested using the Kolmogorov-Smirnov (KS) test [Massey Jr, 1951]. The KS test compares a sample distribution to a theoretical Gaussian distribution with sample mean and variance. It then tests the null hypothesis (H0) that the sample distribution at a specific time is Gaussian. The KS test calculates a p-value, which is a probability that defines how likely it is that the tested distribution is Gaussian. In this study, a significance level of 0.05 (or 5%) will be used to test the null hypothesis. Meaning if the p-value of the KS test is smaller than 0.05, H0 is rejected and the distribution is concluded to be non-Gaussian.

3.6. Performance metrics

In order to optimize the model framework for cislunar space, it is important to analyze a suitable set of performance metrics. The goal is to find a sufficiently accurate model framework with reasonable computational load. It should be noted that accuracy is generally prioritized over computational load during model development.

Firstly, parameters are estimated during orbit determination that most accurately represent a set of optical observations, as described in subsection 3.3. In theory, one can add infinitely many model parameters to perfectly reproduce the observation set. This will cause overfitting, where the algorithm gets overconfident on the dataset it is trained on and performs poorly on data it has not seen before [Ying, 2019]. This can be prevented by not only optimizing the model performance on the training or in-sample (IS) data, but also validating the model on new data. This is called test or out-of-sample (OOS) data. By also optimizing on out-of-sample performance, one ensures that the model is a better representation of reality.

In this research, in-sample and out-of-sample accuracy is analyzed using the total residual of the estimated orbit. The total residual is calculated from right ascension (α) and declination (δ) residuals w.r.t. actual observations. A correction is required for the right ascension residual, and therefore indirectly also for the total residual. This is because the actual angular distance of a change in right ascension becomes smaller, the closer one gets to the celestial pole. The corrected right ascension residual and total residual are calculated as follows [Meeus, 1991]:

$$\Delta\alpha_{corr} = \sqrt{d^2 - \Delta\delta^2} \tag{2.18}$$

$$d = 2 \cdot \arcsin\left(\sqrt{\sin^2\left(\frac{\Delta\delta}{2}\right) + \cos(\delta_1)\cos(\delta_2)\sin^2\left(\frac{\Delta\alpha}{2}\right)}\right) \tag{2.19}$$

Here d is the total residual, $\Delta\alpha$ and $\Delta\delta$ are the original right ascension and declination residuals, δ_1 and δ_2 are the observed and estimated declinations, and $\Delta\alpha_{corr}$ is the corrected right ascension residual.

To analyze the average performance in-sample and out-of-sample of the model, the Root-Mean-Square Error (RMSE) of the total residual is used [Chai et al., 2014]. RMSE is suitable when residuals are Gaussian. This is assumed to be the case, as explained in subsection 3.3.

An important goal of this model framework is to be able to accurately propagate orbits of cislunar space debris as long as possible into the future (out-of-sample), maintaining reasonable computational load. This creates the need for a threshold that defines whether the propagated orbit is still sufficiently accurate or not. A new metric has been developed to be able to define this, called the Period of Sufficient Accuracy (PSA). PSA defines the period in which the residual of the propagated orbit is within Field Of View (FOV) of an average professional observer. Using PSA, one can draw conclusions on when an object might move away from its predicted location and when new observations are required from observers to perform another orbit estimation. The accuracy threshold for PSA is derived from the FOV of an average observer. This can be computed as follows [Rowlands, 2017]:

$$\text{FOV} = 2 \cdot \arctan\left(\frac{S}{2f}\right) \tag{2.20}$$

Here S is the width of the sensor that captures the image and f is the focal length of the telescope. To calculate a FOV for the PSA threshold, the QHY600PH camera is taken. This is a widely used piece of instrumentation in telescopes of average professionals. S for the QHY600PH is 35 mm[12]. Assuming standard telescopes have an F-ratio of F/8 with a lens diameter of 50 cm or larger, an average focal length is estimated to be $8 \cdot 0.5 = 4$ m [W. Smith, 2000]. This leads to a FOV of 0.5° or approximately 1800 arcseconds. Since the space debris being tracked should always be in the FOV of the telescope directed at the predicted state of the space debris, the accuracy threshold should be half the FOV, being $\frac{1800}{2} = 900$ arcseconds. It

[12] https://www.qhyccd.com/astronomical-camera-qhy600/

should be noted that FOV of large observatories, tends to be much bigger [Valsecchi, 2019]. But a more strict FOV is used for the threshold to allow any professional observer to find the objects reliably based on the propagated states. To summarize, PSA is the out-of-sample period in which the propagated residual remains below 900 arcseconds. As will be seen in the results (section 4), PSA will often be a period range (e.g. PSA = 369-418 days). This is caused by the sparse distribution of observations over time, from which residuals and thus also PSA is calculated. The lower bound of the range is always within the accuracy threshold, the higher bound is not.

3.7. Model framework development process

The orbit determination and propagation model framework, described in the previous sections, has been developed for the use cases presented in section 2. The corresponding development process is presented here.

Various design choices should be made in the development process of an orbit determination and propagation model framework for cislunar space debris. Some design choices are made preliminarily and others through extensive testing. Each design choice has a different impact on orbit estimation quality, propagation accuracy and computational load. The development process has therefore been divided into three 3 main steps, which are both summarized and described in more detail below:

1. Make preliminary design choices in model framework based on literature.
2. Make design choices for dynamical model and integrator configuration by testing various configurations on propagation accuracy and computational load. The goal is to find a dynamical model and integrator configuration that achieves sufficient accuracy for all estimation windows while maintaining a reasonable computational load. The corresponding model framework is henceforth called 'generic model framework'. Results are shown in subsection 4.1.
3. Tailor generic model framework on individual estimation windows to further improve propagation accuracy, henceforth called 'tailored model framework'. Tailoring is done by adding C_r and/or ϵ_{obs} to the estimation, applying an a priori covariance matrix or using realistic observation weights. Results are shown in subsection 4.2.

The first step in the development process is making preliminary design choices for the model framework based on literature. This is done to limit the number of model configurations tested during step 2 (generic model framework development). The preliminary choices are made to be robust, easy to use and have sufficient accuracy. Preliminary design choices are made for the orbit determination algorithm (Weighted Least-Squares), estimated parameters (initial state), propagation scheme (Cowell), planetary body ephemeris and rotation models (Spice kernels). Additional reasoning behind the propagation scheme and orbit determination algorithm can be found in subsection 3.2 and subsection 3.3. The Spice kernels contain sufficiently accurate ephemeris and rotation models, which are applied by default in Tudat[13] and are therefore also used in this model framework. Having defined these model framework components, two main design choices remain in the model framework. These are the dynamical model and integrator configurations.

This leads to the second step in the development process, developing the generic model framework based on dynamical model and integrator testing. The dynamical model and integrator configuration are found to have a large influence on model performance. A large variety of dynamical model and integrator configurations are therefore tested to find the configurations that are both sufficiently accurate and computationally efficient for all 13 use cases (described in subsection 2.2). In this process of finding sufficiently accurate model configurations, the effect of epistemic uncertainties over time is greatly reduced. This leads to the 'generic model framework', which is designed to be applicable for any cislunar space debris orbit (but is not necessarily the most accurate configuration achievable). The dynamical model and integrator configurations are tested as follows.

A model framework is defined that has a highly accurate dynamical model and integrator configuration, but also has (unnecessarily) high computational load. This is called the baseline model framework, it can be found in Table B.1 of the appendix. It is used as a benchmark, against which the impact of each individual dynamical model and integrator design choice can be tested and analyzed. For example, one can decrease the Spherical Harmonics (SH) model degree/order for Earth from the baseline (10, 10) to (5, 5) and compare the performance metrics of the resulting orbit propagation with the metrics of the baseline. If in-sample and out-of-sample RMSE as well as PSA (Period of Sufficient Accuracy) are similar to the metrics achieved by the baseline but computational load decreases by 50%, then SH(5, 5) is found to be more suitable for the generic model framework. An overview of the dynamical model and integrator testing plan can be found in Table B.2 of the appendix. Over 20 different dynamical model and integrator configurations have been tested, the resulting generic model framework is presented in subsection 4.1 of the results. Finally, the generic model framework is designed to be applicable for any cislunar space debris orbit, but this is not the highest propagation accuracy achievable.

This is where the third step of the development process comes in, model framework tailoring. In this step, the generic model framework is tailored on each individual estimation window to improve orbit determination quality and accuracy of the propagated orbit. These design choices are made individually (not generally) because they have varying impact on performance for each specific use case and no single tailored model configuration achieves high performance for all windows. Design choices in the tailoring phase are, estimation of unknown parameters like radiation pressure coefficient (C_r) and observation

[13] https://docs.tudat.space/en/latest/_src_user_guide/state_propagation/environment_setup/default_env_models.html?highlight=default

bias (ϵ_{obs}), including an a priori covariance matrix to prevent overfitting of parameters [Montenbruck et al., 2000], and adding realistic observation weights to represent the random noise. The observation bias is estimated in terms of right ascension and declination for each ground station individually. Similar to step 2, each tailoring design choice is tested individually to analyze its effect on the performance metrics compared to the generic model framework. This leads to the tailored model framework configurations described in subsection 4.2 of the results, which show improvements compared to the generic model framework for 8 out of 13 estimation windows. The resulting model framework (after tailoring), is therefore used in all subsequent results and analyses presented in subsection 4.3, subsection 4.4 and subsection 4.5.

To summarize, the generic model framework is designed to be applied in future research as a starting point which achieves sufficient accuracy for any cislunar space debris orbit. Afterwards, the previously described tailoring process should be applied by researchers to ensure that the best estimation quality and propagation accuracy is realized for each individual use case.

4. Results & discussion

In this section, the main research results are presented and discussed. First, the generic model framework and its performance is presented in subsection 4.1. Afterwards in subsection 4.2, the results of model framework tailoring on each individual estimation window and the corresponding performance improvements are shown. The tailored model framework is used for all subsequent results and analyses. A detailed analysis on the effect of various orbit and estimation window characteristics on tailored model framework performance is shown in subsection 4.3. In subsection 4.4, the effect of close approaches to the Moon is discussed. Finally, results on uncertainty propagation using Monte Carlo simulation are presented in subsection 4.5.

4.1. Generic model framework

Using steps 1 and 2 of the development process described in subsection 3.7, a generic model framework is developed that achieves sufficient accuracy with reasonable computational load for all 13 use cases. Various design choices have been considered, of which dynamical model and integrator configurations are tested extensively. This section presents the corresponding findings. The generic model framework with the best overall performance is shown in Table 4. The corresponding in-sample and out-of-sample (total) propagation residuals are shown in the top plots of Figure 5 and Figure 6. To allow comparison between use cases, all residuals are plotted with respect to the end of the estimation window (t_E), also called the in-sample (IS) to out-of-sample (OOS) split. Thus, negative time is in-sample and positive time is out-of-sample.

When analyzing the generic model framework, one finds that a relatively simple framework can already achieve good accuracy with reasonable computational load for all use cases. As can be seen in Table 5, PSA ranges from 50 days to 2 years. The found dynamical model configuration is similar to the one used for orbit determination and propagation of the IBEX [Carrico, Policastri, and Loucks, 2009][Carrico, Policastri, and Lutz, 2018] and Spektr-R [Zakhvatkin et al., 2020][Zhamkov et al., 2016] missions, both having orbited in cislunar space. The main difference with the IBEX and Spektr-R dynamical model is that, for these missions aerodynamic drag is included to address close Earth approaches. This effect is not found significant for the Chang'e 2 and Chang'e 3 boosters. For IBEX, the researchers found a comparable RKF7(8) integrator to be effective as well. These findings further confirm the validity of the generic model framework.

Table 4: Generic model framework

Dynamical models	Spherical harmonics: - Earth (SH[5,5])
	Point-mass: - Sun, Moon, Jupiter
	SRP (cannonball) - C_r = 1.2 - A = 37.14 m^2 - Shadow Earth
Integrator	RKF7(8) variable dt - Error tolerance = 10^{-10} - Initial step-size = 1.0 s - Max step-size = 10000 s
Propagator	Cowell
Ephemeris models	Spice kernels (Tudat default)
Rotation models	Spice kernels (Tudat default)
Estimated param.	Initial state

Analyzing the total residual plots in Figure 5 and Figure 6 and Periods of Sufficient Accuracy (PSA's) in Table 5, it is found that the generic model framework works surprisingly well for the estimation windows of the Chang'e 3 booster that are unstable out-of-sample (windows 5-6). On the other hand, performance for unstable windows of the Chang'e 2 booster (windows 5-7) is relatively poor. As discussed in subsection 2.2, these Chang'e 2 windows include close Moon approaches in-sample and out-of-sample. Especially for the windows that encounter the Moon out-of-sample (windows 5-6), it can be seen in Figure 5 that this causes rapid increases in residual out-of-sample. This effect and possible solutions will be discussed in greater detail in subsection 4.4.

Because the generic model framework is meant as a general framework for cislunar space debris, it is also limited in performance. For example, it only estimates initial state for each specific case. Furthermore, the observations are not weighted for each use case specifically and no a priori covariance matrix is applied. Estimating parameters like the radiation pressure coefficient (C_r) of the object and observation bias (ϵ_{obs}) per ground station in orbit determination or adding observation weights, will cause the model framework to be more complex but can also increase accuracy. Overfitting issues can be mitigated using the a priori covariance matrix. Since each use case reacts differently to the mentioned additions, model tailoring using these components has to be performed for each use case individually.

4.2. Tailored model framework

As explained in step 3 of the development process described in subsection 3.7, the generic model framework should be tailored (if possible) on individual use cases by estimating C_r or ϵ_{obs}, including an a priori covariance matrix, or adding realistic observation weights (equal to the random noise). Observation bias (ϵ_{obs}) is estimated in terms of both right ascension and declination for each ground station individually. No single tailored model configuration achieves high performance for all estimation windows, therefore different tailoring settings are applied on each use case.

It should be noted that for 8 out of 13 use cases, significant performance improvements have been achieved using model tailoring. For the other 5, the generic model framework still performed best. Each estimation window reacts differently to model tailoring. This is caused by differences in both orbit (e.g. stability) and observation characteristics (e.g. window length), which will also be analyzed in this section. The results of tailoring on each use case are presented below.

The tailored model framework configurations that result in the best performance for each individual use case, are shown in Table 6. The overall performance metrics of the tailored model framework per use case are presented in Table 7. Moreover, a comparison between performance of the generic versus the tailored model framework is made in Table 5, which clearly demonstrates the improvements realized from tailoring for 8 out of 13 estimation windows. The tables apply the coding system described in subsection 2.2 (e.g. 1-SSS), to identify the characteristics of each estimation window.

One important note, out-of-sample RMSE of total residual can not be simply compared between use cases. This is because the out-of-sample period and times of observation can vary between use cases. One can compare out-of-sample RMSE between estimation windows of an object ending on the same date, for example comparing estimation windows 1-4 of the Chang'e 2 upper stage is possible. PSA can be compared between use cases. Though one should keep in mind that the broader the PSA range, the harder it is to approximate the actual PSA of that estimation window realistically.

A more straightforward method of performance comparison between estimation windows is analyzing the total propagation residuals over time. Total residuals of the tailored model framework are compared between use cases in the bottom plots of Figure 5 and Figure 6. These plots are also used to compare the generic (top) to the tailored model framework (bottom), allowing for visual analysis of the improvements realized by tailoring.

Table 5: Generic (Gen.) vs tailored model performance

		Gen.	**Tailored**		
Obj.	*Window*	*PSA [d]*	*PSA [d]*	*ΔRMSE IS*	*ΔRMSE OOS*
CH2	1-SSS	145-228	145-228	-83,4%	-67,9%
	2-MSS	145-228	228-231	-75,6%	-50,2%
	3-LSS	369-418	369-418	0,0%	0,0%
	4-LSS	369-418	369-418	0,0%	0,0%
	5-SSU	49-163	49-163	-24,8%	-26,8%
	6-MSU	47-49	49-163	-21,4%	-35,8%
	7-MUS	90-91	182-201	-31,1%	-63,5%
CH3	1-SSS	>672	>672	0,0%	0,0%
	2-MSS	248-295	457-536	-71,6%	-21,1%
	3-LSS	434-436	>672	-76,7%	-94,8%
	4-LSS	434-436	>672	-78,3%	-93,9%
	5-SSU	676-678	676-678	0,0%	0,0%
	6-SUU	501-609	501-609	0,0%	0,0%

Comparing the generic to tailored performance metrics in Table 5, it is seen that especially use cases with relatively low PSA in the generic model framework can improve a lot from tailoring. Moreover, it is found that not only out-of-sample performance improves but also the in-sample residuals are greatly reduced (also seen in Figure 5 and Figure 6). This indicates a better and more realistic fit on the observations. Additionally, it has been found that the in-sample residual Probability Density Functions (PDFs) are closer to a Gaussian distribution after tailoring, which further justifies the finding that tailoring tends to improve the fit (as explained in subsection 3.3). The PDFs of the tailored model framework can be found in Figure B.2 and Figure B.3 of the appendix. The improved propagation accuracy and better fit are why the tailored model framework will be used for the analyses described in subsection 4.3, subsection 4.4 and subsection 4.5.

Table 6: Best performing tailored model framework configurations for all use cases

Object	**Window**	**Parameters**	**Obs. weight**	**A priori cov.**	**A priori input**
Chang'e 2	1-SSS	Initial state, $C_r + \epsilon_{obs}$	1.0 arcsec	Yes	$\sigma_{C_r} = 0.1$, $\sigma_{\epsilon_{obs}} =$ 1E-5
	2-MSS	Initial state, $C_r + \epsilon_{obs}$	1.5 arcsec	Yes	$\sigma_{\epsilon_{obs}} =$ 1E-5
	3-LSS	Initial state	1.0 arcsec	No	-
	4-LSS	Initial state	1.0 arcsec	No	-
	5-SSU	Initial state, C_r	1.0 arcsec	No	-
	6-MSU	Initial state, C_r	1.0 arcsec	No	-
	7-MUS	Initial state, $C_r + \epsilon_{obs}$	1.0 arcsec	Yes	$\sigma_{C_r} = 0.1$, $\sigma_{\epsilon_{obs}} =$ 1E-6
Chang'e 3	1-SSS	Initial state	1.0 arcsec	No	-
	2-MSS	Initial state, $C_r + \epsilon_{obs}$	1.0 arcsec	Yes	$\sigma_{C_r} = 0.1$, $\sigma_{\epsilon_{obs}} =$ 1E-6
	3-LSS	Initial state, $C_r + \epsilon_{obs}$	1.0 arcsec	No	-
	4-LSS	Initial state, $C_r + \epsilon_{obs}$	1.0 arcsec	No	-
	5-SSU	Initial state	1.0 arcsec	No	-
	6-SUU	Initial state	1.0 arcsec	No	-

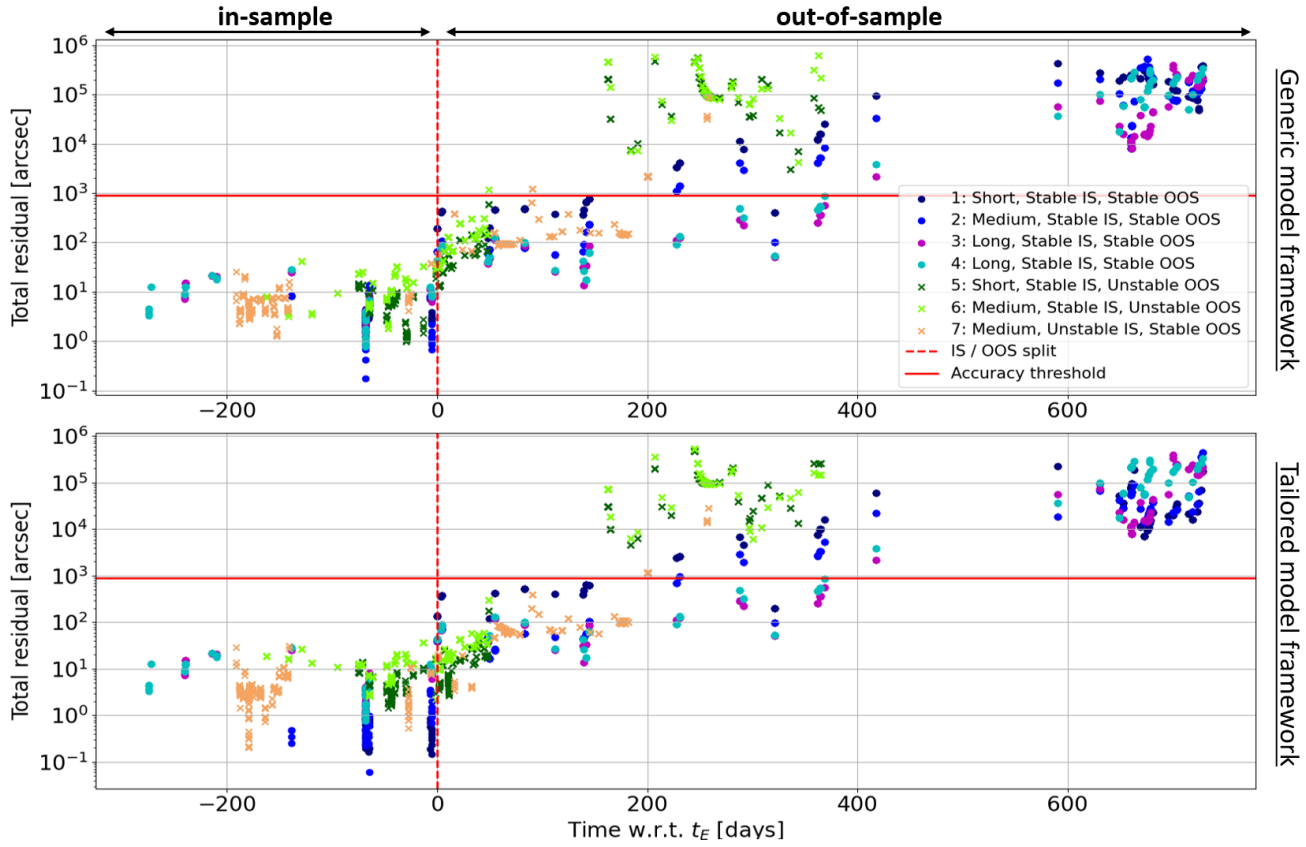


Figure 5: In-sample and out-of-sample residuals comparison of generic (top) and tailored model framework (bottom) for Chang'e 2 upper stage

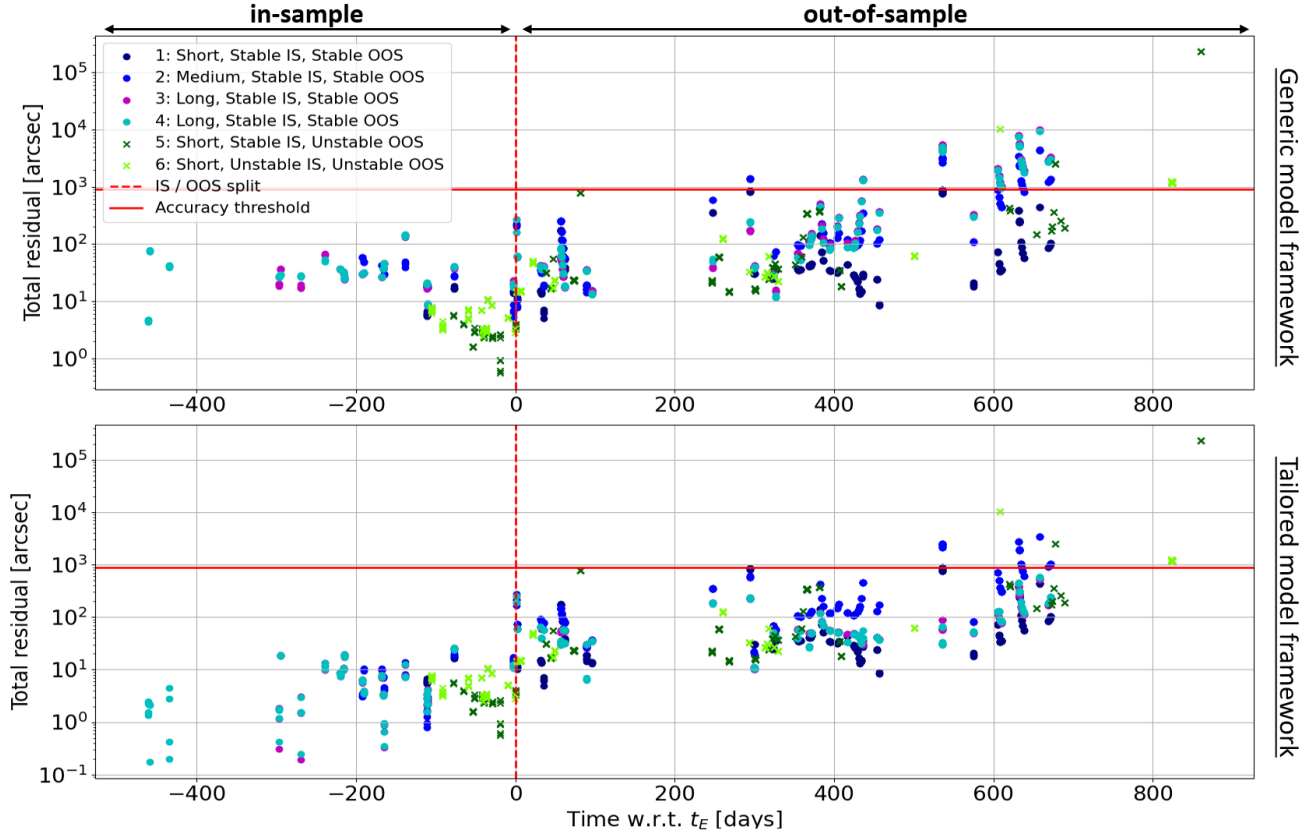


Figure 6: In-sample and out-of-sample residuals comparison of generic (top) and tailored model framework (bottom) for Chang'e 3 upper stage

Table 7: Tailored model framework performance (*do not compare RMSE OOS between estimation windows ending on different dates)

Object	Window	Comp. time [s]	RMSE IS [arcsec]	RMSE OOS* [arcsec]	PSA [days]	C_r [-]
Chang'e 2	1-SSS	10,7	0,6	53432,8	145-228	1,80
	2-MSS	14,9	1,4	72011,7	228-231	1,58
	3-LSS	15,7	11,8	95774,9	369-418	1,20
	4-LSS	18,4	11,7	134097,0	369-418	1,20
	5-SSU	7,8	6,0	124351,4	49-163	1,55
	6-MSU	13,6	13,4	136985,1	49-163	1,67
	7-MUS	24,1	6,2	3565,2	182-201	1,53
Chang'e 3	1-SSS	9,5	11,5	242,9	>672	1,20
	2-MSS	18,8	10,1	910,6	457-536	1,62
	3-LSS	17,3	10,7	126,7	>672	1,50
	4-LSS	33,0	10,0	138,8	>672	1,50
	5-SSU	12,0	2,8	32149,8	676-678	1,20
	6-SUU	13,1	6,1	1699,7	501-609	1,20

Interesting findings are made from the tailoring process itself. Table 8 shows the effect of C_r and/or ϵ_{obs} estimation on performance w.r.t. the generic model framework (which only estimates initial state). No a priori covariance is applied yet in these tests. The varying effects of C_r and/or ϵ_{obs} estimation, shows the need for individual tailoring on use cases and are analyzed in detail below.

Firstly, it is found from Table 8 that adding C_r and/or ϵ_{obs} to the estimated parameter set, easily leads to overfitting. Adding an a priori covariance matrix is in some cases able to solve this, clearly shown for several windows of the Chang'e 2 (windows 1, 2 & 7) and Chang'e 3 booster (window 2) in Table 5 and Table 6. Secondly, adding only ϵ_{obs} to the estimation never leads to better out-of-sample performance than estimating C_r, or C_r and ϵ_{obs}. This is probably because the cannonball model induces a significant use case specific error in the state, which is partly solved by C_r estimation. Thirdly, estimating C_r on short windows (<4 months) can lead to unrealistic values like 3.08, 14.19 or -0.89 (see window 1), which results in poor out-of-sample performance and is therefore undesirable. Adding an a priori covariance matrix does not fix this, as the unmodeled variance in the observations seems to pull harder on the parameter increment per Least-Squares iteration than any a priori covariance matrix can. Since solar radiation pressure has a compounding effect on the orbit [Casanova et al., 2015], less than 4 months of observations is likely too short to find a reliable estimate of C_r.

For use cases in which C_r estimation does lead to better out-of-sample performance, C_r values between 1.5 and 1.8 are estimated for both the Chang'e 2 and 3 booster (shown in Table 7). But for other windows, the original C_r guess of 1.2 results in the best out-of-sample performance. This further justifies that model tailoring should be done on each use case individually, as it is significantly influenced by orbit and observation characteristics of the specific case.

Additionally, Table 9 presents the estimated right ascension and declination observation biases (ϵ_α and ϵ_δ) per ground station (GS). It is found that the estimation does not converge to one bias for each ground station. Furthermore, the biases tend to be much higher than expected from literature [Vereš et al., 2017], exceeding the MPC constraint of 1 arcsecond (see subsection 3.4). This is likely because the estimated bias includes currently unmodeled dynamics in its magnitude, which can be justified from analyzing in-sample right ascension and declination residuals per ground station. An example analysis for window 3 of the Chang'e 3 booster can be found in Appendix B. It should be noted that the bias in an observation set for one ground station does also vary based on the location of the tracked object (due to the time bias). Estimating a bias for each streak of observations (instead of per ground station) should solve this. More realistic observation bias estimation could be achieved by adding empirical accelerations, which would capture unmodeled effects [Montenbruck et al., 2000] (as explained in subsection 3.1). To summarize, one can use ϵ_{obs} estimation to improve out-of-sample performance, but should not draw conclusions on its magnitude for a specific ground station.

Table 9: Estimated observation bias in tailored model framework

Obj.	Window						
CH2	1-SSS	GS	703	C94	G68	K95	Q65
		ϵ_α [arcsec]	-12,2	-7,6	8,4	9,8	-9,8
		ϵ_δ [arcsec]	-6,7	-7,6	10,4	14,5	-1,1
	2-MSS	GS	703	C94	G68	K95	Q65
		ϵ_α [arcsec]	2,2	2,3	-7,4	-9,0	0,1
		ϵ_δ [arcsec]	0,0	-1,4	9,2	14,3	5,2
	7-MUS	GS	703	C95	G96		
		ϵ_α [arcsec]	13,8	15,8	17,7		
		ϵ_δ [arcsec]	-0,5	-4,3	2,3		
CH3	2-MSS	GS	T05	T08			
		ϵ_α [arcsec]	-9,2	-11,1			
		ϵ_δ [arcsec]	-1,4	-1,5			
	3-LSS	GS	I41	T05	T08		
		ϵ_α [arcsec]	-34,6	-29,1	-28,6		
		ϵ_δ [arcsec]	-13,9	-2,1	-1,6		
	4-LSS	GS	703	I41	K88	T05	T08
		ϵ_α [arcsec]	-29,4	-34,3	-14,7	-29,1	-28,6
		ϵ_δ [arcsec]	-2,5	-14,1	9,8	-2,1	-1,6

To ensure a good fit, the in-sample residual distributions should be sufficiently Gaussian for all use cases. It has been found that multiple in-sample residual PDFs are not sufficiently Gaussian (even after tailoring), which can be seen in Figure B.2 and Figure B.3 of the appendix.

Table 8: Effect of various parameter estimations on RMSE w.r.t. generic model framework for all windows, and corresponding C_r estimation

		Initial state & ϵ_{obs}		**Initial state & C_r**			**Initial state, C_r & ϵ_{obs}**		
Obj.	*Window*	*ΔRMSE IS*	*ΔRMSE OOS*	*C_r [-]*	*ΔRMSE IS*	*Δ RMSE OOS*	*C_r [-]*	*ΔRMSE IS*	*ΔRMSE OOS*
CH2	1-SSS	-86,5%	31,3%	3,08	-36,4%	28,3%	14,19	-86,6%	30,1%
	2-MSS	-91,0%	-10,4%	1,90	-14,4%	29,7%	0,71	-91,3%	7,6%
	3-LSS	-62,6%	46,4%	1,88	-24,2%	55,3%	1,17	-62,6%	70,9%
	4-LSS	-64,1%	17,0%	1,57	-18,9%	-10,8%	1,18	-64,1%	-2,9%
	5-SSU	-22,6%	-23,9%	1,55	-24,8%	-26,8%	1,42	-26,8%	-20,9%
	6-MSU	-42,8%	-10,0%	1,67	-21,4%	-35,8%	1,35	-45,4%	-28,1%
	7-MUS	-36,0%	-49,4%	1,64	-36,0%	-49,4%	1,68	-44,8%	-52,8%
CH3	1-SSS	-86,8%	4613,3%	-0,89	-8,8%	1519,6%	2,26	-87,5%	4665,0%
	2-MSS	-66,5%	64,3%	1,68	-68,4%	0,0%	1,48	-78,9%	-43,2%
	3-LSS	-19,7%	-33,9%	1,48	-60,3%	-79,2%	1,50	-76,7%	-94,8%
	4-LSS	-25,0%	-31,6%	1,48	-61,5%	-72,2%	1,50	-78,3%	-93,9%
	5-SSU	-48,0%	41,2%	2,86	-48,0%	41,2%	2,86	-48,0%	41,2%
	6-SUU	-86,0%	14861,6%	1,57	-14,5%	67,5%	1,38	-91,0%	1453,5%

Therefore, it becomes clear that the current tailored model framework leaves room for improvement. Firstly, empirical accelerations can be added to the dynamical model to capture any unmodeled effects [Montenbruck et al., 2000], allowing for more accurate observation bias estimation. Moreover, estimating overall observation bias per streak of observations (not per ground station) should allow for differentiation between the constant observation bias in the instrumentation and the varying time bias. Secondly, as is confirmed from C_r estimation, a large uncertainty is introduced by the cannonball approach. Especially because the objects are likely tumbling, violating the assumption described in subsection 3.3 that C_r is sufficiently constant. Optical evidence of the tumbling behaviour is shown in Figure 7, seen from the varying brightness of the observations. Using the panelled approach described in subsection 3.1, or estimating a variational C_r to incorporate tumbling of the objects, could allow for a more accurate representation of the (solar) radiation pressure over time. Though, it should be investigated whether this is worth the effort. Finally, multi-arc orbit estimation [Serra et al., 2018] should improve the estimated model parameters (e.g. C_r) for unstable periods, further discussed in subsection 4.4.

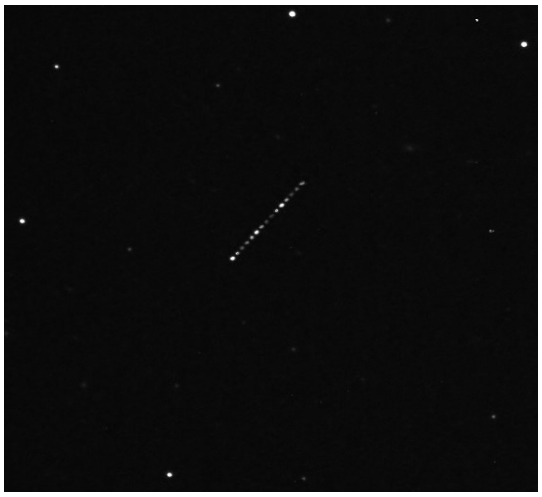


Figure 7: Observation stack containing 15 observations of the Chang'e 3 upper stage, made in 2015 [Langbroek, 2023]

4.3. Effect of orbit and estimation window characteristics on model performance

The tailored model framework has been developed for the Chang'e 2 and 3 upper stages. To make the findings of the framework applicable to cislunar space debris as a whole, it is important to analyze the effect that different orbit and estimation window characteristics have on performance. An overview of orbit and estimation window characteristics can be seen in Table 1. Window codes (e.g. 1-SSS) are used in tables to link results to use case characteristics.

The performance metrics of the tailored model framework for various use cases are shown in Table 7. The total propagation residual over time can be found in the bottom plots of Figure 5 and Figure 6. Analyzing the performance metrics, interesting behaviour is observed. Firstly, out-of-sample performance of the model framework is much worse for the orbit of the Chang'e 2 booster, than it is for the orbit of the Chang'e 3 booster. This confirms the initial findings, described in subsection 2.1, that the orbit of the Chang'e 2 booster is inherently more unstable or non-linear. Higher non-linearity is known to be more challenging for accurate parameter estimation using single-arc orbit determination [Serra et al., 2018][Lari et al., 2019].

Comparing Period of Sufficient Accuracy (PSA) of the 4 stable windows for each object (window 1-4), it is found that estimating on more than 6 months of observations consistently results in the best out-of-sample propagation accuracy. PSA defines how long an orbit propagation is sufficiently accurate, being a residual smaller than 900 arcseconds. Though, estimating on less than 4 months of data with only 18 observations can in some cases lead to high accuracy as well, shown by estimation window 1 of the Chang'e 3 upper stage. Stable estimation windows 1 and 2 of the Chang'e 2 upper stage have lower out-of-sample RMSE but also lower PSA than stable windows 3 and 4. This can be explained when investigating the out-of-sample residuals in Figure 5. Windows 1 and 2 cross the accuracy threshold much sooner, but over longer periods plateau at lower errors than windows 3 and 4. Comparing the long, stable windows for the Chang'e 2 and 3 booster (window 3-4), it is found that increasing

estimation window beyond 8 or 10 months of observations respectively, has no added benefit. This only increases computation time and even tends to make out-of-sample performance slightly worse. Which is likely caused by overconfidence of the estimation due to the (excessively) long windows, making the estimator blind to new orbital behaviour. Thus, it is found that estimating on <4 months of observations is feasible in this model framework but can be unreliable. Less than 4 months tends to be challenging because the estimator has a hard time finding a realistic fit of the orbital behaviour, due to the limited time span and observations. For orbits of cislunar space debris it is found that estimating on a 7-10 month window seems to lead to the most robust performance. This time span seems to be sufficiently large to capture all effects reliably. It is also sufficiently small to prevent overconfidence (or overfitting) and still is efficient.

The model framework works surprisingly well for estimating on observation sets that are unstable in-sample, or in other words experience non-linear behaviour. Analyzing Figure 1, the unstable in-sample windows of the Chang'e 2 (window 7) and Chang'e 3 booster (window 6) experience large semi-major axis fluctuations in-sample. Chang'e 2, window 7 also includes an inclination drop of 60 degrees due to a close Moon approach (seen in Figure A.1). Window 7 of the Chang'e 2 booster has similar residual over time behaviour as stable windows 1 and 2, which is seen in Figure 5. Window 6 of the Chang'e 3 booster reaches a PSA of nearly 2 years. This indicates that the model framework is able to stabilize for large in-sample orbital elements fluctuations, which is further justified from analyses of the in-sample residual for both windows (see bottom plots of Figure 5 and Figure 6). It is observed that the residual is actually relatively low in-sample for the unstable in-sample windows, and no major residual spikes are seen. Thus, the fit does not seem to suffer a lot from non-linearity in the orbit.

Investigating performance of the unstable out-of-sample windows of the Chang'e 2 booster (windows 5-6), one finds that the tailored model framework has a hard time estimating large out-of-sample orbital elements fluctuations (seen in Figure 1). The fluctuations are caused by close approaches to the Moon. A more in-depth analysis on this phenomenon will be performed in subsection 4.4.

On the other hand, the model framework achieves similar performance for unstable out-of-sample window 5 of the Chang'e 3 booster compared to stable out-of-sample windows 1-4. This proves that, though the model framework has issues with propagating over close Moon approaches, it does have the ability to accurately propagate slightly less non-linear orbits. The chaotic trajectory out-of-sample of window 5 is more clearly visualized in the three-dimensional orbit propagation, found in Figure A.4. Investigating the out-of-sample residuals in Figure 6, one does find that after nearly 2 years the residual of window 5 increases sharply. The large increase in residual can be explained by taking a closer look at the Keplerian elements in Figure A.2. After approximately 2 years, the Chang'e 3 upper stage encounters a close approach to the Moon which seems to induce large orbital elements fluctuations. After a while the orbit stabilizes in a more stable and predictable movement around the Earth. Similar to the unstable windows of Chang'e 2 (windows 5-6), the model framework is currently not able to propagate over close Moon approaches accurately.

Finally, a large challenge for estimating on observations of cislunar space debris, especially in unstable orbits, is the lack of and necessity for a good initial state guess. Currently, an initial state guess is made from ephemerides retrieved from Find_Orb orbit estimations. Find_Orb, similar to Tudat, also has challenges with orbit estimation for unstable orbits and the corresponding ephemerides are thus not always sufficiently accurate. In the process of finding a suitable selection of unstable use cases (described in subsection 2.2), it is found that it requires much tweaking of the starting date (and corresponding Find_Orb initial state guess) to find estimation windows leading to converging orbit estimations. The issue occurred especially for unstable windows 5 and 6 of the Chang'e 3 booster. This is less apparent for stable orbits, as initial state guesses tend to be more accurate. Stable orbit estimations are also more forgiving for errors in the initial state guess.

4.4. Effect of close approach to the Moon

As explained in subsection 2.2, several estimation windows include close Moon approaches in-sample or out-of-sample. In-sample Moon encounters have been shown in subsection 4.3 to be balanced out during orbit determination and do not necessarily result in poor out-of-sample propagation accuracy. But the effect of out-of-sample close Moon encounters is still a challenge. For both windows 5 and 6 of the Chang'e 2 booster and window 5 of the Chang'e 3 booster, the Moon encounter seems to result in a sharp residual increase (also seen in Figure 5 and Figure 6) and thus can cause observers to lose the object from their sight. Using estimation window 6 of the Chang'e 2 booster as a representative case study, investigation will be performed on this effect and solutions are proposed.

Starting with the residual itself, plotted against distance to the Moon in Figure 8. It can be seen that the Chang'e 2 booster has two close Moon encounters at a distance of approximately 20.000 km on the 19th of April and 12th of June 2019. The in-sample encounter (19th of April) is balanced out by the orbit determination algorithm. But the out-of-sample close approach (12th of June) induces an immediate increase in residual after which the object is lost to observers for quite some time. After 4 months, new observations are made of the object but the residual is now 100 times larger w.r.t. the original orbit. The hypothesis for this exponential increase in residual is that, any errors in the orbit propagation at time of close encounter are enhanced significantly by the large velocity and acceleration increase caused by the gravitational force of the Moon.

This hypothesis is further justified when analyzing the effect that the out-of-sample Moon encounter has on the position difference between 200 Monte Carlo (MC) propagations and the tailored model propagation, shown in Figure 9. The simulation of the 200 MC samples and MC propagations will be elaborated upon in subsection 4.5. It is observed that the position difference for nearly all

MC propagations increases 10 to 1000 times due to the encounter, though their position difference before the close Moon approach is relatively small. Only one MC propagation stays stays near the original propagation. Similar findings have been made in studies on close encounters of JUICE with Jupiter's Moons [Lari et al., 2019]. The encounters greatly enhance the errors present in the propagated orbit, making it hard to perform accurate out-of-sample orbit propagation using the current single-arc estimation approach.

Dynamical model testing for generic model development has shown that using a high order and degree spherical harmonics model instead of a point-mass model for the Moon's gravity, does not improve out-of-sample RMSE for windows 5 and 6 of the Chang'e 2 booster. So that would not solve the sharp residual increase caused by the Moon encounter. One could try to reduce the residual before the close approach by applying the model improvements described at the end of subsection 4.2, but this is not expected to fully solve the error enhancement issue.

A better solution would be to perform multi-arc orbit estimation [Serra et al., 2018] in periods where close Moon approaches are expected. This has been shown effective for orbit determination of chaotic orbits around Jupiter's Moons [Lari et al., 2019]. One would perform separate orbit estimations for the more stable orbit leading up to the close Moon approach, for the highly unstable orbit during the close Moon approach, and for the (stable) orbit after the encounter. This way initial state and model parameters can be estimated more accurately for the behaviour in each arc. Unfortunately, observations following a close Moon approach tend to be scarce because the object is often lost to observers, as can also be seen for window 6 of the Chang'e 2 booster in Figure 8. This makes accurate orbit determination challenging for the short chaotic arc around the encounter.

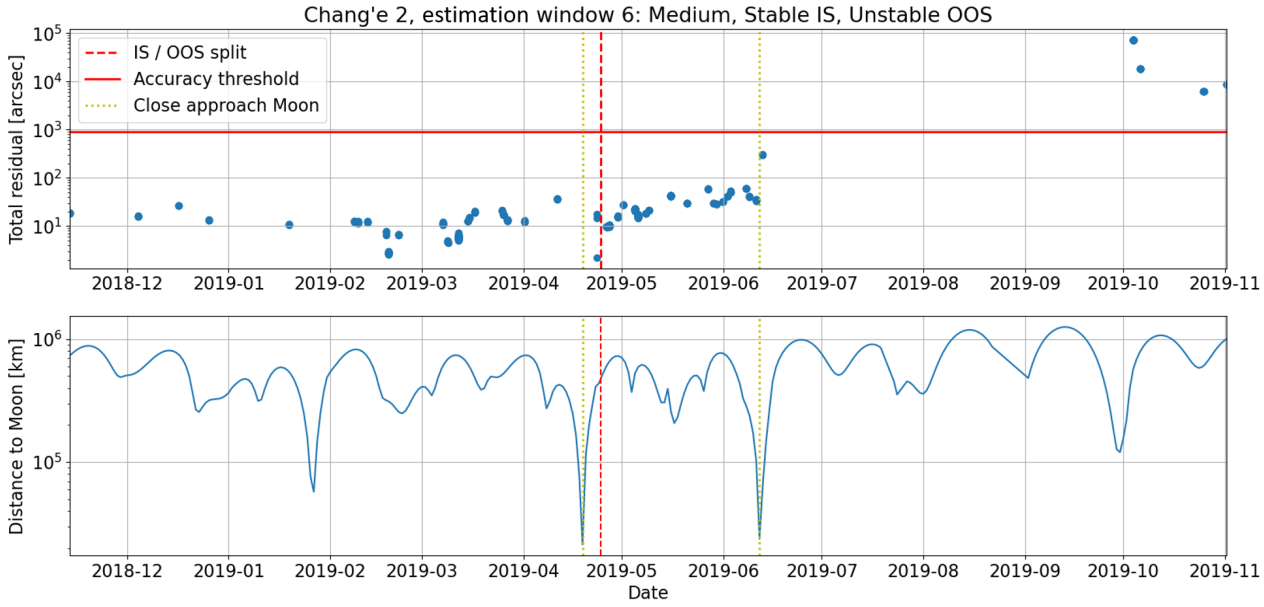


Figure 8: Effect of close approach Moon on residuals of Chang'e 2, estimation window 6

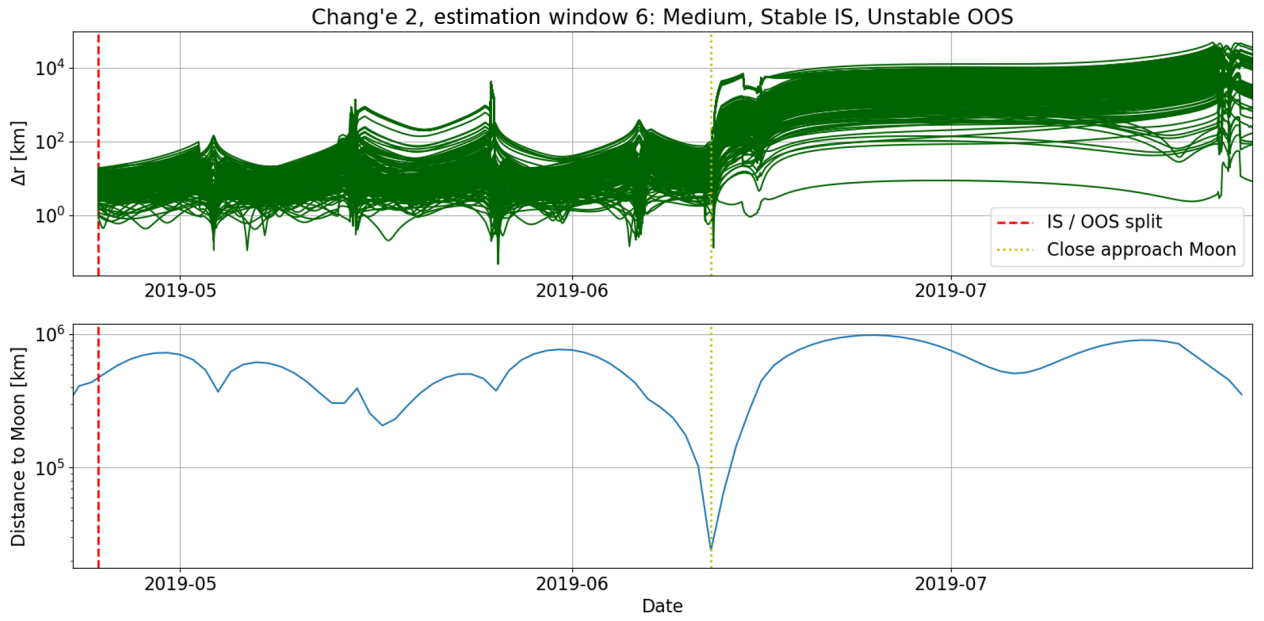


Figure 9: Effect of close approach Moon on position difference in MC propagation of Chang'e 2, estimation window 6

Observation scarcity following a Moon encounter can be tackled on the observer's side. Knowledge on when a close Moon approach is expected, based on orbit propagation, can be passed on to observers. This information allows the observers to proactively focus their instrumentation on the object for a sufficient period before, during and especially after the close encounter. This would ensure plenty coverage of the object, both preventing the object from getting lost to observers and allowing for accurate multi-arc orbit estimation on this highly unstable arc. Though, it should be considered that the Moon can cause interference in optical observations during close encounters.

4.5. Uncertainty propagation

As described in subsection 3.5, epistemic and aleatory uncertainties are present in the cislunar space debris problem. The effect of epistemic uncertainty, caused by model simplifications, has been minimized during development of the generic model framework using extensive testing and comparisons of various model configurations. Thus, there is mainly aleatory uncertainty or imperfect knowledge uncertainty left in the estimated state. The effect over time of this uncertainty in the estimated parameter set is analyzed using Monte Carlo (MC) simulation. Aleatory uncertainty is expected to be caused mainly by uncertainty in the radiation pressure model and uncertainty introduced by observation bias. To analyze its effect over time, 200 MC propagations are performed per use case.

As explained in subsection 3.5, a mean and covariance of the parameters at t_E is required for generating 200 MC samples at that time. The true covariance matrix is calculated by propagating the estimated covariance at t_0 to t_E, and applying Equation 2.17. Investigating k-ratio over time, it is found that averaging the k-ratio in the first 3 months out-of-sample using Equation 2.16 is most suitable. This period is chosen because all use cases are still sufficiently accurate and the k-ratio is relatively constant, indicating linearity. An overview of the average k-ratios can be found in Table 10.

Table 10: Average k-ratio, 3 months out-of-sample

Object	Window	k-ratio (avg.) [-]
Chang'e 2	1-SSS	10,3
	2-MSS	1,7
	3-LSS	54,4
	4-LSS	64,5
	5-SSU	10,4
	6-MSU	14,5
	7-MUS	98,0
Chang'e 3	1-SSS	26,2
	2-MSS	94,2
	3-LSS	101,3
	4-LSS	93,3
	5-SSU	7,2
	6-SUU	7,9

As explained in subsection 3.5, a large true-to-formal error ratio is caused by significant imperfections in the dynamical model or unmodeled observation noise. Comparing the k-ratios to the tailored model configurations found in Table 6, shows that estimating C_r and ϵ_{obs} does not consistently result in lower true-to-formal error ratio. This is likely caused by parameter errors due to the cannonball simplification as well as unmodeled effects being estimated in the observation bias. Applying the model recommendations made at the end of subsection 4.2, formal error is expected to become more realistic and k-ratio should become smaller.

Before analyzing the uncertainty propagation results, first the linear uncertainty propagation assumptions are analyzed in-sample and out-of-sample. It is also vital to test normality of the initial state sample, which is simulated using a multivariate Gaussian Random Number generator. Since the sample size of 200 is limited, one needs to be careful if this randomly generated normal distribution is actually Gaussian. If the initial distribution is non-Gaussian, then the overall MC method is not valid. Normality is tested using the Kolmogorov-Smirnov (KS) test, which tests the null hypothesis (H0) that a distribution is Gaussian. Results of the KS tests on x, y and z-distributions are shown in Table 11. The closer the p-value is to 1, the higher the likelihood the distribution is Gaussian. In case all initial state sample distributions are Gaussian, it is found that the moment of H0 rejection (distribution is considered non-Gaussian) differs significantly between use cases. From this moment uncertainty propagation can no longer be assumed linear. As expected, uncertainty can be linearly propagated for much longer for stable windows (1-4) than for the unstable windows (5,6,7). For the stable Chang'e 2 windows, linear uncertainty propagation holds for 162-262 days. For the stable Chang'e 3 windows, it holds for 55-110 days. For the unstable windows, uncertainty can be linearly propagated for only 1-3 weeks after which the distribution becomes non-Gaussian. Since the k-ratio approach assumes linear uncertainty propagation in the first three months, this indicates that the true covariance and therefore the sampling might not be an accurate representation of the actual initial uncertainty distribution for unstable windows.

One important note. During KS testing, it has been found that the test is sensitive to the mean given to the theoretical Gaussian distribution. Theoretically, the MC propagations should have a zero mean for them to be Gaussian and thus be perfectly distributed around the tailored propagation. But since the sample size of 200 is still relatively small, the mean is never really zero. The KS test easily concludes distributions with non-zero mean to be non-Gaussian [Drezner et al., 2008], even occurring sometimes for normally generated initial sample distributions. Therefore, the mean of the sample distribution itself is used in the KS test and it has been verified that this mean remains sufficiently zero by ensuring that it is smaller than the standard error of the mean over time ($\sigma_\mu = \frac{\sigma}{\sqrt{N}}$) [Lee et al., 2015]. The standard error of the mean is a measure for the difference between the sample mean and the mean of the actual distribution. Even when applying the sample mean in the KS test, it can be seen in Table 11 that the initial distributions are almost never fully Gaussian (p=1). Recommendations will be made at the end of this section.

Now that it is known for how long linear uncertainty propagation is valid out-of-sample, it is important to verify whether the propagated covariance used for MC sample generation at t_E is correct (assuming a sufficiently accurate dynamical model). To verify the covariance propagation step, MC propagations from t_E are compared to MC propagations from t_0. For nearly all use cases, the out-of-sample residual distribution from t_E and from t_0 is highly similar. This indicates that the covariance propagation step is valid. The one exception is the window of the Chang'e 2 booster that is unstable in-sample (window 7), of which the comparison is shown in Figure 10 (t_E is equal to the IS / OOS split). It can be seen that the residual distributions out-of-sample when propagated from t_0 are nearly twice as wide compared to when propagated from t_E. This indicates that the propagated true covariance at t_E underestimates the uncertainty distribution at that time. This is caused by a close Moon approach in-sample of window 7 resulting in highly non-linear behaviour and thus causing errors during covariance propagation. Additionally, it has been found that the propagated covariance matrix at t_E has become slightly asymmetric, while it should always remain symmetric, indicating numerical errors during covariance propagation due to non-linearity.

Finally, the effect of uncertainty over time is analyzed. Since stable windows 1-4 of both objects allow sufficient linear uncertainty propagation out-of-sample required for the average k-ratio approach and the initial covariance is therefore likely more realistic, the uncertainty distributions over time are analyzed between them. A comparison of the uncertainty distribution in total residual over time is shown in Figure 11, the first time that the x, y or z-distributions of the 200 samples become non-Gaussian is also plotted. It is seen that uncertainty of the propagated orbit over time for the Chang'e 2 booster is much larger than for the Chang'e 3 booster. After 8-9 months the uncertainty distribution of the Chang'e 2 booster can be up to 10x wider. This can be explained by the inherently higher instability of the Chang'e 2 booster orbit, described in subsection 2.1. Moreover, uncertainty over time is consistently highest for short windows (window 1). Further justifying the findings in subsection 4.3 that estimating on less than 4 months of data is less robust and can result in a poor fit. This is also confirmed when considering that C_r and ϵ_{obs} are estimated for the tailored model configuration of window 1 (2.5 months) and window 2 (5 months) of the Chang'e 2 booster. Both windows experience similarly stable orbital behaviour, but uncertainty is significantly higher for window 1. This proves the findings that reliably estimating C_r for short windows is challenging and uncertainty remains high. Lastly, it is observed that though PSA for some use cases may be nearly 2 years (e.g. window 1-4 of Chang'e 3), a large tail of the uncertainty distribution can quickly start violating the accuracy threshold. For these use cases this already starts happening from 8 months onwards. Since the Monte Carlo approach for these windows is found to be realistic, this proves that the effect of uncertainty in orbit propagation of cislunar space debris is large and should be taken into account for predicting when an object might lose our sight. Though, it is expected that the model recommendations, made at the end of subsection 4.2, should reduce the uncertainty over time.

A couple of recommendations are made for the current uncertainty propagation approach. Firstly, to get more realistic KS test results and uncertainty distributions, sample size needs to be increased (N=1000-10000). Due to equipment and time constraints, this has not been performed yet. Core splitting could be a good method to realize this. Furthermore, it has been found that using 200 samples is already time consuming (~6 hours to run all use cases). One could try using more efficient non-linear uncertainty propagators like the Polynomial Chaos Expansion (PCE), which has already been applied in cislunar context [Giordano et al., 2023]. Lastly, instead of approximating a covariance matrix at t_E (using the k ratio), generating 200 MC samples and propagating them, one could also use bootstrapping to perform uncertainty propagation [Desmars, J. et al., 2009]. Bootstrapping resamples the observations N times, performing new orbit determination on each resampled observation set, and propagating an uncertainty distribution over time. The advantage of bootstrapping is that errors can be non-Gaussian.

Table 11: Kolmogorov-Smirnov test results for initial distribution (x_0, y_0 and z_0), and first moment of null hypothesis rejection (H0 rej.) indicating non-Gaussian x, y and z-distributions

Object	Window	x_0 distribution	y_0 distribution	z_0 distribution	x - H0 rej.	y - H0 rej.	z - H0 rej.
Chang'e 2	1-SSS	Gaus. | p=0,75	Gaus. | p=0,98	Gaus. | p=1	162,1 days	239,3 days	242,6 days
	2-MSS	Gaus. | p=0,76	Gaus. | p=0,91	Gaus. | p=0,6	261,8 days	255,5 days	257,8 days
	3-LSS	Gaus. | p=0,54	Gaus. | p=0,87	Gaus. | p=0,7	233,3 days	239,3 days	257,5 days
	4-LSS	Gaus. | p=0,77	Gaus. | p=0,91	Gaus. | p=0,72	233,4 days	239,4 days	257,7 days
	5-SSU	Gaus. | p=0,82	Gaus. | p=0,85	Gaus. | p=0,61	8,9 days	8,9 days	8,9 days
	6-MSU	Gaus. | p=0,8	Gaus. | p=0,74	Gaus. | p=0,72	8,7 days	8,9 days	8,9 days
	7-MUS	Gaus. | p=0,34	Gaus. | p=0,18	Gaus. | p=0,15	15,4 days	15,4 days	15,4 days
Chang'e 3	1-SSS	Gaus. | p=0,95	Gaus. | p=0,85	Gaus. | p=0,97	56,6 days	56,6 days	56,9 days
	2-MSS	Gaus. | p=0,89	Gaus. | p=0,53	Gaus. | p=0,78	84,1 days	84,1 days	84,1 days
	3-LSS	Gaus. | p=0,99	Gaus. | p=0,83	Gaus. | p=0,94	108,5 days	108,5 days	108,5 days
	4-LSS	Gaus. | p=0,54	Gaus. | p=0,51	Gaus. | p=0,98	56,4 days	56,4 days	56,9 days
	5-SSU	Gaus. | p=0,88	Gaus. | p=0,99	Gaus. | p=0,87	19,6 days	5,3 days	5,3 days
	6-SUU	Gaus. | p=0,89	Gaus. | p=0,9	Gaus. | p=0,95	4,1 days	4,2 days	4,2 days

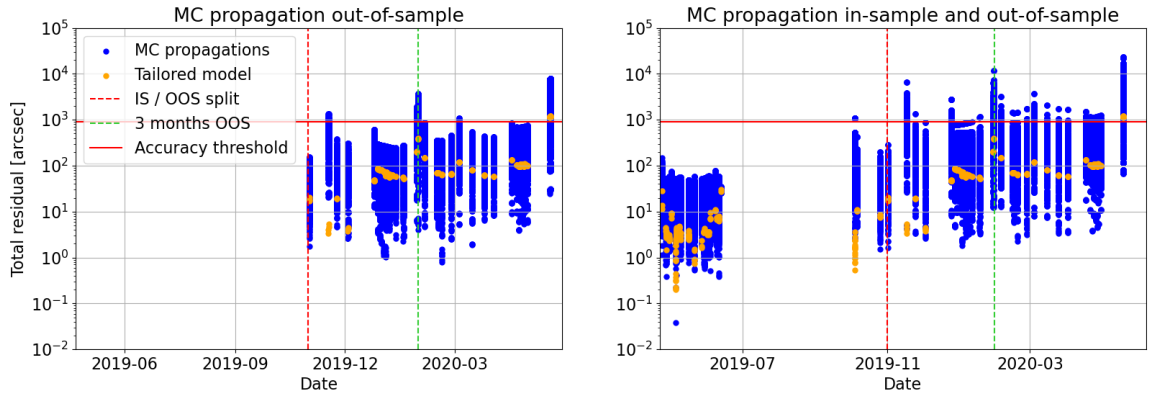


Figure 10: Uncertainty distribution in total residual over time propagated from t_E vs. from t_0 for window 7 of the Chang'e 2 booster

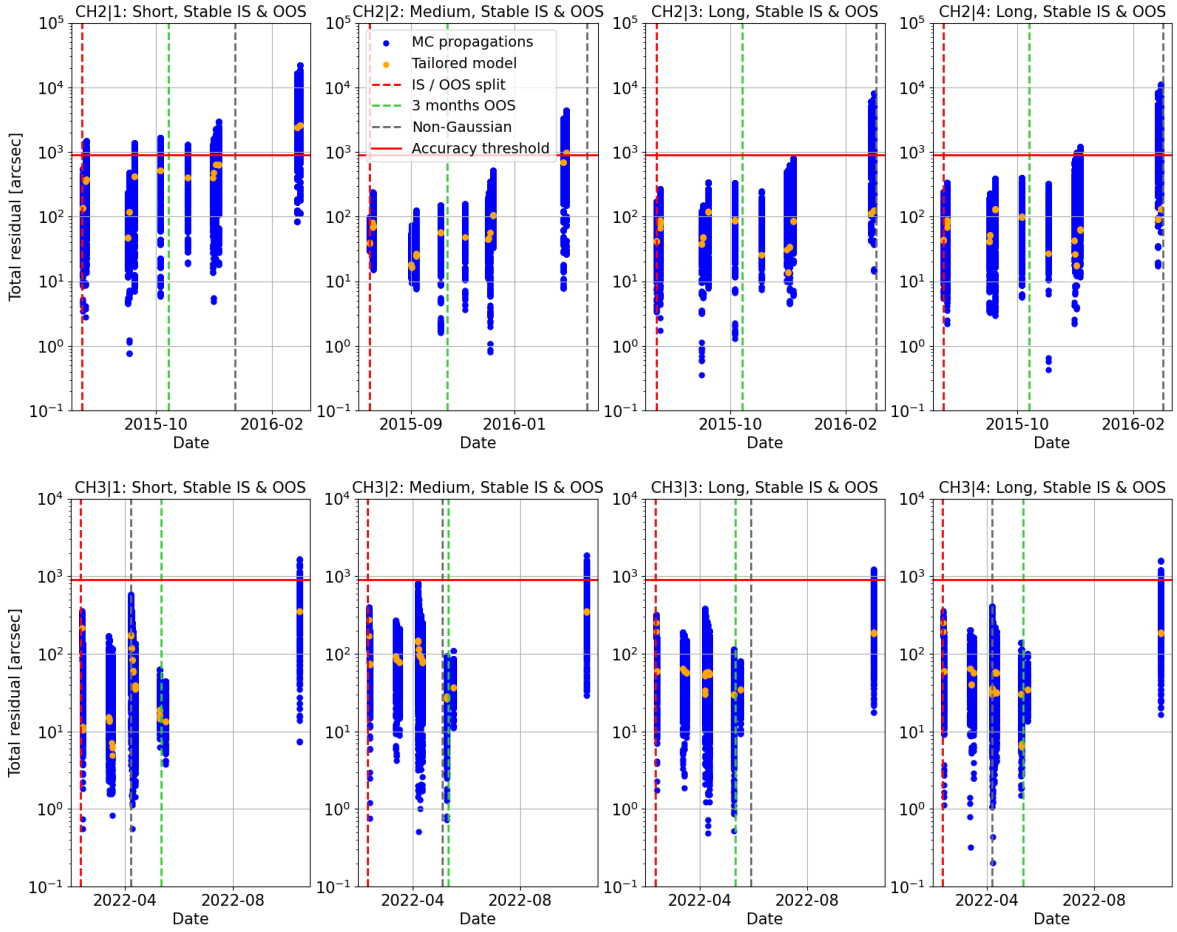


Figure 11: Uncertainty distribution in total residual over time for windows 1-4 of the Chang'e 2 (top) and Chang'e 3 booster (bottom)

5. Conclusions & recommendations

This works describes the development of a generic model framework for high accuracy orbit determination and propagation of cislunar space debris, and additional model tailoring on individual use cases. The model framework is developed utilizing the open-source Tudat Python package and applied on 13 estimation windows of the Chang'e 2 and 3 upper stage orbits. The source code of this research has been made available[14]. A new performance metric is introduced called Period of Sufficient Accuracy (PSA). PSA indicates the time that an orbit can be propagated with sufficient accuracy, or below a residual threshold of 900 arcseconds. Analysis on the effect of uncertainty over time for cislunar space debris has been performed using Monte Carlo simulation. This framework can be used as a tool for cislunar space debris tracking and collision prevention.

The goal of generic model framework development has been to find a model framework that achieves high accuracy with reasonable computational load for all 13 cislunar use cases, thus making it applicable for other cislunar objects. This model framework utilizes a single-arc Weighted Least-Squares algorithm to estimate the initial state. Using extensive testing, the dynamical model and integrator configurations have been optimized for the generic model

[14]GitHub repository: `https://github.com/daanwitte99/Thesis/tree/dev`

framework. Overall, it is found that a relatively simple dynamical model can achieve high performance for all use cases. This includes a spherical harmonics model (5,5) for Earth's gravity, point-mass gravity model for the Sun, Moon and Jupiter, and a cannonball solar radiation pressure model. The popular variable step-size RKF7(8) integrator with an error tolerance of 10^{-10} achieves the best performance (out of the available integrators). In general, the model framework performs better for the Chang'e 3 upper stage, caused by higher non-linearity in the orbit of the Chang'e 2 upper stage. The generic model framework is able to achieve PSAs up to 2 years, depending on the stability and window length of the use case. Though, orbit estimation quality and propagation accuracy can still be improved significantly.

Therefore, the generic model framework has been tailored on each individual estimation window. No single tailored model configuration achieves high performance for all use cases. Estimating radiation pressure coefficient (C_r) and/or observation bias (ϵ_{obs}), including an a priori covariance matrix or applying realistic observation weights, has been shown to increase model performance significantly for 8 out of 13 estimation windows. For these windows, tailoring decreased in-sample RMSE between 25-80%, out-of-sample RMSE between 20-95% and caused PSA to increase up to 250 days. A couple of important findings are made. Estimating C_r or ϵ_{obs} easily leads to overfitting, which can sometimes be solved using an a priori covariance matrix. It is found that estimating C_r is an important factor in performance improvement, caused by the large errors introduced from the cannonball simplification. Though, estimating C_r on less than 4 months of data is undesirable as it easily leads to unrealistic magnitudes, even with an a priori covariance matrix. Estimating ϵ_{obs} per ground station can improve out-of-sample performance, but does result in unreasonably high bias magnitude caused by unmodeled effects being attributed to the observation bias during estimation. This could potentially be solved by adding empirical accelerations to the model framework. After tailoring, PSAs for the Chang'e 2 upper stage range between 50 days and 1 year, the lower bound being caused by a close Moon approach. All windows of the Chang'e 3 upper stage achieve PSAs of approximately 1.5 to 2 years.

Comparing the performance of the tailored model framework between use cases, important findings are made for cislunar space debris. Firstly, estimating on 7-10 months seems to lead to the most robust out-of-sample performance. The number of observations in this window does not seem to be a very significant factor on performance, a couple of observations per month looks to be sufficient. The model framework is surprisingly competitive for unstable in-sample windows, it is capable of balancing out large orbital element fluctuations. It is also found that the model framework is able to propagate non-linear out-of-sample orbits with high PSA. The only behaviour the model framework still struggles with, are out-of-sample close Moon approaches. This always seems to result in a sharp residual increase, which quickly violates the accuracy threshold.

To find cause and solution for this issue, in-depth investigation is performed on estimation window 6 of the Chang'e 2 upper stage, which encounters an out-of-sample close Moon approach at a distance of 20.000 km. Propagating 200 Monte Carlo samples over the close Moon approach, it is found that the position difference with respect to the original propagation increases an order 10-1000 in a matter of days. Only one MC propagation stays at a similar position difference as before the encounter. This provides evidence that the current single-arc orbit determination approach is unsuitable for close Moon approaches. Moreover, tests have show that using a high order and degree spherical harmonics model for the Moon's gravity or integrators with higher error tolerances, does not solve this issue or even decrease out-of-sample RMSE. Therefore, multi-arc orbit determination is proposed specifically for the parts of the orbits with close Moon approaches. This allows for more accurate parameter estimation corresponding to the behaviour in a specific arc. Furthermore, it is recommended that the time of close approach, retrieved from orbit propagation, is communicated to observers. This should improve optical coverage after the encounter, allowing for accurate multi-arc orbit determination and preventing the object from getting lost to observers.

Interesting findings have been made from uncertainty propagation. Firstly, for unstable in-sample or out-of-sample windows, uncertainty propagation quickly becomes non-linear. This makes it hard to validate whether the initially generated samples are a good representation of the true uncertainty distribution, as this violates the assumptions of the current k-ratio approach. While for stable windows, uncertainty can usually be propagated linearly for a couple months to half a year, and thus the initial uncertainty distribution is more likely to be accurate. Furthermore, if the in-sample orbit is highly non-linear, due to for example a close Moon approach, the current covariance propagation approach tends to become invalid. The non-linearity induces significant errors in the propagated covariance, as seen when comparing MC propagation from t_E versus from t_0. Regarding the effect of uncertainty over time, orbit stability has been found to be a large factor in uncertainty magnitude. Propagating the stable orbits of windows 1-4 of the inherently more unstable Chang'e 2 booster, leads to up to 10x higher uncertainty over time compared to windows 1-4 of the Chang'e 3 booster. Moreover, effect of uncertainty over time is highest for short estimation windows (<4 months) and significantly lower for long windows (>6 months). This further justifies the conclusion that orbit estimation on 7-10 months of data seems most robust for cislunar space debris. In general, it is found that the effect of uncertainty for cislunar space debris orbits is significant, even for the tailored model framework. Use cases with good out-of-sample performance, can still quickly see a significant portion of the samples violating the accuracy threshold. Although the previously mentioned model improvements should reduce the uncertainty over time, the current PSAs are likely slightly overconfident compared to the actual time that objects can reliably be found by observers. An additional metric on the chance that the object will be lost to observers (based on propagated uncertainty), should solve this.

Several recommendations are made for future research. Firstly, some model improvements can be made. The cannonball simplification, even with C_r estimation, tends to cause large state errors over time. This is not only due to errors in C_r estimation, but also because the Chang'e 2 and 3 boosters are tumbling. This can be solved by applying the panelled radiation pressure approach, or by estimating a variational component in the radiation pressure coefficient to represent the tumbling behaviour. The error caused by other unmodeled effects can be reduced by including empirical accelerations in the model framework, which would also result in more realistic observation bias estimation. Estimating bias per streak of observations (not per ground station) would allow for differentiation between the constant instrumentation bias and variational time bias components, allowing for more realistic analysis on the magnitude of the overall observation bias. Furthermore, multi-arc orbit determination is expected to solve the shortcomings of this model framework regarding close Moon approaches. Finally, the number of samples used in Monte Carlo simulation should be increased significantly to make it more robust (N = 1000-10000). Effectiveness of other uncertainty propagation methods, like Polynomial Chaos Expansions or bootstrapping, could also be investigated for uncertainty propagation of cislunar space debris.

References

Blake, J. A. (2021). "Optical imaging of space debris in low and high altitude orbits". PhD thesis. University of Warwick (cit. on p. 6).

Boone, N. and Bettinger, R. A. (2021). "Cislunar debris propagation following a catastrophic spacecraft mishap". In: *AIAA SciTech 2021 Forum*, pp. 01–02 (cit. on p. 6).

Cano, A. et al. (2022). "Improving Orbital Uncertainty Realism Through Covariance Determination in GEO". In: *The Journal of the Astronautical Sciences* 69.5, pp. 1394–1420 (cit. on p. 12).

Carrico, J., Policastri, L., and Loucks, M. (2009). "Prelaunch Orbit Determination Design and Analysis for the IBEX Mission". In: *19th AAS/AIAA Spaceflight Mechanics Meeting*, pp. 09–133 (cit. on p. 16).

Carrico, J., Policastri, L., and Lutz, S. (2018). "IBEX Revisited: Operational Results for Long-Term Cislunar Orbit Propagation". In: *Paper No. AAS*, pp. 18–274 (cit. on pp. 9, 16).

Casanova, D., Petit, A., and Lemaıtre, A. (2015). "Long-term evolution of space debris under the J 2 effect, the solar radiation pressure and the solar and lunar perturbations". In: *Celestial Mechanics and Dynamical Astronomy* 123.2, pp. 223–238 (cit. on pp. 9, 19).

Celletti, A. et al. (2017). "Dynamical models and the onset of chaos in space debris". In: *International Journal of Non-Linear Mechanics* 90, pp. 147–163 (cit. on pp. 6–9).

Cen, Zheng (2011). "LM-3A Series Launch Vehicles User's Manual Issue". In: (cit. on p. 6).

Chai, Tianfeng and Draxler, R.R. (2014). "Root mean square error (RMSE) or mean absolute error (MAE)?– Arguments against avoiding RMSE in the literature". In: *Geoscientific Model Development* 7, pp. 1247–1250 (cit. on pp. 6, 14).

Desmars, J. et al. (2009). "Estimating the accuracy of satellite ephemerides using the bootstrap method". In: *AA* 499.1, pp. 321–330 (cit. on p. 24).

Dirkx, D. et al. (2017). "On the contribution of PRIDE-JUICE to Jovian system ephemerides". In: *Planetary and Space Science* 147, pp. 14–27. ISSN: 0032-0633 (cit. on p. 13).

Dirkx, Dominic, Fayolle, Marie, et al. (2022). "The open-source astrodynamics tudatpy software–overview for planetary mission design and science analysis". In: *EPSC2022* EPSC2022-253 (cit. on p. 6).

Dirkx, Dominic, Mooij, Erwin, and Root, Bart (2019). "Propagation and estimation of the dynamical behaviour of gravitationally interacting rigid bodies". In: *Astrophysics and Space Science* 364, pp. 1–22 (cit. on p. 6).

Drezner, Zvi, Turel, Ofir, and Zerom, Dawit (2008). "A Modified Kolmogorov-Smirnov Test for Normality". In: *University Library of Munich, Germany, MPRA Paper* 39 (cit. on p. 23).

Fayolle, M. et al. (2022). "Decoupled and coupled moons' ephemerides estimation strategies application to the JUICE mission". In: *Planetary and Space Science* 219, p. 105531. ISSN: 0032-0633 (cit. on p. 13).

Fehlberg, E. (1968). *Classical fifth-, sixth-, seventh-, and eighth-order Runge—Kutta formulas with stepsize control*. Tech. rep. (cit. on p. 9).

Frueh, C., Howell, K., DeMars, K., et al. (2021). *Cislunar space traffic management: Surveillance through earth-moon resonance orbits* (cit. on pp. 5, 12).

Frueh, C., Howell, K., DeMars, K. J., et al. (2021). "Cislunar space situational awareness". In: *31st AIAA/AAS Space Flight Mechanics Meeting*, pp. 6–7 (cit. on p. 13).

Garcia, M. (2021). *Space Debris and Human Spacecraft*. URL: `https://www.nasa.gov/mission_pages/station/news/orbital_debris.html` (cit. on p. 5).

Giordano, C., Topputo, F., et al. (2023). "Analysis and Optimization of Robust Trajectories in Cislunar Environment with Application to the LUMIO Cubesat". In: *33rd AAS/AIAA Space Flight Mechanics Meeting*, pp. 1–19 (cit. on p. 24).

Hoffman, F. O. and Hammonds, J. S. (1994). "Propagation of uncertainty in risk assessments: the need to distinguish between uncertainty due to lack of knowledge and uncertainty due to variability". In: *Risk analysis* 14.5, pp. 707–712 (cit. on pp. 12, 13).

Hofsteenge, R. (2013). *Computational Methods for the Long-Term Propagation of Space Debris Orbits* (cit. on p. 6).

Holzinger, M. J., Chow, C. C., and Garretson, P. (2021). *A primer on cislunar space* (cit. on pp. 5–8).

Hubaux, C. H. and Lemaıtre, A. (2013). "The impact of Earth's shadow on the long-term evolution of space debris". In: *Celestial Mechanics and Dynamical Astronomy* 116, pp. 79–95 (cit. on p. 9).

ISO, 27852 (2016). *Space systems — Estimation of orbit lifetime* (cit. on p. 9).

Johnson, Sandra K et al. (2021). "Gateway–a communications platform for lunar exploration". In: *38th International Communications Satellite Systems Conference (ICSSC 2021)*. Vol. 2021. IET, pp. 9–16 (cit. on p. 5).

Kennewell, J. A. and Vo, B. N. (2013). "An overview of space situational awareness". In: *Proceedings of the 16th International Conference on Information Fusion*. IEEE, pp. 1029–1036 (cit. on p. 5).

Kessler, D. J. et al. (2010). "The kessler syndrome: implications to future space operations". In: *Advances in the Astronautical Sciences* 137.8, p. 2010 (cit. on p. 5).

Kjeldsen, H. and Frandsen, S. (1992). "High-precision time-resolved CCD photometry". In: *Publications of the Astronomical Society of the Pacific* 104.676 (cit. on p. 12).

Klinkrad, H. (2006). *Space Debris: Models and Risk Analysis*. ISBN: 978-3-540-25448-5 (cit. on pp. 11, 12).

Langbroek, M. (2023). *priv. com.* (Cit. on pp. 12, 20).

Lari, Giacomo and Milani, Andrea (2019). "Chaotic orbit determination in the context of the JUICE mission". In: *Planetary and Space Science* 176, p. 104679. ISSN: 0032-0633 (cit. on pp. 20, 22).

Lee, Dong Kyu, In, Junyong, and Lee, Sangseok (2015). "Standard deviation and standard error of the mean". In: *Korean journal of anesthesiology* 68.3, p. 220 (cit. on p. 23).

Lemaître, A. (2019). "Space Debris: From LEO to GEO". In: *Satellite Dynamics and Space Missions*. Ed. by G. Baù et al. Cham: Springer International Publishing, pp. 115–157. ISBN: 978-3-030-20633-8 (cit. on p. 6).

Lhotka, C., Celletti, A., and Galeş, C. (2016). "Poynting–Robertson drag and solar wind in the space debris problem". In: *Monthly Notices of the Royal Astronomical Society* 460.1, pp. 802–815 (cit. on p. 9).

Li, C. et al. (2021). "Overview of the Chang'e-4 mission: Opening the frontier of scientific exploration of the lunar far side". In: *Space Science Reviews* 217, pp. 1–32 (cit. on p. 5).

Luo, Y. and Yang, Z. (2017). "A review of uncertainty propagation in orbital mechanics". In: *Progress in Aerospace Sciences* 89, pp. 23–39 (cit. on pp. 6, 12, 13).

Marshall, J. A. et al. (1992). *Modeling radiation forces acting on TOPEX/Poseidon for precision orbit determination*. Tech. rep. (cit. on p. 9).

Massey Jr, Frank J (1951). "The Kolmogorov-Smirnov test for goodness of fit". In: *Journal of the American statistical Association* 46.253, pp. 68–78 (cit. on p. 14).

McCarthy, D. D. and Petit, G. (2004). *IERS conventions (2003)*. Tech. rep. International Earth Rotation and Reference Systems Service (Iers)(Germany) (cit. on p. 9).

Meeus, Jean H. (1991). *Astronomical Algorithms*. Willmann-Bell, Incorporated. ISBN: 0943396352 (cit. on p. 14).

Micheli, Marco et al. (2018). "The observing campaign on the deep-space debris WT1190F as a test case for short-warning NEO impacts". In: *Icarus* 304. Asteroids and Space Debris, pp. 4–8. ISSN: 0019-1035 (cit. on pp. 6, 11).

Montenbruck, O. and Gill, E. (2000). *Satellite Orbits*. Vol. 1. ISBN: 978-3-540-67280-7 (cit. on pp. 6, 9–12, 16, 19, 20).

Moyer, Theodore D (2005). *Formulation for observed and computed values of Deep Space Network data types for navigation*. John Wiley & Sons (cit. on p. 12).

Murawiecka, M. and Lemaitre, A. (2018). "Yarkovsky-Schach effect on space debris motion". In: *Advances in Space Research* 61.3, pp. 935–940 (cit. on p. 9).

NASA (2020). *Memorandum of Understanding between the National Aeronautics and Space Administration and the United States Space Force*. URL: `https://www.nasa.gov/sites/default/files/atoms/files/nasa_ussf_mou_21_sep_20.pdf` (cit. on p. 5).

Picone, J. M. et al. (2002). "NRLMSISE-00 empirical model of the atmosphere: Statistical comparisons and scientific issues". In: *Journal of Geophysical Research: Space Physics* 107.A12, SIA 15-1-SIA 15–16 (cit. on p. 9).

Rowlands, Andy (2017). *Physics of Digital Photography*. 2053-2563. IOP Publishing. ISBN: 978-0-7503-1242-4 (cit. on p. 14).

Scheeres, D., Rosengren, A., and McMahon, J. (2011). "The dynamics of high area-to-mass ratio objects in earth orbit: The effect of solar radiation pressure". In: *Proceedings of the AAS/AIAA Space Flight Mechanics Meeting, number AAS*, pp. 11–178 (cit. on p. 9).

Serra, Daniele, Spoto, Federica, and Milani, Andrea (2018). "A multi-arc approach for chaotic orbit determination problems". In: *Celestial Mechanics and Dynamical Astronomy* 130.11, 75, p. 75 (cit. on pp. 20, 22).

Smith, M. et al. (2020). "The artemis program: An overview of nasa's activities to return humans to the moon". In: *2020 IEEE Aerospace Conference*, pp. 1–10 (cit. on p. 5).

Smith, Warren (2000). *Modern Optical Engineering: The Design of Optical Systems* (cit. on p. 14).

Valsecchi, Giovanni B. (2019). "Comparing NEA surveys". In: *1st NEO and Debris Detection Conference_ESA2019*, 29, p. 29 (cit. on p. 15).

Veis, G. (1963). "Optical tracking of artificial satellites". In: *Space Science Reviews* 2, pp. 250–296 (cit. on p. 11).

Ventura-Traveset, J. and Endere, W. (2021). *ESA Lunar PNT on-going activities: The ESA Moonlight Programme* (cit. on p. 5).

Vereš, P. et al. (2017). "Statistical analysis of astrometric errors for the most productive asteroid surveys". In: *Icarus* 296, pp. 139–149 (cit. on pp. 12, 19).

Virgili, B. B. et al. (2021). "Uncertainty propagation meeting space debris needs". In: *8th European Conference on Space Debris* (cit. on p. 13).

Wakker, K. F. (2015). *Fundamentals of astrodynamics*. TU Delft (cit. on pp. 8, 10).

Wilmer, A. P., Boone, N. R., and Bettinger, R. A. (2021). *Artificial Debris Propagation in Cislunar Periodic Orbits*. Tech. rep. (cit. on p. 6).

— (2022). "Debris Propagation and Spacecraft Survivability Assessment for Catastrophic Mishaps Occurring in Cislunar Periodic Orbits". In: *Journal of Space Safety Engineering* 9 (2), pp. 207–222. ISSN: 24688967 (cit. on p. 6).

Xu, Lin and Ouyang, Ziyuan (2014). "Scientific Progress in China's Lunar Exploration Program". In: *Chinese Journal of Space Science* 34.5, p. 525 (cit. on p. 6).

Ying, Xue (2019). "An Overview of Overfitting and its Solutions". In: *Journal of Physics: Conference Series* 1168, p. 022022 (cit. on p. 14).

Zakhvatkin, M. V. et al. (2020). "RadioAstron orbit determination and evaluation of its results using correlation of space-VLBI observations". In: *Advances in Space Research* 65.2, pp. 798–812 (cit. on p. 16).

Zhamkov, A. S. and Zharov, V. E. (2016). "Improvement of the orbit of the Spektr-R spacecraft in the RadioAstron mission on the basis of radio range and Doppler measurements". In: *Moscow University Physics Bulletin* 71, pp. 299–308 (cit. on p. 16).

Appendix A. Use case visualization

The 13 use cases selected in section 2, have been designed using the Keplerian elements plots shown in Figure A.1 and Figure A.2. The Keplerian elements of both objects are plotted on the observation times and against the distance of the objects to the Moon. As described in the paper, the Keplerian elements have been derived from Two-Line Elements created using the Find_Orb software.

Three-dimensional orbit visualizations are shown for all 13 use cases in Figure A.3 and Figure A.4. The orbits are combinations of in-sample and out-of-sample propagations of the tailored model framework. To ensure that the orbits are accurate representations of reality, multiple use case estimations have sometimes been combined to ensure that all plotted orbits are sufficiently accurate (residual <900 arcsec). For example, windows 5, 6 and 7 of the Chang'e 2 booster overlap with each other. Out-of-sample propagations of estimation window 5 and 6 for the Chang'e 2 booster have poor accuracy, the residual explodes due to a close Moon approach. The out-of-sample propagation of these windows are therefore replaced by in-sample propagations of window 7, which are known to be sufficiently accurate. The three-dimensional plots can be used as a visual reference for understanding any behaviour discussed in section 4.

Appendix B. Model development and tailoring

During development of an orbit determination and propagation model framework, described in subsection 3.7, many steps have been taken. This appendix provides additional material on these steps.

The baseline model framework used for comparison during generic model framework development is shown in Table B.1, it includes nearly all dynamics and a RKF7(8) variable step-size integrator with an error tolerance of 10^{-10}. The corresponding testing plan can be found in Table B.2. These model tests have lead to the generic model framework described in the paper in Table 4. The performance metrics of the generic model framework are shown in Table B.3.

Some additional material on the tailored model framework is presented. The estimated initial state (and C_r) of the tailored model framework is shown for all respective use cases in Table B.4. Moreover, the Probability Density Functions (PDFs) of the right ascension and declination in-sample residuals of the tailored model configurations for each use case are shown in Figure B.2 and Figure B.3.

Finally, an example analysis on the effect of model tailoring on the in-sample right ascension (α) and declination (δ) per ground station of window 3 of the Chang'e 3 booster is shown in Figure B.1. For this window, C_r and ϵ_{obs} is estimated during tailoring, to significantly improve in-sample and out-of-sample performance. This plot justifies the findings in subsection 4.2 that unmodeled effects are estimated into the observation bias, causing the unrealistic values shown in Table 9 of the paper. It is seen that the residuals before tailoring are very large and not normally distributed. The unmodeled effects are much larger than only radiation pressure discrepancies, and observation bias captures them. After tailoring, the in-sample residual distributions are much more Gaussian, also shown in Figure B.3.

Table B.1: Baseline model framework

Dynamical models	Spherical harmonics: - Earth (SH[10,10]) - Moon (SH[10,10])
	Point-mass: - Sun - Mercury - Venus - Mars - Jupiter - Saturn - Uranus - Neptune
	SRP (cannonball) - Shadow Earth Earth RP (cannonball)
Integrator	RKF7(8) variable dt - Error tolerance = 10^{-10} - Initial step-size = 1.0 s - Max step-size = 10000 s
Propagator	Cowell
Ephemeris models	Default (Spice kernels)
Rotation models	Default (Spice kernels)
Estimated param.	Initial state

Table B.2: Overview of testing plan (it=iterative)

Test nr.	Test description
Dyn. test 1 (it)	Increase SH Earth
Dyn. test 2 (it)	Decrease SH Earth
Dyn. test 3 (it)	Increase SH Moon
Dyn. test 4 (it)	Decrease SH Moon
Dyn. test 5 (it)	Test PM Moon
Dyn. test 6	Remove PM Mercury
Dyn. test 7	Remove PM Venus
Dyn. test 7	Remove PM Mars
Dyn. test 8	Remove PM Jupiter
Dyn. test 9	Remove PM Saturn
Dyn. test 10	Remove PM Uranus
Dyn. test 11	Remove PM Neptune
Dyn. test 12	Remove Earth rad. pressure
Dyn. test 13	Remove aerodynamics
Dyn. test 14	Remove relativistic effects
Int. test 1 (it)	RKF7(8) - error tolerance / 10
Int. test 2 (it)	RKF7(8) - error tolerance * 10
Int. test 3 (it)	DOPRI8(7) - error tolerance / 10
Int. test 4 (it)	DOPRI8(7) - error tolerance * 10
Int. test 5	RKF7(8) - fixed step size
Int. test 6	RK4
Int. test 7	ABM (order 6-11)
Int. test 8	BS6

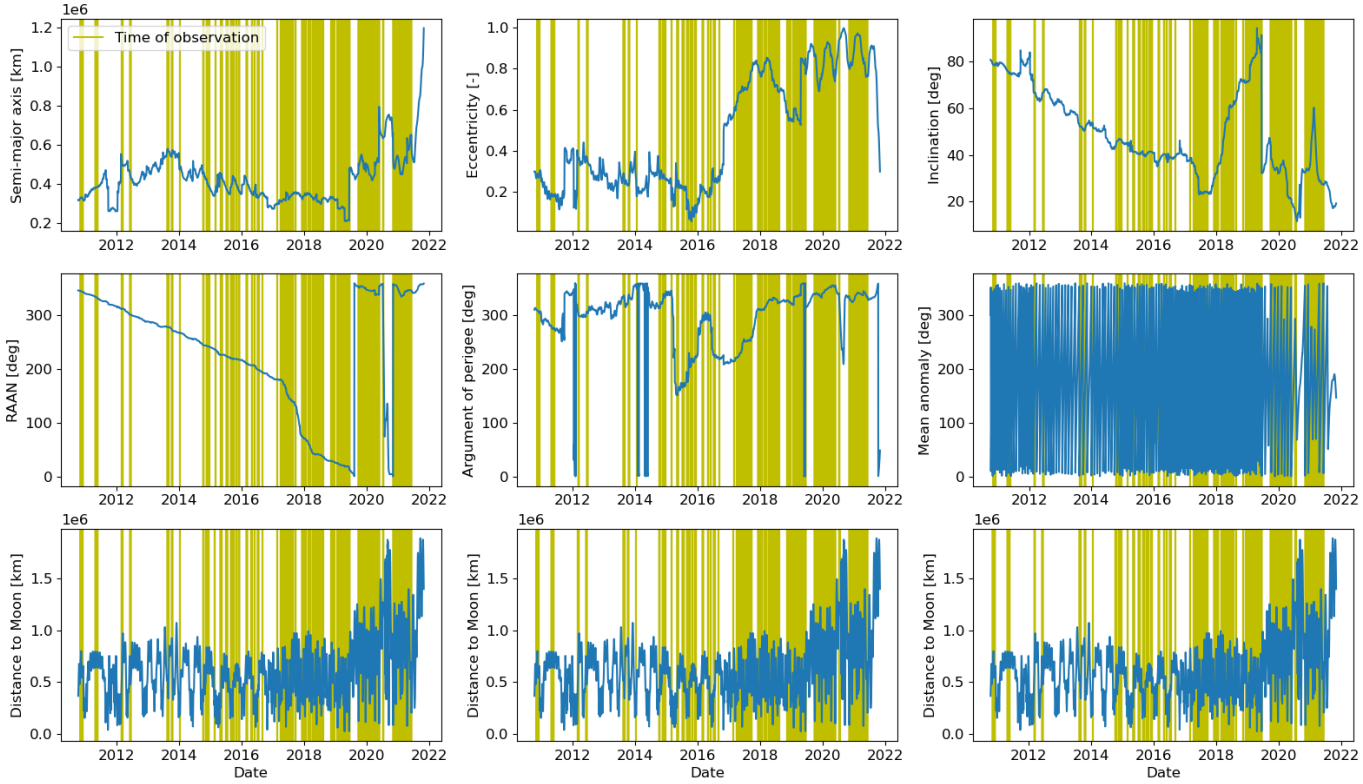


Figure A.1: TLE Keplerian elements for Chang'e 2 upper stage vs. distance to Moon

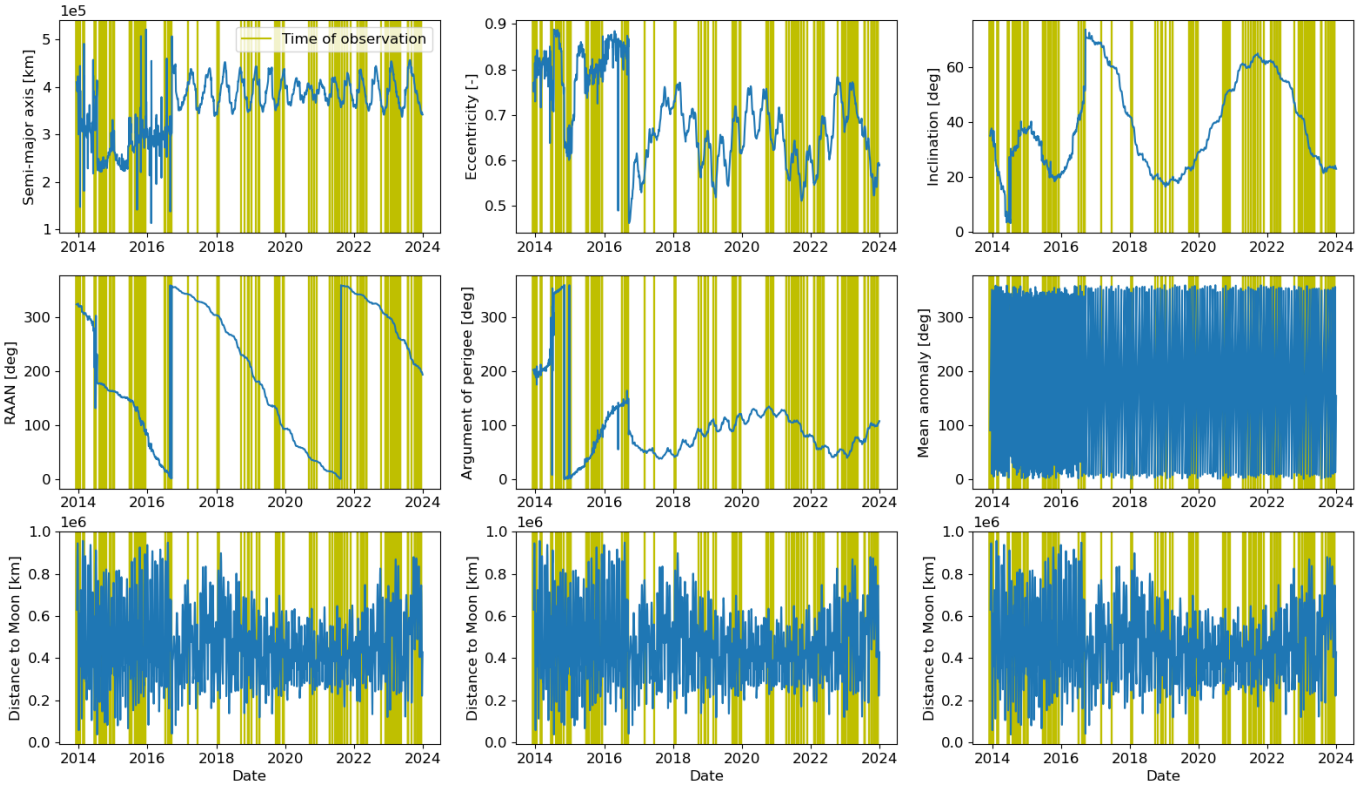


Figure A.2: TLE Keplerian elements for Chang'e 3 upper stage vs. distance to Moon

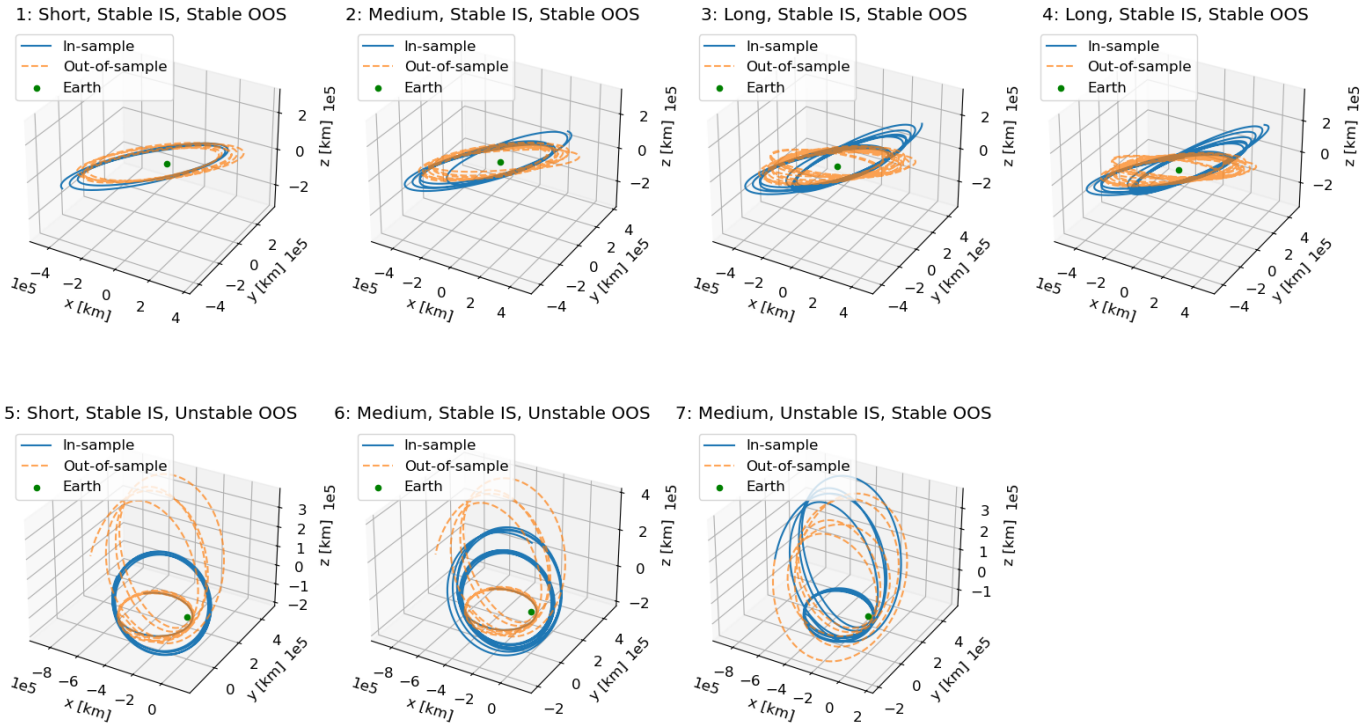


Figure A.3: Three-dimensional orbit visualizations for use cases of Chang'e 2 upper stage based on tailored model estimations

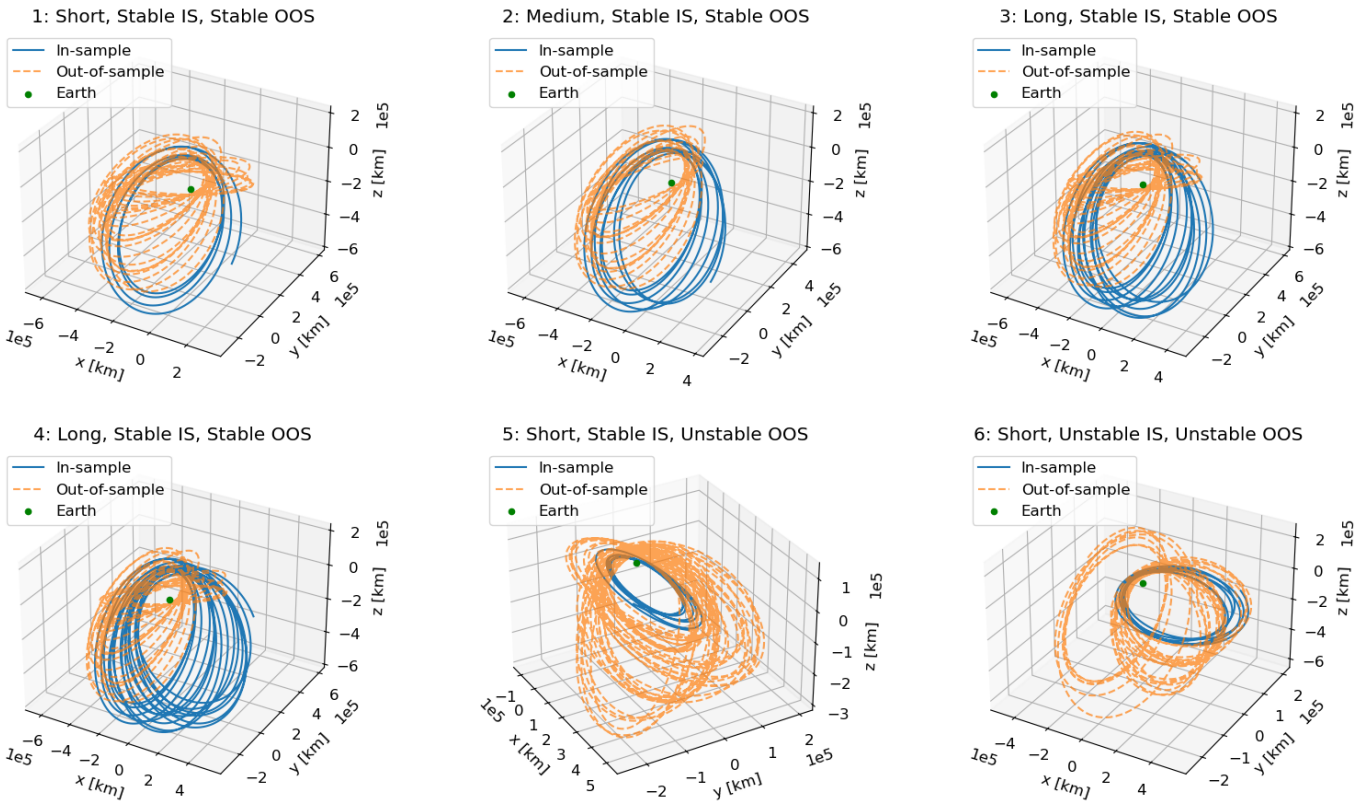


Figure A.4: Three-dimensional orbit visualizations for use cases of Chang'e 3 upper stage based on tailored model estimations

Table B.3: Generic model performance (*do not compare RMSE OOS between estimation windows ending on different dates)

Object	Window	Comp. time [s]	RMSE IS [arcsec]	RMSE OOS* [arcsec]	PSA [days]
Chang'e 2	1-SSS	8,1	3,7	166553,1	145-228
	2-MSS	13,4	5,9	144565,0	145-228
	3-LSS	15,9	11,8	95774,9	369-418
	4-LSS	19,5	11,7	134097,0	369-418
	5-SSU	8,7	8,0	169836,6	49-163
	6-MSU	15,8	17,0	213514,8	47-49
	7-MUS	100,8	9,0	9763,7	90-91
Chang'e 3	1-SSS	8,7	11,5	242,9	>672
	2-MSS	13,3	35,5	1154,7	248-295
	3-LSS	27,0	45,9	2437,7	434-436
	4-LSS	32,3	46,0	2268,9	434-436
	5-SSU	12,3	2,8	32149,8	676-678
	6-SUU	13,0	6,1	1699,7	501-609

Table B.4: Estimated initial state and C_r of tailored model framework

Obj.	Window	$x(t_0)$ [m]	$y(t_0)$ [m]	$z(t_0)$ [m]	$V_x(t_0)$ [m/s]	$V_y(t_0)$ [m/s]	$V_z(t_0)$ [m/s]	C_r [-]
CH2	1-SSS	-4,59E+08	-2,77E+08	-1,23E+08	236,5	-536,4	473,9	1,80
	2-MSS	3,32E+08	1,91E+08	1,65E+08	-86,5	734,0	-491,0	1,58
	3-LSS	3,72E+08	3,35E+08	1,91E+08	-250,5	508,4	-478,7	1,20
	4-LSS	4,48E+08	2,14E+08	2,99E+08	-70,6	630,7	-324,3	1,20
	5-SSU	-3,77E+08	-1,19E+08	1,92E+08	-220,0	-199,6	-518,4	1,55
	6-MSU	-6,69E+07	7,90E+07	3,08E+08	-907,0	-287,7	571,7	1,67
	7-MUS	-1,45E+08	-4,91E+07	-1,02E+08	1477,9	493,0	49,7	1,53
CH3	1-SSS	3,04E+08	-1,39E+08	-2,50E+08	176,3	423,9	860,8	1,20
	2-MSS	3,54E+08	-1,76E+08	-3,48E+08	197,8	298,5	571,1	1,62
	3-LSS	1,77E+08	1,13E+08	1,20E+08	-1380,9	-2,4	593,7	1,50
	4-LSS	4,22E+08	1,88E+08	-9,38E+07	-525,2	210,8	576,0	1,50
	5-SSU	3,61E+08	-1,49E+08	6,97E+07	442,6	153,6	-102,1	1,20
	6-SUU	4,56E+08	3,70E+07	-1,36E+08	-177,7	447,1	-105,4	1,20

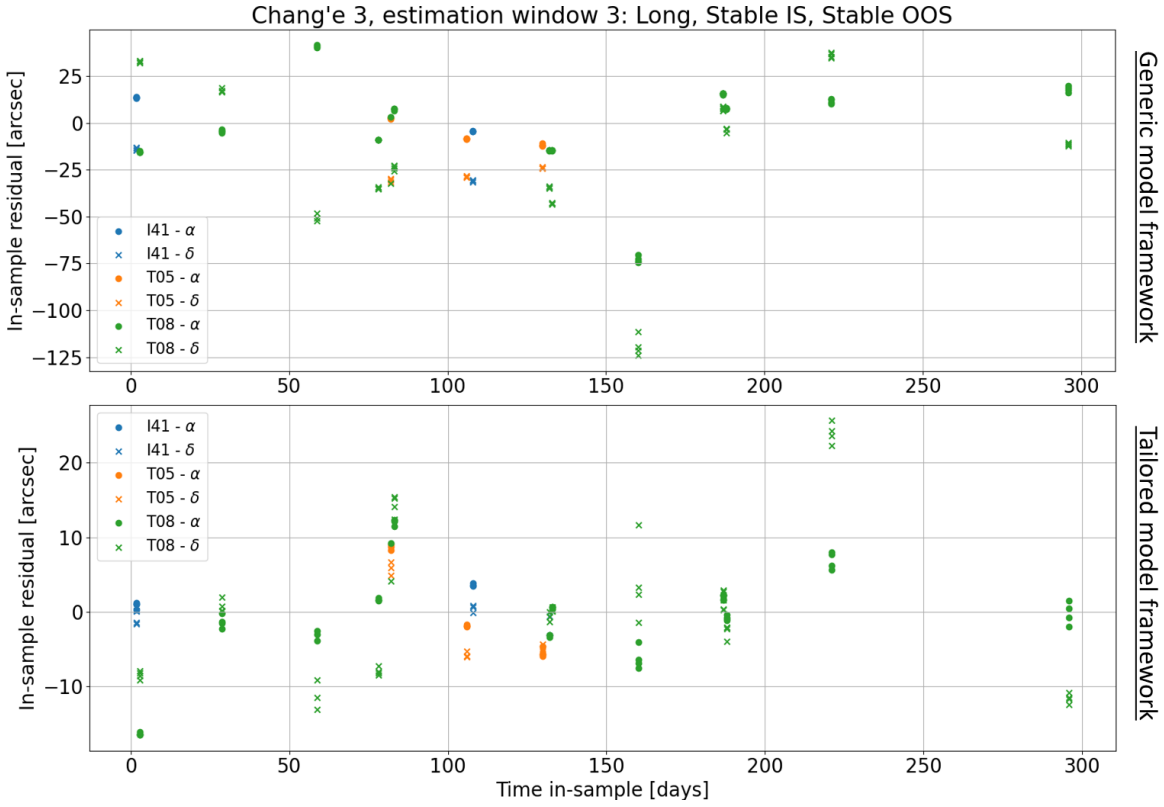


Figure B.1: In-sample α and δ residual (per ground station) for generic vs. tailored model framework of window 3 of the Chang'e 3 booster

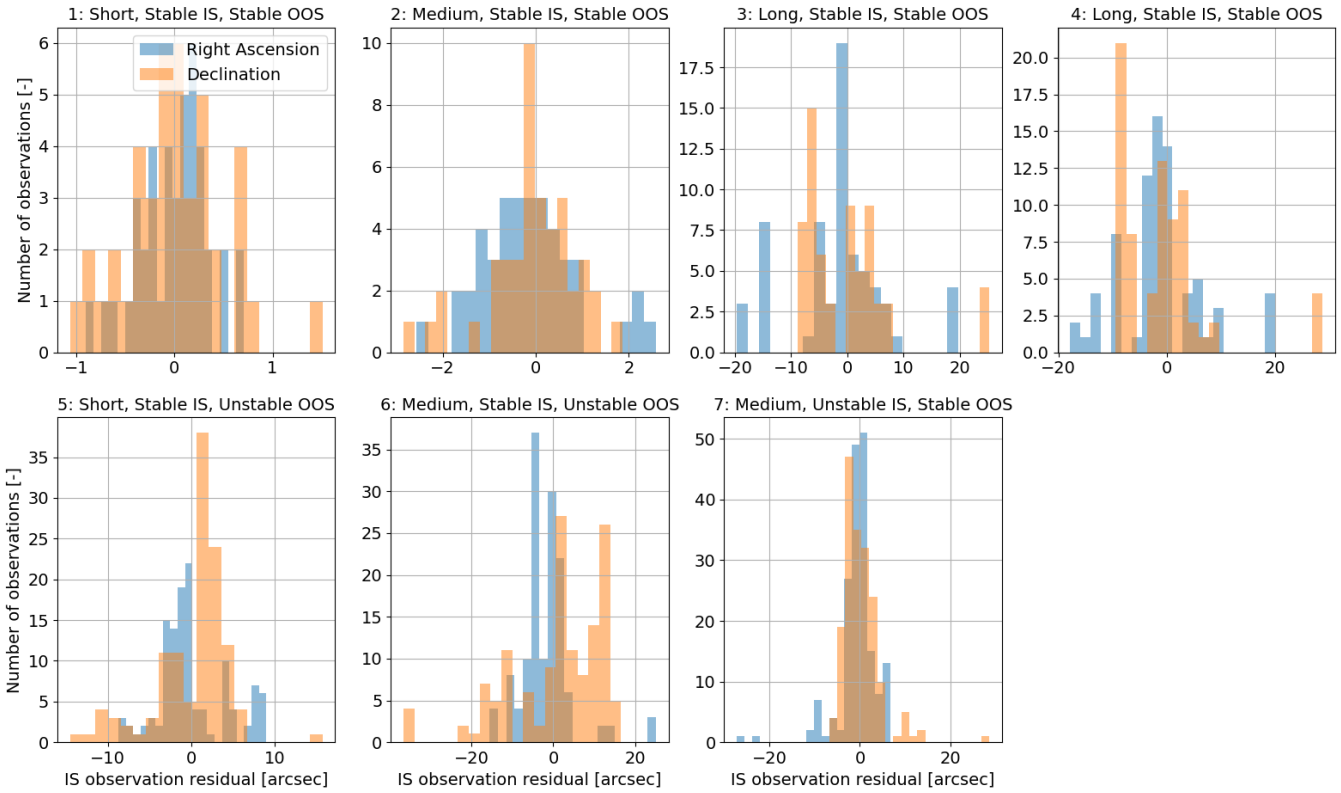


Figure B.2: PDFs of tailored model framework in-sample residuals per use case of Chang'e 2 upper stage

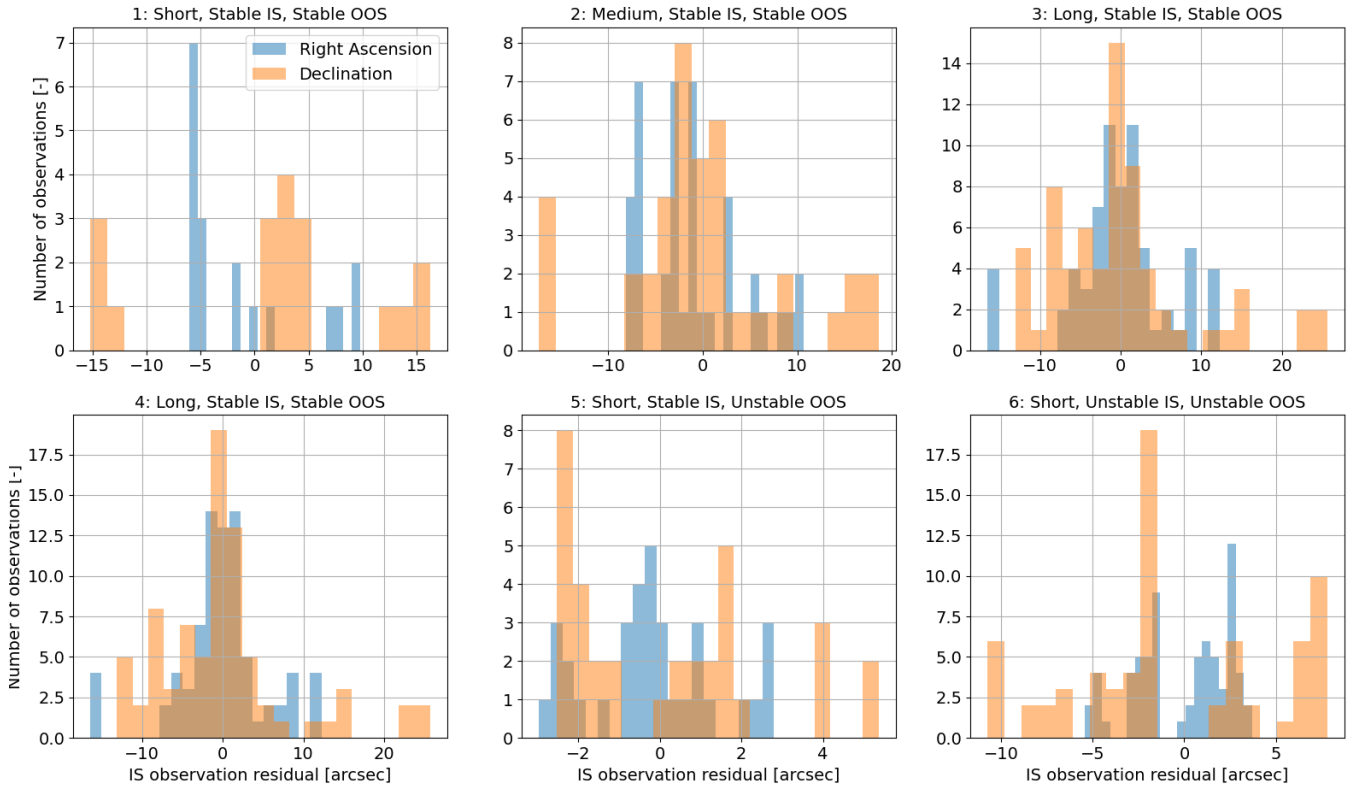


Figure B.3: PDFs of tailored model framework in-sample residuals per use case of Chang'e 3 upper stage

3

Conclusions & Recommendations

This chapter describes the main conclusions and recommendations of this research. Conclusions are drawn in section 3.1 and recommendations for future research are made in section 3.2.

3.1 Conclusions

The research questions, presented in chapter 1, will be answered in this section. The underlying sub-questions are answered first, after which the main research question is addressed. The results supporting these conclusions can be found in section 4 of the paper and in the appendices.

Sub-question 1: *What model framework configuration achieves the best propagation accuracy with reasonable computational load, for any cislunar space debris orbit?*
Many different design choices can be made in the model framework. Some preliminary design choices are made before model framework development using literature, and others are made based on extensive testing during development. For the generic model framework, Weighted Least-Squares single-arc orbit determination is used to estimate the initial state. This has been defined beforehand. The dynamical model and integrator configuration has been tested extensively during development of the generic model framework. It is found that a relatively simple dynamical model configuration provides the best accuracy to computational load trade-off. This includes a degree / order 5 spherical harmonics model for Earth's gravitational acceleration, point-mass gravity models for the Sun, Moon and Jupiter, and a cannonball solar radiation pressure model. The optimal integrator configuration is found to be a variable step-size RKF7(8) integrator with an error tolerance of 10^{-10}. The generic model framework is designed to achieve good accuracy with reasonable computational load for all 13 use cases. It achieves Periods of Sufficient Accuracy (PSA's) ranging from 50 days to 2 years, depending on the use case. PSA indicates for how long the total residual is lower than 900 arcseconds and will be elaborated upon in sub-question 2. Tailoring of the model framework on each individual use case, can in many cases improve performance even further.

Sub-question 2: *What is the theoretical accuracy that the model framework should achieve to be effective for cislunar Space Situational Awareness and what is the maximum propagation duration for which the model framework can maintain this accuracy?*
The theoretical accuracy is based on the maximum allowable error in the propagated orbit, for which an observer can still reliably find the object from the predicted location. It is found that as long as the total residual of the propagated orbit is below 900 arcseconds, that the propagated orbit from the model framework is sufficiently accurate to be applied for cislunar Space Situational Awareness (SSA). The accuracy threshold has been approximated from the field of view of the instrumentation from an average professional observer, being 0.5 degrees or 1800 arcseconds. Since the propagated orbit should always be in the field of view of an observer, a maximum residual of 900 arcseconds is defined. Using this threshold, a metric has been defined called the Period of Sufficient Accuracy (PSA). This defines the out-of-sample period for which the propagated orbit is below the accuracy threshold.

Model tailoring on individual use cases, also mentioned in sub-question 1, has allowed for significant improvements in PSA (up to 250 days longer) compared to the generic model framework. The

various tailoring settings (C_r and/or ϵ_{obs} estimation, a priori covariance, observation weights) have varying effects on performance for different use cases. Not one single tailoring configuration results in performance improvements for all use cases, thus individual tailoring configurations are applied to achieve the best out-of-sample performance (PSA) for each estimation window. Still, the PSA is very use case dependent. For the Chang'e 2 upper stage, PSAs range from 50 days to 1 year between the different windows. The low PSA of 50 days is caused by a close Moon approach. For the Chang'e 3 upper stage, the tailored model framework achieves a PSA of 1.5 to 2 years for all windows.

Sub-question 3: *How is the orbit estimation quality and propagation accuracy affected by various cislunar orbit and estimation window characteristics?*
The quality of cislunar orbit estimation depends on both the orbit behaviour and estimation window characteristics. It has been found that estimating on 7-10 months of observations seems to result in the most robust orbit estimation and best propagation accuracy. Estimating on less than 4 months of data can in some cases be competitive, but can also lead to a poor fit. The number of observations does not seem to have a large influence on orbit estimation, though at least a couple of observations per month are required. The tailored model framework performs surprisingly well for windows that are unstable in-sample, such as window 7 of the Chang'e 2 booster and window 6 of the Chang'e 3 booster, where the former even includes a close Moon approach in-sample. The orbit determination algorithm seems to balance out the large in-sample orbital elements fluctuations, resulting in competitive propagation accuracy compared to other use cases. Moreover, the model framework is capable of accurately propagating out-of-sample orbits which are non-linear. This has been proven by both windows 5 and 6 of the Chang'e 3 booster, which achieve similar out-of-sample performance as the first 4 stable windows of the object. The only orbit characteristic that the model framework struggles with, are close Moon approaches out-of-sample.

Out-of-sample close encounters with the Moon occur for windows 5 and 6 of both the Chang'e 2 and 3 booster, reaching a distance of 20.000 and 40.000 km to the Moons center of mass respectively. The effect has become especially clear in a deep-dive on window 6 of the Chang'e 2 booster, where the orbit reaches a distance of 20.000 km. The encounter results in a sharp increase of the propagation residual. It usually only takes a short period before the orbit crosses the accuracy threshold of 900 arcseconds. This is caused by the sudden increase in acceleration and velocity due to the gravitational pull of the Moon, this significantly enhances the errors in the state right before the encounter. This has further been confirmed from Monte Carlo (MC) simulation, where the position difference between the tailored model propagation and nearly all 200 MC propagations increases 10-1000 times, though their position difference before the close Moon approach is relatively small. Only one MC propagation stays in the vicinity of the original propagation. It has been found during dynamical model testing that increasing accuracy of the Moon's gravity by applying a high order and degree spherical harmonics model does not solve this or even result in improved out-of-sample RMSE. Multi-arc orbit determination is likely the only solution to improve propagation accuracy after a close Moon approach, which would probably require better observational coverage after the encounter. This could be realized by predicting close Moon approaches based on the propagated orbit and passing this information to the observers.

Sub-question 4: *What is the effect of estimating the radiation pressure coefficient and observation bias on orbit estimation quality and propagation accuracy of cislunar space debris?*
During model tailoring on individual use cases, both the radiation pressure coefficient (C_r) and the observation bias (ϵ_{obs}) are added to parameter estimation and tested. The effect that this has on orbit determination quality and propagation accuracy is very use case specific due to the different orbit (e.g. stability) and observation (e.g. window length) characteristics. But on average it is found to have a positive effect on both orbit determination quality and propagation accuracy. During tailoring, it has been seen that only adding ϵ_{obs} does not achieve better performance compared to adding C_r or C_r & ϵ_{obs} to parameter estimation (which also contains initial state). This is caused by the large state errors over time resulting from the cannonball simplification, which are partly addressed in C_r estimation. Furthermore, estimating C_r and/or ϵ_{obs} easily results in overfitting of the orbit estimation. This can sometimes be solved by adding an a priori covariance matrix, but not always. When estimating on windows that contain less than 4 months of data, C_r estimation becomes unreliable and can result in unrealistic values like 14, or -1. This is because the compounding effect of the radiation pressure over time is not significant enough for the orbit determination algorithm to find a reliable estimation. Finally,

it is found that estimating ϵ_{obs} often results in unrealistic biases of over 10 arcseconds, which does result in improved propagation accuracy. This is partly caused by other unmodeled effects being accounted for in the observation bias. Adding empirical accelerations to the model framework is expected to solve this issue and result in more realistic observation bias estimation. Though, one must be careful with overfitting when applying empirical accelerations. Furthermore, the estimation does not differentiate between the constant observation bias and varying time bias, this could be solved by performing ϵ_{obs} estimation per observation streak instead of per ground station. Thus, currently one can not draw conclusions on the estimated bias magnitude for specific ground stations but can use ϵ_{obs} estimation to improve out-of-sample performance.

Sub-question 5: *What is the effect of model and observation uncertainty over time for cislunar space debris orbits?*
Analyzing uncertainty propagation results for various use cases, interesting findings are made. Firstly, linearity of the orbit is a significant contributor to uncertainty of the orbit propagation over time. This becomes evident from the fact that uncertainty of the Chang'e 2 upper stage's stable windows (window 1-4) has been found to be up to 10x higher over time compared to the Chang'e 3 upper stage's stable windows (window 1-4). Which is caused by the inherently more instability present in the orbit of the Chang'e 2 upper stage. Comparing the first four windows of the individual objects, it is found that for short windows (<4 months) the effect of uncertainty over time is significantly higher. This has further justified the conclusions described in sub-question 3, that estimating on 7-10 months of data seems most robust. Moreover, it is concluded that estimating C_r and ϵ_{obs} during model tailoring does not consistently result in fits that significantly reduce uncertainty. This is likely caused by the cannonball simplification as well as poorly estimated observation biases. In general, it is found that the effect of uncertainty over time for cislunar space debris is large for the current tailored model configurations. Even for use cases with good propagation accuracy, like the Chang'e 3 windows with PSAs up to 2 years, uncertainty distributions quickly have significant tails lying above the accuracy threshold. This usually happens in a matter of months. Realistic uncertainty propagation is thus vital for making sure that cislunar space debris objects do not lose our sight. Though, improvements in solar radiation pressure modelling, observation bias estimation (by adding empirical accelerations) and proper time bias quantification are expected to reduce the effect of uncertainty over time significantly.

Main research question: *How can long-term cislunar space debris orbits be estimated and propagated with sufficient accuracy, and how does the uncertainty of the resulting orbits evolve over time?*
A generic model framework has been found that is able to reach sufficient propagation accuracy with reasonable computational load for all 13 cislunar use cases and can be applied as a foundation for other cislunar space debris. This framework applies single-arc Weighted Least-Squares orbit determination to estimate the initial state. A relatively simple dynamical model configuration is found to be sufficient, including a SH[5,5] model for Earth's gravity, point-mass gravity models for the Sun, Moon and Jupiter, and a cannonball solar radiation pressure model. Furthermore, a variable step-size RKF7(8) integrator with error tolerance 10^{-10} is optimal. It has been found that additional tailoring of the generic model framework on each individual use case is vital. By estimating radiation pressure coefficient and/or observation bias, adding an a priori covariance matrix or applying realistic observation weights, both the orbit determination quality and propagation accuracy can be improved significantly. For 8 out of 13 use cases, tailoring resulted in out-of-sample RMSE improvements of 20-95% and increases in PSA of up to 250 days. The tailored model framework reaches PSAs up to 2 years, depending on the stability of the orbit. Comparing various use cases, it has been found that estimating on 7-10 months of observations results in highest orbit determination quality and propagation accuracy for these cislunar use cases. Estimating on less than 4 months of data is sometimes competitive, but seems much less robust.

Analyzing the effect that uncertainty has over time, interesting conclusions are drawn. The effect of uncertainty over time is significant for orbit propagation of cislunar space debris with the tailored model configurations. Uncertainty is especially high for short windows (<4 months) and the effect of non-linearity on uncertainty magnitude is clearly observed. For all cislunar space debris use cases, large portions of the uncertainty distributions will quickly start violating the accuracy threshold of 900 arcseconds with the current framework. When propagating orbits into the unseen future, this indicates that the calculated PSA metric is likely a slight overestimation of the actual period from which objects will start losing our sight.

3.2 Recommendations for future work

Based on the results and conclusions drawn in this research, several recommendations can be made for future researchers investigating orbit determination and propagation of cislunar space debris.

Firstly, though achieving PSAs of up to 2 years, the tailored model framework does leave room for improvement. From analysis of in-sample residual distributions as well as uncertainty propagations, it is found that the fit can be improved for several use cases and uncertainty over time could be reduced. The main changes that would increase the orbit estimation quality and propagation accuracy of the model framework, is changing the way radiation pressure is modeled and adding empirical accelerations. The simplified cannonball approach used in radiation pressure modelling has been found to create compounding state errors over time. Since this approach does not take into account shape and tumbling of the object, its accuracy is limited. To improve the radiation pressure model accuracy, one can either use the panelled radiation pressure approach described in subsection 3.1 of the paper, or add a variational component to the estimated C_r which accounts for tumbling. This should significantly reduce the largest model uncertainty left in the system. Any other unmodeled effects can be accounted for using empirical accelerations. In Tudat, one can choose to estimate either a constant or full empirical acceleration[1].

Secondly, the largest challenge that the current model framework has, is accurately propagating orbits out-of-sample over close Moon approaches. It is found that the current single-arc orbit determination method is unsuitable and will likely never be able to achieve this reliably. Therefore, a multi-arc orbit determination method is proposed. Here the orbit is split in several arcs, depending on the linearity and behaviour of that arc. Separate orbit determination is then performed for each arc. A small arc would be used around the close Moon approaches, allowing for more accurate parameter estimation which in turn improves propagation accuracy. Multi-arc orbit determination and propagation of cislunar space debris during close Moon approaches, would be a valuable research topic.

Furthermore, the current model framework is developed for 13 estimation windows divided amongst the Chang'e 2 and 3 upper stages. These use cases are hypothesized to be sufficiently diverse, to allow for the conclusions to be generally applicable to cislunar space debris as a whole. This can be validated in future work by applying the generic model framework and tailoring process to other cislunar space debris objects, and comparing the findings with the ones in this research. Some recommended objects to validate on are, the Chang'e 4 booster (2018-103B), Chang'e 5 booster (2020-087B) and Queqiao booster (2018-045B). Optical observations of these objects are stored by Project Pluto[2]. Ephemerides to be used for initial state guesses are also available from Find_Orb estimations.

Afterwards, several recommendations can be made for the way uncertainty is currently propagated, as well as how it can be applied in an actual orbit prediction tool. Firstly, sample size needs to be increased greatly (N = 1000-10000), as found from the Kolmogorov-Smirnov testing process. Since this will be computationally intensive, core splitting could be used to achieve this or one could choose to use an external server to perform the computations. Research on the application of more efficient non-linear uncertainty propagators like the Polynomial Chaos Expansion (PCE) [Giordano et al., 2023] would also be valuable to solve the computational load issue. Additionally, future researchers could look into uncertainty propagation using the bootstrapping approach [Desmars, J. et al., 2009]. Bootstrapping resamples the observations N times, performing orbit determination and propagation on each new observation set to create an uncertainty distribution over time. The main advantage of bootstrapping is that errors are allowed to be non-Gaussian.

Finally, since the PSA metric currently does not take into account the effect of uncertainty, future work on a probabilistic PSA metric is expected to be valuable. Combining orbit and uncertainty propagation, one could calculate a statistical risk that the object will be lost to observers at a certain time and set a threshold on this risk to determine when new observations are required. Moreover, the PSA metric applies an accuracy threshold of 900 arcseconds to make sure an object does not get lost. 900 arcseconds is likely not strict enough to tell something about how long a propagated orbit can be efficiently applied for collision avoidance. This does not mean that the propagated orbit with a residual of 900 arcseconds can not be applied for collision avoidance, but it would require the application of much higher risk margins for when maneuvers are needed and thus is impractical. A new (probabilistic) metric can be defined in follow-up research that defines the period for which the propagated orbit is sufficiently accurate for collision avoidance.

[1] https://py.api.tudat.space/en/latest/parameter.html
[2] https://projectpluto.com/pluto/mpecs/pseudo.htm

References

Administrator-NASA (2022). *Apollo 13*. URL: https://www.nasa.gov/mission_pages/apollo/missions/apollo13.html (visited on 06/28/2023) (cit. on p. 1).

Blake, J. A. (2021). "Optical imaging of space debris in low and high altitude orbits". PhD thesis. University of Warwick (cit. on p. 1).

Boone, N. and Bettinger, R. A. (2021). "Cislunar debris propagation following a catastrophic spacecraft mishap". In: *AIAA SciTech 2021 Forum*, pp. 01–02 (cit. on p. 2).

Boone, Nathan R. and Bettinger, Robert A. (2023). "Debris Collision Risk Analysis Following Simulated Cislunar Spacecraft Explosions". In: *Journal of Spacecraft and Rockets* 60.2, pp. 668–684 (cit. on p. 2).

Braun, V. et al. (2017). "Analysis of breakup events". In: *Proceedings of the 7th European Conference on Space Debris*. Vol. 7 (cit. on p. 1).

Cano, A. et al. (2022). "Improving Orbital Uncertainty Realism Through Covariance Determination in GEO". In: *The Journal of the Astronautical Sciences* 69.5, pp. 1394–1420 (cit. on p. 67).

Celletti, A. et al. (2017). "Dynamical models and the onset of chaos in space debris". In: *International Journal of Non-Linear Mechanics* 90, pp. 147–163 (cit. on p. 2).

Cen, Zheng (2011). "LM-3A Series Launch Vehicles User's Manual Issue". In: (cit. on pp. 40, 83).

Chow, C. C. et al. (2022). "Cislunar Orbit Determination Behavior: Processing Observations of Periodic Orbits with Gaussian Mixture Model Estimation Filters". In: *The Journal of the Astronautical Sciences* 69.5, pp. 1477–1492 (cit. on pp. 66, 67).

DeMars, K. J., Cheng, Y., and Jah, M. K. (2014). "Collision probability with gaussian mixture orbit uncertainty". In: *Journal of Guidance, Control, and Dynamics* 37.3, pp. 979–985 (cit. on p. 67).

DeMars, K. J., Jah, M., et al. (2009). "Orbit determination performance improvements for high area-to-mass ratio space object tracking using an adaptive Gaussian mixtures estimation algorithm". In: *21st International Symposium on Space Flight Dynamics* (cit. on p. 67).

Desmars, J. et al. (2009). "Estimating the accuracy of satellite ephemerides using the bootstrap method". In: *AA* 499.1, pp. 321–330 (cit. on p. 37).

Frueh, C., Howell, K., DeMars, K., et al. (2021). *Cislunar space traffic management: Surveillance through earth-moon resonance orbits* (cit. on p. 1).

Frueh, C., Howell, K., DeMars, K. J., et al. (2021). "Cislunar space situational awareness". In: *31st AIAA/AAS Space Flight Mechanics Meeting*, pp. 6–7 (cit. on p. 67).

Garcia, M. (2021). *Space Debris and Human Spacecraft*. URL: https://www.nasa.gov/mission_pages/station/news/orbital_debris.html (cit. on p. 1).

Geller, D. K. (2006). "Linear covariance techniques for orbital rendezvous analysis and autonomous onboard mission planning". In: *Journal of Guidance, Control, and Dynamics* 29.6, pp. 1404–1414 (cit. on p. 67).

Giordano, C., Topputo, F., et al. (2023). "Analysis and Optimization of Robust Trajectories in Cislunar Environment with Application to the LUMIO Cubesat". In: *33rd AAS/AIAA Space Flight Mechanics Meeting*, pp. 1–19 (cit. on pp. 37, 66, 67).

Hofsteenge, R. (2013). *Computational Methods for the Long-Term Propagation of Space Debris Orbits* (cit. on pp. 2, 47).

Holzinger, M. J., Chow, C. C., and Garretson, P. (2021). *A primer on cislunar space* (cit. on p. 2).

Hoogendoorn, R., Mooij, E., and Geul, J. (2018). "Uncertainty propagation for statistical impact prediction of space debris". In: *Advances in Space Research* 61.1, pp. 167–181 (cit. on p. 67).

Jones, B. A. and Doostan, A. (2013). "Satellite collision probability estimation using polynomial chaos expansions". In: *Advances in Space Research* 52.11, pp. 1860–1875 (cit. on p. 67).

Jones, B. A., Doostan, A., and Born, G. H. (2013). "Nonlinear propagation of orbit uncertainty using non-intrusive polynomial chaos". In: *Journal of Guidance, Control, and Dynamics* 36.2, pp. 430–444 (cit. on p. 67).

Jones, B. A., Parrish, N., and Doostan, A. (2015). "Postmaneuver collision probability estimation using sparse polynomial chaos expansions". In: *Journal of Guidance, Control, and Dynamics* 38.8, pp. 1425–1437 (cit. on p. 67).

Kessler, D. J. et al. (2010). "The kessler syndrome: implications to future space operations". In: *Advances in the Astronautical Sciences* 137.8, p. 2010 (cit. on p. 1).

Klinkrad, H. (2006). *Space Debris: Models and Risk Analysis*. ISBN: 978-3-540-25448-5 (cit. on p. 1).

Laurini, K. et al. (2018). *The global exploration roadmap* (cit. on p. 1).

Lee, Dong Kyu, In, Junyong, and Lee, Sangseok (2015). "Standard deviation and standard error of the mean". In: *Korean journal of anesthesiology* 68.3, p. 220 (cit. on p. 75).

Lemaître, A. (2019). "Space Debris: From LEO to GEO". In: *Satellite Dynamics and Space Missions*. Ed. by G. Baù et al. Cham: Springer International Publishing, pp. 115–157. ISBN: 978-3-030-20633-8 (cit. on p. 2).

Li, C. et al. (2021). "Overview of the Chang'e-4 mission: Opening the frontier of scientific exploration of the lunar far side". In: *Space Science Reviews* 217, pp. 1–32 (cit. on pp. 1, 2).

Micheli, Marco et al. (2018). "The observing campaign on the deep-space debris WT1190F as a test case for short-warning NEO impacts". In: *Icarus* 304. Asteroids and Space Debris, pp. 4–8. ISSN: 0019-1035 (cit. on p. 1).

Montenbruck, O. and Gill, E. (2000). *Satellite Orbits*. Vol. 1. ISBN: 978-3-540-67280-7 (cit. on pp. 46–48).

Smith, M. et al. (2020). "The artemis program: An overview of nasa's activities to return humans to the moon". In: *2020 IEEE Aerospace Conference*, pp. 1–10 (cit. on p. 1).

Temming, M. (2019). "Israel's First Moon Mission Lost Moments before Landing". In: URL: `https://www.sciencenews.org/article/israel-moon-mission-spacecraft-crash` (visited on 06/20/2023) (cit. on p. 1).

Virgili, B. B. et al. (2021). "Uncertainty propagation meeting space debris needs". In: *8th European Conference on Space Debris* (cit. on p. 67).

Wilmer, A. P., Boone, N. R., and Bettinger, R. A. (2021). *Artificial Debris Propagation in Cislunar Periodic Orbits*. Tech. rep. (cit. on p. 2).

— (2022). "Debris Propagation and Spacecraft Survivability Assessment for Catastrophic Mishaps Occurring in Cislunar Periodic Orbits". In: *Journal of Space Safety Engineering* 9 (2), pp. 207–222. ISSN: 24688967 (cit. on p. 2).

Xu, Lin and Ouyang, Ziyuan (2014). "Scientific Progress in China's Lunar Exploration Program". In: *Chinese Journal of Space Science* 34.5, p. 525 (cit. on p. 40).

Zakhvatkin, M. V. et al. (2020). "RadioAstron orbit determination and evaluation of its results using correlation of space-VLBI observations". In: *Advances in Space Research* 65.2, pp. 798–812 (cit. on p. 83).

Zhamkov, A. S. and Zharov, V. E. (2016). "Improvement of the orbit of the Spektr-R spacecraft in the RadioAstron mission on the basis of radio range and Doppler measurements". In: *Moscow University Physics Bulletin* 71, pp. 299–308 (cit. on p. 83).

A

Use cases

The use cases used for model framework development have been shortly described in section 2 of the paper, but information on the decision process and detailed visualizations are limited there. This chapter presents a detailed description of the use case selection process in section A.1 and compares the different use cases visually using three-dimensional orbits in section A.2.

A.1 Use case selection process

As described in the paper, two pieces of cislunar space debris have been chosen for model framework development. These are the Chang'e 2 and Chang'e 3 upper stages, which have object identifiers 2010-050B and 2013-070B respectively and are remnants of the Chinese Lunar Exploration Program [Xu et al., 2014]. The Chang'e 2 and 3 upper stages are boosters part of a Long March 3C and 3B rocket respectively[1][2]. Each booster has a maximum surface area of 37.14 m^2 and estimated dry mass of 5000 kg [Cen, 2011].

The objects have been chosen because their observation sets both span around 10 years and contain sufficient observations throughout this period (as can be seen in Figure A.3). Both objects orbit Earth, and contain several close Moon approaches. In 2021, the Chang'e 2 booster was excited into heliocentric orbit and lost to observers. The Chang'e 3 booster still orbits Earth and can be observed. To identify the orbit characteristics of both objects over time, Keplerian elements over time are identified from Two-Line Elements (TLEs). These have been created using the Find_Orb software package[3] and stored in an open GitHub repository[4]. Find_Orb is a separate orbit determination software developed by Project Pluto. Plotting the Keplerian elements over time against the distance the object has to the Moon in Figure A.1 and Figure A.2, different characteristics can be identified.

Firstly, it can be seen from the patterns in the Keplerian elements, that the Chang'e 2 upper stage is in a more unstable orbit than the Chang'e 3 upper stage is (at least for most of the 10 year span).

Secondly, interesting points in time are the 12th of June 2019 for the Chang'e 2 booster and 17th of September 2016 for the Chang'e 3 booster. At these moments, the objects experience large changes in orbital elements due to a close Moon approach. The Chang'e 2 booster gets as close as 20.000 km, while the Chang'e 3 one approaches the Moon at a distance of 40.000 km. For the Chang'e 2 booster, this results in a highly non-linear and unpredictable orbit after the encounter, ultimately causing it to be excited into heliocentric orbit. For the Chang'e 3 booster, the opposite happens and it enters a more stable and predictable orbit. This further indicates the challenge of orbit determination and propagation in cislunar space, also analyzed in the results of the paper. These rapid changes in orbital elements due to close approaches are highly unpredictable.

[1] https://www.nasaspaceflight.com/2010/10/live-long-march-3c-launch-change-2-probe/
[2] https://nssdc.gsfc.nasa.gov/nmc/spacecraft/display.action?id=2013-070A
[3] https://www.projectpluto.com/find_orb.htm
[4] https://github.com/Bill-Gray/tles

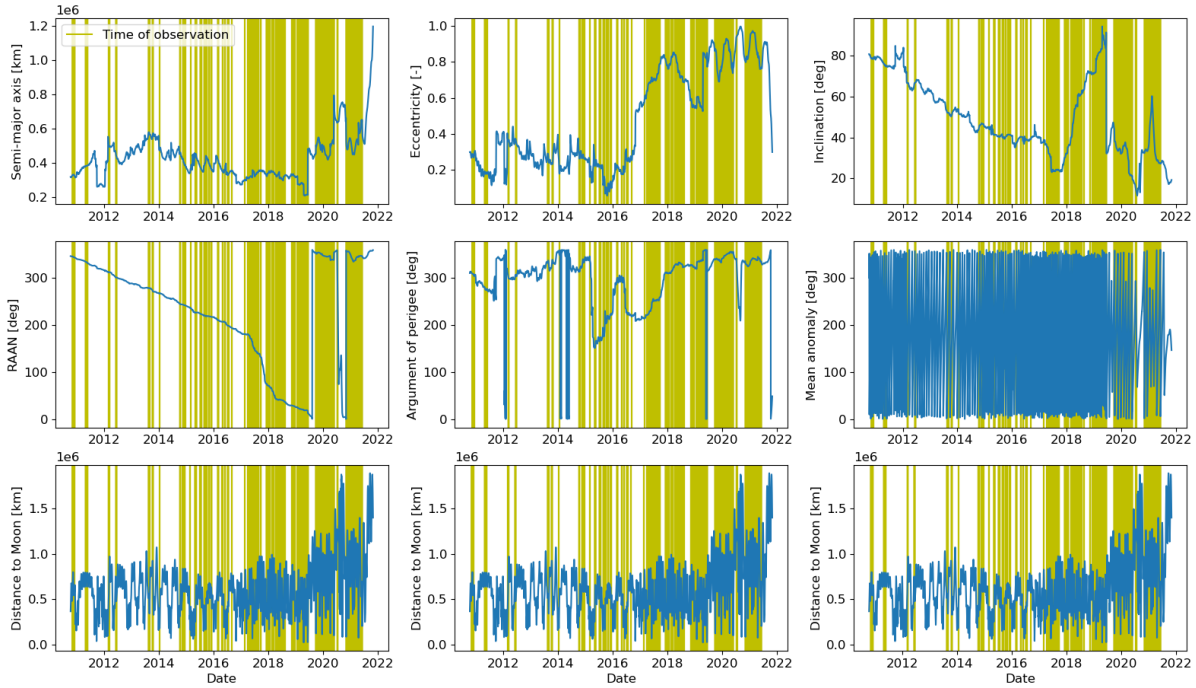


Figure A.1: TLE Keplerian elements for Chang'e 2 upper stage vs. distance to Moon

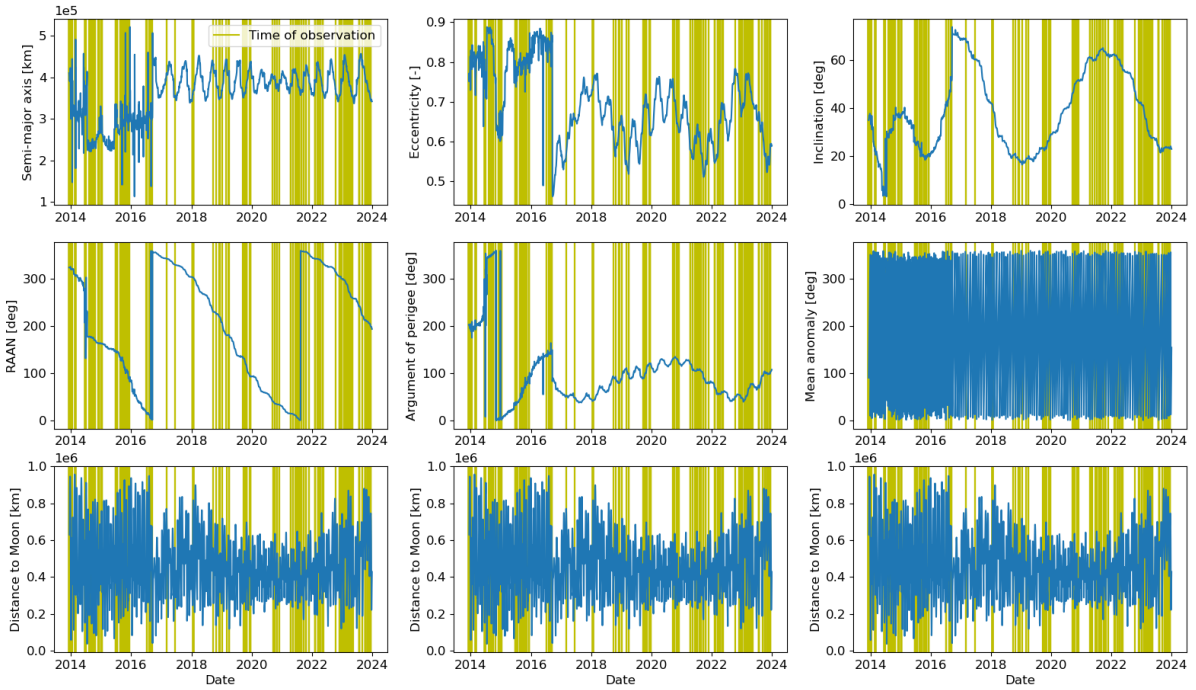


Figure A.2: TLE Keplerian elements for Chang'e 3 upper stage vs. distance to Moon

Using the information on orbit characteristics over time and the times of observation for both objects, a wide variety of estimation windows can be chosen that have various characteristics for which the model framework will be developed. The goal is to be able to draw conclusions on the effect of orbit and estimation windows characteristics on model performance. Mainly because the findings of this model framework should not only be applicable to the Chang'e 2 and 3 boosters, but to all cislunar space debris.

To find the optimal set of estimation windows, several important characteristics are identified. These

are estimation window length, number of observations in window, stability of orbit in-sample and stability of orbit out-of-sample. Window length is characterized using short (<4 months), medium (4-6 months) and long (>6 months). These buckets have been found to be effective after several use case iterations. In-sample and out-of-sample stability is characterized using stable and unstable. These are found visually from the patterns shown in the Keplerian elements plots. Finally, the previously mentioned close approaches to the Moon should at least be in some of the windows, to allow for analysis of this effect on model performance.

This resulted in the 13 windows (7 for the Chang'e 2 booster, 6 for the Chang'e 3 booster) described in Table A.1 and visualized on the TLE-derived semi-major axis in Figure A.3. It can be seen that all windows with the same in-sample and out-of-sample stability characteristics for one object end on the same date. This has been done on purpose. Because observations of the object are not continuous with constant spacing, the times of observation out-of-sample tend to vary a lot. To be able to analyze the effect that different estimation window length have on out-of-sample RMSE, one needs to be able to analyze the residuals on the exact same out-of-sample observations for one stability configuration. Constraining all windows with the same stability configuration to the same end-date, solves this. This also means that one can not simply compare out-of-sample RMSE between windows with different end-dates. Thus one can compare out-of-sample RMSE between windows 1-4 of Chang'e 2, but not between window 4 and 5 of Chang'e 2. The same goes for Chang'e 3, or for inter-object comparisons.

As explained in the paper, each estimation window receives a code that identifies the window length, the in-sample stability and out-of-sample stability. An example, 5-SSU would be window 5, (S)hort length, (S)table in-sample and (U)stable out-of-sample. These codes are only applied in tables containing results on the 13 windows. The codes are applied because they save space but do still allow a reader to understand the characteristics of a specific window. In the text itself, the codes are not applied but the characteristics of the windows are simply written out for additional clarity.

Table A.1: Overview of estimation windows

Object	Nr.	Window period	Nr. of obs.	Wind. length	IS stability	OOS stability	Code
Chang'e 2	1	2015-5-6 to 2015-7-14	42	Short	Stable	Stable	1-SSS
	2	2015-2-24 to 2015-7-14	46	Medium	Stable	Stable	2-MSS
	3	2014-11-15 to 2015-7-14	65	Long	Stable	Stable	3-LSS
	4	2014-10-11 to 2015-7-14	75	Long	Stable	Stable	4-LSS
	5	2019-2-7 to 2019-4-24	121	Short	Stable	Unstable	5-SSU
	6	2018-11-11 to 2019-4-24	138	Medium	Stable	Unstable	6-MSU
	7	2019-4-21 to 2019-10-31	181	Medium	Unstable	Stable	7-MUS
Chang'e 3	1	2021-10-19 to 2022-2-9	18	Short	Stable	Stable	1-SSS
	2	2021-7-31 to 2022-2-9	41	Medium	Stable	Stable	2-MSS
	3	2021-4-17 to 2022-2-9	68	Long	Stable	Stable	3-LSS
	4	2020-11-4 to 2022-2-9	79	Long	Stable	Stable	4-LSS
	5	2014-8-5 to 2014-10-23	34	Short	Stable	Unstable	5-SSU
	6	2015-7-1 to 2015-10-18	70	Short	Unstable	Unstable	6-SUU

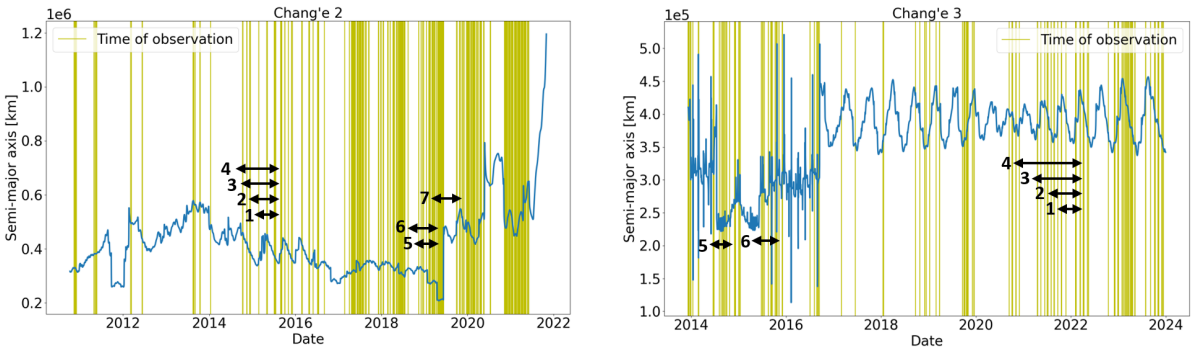


Figure A.3: Estimation window visualization on semi-major axis of the Chang'e 2 and 3 boosters

The process of finding these 13 windows, has not been as straightforward as described in the paper. It required many iterations, and contained several challenges. These challenges should be known for any future researchers creating orbit determination and propagation model frameworks for cislunar objects.

Firstly, observations seem abundant but are still a limiting factor in the selection process. One wants to find a estimation window that starts right before an optical observation (~1 day before) and ends between two observations that are close together (~1 week). The spacing at the start is important for prevention of extrapolation errors at the first observation (since start if window is also when initial state guess is made). The end spacing is important because one wants to know the model performance right before and right after the end of the training set. If the estimation window ends right after an observation, after which out-of-sample does not contain an observation for several months, then it is impossible to analyze the development of the residual at the beginning out-of-sample and the orbit determination algorithm can not take into account any orbit behaviour in that period. Now considering that different estimation windows are required, this significantly reduces the number of estimation window options that can be chosen from.

Secondly, in the process of finding 13 suitable estimation windows, it has been found that not every starting date leads to a converging Weighted Least-Squares solution. This is especially the case when the orbit is unstable. The reason is that in order to initialize the orbit determination algorithm, an approximate initial state (position and velocity) of the object at that time is required. As explained in the paper, this is obtained from ephemeris generated by orbit estimations of the Find_Orb software. Similar to this model framework in Tudat, Find_Orb has trouble with finding highly accurate orbit estimations for unstable orbits. For this reason, the ephemeris used for the initial state guess is not sufficiently accurate. This causes the Least-Squares solution to diverge exponentially away from the initial state guess and not find a realistic initial state and parameter estimation. This problem especially occurred when finding working estimation windows in 2014 to 2016 for the Chang'e 3 booster, as apparently errors in the Find_Orb estimation are often too large to use as an initial guess. It took many tries before finding converging estimation for unstable windows 5 and 6, which could be included in the list of use cases. This issue is further elaborated upon in section D.2.

In conclusion, effort is required for finding a suitable set of use cases to develop a cislunar space debris model framework on. One needs to put careful consideration in designing the selection of use cases and choosing the corresponding estimation windows is not always straightforward. It is therefore important to take this into account for the design of future cislunar space debris studies.

A.2 Use case visualization

This sections presents three-dimensional plots of the Chang'e 2 and Chang'e 3 booster orbits for the 13 estimation windows shown in section A.1. Both the in-sample and out-of-sample orbits are plotted to visualize the stability characteristics described in Table A.1. The 3D orbits for the various use cases are shown in Figure A.4 and Figure A.5, and can be used as a visual reference for understanding the behaviour seen in all results presented in the paper.

The orbits visualizations are based on orbit propagations of the tailored model framework that are known to be sufficiently accurate, thus having a total residual smaller than 900 arcsec. For estimation of unstable windows 5 and 6 of the Chang'e 2 booster, out-of-sample residual of the propagation quickly diverges and thus is not suitable for visualizing out-of-sample orbit. This is caused by a close Moon approach, as shown in subsection 4.4. Therefore, the sufficiently accurate propagated in-sample orbit of estimation window 7, which overlaps with the out-of-sample orbit of windows 5 and 6, is used as a replacement for the out-of-sample orbit of windows 5 and 6.

Analyzing the behaviour of the three-dimensional orbits, it can be observed that windows 1-4 of both objects are quite predictable and do not undergo any sudden changes in orbital elements. On the other hand, in window 5 to 7 of the Chang'e 2 booster, the sudden change in orbital elements (due to the close Moon approach) can be clearly observed. For windows 5 and 6 of the Chang'e 3 booster, the out-of-sample instability is apparent.

One important note, is that the in-sample estimated orbit for (unstable in-sample) window 6 of the Chang'e 3 booster does not look very unstable. Which raises questions on the validity of the accuracy of the Keplerian elements retrieved from Find_Orb TLEs (as well as the Find_Orb orbit estimations) in this period for Chang'e 3 booster. Further suspicion is also raised by the surprisingly low out-of-sample uncertainty for window 6 in Figure C.6. The inaccuracy of the Find_Orb estimations and TLEs for this

period, is partly confirmed from comparison between tailored model framework orbit propagations and Find_Orb estimations in section D.2. To fully confirm that the in-sample orbit is actually stable and not unstable, residuals of the Find_Orb estimations and TLEs should be analyzed to validate whether they have poor accuracy in this period. For now, estimation window 6 of the Chang'e 3 booster will still be characterized to be unstable in-sample.

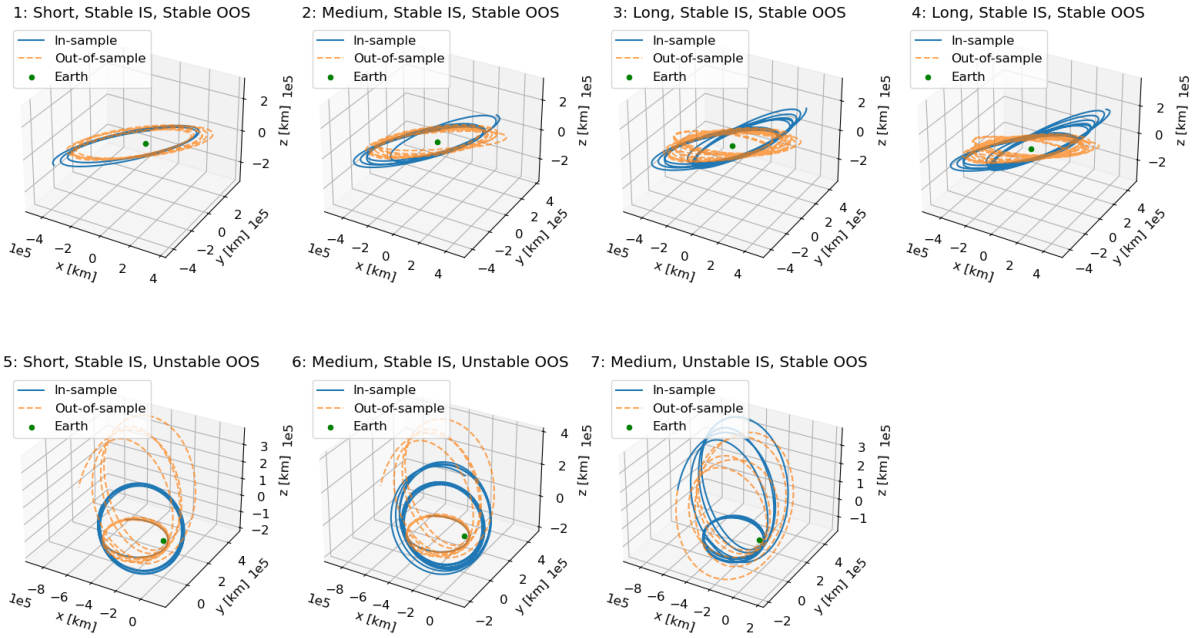


Figure A.4: 3D orbit visualizations for Chang'e 2 use cases using tailored model framework estimations

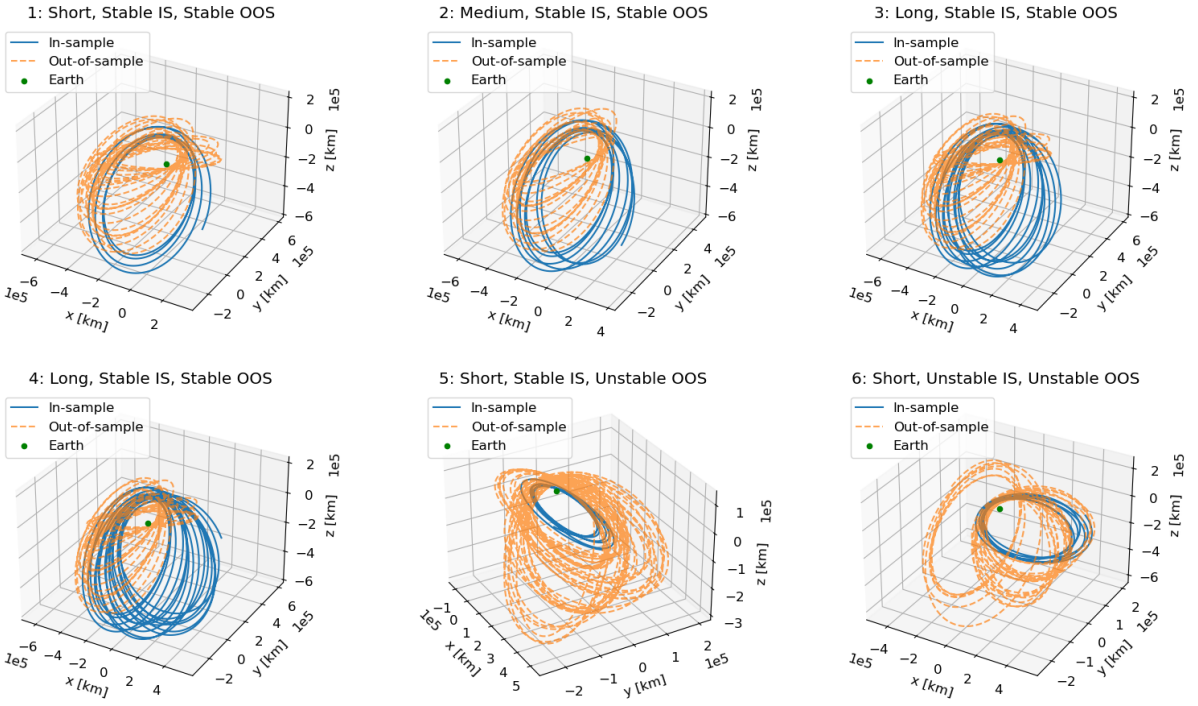


Figure A.5: 3D orbit visualizations for Chang'e 3 use cases using tailored model framework estimations

B

Generic model design, development and tailoring

In section 3 and section 4 of the paper, the design choices, development process and results of model framework development are summarized. This chapter, describes in great detail what this process looks like and what results have lead to the final configurations of the generic and tailored model framework.

First, in section B.1, an overview of all preliminary design choices as well as their characteristics (made during the literature review phase) is shown. Afterwards, the corresponding testing plan is described in detail in section B.2. The results of the dynamical model and integrator tests are shown in section B.3 and section B.4 respectively. In section B.5, the generic model framework for all use cases is presented along with its performance in various visualizations. Finally, in section B.6, the results of tailoring on each use case individually are shown, including any findings made in intermediate steps.

B.1 Preliminary design choices

In section 3, an overview is shown of all design choices required in model framework development. These include dynamical models, integrators, propagators, orbit determination algorithm, uncertainty propagator, rotation models, ephemeris models and much more. Optimizing all these settings concurrently, can be very challenging and might not add too many value. Especially because some design choices expect to contribute more to performance (accuracy vs computational load) than others.

The paper describes that different dynamical model and integrator configurations are tested to develop the generic model framework. All other design choices are made beforehand. Though, it is important to note that during the literature phase of this research, these other design choices have been analyzed extensively to make a reasonable decision on which are most suitable for the cislunar space debris problem. For example, the standard Spice kernels used in Tudat have been found to be suitable for the ephemeris and rotation models preliminary. Furthermore, even the dynamical model and integrator tests (described in section B.2) are purposely limited to only candidates with high potential. The literature review of all design choices including characteristics can be found in this section.

Firstly, a more elaborate list of dynamical model choices (compared to Table 2 of the paper) is shown in Table B.1. Here, not only their availability in Tudat, but also their characteristics are summarized. Precision level indicates the level of accuracy the dynamical model has compared to reality, impact is a measure of the effect of the corresponding acceleration on cislunar space debris and comp. time describes the computational load that the dynamical model adds to the model framework. The characteristics have been found from an extensive literature review. It has been used as a guide to identify which dynamical models should be included in the testing phase shown in Table B.6, and which are likely not worth the effort.

Furthermore, the characteristics of the various numerical integrators, most of which have been tested, are shown in Table B.3. Besides their availability in Tudat, the advantages and disadvantages are described for each integrator. These characteristics can be used to further justify the test results shown in section B.4.

Most design choices presented here for the dynamical model and integrator configuration, have actually been tested during model development. But as mentioned, some preliminary design choices have also been made based on comparison using literature. More specifically a literary trade-off is made for the orbit determination algorithm and orbit propagator. As is discussed in section 3, the Weighted Least-Squares algorithm and Cowell propagator have been picked. This is mainly because they are robust and easy to use, which are useful properties for orbit determination and propagation of cislunar space debris. For the rotation and ephemeris models of the bodies, the default Spice kernels in Tudat have been found to be most suitable. But other design choices considered in this process, are summarized with advantages and disadvantages in Table B.2 and Table B.4.

Table B.1: Overview of dynamical models for cislunar space, the characteristics and weights have been determined from the qualitative analysis described in the following sections (RP = Radiation Pressure, SRP = Solar Radiation Pressure)

Accelerations	**Precision level**	**Impact**	**Comp. time**	**In Tudat?**
Gravity - Point-mass	Low	High	Low	Yes
Gravity - Spherical harmonics	High	High	Medium - high	Yes
Gravity - Relativistic effects	High	Medium	Medium - high	Yes
Gravity - Ocean tides	High	Low	Medium	No
Gravity - Terrestrial tides	High	Medium	Medium	Yes
SRP - Cannonball model SRP	Low - medium	High	Low	Yes
SRP - Panelled model SRP	High	High	High	No
SRP - Shadow effects	High	Medium	Low	Yes
SRP - Yarkovski effect	High	Low	High	No
SRP - Poynting-Robertson drag	High	Low	High	No
RP - Earth/Moon radiation pressure	Low - medium	Low	Low	Yes
Other - Atmospheric drag (US76 model)	Medium	Uncertain	Low	Yes
Other - Empirical accelerations	High	Medium	Medium	Yes

Table B.2: Characteristics of orbit determination methods [Montenbruck et al., 2000]

	Advantages	**Disadvantages**	**In Tudat?**
Weighted Least-Squares	- Robust and easy to use - Good at handling bad datapoints - Divergence between iterations rarely occurs - Competitive for small datasets	- Multiple iterations required, longer computation time - Requires more storage	Yes
(Extended) Kalman Filter	- Little storage required - Faster computation time - Process noise accounts for unmodeled dynamics - Process noise allows higher weighting for recent observations	- Divergence occurs more easily, without process noise - Problems may occur, for large initial uncertainties - Harder to interpret and backtrace	No

Table B.3: Characteristics of integration methods [Montenbruck et al., 2000][Hofsteenge, 2013]

Integrator	Advantages	Disadvantages	In Tudat?
Euler integrator	Easy to use	Poor performance for high eccentricity	Yes
Multi-stage, fixed step-size: RK4	Easy to use Applicable for wide range of problems	Poor performance for high eccentricity	Yes
Multi-stage, variable step-size: RKFp(q), DOPRIp(q)	Higher order equals higher efficiency DOPRI8(7) useful when velocity determines acceleration No storage required Useful for long-term propagation RKF7(8) and DOPRI8(7) work best for high-eccentricities	Performs poorly if time between output is smaller than step size DOPRI shows odd behaviour for low tolerances	Yes
Multi-step, fixed step-size: AB(p)	Multi-step increases efficiency	Poor performance for high eccentricity Worse performance than ABM(p) integrators Needs to be initialized Require storage of past data points Can't handle discontinuities in dynamics, errors from SRP eclipse Can get stuck at small step sizes	Yes
Multi-step, variable step-size: ABM(p)	Highly efficient Applicable for eccentric and non-eccentric orbits Computation time relatively lower for eccentric orbits Can have small distance between output points Optimal for low to moderate eccentricities	Needs to be initialized Require storage of past data points Can't handle discontinuities in dynamics, errors from SRP eclipse Can get stuck at small step sizes	Yes
Extrapolation: BS(p), ODEX2	Useful for long-term propagation Works best for high-eccentricities	Efficiency decreases if time between output smaller than step size Extrapolators provide sparse output	Yes
Symplectic integrators	Useful for long-term propagations Good for Hamiltonian systems	Most methods are fixed step-size Energy assumption does not work for high eccentricity	No

Table B.4: Characteristics of propagation methods [Montenbruck et al., 2000]

Propagator	Advantages	Disadvantages	In Tudat?
Cowell	Straightforward, simple EOM Robust propagator Formulation does not depend on behaviour assumptions Works well for highly perturbed orbits (entry)	Accuracy can be a bit low in some cases Large state derivative ->chance of large numerical errors Large state derivative variations ->hard to adapt time step	Yes
Encke	Small state derivative Small state derivative variations	Poor long-term performance Differential equations are more complex Faster variations in state elements Additional source of error from solving Kepler's equation	Yes
Gauss-Kepler	Easy to use Works well for lightly perturbed motions, far from singularities	Conversion from Keplerian to Cartesian hard for large r Singularity for orbits close to hyperbolic orbits	Yes
Gauss-MEE	Works well for all lightly perturbed motions	Singularities left at inclination of 0 and 180 degrees	Yes
USM - Quaternions	No singularities Works well for all lightly perturbed motions	Doesn't work if angular momentum is close to 0	Yes
USM - Rodrigues Parameters	No singularities Works well for all lightly perturbed motions	Doesn't work if angular momentum is close to 0	Yes
USM - Exponential Map	No singularities Works well for all lightly perturbed motions	Doesn't work if angular momentum is close to 0	Yes
(Semi-)analytical: Lambert targeter, Two-Line Elements etc.	Useful when accuracy is not very important	Poor performance compared to other propagators	No

B.2 Testing plan

As explained in section B.1, different design choices are available for the dynamical model and integrator configuration. A quick summary on the testing and development process is given in subsection 3.7 of the paper, but the actual testing plan is not shown.

In order to find the best model configuration for the generic model framework, it is decided to start with a highly accurate model framework including all relevant dynamics and an accurate integrator (being RKF7(8) with an error tolerance of 10^{-10}). This baseline framework is shown in Table B.5. Afterwards, the influence of each individual dynamical model is tested by removing it from the dynamical model configuration and testing the new model performance compared to the highly accurate model framework. The same is done for different integrator configurations. Either a setting is changed or a new integrator is tried, and the model performance is compared against the highly accurate framework. This has lead to the testing plan shown in Table B.6.

Once all tests are performed, the effect of each individual change is analyzed using the performance metrics described in subsection 3.6. This way a generic model framework is made with design choices that contribute to high accuracy, but maintain reasonable computational load. The results of the dynamical model and integrator tests can be found in section B.3 and section B.4. As an extra validation step of this individual model testing plan, in the end the generic model framework is tested in its entirety against the baseline. The results are analyzed for confirmation of the generic framework, and shown in Table B.12.

Table B.5: Baseline model framework used for comparison during model testing

Dynamical models	Spherical harmonics: - Earth (SH[10, 10]) - Moon (SH[10, 10])
	Point-mass: - Sun - Mercury - Venus - Mars - Jupiter - Saturn - Uranus - Neptune
	Cannonball model: - Solar radiation pressure (incl. conical shadow) - Earth radiation pressure
	Aerodynamic drag
	Earth relativistic correction (Schwarzschild)
Integrator	RKF7(8) variable dt - Error tolerance = 10^{-10} - Initial step-size = 1.0 s - Max step-size = 10000 s
Propagator	Cowell
Ephemeris model	Spice kernels (Tudat default)
Rotation model	Spice kernels (Tudat default)
Estimated parameters	Initial state

Table B.6: Overview of testing plan (it=iterative)

Test number	Test description
Dyn. test 1 (it)	Increase SH Earth
Dyn. test 2 (it)	Decrease SH Earth
Dyn. test 3 (it)	Increase SH Moon
Dyn. test 4 (it)	Decrease SH Moon
Dyn. test 5	Remove Mercury
Dyn. test 6	Remove Venus
Dyn. test 7	Remove Mars
Dyn. test 8	Remove Jupiter
Dyn. test 9	Remove Saturn
Dyn. test 10	Remove Uranus
Dyn. test 11	Remove Neptune
Dyn. test 12	Remove Earth radiation pressure
Dyn. test 13	Remove aerodynamic drag
Dyn. test 14	Remove Schwarzschild correction

Test number	Test description
Int. test 1 (it)	RKF7(8) - error tolerance / 10
Int. test 2 (it)	RKF7(8) - error tolerance * 10
Int. test 3	DOPRI8(7) - error tolerance / 10
Int. test 4	DOPRI8(7) - error tolerance * 10
Int. test 5	RKF7(8) - fixed step size
Int. test 6	RK4
Int. test 7	ABM (order 6-11)
Int. test 8	BS6

B.3 Dynamical model test results

In Table B.7 and Table B.8, the results of all dynamics tests (shown in Table B.6) are presented. The tables can be used as a reference to explain the dynamical model design choices made for the generic model framework, described in Table B.10.

The table should be read as follows. For each test, the effect of the individual dynamical model on performance metrics is shown with respect to the baseline framework with all dynamics (shown in Table B.5). The 'Baseline' column shows the performance of the baseline model framework in terms of computation time, in-sample (IS) RMSE and out-of-sample (OOS) RMSE. The following columns present the percentage change in performance due to the individual model test with respect to this baseline.

B.4 Integrator test results

Similarly to the dynamic test results tables, in Table B.9 the performance of the various integrator tests (shown in Table B.6) with respect to the baseline model framework is shown. The baseline model framework uses a variable step-size RKF7(8) integrator with error tolerance of 10^{-10}. Comparing all integrator configurations, it becomes apparent that this baseline integrator configuration (RKF7(8) 10^{-10}) is optimal considering the high accuracy and reasonable computational load requirements. Higher error tolerances give unnecessary computational load, lower error tolerances deteriorate out-of-sample RMSE, and other integrators also do not result in improvements.

Table B.7: Dynamical model test results per use case (part 1)

Object	Window	Metric	Baseline	Earth SH[20,20]	Earth SH[5,5]	Earth SH[2,2]	Moon SH[20,20]	Moon SH[5,5]	Moon SH[2,2]	Moon PM
Chang'e 2	1-SSS	Comp. time (s)	34,8	2,2%	-6,3%	-6,0%	2,1%	-13,3%	-16,1%	-19,4%
		RMSE IS (arcsec)	3,8	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
		RMSE OOS (arcsec)	210558,1	0,0%	0,0%	0,1%	0,0%	0,0%	0,0%	-15,7%
	2-MSS	Comp. time (s)	72,6	7,8%	-2,3%	-2,9%	16,3%	-3,4%	-4,8%	-9,4%
		RMSE IS (arcsec)	5,9	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
		RMSE OOS (arcsec)	137524,5	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	2,4%
	3-LSS	Comp. time (s)	91,5	8,6%	-1,5%	-2,3%	17,5%	-3,6%	-6,0%	-9,5%
		RMSE IS (arcsec)	11,8	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
		RMSE OOS (arcsec)	95748,9	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
	4-LSS	Comp. time (s)	116,5	8,6%	-2,6%	1,8%	15,9%	-5,4%	-5,3%	-8,6%
		RMSE IS (arcsec)	11,7	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
		RMSE OOS (arcsec)	133079,5	0,0%	0,0%	-0,2%	0,0%	0,0%	0,0%	-1,6%
	5-SSU	Comp. time (s)	54,8	9,2%	0,6%	-10,6%	17,2%	-43,0%	-5,3%	-14,7%
		RMSE IS (arcsec)	7,9	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,1%
		RMSE OOS (arcsec)	168928,7	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,4%
	6-MSU	Comp. time (s)	101,0	5,1%	-0,7%	-1,4%	17,7%	-9,1%	-8,7%	-7,3%
		RMSE IS (arcsec)	17,0	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
		RMSE OOS (arcsec)	213435,5	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
	7-MUS	Comp. time (s)	143,1	10,5%	-8,8%	-9,2%	6,4%	-12,6%	-14,9%	-20,5%
		RMSE IS (arcsec)	9,0	0,0%	0,0%	-1,5%	0,0%	0,0%	0,0%	-0,1%
		RMSE OOS (arcsec)	9937,5	0,1%	0,0%	2,4%	0,1%	0,0%	-0,1%	-1,5%
Chang'e 3	1-SSS	Comp. time (s)	66,0	7,9%	8,5%	8,3%	19,0%	3,8%	6,3%	-35,2%
		RMSE IS (arcsec)	11,6	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
		RMSE OOS (arcsec)	277,0	0,0%	0,0%	-0,8%	0,0%	0,0%	0,0%	0,0%
	2-MSS	Comp. time (s)	113,1	6,7%	-5,4%	-7,0%	13,5%	-8,9%	-9,6%	-12,7%
		RMSE IS (arcsec)	35,4	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
		RMSE OOS (arcsec)	1150,6	0,0%	0,0%	-0,3%	0,0%	0,0%	0,0%	0,0%
	3-LSS	Comp. time (s)	164,1	8,9%	-36,6%	-2,6%	16,1%	-4,1%	-8,7%	-9,8%
		RMSE IS (arcsec)	45,6	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
		RMSE OOS (arcsec)	2420,1	0,0%	0,0%	0,1%	0,0%	0,0%	0,0%	0,0%
	4-LSS	Comp. time (s)	252,3	7,2%	-5,6%	-2,8%	20,7%	-2,9%	-39,6%	-33,6%
		RMSE IS (arcsec)	45,7	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
		RMSE OOS (arcsec)	2251,8	0,0%	0,0%	0,1%	0,0%	0,0%	0,0%	0,0%
	5-SSU	Comp. time (s)	80,9	3,4%	-14,4%	-6,8%	9,3%	-25,0%	-16,7%	-22,8%
		RMSE IS (arcsec)	2,8	0,0%	0,0%	0,1%	0,0%	0,0%	0,0%	0,0%
		RMSE OOS (arcsec)	32157,0	0,0%	0,2%	156,3%	0,0%	0,0%	0,0%	-0,2%
	6-SUU	Comp. time (s)	80,4	11,6%	0,1%	-2,7%	17,6%	-4,8%	-6,6%	-10,7%
		RMSE IS (arcsec)	6,0	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
		RMSE OOS (arcsec)	1795,6	-0,1%	0,0%	-7,4%	0,0%	0,0%	0,0%	-0,2%

Table B.8: Dynamical model test results per use case (part 2)

Object	Window	Metric	Baseline	Mercury	Venus	Mars	Jupiter	Saturn	Uranus	Neptune	Earth RP	Drag	Schwarz.
Chang'e 2	1-SSS	Comp. time (s)	34,8	-12,2%	-12,3%	-11,4%	-14,4%	-14,5%	-13,6%	-15,1%	-43,7%	-19,6%	-12,4%
		RMSE IS (arcsec)	3,8	0,0%	-0,5%	0,0%	-0,3%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
		RMSE OOS (arcsec)	210558,1	-15,6%	-21,1%	0,6%	-18,0%	-23,4%	0,1%	-0,4%	-0,8%	-10,5%	0,1%
	2-MSS	Comp. time (s)	72,6	-1,4%	-1,9%	-3,1%	-2,4%	-2,6%	-1,3%	-2,3%	-48,1%	-17,9%	-0,5%
		RMSE IS (arcsec)	5,9	0,0%	-0,2%	0,0%	0,1%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
		RMSE OOS (arcsec)	137524,5	-12,1%	4,0%	0,1%	-5,0%	-9,7%	0,0%	0,0%	0,0%	-2,1%	0,0%
	3-LSS	Comp. time (s)	91,5	2,1%	2,9%	2,3%	2,0%	2,1%	3,1%	3,3%	-49,1%	-17,8%	-0,9%
		RMSE IS (arcsec)	11,8	0,0%	-0,1%	0,0%	-0,4%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
		RMSE OOS (arcsec)	95748,9	0,0%	0,0%	0,0%	4,2%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
	4-LSS	Comp. time (s)	116,5	0,0%	-3,7%	-7,3%	-2,5%	1,6%	-3,7%	2,1%	-49,0%	-16,2%	-2,6%
		RMSE IS (arcsec)	11,7	0,0%	-0,1%	0,0%	-0,5%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
		RMSE OOS (arcsec)	133079,5	0,0%	-1,6%	0,1%	-6,8%	5,9%	0,3%	0,0%	-0,5%	-1,8%	-0,2%
	5-SSU	Comp. time (s)	54,8	-2,9%	-3,4%	-4,4%	-5,5%	-28,3%	-3,6%	1,0%	-44,7%	-12,3%	-2,3%
		RMSE IS (arcsec)	7,9	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
		RMSE OOS (arcsec)	168928,7	0,0%	0,1%	0,0%	-0,4%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
	6-MSU	Comp. time (s)	101,0	1,6%	-4,3%	1,4%	0,8%	-4,0%	-0,5%	-2,1%	-57,5%	-43,4%	6,9%
		RMSE IS (arcsec)	17,0	0,0%	-0,1%	0,0%	-0,1%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
		RMSE OOS (arcsec)	213435,5	0,0%	0,0%	0,0%	-0,1%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
	7-MUS	Comp. time (s)	143,1	-10,9%	-12,8%	-21,0%	-13,3%	-11,0%	-10,5%	-8,3%	-61,0%	-26,6%	-6,7%
		RMSE IS (arcsec)	9,0	-0,1%	0,0%	0,0%	-1,2%	-0,1%	0,0%	0,0%	0,0%	0,0%	0,0%
		RMSE OOS (arcsec)	9937,5	-0,2%	0,6%	0,1%	-2,7%	-0,5%	0,1%	0,0%	0,1%	0,0%	0,1%
Chang'e 3	1-SSS	Comp. time (s)	66,0	-29,3%	10,8%	8,7%	8,5%	2,5%	-12,6%	9,4%	-59,0%	-5,0%	-16,2%
		RMSE IS (arcsec)	11,6	0,0%	-0,4%	0,0%	-0,1%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
		RMSE OOS (arcsec)	277,0	0,0%	-12,8%	0,9%	-20,0%	-0,5%	0,0%	0,0%	0,2%	0,0%	0,0%
	2-MSS	Comp. time (s)	113,1	1,6%	-3,4%	-6,5%	-6,8%	-4,1%	-5,7%	0,8%	-61,3%	-15,0%	-3,0%
		RMSE IS (arcsec)	35,4	0,0%	0,4%	0,0%	1,2%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
		RMSE OOS (arcsec)	1150,6	0,0%	0,1%	0,3%	-5,6%	-0,1%	0,0%	0,0%	0,1%	0,0%	0,0%
	3-LSS	Comp. time (s)	164,1	-36,3%	-0,5%	-1,3%	-1,7%	-0,4%	-0,6%	0,9%	-49,9%	-12,6%	2,4%
		RMSE IS (arcsec)	45,6	0,0%	0,7%	0,0%	2,0%	0,1%	0,0%	0,0%	0,0%	0,0%	0,0%
		RMSE OOS (arcsec)	2420,1	-0,1%	0,8%	-0,1%	7,6%	0,2%	0,0%	0,0%	0,0%	0,0%	0,0%
	4-LSS	Comp. time (s)	252,3	-2,3%	-1,5%	1,0%	-36,1%	-0,1%	0,0%	2,1%	-70,1%	-7,8%	4,3%
		RMSE IS (arcsec)	45,7	0,0%	0,6%	0,0%	1,5%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
		RMSE OOS (arcsec)	2251,8	-0,1%	0,8%	-0,2%	8,2%	0,2%	0,0%	0,0%	0,0%	0,0%	0,0%
	5-SSU	Comp. time (s)	80,9	-22,7%	-20,8%	-9,4%	-10,2%	-13,0%	-10,9%	-11,1%	-56,1%	-21,7%	-10,5%
		RMSE IS (arcsec)	2,8	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
		RMSE OOS (arcsec)	32157,0	0,0%	0,5%	0,0%	-2,5%	0,2%	0,0%	0,0%	-0,6%	-0,1%	0,0%
	6-SUU	Comp. time (s)	80,4	-1,3%	-3,0%	-2,2%	-3,6%	-18,7%	-2,7%	-1,3%	-52,1%	-13,3%	-1,1%
		RMSE IS (arcsec)	6,0	0,0%	2,3%	0,0%	0,2%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
		RMSE OOS (arcsec)	1795,6	0,1%	-5,1%	-0,1%	1,2%	0,1%	0,0%	0,0%	-0,1%	0,0%	0,0%

Table B.9: Integrator test results per use case

Object	Window	Metric	Baseline	RKF7(8) e-11	RKF7(8) e-9	DOPRI8(7) e-10	DOPRI8(7) e-9	DOPRI8(7) e-8	RK4	BS(6)
Chang'e 2	1-SSS	Comp. time (s)	34,5	-8,2%	-9,4%	75,3%	-4,0%	-9,2%	653,3%	240,2%
		RMSE IS (arcsec)	3,8	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
		RMSE OOS (arcsec)	210558,1	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
	2-MSS	Comp. time (s)	70,4	0,5%	8,5%	254,4%	16,5%	8,0%	764,5%	310,5%
		RMSE IS (arcsec)	5,9	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
		RMSE OOS (arcsec)	137524,5	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
	3-LSS	Comp. time (s)	88,5	1,6%	-1,4%	278,6%	5,8%	-1,8%	751,2%	333,3%
		RMSE IS (arcsec)	11,8	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
		RMSE OOS (arcsec)	95748,9	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
	4-LSS	Comp. time (s)	113,8	1,4%	-1,6%	275,8%	8,7%	-0,6%	759,3%	332,6%
		RMSE IS (arcsec)	11,7	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
		RMSE OOS (arcsec)	133079,5	0,1%	0,1%	1,3%	0,7%	1,3%	-1,6%	0,0%
	5-SSU	Comp. time (s)	53,3	-34,7%	-6,3%	-2,8%	-0,2%	-5,9%	320,0%	248,2%
		RMSE IS (arcsec)	7,9	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
		RMSE OOS (arcsec)	168928,7	0,0%	-0,1%	0,1%	0,0%	0,0%	0,9%	0,0%
	6-MSU	Comp. time (s)	99,7	8,7%	-1,4%	54,8%	2,8%	-3,5%	441,3%	304,1%
		RMSE IS (arcsec)	17,0	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
		RMSE OOS (arcsec)	213435,5	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
	7-MUS	Comp. time (s)	139,1	14,9%	-3,7%	560,7%	35,6%	-2,4%	544,8%	462,4%
		RMSE IS (arcsec)	9,0	0,0%	0,0%	0,0%	0,0%	0,0%	-0,3%	0,0%
		RMSE OOS (arcsec)	9937,5	-0,3%	10,3%	-0,2%	0,8%	35,3%	-0,9%	-0,4%
Chang'e 3	1-SSS	Comp. time (s)	67,6	-12,7%	-26,0%	-4,3%	9,0%	7,3%	450,6%	220,2%
		RMSE IS (arcsec)	11,6	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
		RMSE OOS (arcsec)	277,0	0,0%	-0,2%	0,1%	0,1%	0,1%	0,1%	0,1%
	2-MSS	Comp. time (s)	116,1	10,5%	-0,7%	90,3%	0,5%	-1,2%	390,4%	267,8%
		RMSE IS (arcsec)	35,4	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
		RMSE OOS (arcsec)	1150,6	0,0%	-0,1%	0,0%	0,0%	0,0%	0,0%	0,0%
	3-LSS	Comp. time (s)	164,9	2,8%	3,7%	168,5%	-1,9%	-2,0%	383,0%	117,3%
		RMSE IS (arcsec)	45,6	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
		RMSE OOS (arcsec)	2420,1	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
	4-LSS	Comp. time (s)	248,5	3,2%	-55,3%	56,9%	0,2%	-3,1%	367,7%	288,8%
		RMSE IS (arcsec)	45,7	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
		RMSE OOS (arcsec)	2251,8	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
	5-SSU	Comp. time (s)	73,4	-3,6%	-8,3%	14,4%	0,6%	-10,2%	429,0%	165,5%
		RMSE IS (arcsec)	2,8	0,0%	0,0%	0,0%	0,0%	0,0%	0,7%	0,0%
		RMSE OOS (arcsec)	32157,0	-0,1%	1,3%	-0,1%	-0,1%	0,4%	9,4%	0,0%
	6-SUU	Comp. time (s)	81,4	8,6%	-5,6%	44,7%	-1,1%	-7,4%	506,6%	239,4%
		RMSE IS (arcsec)	6,0	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
		RMSE OOS (arcsec)	1795,6	0,1%	-1,7%	0,2%	0,1%	-0,1%	24,5%	0,3%

B.5 Generic model framework

All steps mentioned in the previous sections, have lead to to the generic model framework shown in Table B.10. This model framework has the highest accuracy for all use cases, while maintaining a reasonable computational load. In the following sections, the performance of the generic model framework for each use case will be presented in different ways.

Table B.10: Generic model configuration

Dynamical models	Spherical harmonics: - Earth (SH[5,5])
	Point-mass: - Sun - Moon - Jupiter
	SRP (cannonball) - C_r = 1.2 - A = 37.14 m2 - Shadow Earth
Integrator	RKF7(8) variable dt - Error tolerance = 10^{-10} - Initial step-size = 1.0 s - Max step-size = 10000 s
Propagator	Cowell
Ephemeris models	Spice kernels (Tudat default)
Rotation models	Spice kernels (Tudat default)
Estimated parameters	Initial state

B.5.1 Performance

The overall performance, based on the metrics discussed in subsection 3.6 of the paper, can be seen in Table B.11. Different insights can already be drawn from this.

Before analyzing, some important notes need to be made on the out-of-sample RMSE values. Firstly, out-of-sample RMSE can only be compared between windows of the same object ending on the same date. This is because the time of observation out-of-sample differs a lot over time. If observations were continuous with constant spacing, out-of-sample RMSE could be compared more easily. Furthermore, the length of the out-of-sample window differs between most use cases. This period is chosen for each use case specifically, making sure to relatively close to the first accuracy threshold violation.

Analyzing the performance, it is seen that the generic model framework (without any use case tailoring) already achieves Period of Sufficient Accuracy (PSA) up to nearly 2 years. It is especially accurate for the Chang'e 3 booster, which can be attributed to the fact that its orbit is analyzed in Appendix A to be less non-linear than the orbit of the Chang'e 2 booster. Even the unstable Chang'e 3 use cases, achieve good out-of-sample performance. Use cases 1-4 of the Chang'e 2 booster, which are stable in-sample and out-of-sample, achieve similar PSA. But windows 5-7, which contain a close approach to the Moon either in-sample or out-of-sample are clearly under performing. Use case specific model tailoring will be performed in section B.6, which allows for accuracy improvements for many of the use cases.

Finally, a comparison is made between the generic and the baseline model framework, shown in Table B.5. The percentage difference in performance metrics of the baseline is shown with respect to those of the generic model. It is observed that the generic model framework is computationally much more efficient (the baseline tends to be 2-8 times slower). Furthermore, accuracy of the generic model is similar to the accuracy of the baseline. For windows 1 of the Chang'e 2 booster and windows 1 & 6 of the Chang'e 3 booster, the out-of-sample RMSE of the baseline is even worse. This is expected to occur because the windows are so short, that errors can occur for certain model configurations in the orbit determination which are not necessarily fixed by adding a more complex dynamical model. In general, it can be said that the generic model configuration (shown in Table B.10) is a significantly more efficient model than the baseline whilst maintaining high accuracy.

Table B.11: Generic model performance (*do not compare RMSE OOS between estimation windows ending on different dates)

Object	Window	Comp. time [s]	RMSE IS [arcsec]	RMSE OOS* [arcsec]	PSA [days]
Chang'e 2	1-SSS	8,1	3,7	166553,1	145-228
	2-MSS	13,4	5,9	144565,0	145-228
	3-LSS	15,9	11,8	95774,9	369-418
	4-LSS	19,5	11,7	134097,0	369-418
	5-SSU	8,7	8,0	169836,6	49-163
	6-MSU	15,8	17,0	213514,8	47-49
	7-MUS	100,8	9,0	9763,7	90-91
Chang'e 3	1-SSS	8,7	11,5	242,9	>672
	2-MSS	13,3	35,5	1154,7	248-295
	3-LSS	27,0	45,9	2437,7	434-436
	4-LSS	32,3	46,0	2268,9	434-436
	5-SSU	12,3	2,8	32149,8	676-678
	6-SUU	13,0	6,1	1699,7	501-609

Table B.12: Baseline model performance vs generic (gen.) model performance

Object	Window	Δ Comp. time	Δ RMSE IS	Δ RMSE OOS	PSA all [d]	PSA gen. [d]
Chang'e 2	1-SSS	283,4%	0,5%	26,4%	145-228	145-228
	2-MSS	445,8%	0,2%	-4,9%	145-228	145-228
	3-LSS	521,6%	0,1%	0,0%	369-418	369-418
	4-LSS	525,7%	0,1%	-0,8%	369-418	369-418
	5-SSU	536,5%	-0,1%	-0,5%	49-163	49-163
	6-MSU	554,5%	0,1%	0,0%	47-49	47-49
	7-MUS	705,7%	0,4%	1,8%	90-91	90-91
Chang'e 3	1-SSS	649,8%	0,4%	14,0%	248-295	>672
	2-MSS	797,8%	-0,4%	-0,4%	248-295	248-295
	3-LSS	626,2%	-0,7%	-0,7%	434-436	434-436
	4-LSS	683,0%	-0,6%	-0,8%	434-436	434-436
	5-SSU	562,2%	0,0%	0,0%	676-678	676-678
	6-SUU	564,1%	-2,2%	5,6%	501-609	501-609

B.5.2 Estimated parameters

In the generic model framework, only the initial state is estimated. The initial state estimations for each use case are summarized in Table B.13. These values can be used as comparison to the parameters estimated for the tailored model configuration in subsection B.6.4.

Table B.13: Estimated parameters in generic model framework

Obj.	Window	$x(t_0)$ [m]	$y(t_0)$ [m]	$z(t_0)$ [m]	$V_x(t_0)$ [m/s]	$V_y(t_0)$ [m/s]	$V_z(t_0)$ [m/s]
CH2	1-SSS	-4,59E+08	-2,77E+08	-1,23E+08	236,5	-536,6	473,9
	2-MSS	3,32E+08	1,91E+08	1,65E+08	-86,7	733,8	-491,1
	3-LSS	3,72E+08	3,35E+08	1,91E+08	-250,5	508,4	-478,7
	4-LSS	4,48E+08	2,14E+08	2,99E+08	-70,6	630,7	-324,3
	5-SSU	-3,77E+08	-1,19E+08	1,92E+08	-220,0	-199,6	-518,4
	6-MSU	-6,69E+07	7,89E+07	3,08E+08	-907,1	-287,6	571,8
	7-MUS	-1,45E+08	-4,92E+07	-1,02E+08	1477,7	493,2	49,7
CH3	1-SSS	3,04E+08	-1,39E+08	-2,50E+08	176,3	423,9	860,8
	2-MSS	3,54E+08	-1,76E+08	-3,47E+08	197,8	298,6	571,1
	3-LSS	1,78E+08	1,12E+08	1,20E+08	-1380,9	-2,1	593,5
	4-LSS	4,22E+08	1,88E+08	-9,38E+07	-525,3	210,9	576,1
	5-SSU	3,61E+08	-1,49E+08	6,97E+07	442,6	153,6	-102,1
	6-SUU	4,56E+08	3,70E+07	-1,36E+08	-177,7	447,1	-105,4

B.5.3 Residuals

Before, only the performance metrics for the generic model have been discussed. But it can be useful to analyze the actual residuals in-sample and out-of-sample, as well as their distributions.

In Figure B.1 and Figure B.2, a comparison is made between the in-sample and out-of-sample total residuals for all use cases of the Chang'e 2 and 3 upper stages. The time axis is measured in days with respect to t_E, being the end of the in-sample estimation window. This is of course also the split between in-sample and out-of-sample. Since tailoring will still be performed on most use cases (as discussed in subsection 3.7 of the paper), no in-depth analysis will be performed on the behaviour of the residuals over time.

What is interesting to analyze, are the Probability Density Functions (PDFs) of the right ascension and declination residuals in Figure B.3 and Figure B.4. The distributions are an indication of how well the current model framework is able to fit the optical observations. The more Gaussian the shape of both the right ascension and declination distribution, the better the fit. The opposite can be said for PDFs which are more randomly distributed over the residual spectrum. Analyzing the PDF histograms, it can be seen that for most use cases, the current model framework does not fit the observations well. It is hypothesized, that this is caused by the fact that only the initial state is currently estimated. While the radiation pressure coefficient and observation biases are large uncertainties. Estimating these parameters is expected to improve the fit, thus improving both performance and residual distribution for use cases that currently have a non-Gaussian distribution.

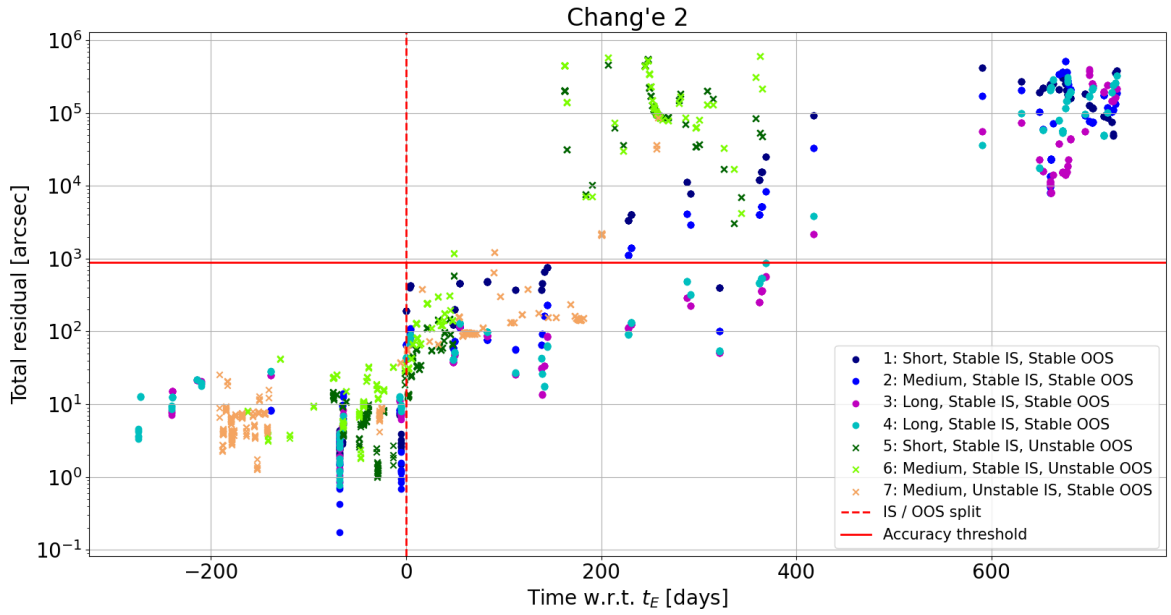


Figure B.1: Total residual of generic model propagation over time per Chang'e 2 use case, left of dashed line are residuals in training set, right are residuals in test set

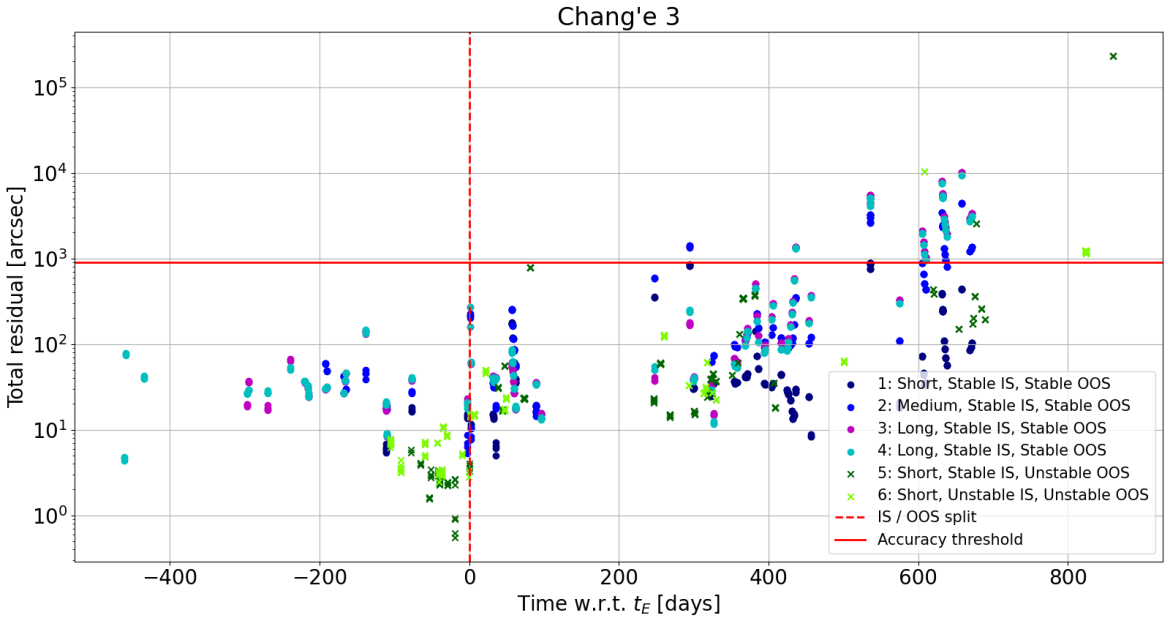


Figure B.2: Total residual of generic model propagation over time per Chang'e 3 use case, left of dashed line are residuals in training set, right are residuals in test set

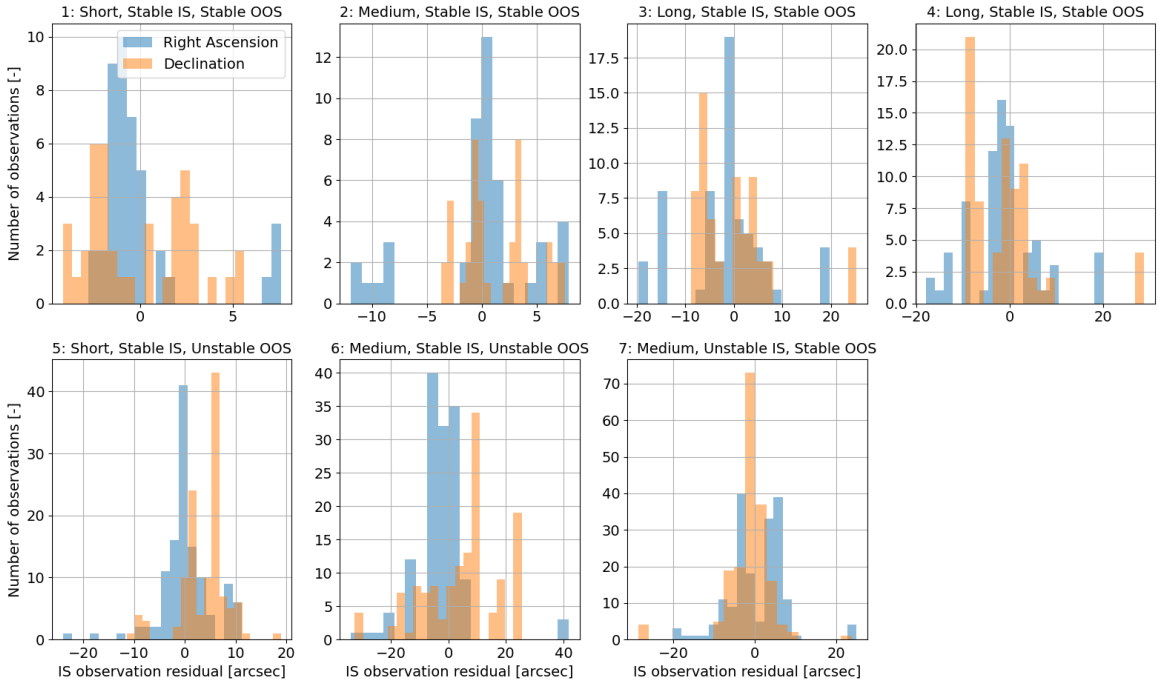


Figure B.3: PDF of generic model framework in-sample residuals per use case of Chang'e 2

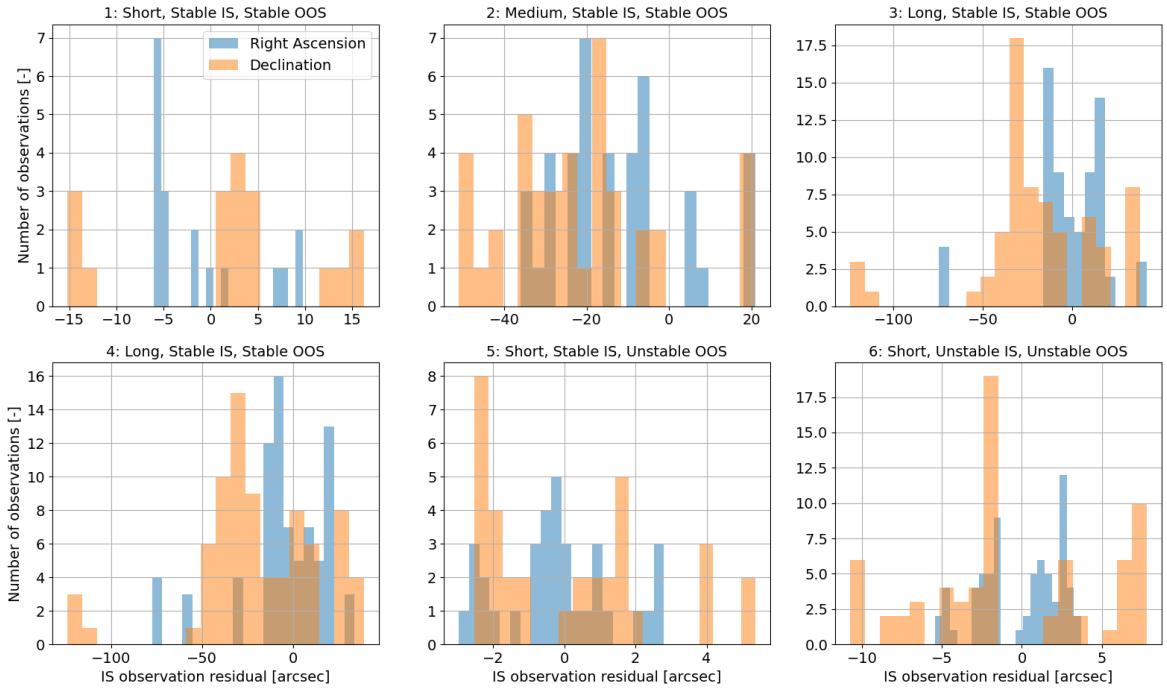


Figure B.4: PDF of generic model framework in-sample residuals per use case of Chang'e 3

B.6 Model tailoring

As mentioned in the previous section, the generic model framework is limited in performance, though being applicable to all use cases. Since only initial state is being estimated and no other information is included in the estimation, the fit can be poor for some use cases.

To solve this, the generic model framework is tailored for each use case individually. This makes it no longer applicable to all cislunar space debris use cases, but is essential for significantly increasing the model framework performance. The possible model improvements that can be made for each use case are, estimating the radiation pressure coefficient (C_r) and/or the observation bias (ϵ_{obs}). These are found to be the largest unknowns, so far the C_r has been guessed at 1.2 and no observation biases are taken into account in the generic model framework. Another way to improve the fit out-of-sample, and especially prevent overfitting, is to include an a priori covariance matrix (explained in subsection 3.3 of the paper). This way one can include information on the accuracy of the initial state and parameter guess used for initialization of the orbit determination algorithm. The values used in the a priori covariance matrix, can be seen as an approximation of the amount that an estimated parameter can change per iteration of the Least-Squares algorithm. This can prevent the algorithm from overfitting. Finally, adding realistic values of the observation weights (being the random noise in the observations), can also improve the fit. So far, Tudat has assumed a random noise of 1 rad, which is clearly an overly loose requirement.

It is found that for C_r an a priori covariance input of 0.1 is realistic and for observation bias an input between 10^{-6} and 10^{-5} radians is reasonable (0.2 - 2 arcsec). Furthermore, a realistic random noise between 1 and 1.5 arcsec is tested with. Testing various configurations of the estimated parameter set, observation weight and a priori covariance matrix, the final tailored model configurations are shown in Table B.14. It becomes apparent that not one set of tailoring settings works for all use cases. The process and findings of individual tailoring will be presented in the following sections.

Table B.14: Tailored model configurations leading to best performance for each use case

Object	Window	Parameters	Obs. weight	A priori cov.	A priori input
Chang'e 2	1-SSS	Initial state, $C_r + \epsilon_{obs}$	1.0 arcsec	Yes	$\sigma_{C_r} = 0.1$, $\sigma_{\epsilon_{obs}} =$ 1E-5
	2-MSS	Initial state, $C_r + \epsilon_{obs}$	1.5 arcsec	Yes	$\sigma_{\epsilon_{obs}} =$ 1E-5
	3-LSS	Initial state	1.0 arcsec	No	-
	4-LSS	Initial state	1.0 arcsec	No	-
	5-SSU	Initial state, C_r	1.0 arcsec	No	-
	6-MSU	Initial state, C_r	1.0 arcsec	No	-
	7-MUS	Initial state, $C_r + \epsilon_{obs}$	1.0 arcsec	Yes	$\sigma_{C_r} = 0.1$, $\sigma_{\epsilon_{obs}} =$ 1E-6
Chang'e 3	1-SSS	Initial state	1.0 arcsec	No	-
	2-MSS	Initial state, $C_r + \epsilon_{obs}$	1.0 arcsec	Yes	$\sigma_{C_r} = 0.1$, $\sigma_{\epsilon_{obs}} =$ 1E-6
	3-LSS	Initial state, $C_r + \epsilon_{obs}$	1.0 arcsec	No	-
	4-LSS	Initial state, $C_r + \epsilon_{obs}$	1.0 arcsec	No	-
	5-SSU	Initial state	1.0 arcsec	No	-
	6-SUU	Initial state	1.0 arcsec	No	-

B.6.1 Intermediate results of tailoring process

In the process of finding the best model configurations presented in Table B.14, various steps have been taken. First, it has been analyzed what effect estimating C_r and/or ϵ_{obs} has on model performance. For this process an observation weight of 1.0 arcsec is used, to make sure the both C_r and ϵ_{obs} can be estimated realistically. The results with respect to the generic model performance is shown in Table B.15, Table B.16 and Table B.17.

It becomes apparent, that simply estimating C_r and/or ϵ_{obs} easily leads to overfitting and its effectiveness is very use case specific. Furthermore, for short windows estimating C_r can lead to unreasonable values of 3.1, -0.9, 14.2 etc. At this point, it is hypothesized that adding an a priori covariance matrix might, at least partially, solve this problem. But simply estimating C_r and/or ϵ_{obs} does not always seem ideal.

Table B.15: Model performance when estimating C_r and initial state

Object	Window	C_r [-]	Δ RMSE IS	Δ RMSE OOS	PSA gen. [days]	PSA new [days]
Chang'e 2	1-SSS	3,1	-36,4%	28,3%	145-228	145-228
	2-MSS	1,9	-14,4%	29,7%	145-228	145-228
	3-LSS	1,9	-24,2%	55,3%	369-418	145-228
	4-LSS	1,6	-18,9%	-10,8%	369-418	145-228
	5-SSU	1,5	-24,8%	-26,8%	49-163	49-163
	6-MSU	1,7	-21,4%	-35,8%	47-49	49-163
	7-MUS	1,6	-36,0%	-49,4%	90-91	182-201
Chang'e 3	1-SSS	-0,9	-8,8%	1519,6%	>672	35-57
	2-MSS	1,7	-68,4%	-36,7%	248-295	457-536
	3-LSS	1,5	-60,3%	-79,2%	434-436	457-536
	4-LSS	1,5	-61,5%	-72,2%	434-436	457-536
	5-SSU	2,9	-48,0%	41,2%	676-678	73-82
	6-SUU	1,6	-14,5%	67,5%	501-609	330-501

Table B.16: Model performance when estimating ϵ_{obs} and initial state

Object	Window	C_r [-]	Δ RMSE IS	Δ RMSE OOS	PSA gen. [days]	PSA new [days]
Chang'e 2	1-SSS	1,2	-86,5%	31,3%	145-228	145-228
	2-MSS	1,2	-91,0%	-10,4%	145-228	145-228
	3-LSS	1,2	-62,6%	46,4%	369-418	145-228
	4-LSS	1,2	-64,1%	17,0%	369-418	145-228
	5-SSU	1,2	-22,6%	-23,9%	49-163	49-163
	6-MSU	1,2	-42,8%	-10,0%	47-49	49-163
	7-MUS	1,2	-36,0%	-49,4%	90-91	182-201
Chang'e 3	1-SSS	1,2	-86,8%	4613,3%	>672	35-57
	2-MSS	1,2	-66,5%	64,3%	248-295	457-536
	3-LSS	1,2	-19,7%	-33,9%	434-436	457-536
	4-LSS	1,2	-25,0%	-31,6%	434-436	457-536
	5-SSU	1,2	-48,0%	41,2%	676-678	73-82
	6-SUU	1,2	-86,0%	14861,6%	501-609	330-501

Table B.17: Model performance when estimating C_r, ϵ_{obs} and initial state

Object	Window	C_r [-]	Δ RMSE IS	Δ RMSE OOS	PSA gen. [days]	PSA new [days]
Chang'e 2	1-SSS	14,2	-86,6%	30,1%	145-228	145-228
	2-MSS	0,7	-91,3%	7,6%	145-228	145-228
	3-LSS	1,2	-62,6%	70,9%	369-418	145-228
	4-LSS	1,2	-64,1%	-2,9%	369-418	145-228
	5-SSU	1,4	-26,8%	-20,9%	49-163	49-163
	6-MSU	1,3	-45,4%	-28,1%	47-49	49-163
	7-MUS	1,7	-44,8%	-52,8%	90-91	182-201
Chang'e 3	1-SSS	2,3	-87,5%	4665,0%	>672	35-57
	2-MSS	1,5	-78,9%	-43,2%	248-295	457-536
	3-LSS	1,5	-76,7%	-94,8%	434-436	457-536
	4-LSS	1,5	-78,3%	-93,9%	434-436	457-536
	5-SSU	2,9	-48,0%	41,2%	676-678	73-82
	6-SUU	1,4	-91,0%	1453,5%	501-609	330-501

B.6.2 Performance of realistic model configuration

Before doing tailoring with a priori covariance matrices for each use case individually, it is first analyzed what the effect of one realistic model configuration with these design choices would be for all use cases.

For the realistic configuration, both C_r and ϵ_{obs} are estimated with an observation weight of 1.0 arcsec and an a priori covariance matrix. The inputs of the a priori covariance matrix are quite strict at 0.1 for C_r and 10^{-6} (0.2 arcsec) for ϵ_{obs}. The results are shown in Table B.18. It can be seen that the adding the a priori covariance matrix does result in some improvements (compared to the results in subsection B.6.1), but are definitely not consistent and sometimes lead to extremely poor results. For unstable estimation window 5 of the Chang'e 3 upper stage, the realistic configuration did not even lead to a converging estimation. Thus maybe not one a priori covariance matrix configuration fits all use cases. This is confirmed when analyzing the performance of the individually tailored model framework, shown in the next section.

Table B.18: Realistic model vs generic model performance

Object	Window	Δ Comp. time	Δ RMSE IS	Δ RMSE OOS	PSA gen. [d]	PSA real. [d]
Chang'e 2	1-SSS	26,9%	-72,1%	-35,0%	145-228	145-228
	2-MSS	17,5%	-62,1%	44,3%	145-228	145-228
	3-LSS	24,6%	-48,8%	146,7%	369-418	231-288
	4-LSS	22,9%	-6,0%	33,6%	369-418	145-228
	5-SSU	21,8%	-26,8%	-27,0%	49-163	49-163
	6-MSU	-1,0%	-39,8%	-35,5%	47-49	49-163
	7-MUS	8,0%	-31,1%	-63,5%	90-91	90-91
Chang'e 3	1-SSS	39,7%	-23,1%	2016,2%	672-0	248-295
	2-MSS	29,0%	-71,6%	-21,1%	248-295	610-632
	3-LSS	1,7%	-75,5%	-92,9%	434-436	434-436
	4-LSS	2,6%	-76,7%	-90,3%	434-436	434-436
	5-SSU	0,0%	0,0%	0,0%	676-678	676-678
	6-SUU	11,8%	3315,8%	16557,5%	501-609	0-1

B.6.3 Performance of tailored model configurations

Investigating the overall performance of the tailored model configuration described in Table B.14, in Table B.19 and Table B.20, it is seen that clear improvements can be made with respect to the generic model. Individual tailoring has lead to significant performance improvements for 8 out of 13 use cases. Significant increases in Period of Sufficient Accuracy (PSA) can also be seen. All windows for the Chang'e 3 booster are now sufficiently accurate for 1.5 - 2 years. Out-of-sample performance for short to medium sized windows of the Chang'e 2 booster has also improved significantly (window 1, 2, 5, 6 an 7).

Table B.19: Tailored (tail.) performance with respect to generic model framework performance (gen.)

Object	Window	Δ Comp. time	Δ RMSE IS	Δ RMSE OOS	PSA gen. [d]	PSA tail. [d]
Chang'e 2	1-SSS	32,49%	-83,39%	-67,92%	145-228	145-228
	2-MSS	10,93%	-75,61%	-50,19%	145-228	228-231
	3-LSS	0,0%	0,00%	0,00%	369-418	369-418
	4-LSS	0,0%	0,00%	0,00%	369-418	369-418
	5-SSU	-9,96%	-24,80%	-26,78%	49-163	49-163
	6-MSU	-13,97%	-21,44%	-35,84%	47-49	49-163
	7-MUS	-76,07%	-31,08%	-63,49%	90-91	182-201
Chang'e 3	1-SSS	0,0%	0,00%	0,00%	>672	>672
	2-MSS	40,60%	-71,62%	-21,14%	248-295	457-536
	3-LSS	-35,88%	-76,65%	-94,80%	434-436	>672
	4-LSS	2,27%	-78,30%	-93,88%	434-436	>672
	5-SSU	0,0%	0,00%	0,00%	676-678	676-678
	6-SUU	0,0%	0,00%	0,00%	501-609	501-609

Table B.20: Tailored model performance (*RMSE OOS can only be compared between estimation windows of an object ending on the same date)

Object	Window	Comp. time [s]	RMSE IS [arcsec]	RMSE OOS* [arcsec]	PSA [days]
Chang'e 2	1-SSS	10,7	0,6	53432,8	145-228
	2-MSS	14,9	1,4	72011,7	228-231
	3-LSS	15,7	11,8	95774,9	369-418
	4-LSS	18,4	11,7	134097,0	369-418
	5-SSU	7,8	6,0	124351,4	49-163
	6-MSU	13,6	13,4	136985,1	49-163
	7-MUS	24,1	6,2	3565,2	182-201
Chang'e 3	1-SSS	9,5	11,5	242,9	>672
	2-MSS	18,8	10,1	910,6	457-536
	3-LSS	17,3	10,7	126,7	>672
	4-LSS	33,0	10,0	138,8	>672
	5-SSU	12,0	2,8	32149,8	676-678
	6-SUU	13,1	6,1	1699,7	501-609

B.6.4 Estimated parameters of tailored model framework

Investigating the estimated parameters of the tailored model framework in Table B.21 and Table B.22, a couple of interesting findings are made.

When estimating C_r leads to better performance, its value seems to lie between 1.5 and 1.8. But for some cases, estimating C_r did not lead to better performance than simply using the initial guess of 1.2.

Furthermore, analyzing the estimated observation biases for the different ground stations, it is seen that the biases are really high. Much higher than the 1 arcsec guideline of the Minor Planet Center (MPC) but also higher than the 3 arcseconds discussed in subsection 3.4 of the paper. Furthermore, the biases for the same ground station between estimation windows with overlapping observation sets, for example window 1 and 2 of Chang'e 2, are also widely different. It is hypothesized that this is caused due to the estimated observation bias catching any other model errors like dynamical model uncertainties in its magnitude. This can be solved by adding empirical accelerations to the model framework. In general, it can be said that the observation bias estimation can be used to improve out-of-sample performance, but should not be used for direct analysis of ground station accuracy.

Finally, comparing the initial state estimations of Table B.21 to Table B.13, it is seen that the initial states do not seem to differ that much. This is a good indication of the fact that at least the initial state estimation can be assumed to be realistic and accurate.

Table B.21: Estimated initial state and C_r of tailored model configurations

Object	Window	$x(t_0)$ [m]	$y(t_0)$ [m]	$z(t_0)$ [m]	$V_x(t_0)$ [m/s]	$V_y(t_0)$ [m/s]	$V_z(t_0)$ [m/s]	C_r [-]
Chang'e 2	1-SSS	-4,59E+08	-2,77E+08	-1,23E+08	236,5	-536,4	473,9	1,80
	2-MSS	3,32E+08	1,91E+08	1,65E+08	-86,5	734,0	-491,0	1,58
	3-LSS	3,72E+08	3,35E+08	1,91E+08	-250,5	508,4	-478,7	1,20
	4-LSS	4,48E+08	2,14E+08	2,99E+08	-70,6	630,7	-324,3	1,20
	5-SSU	-3,77E+08	-1,19E+08	1,92E+08	-220,0	-199,6	-518,4	1,55
	6-MSU	-6,69E+07	7,90E+07	3,08E+08	-907,0	-287,7	571,7	1,67
	7-MUS	-1,45E+08	-4,91E+07	-1,02E+08	1477,9	493,0	49,7	1,53
Chang'e 3	1-SSS	3,04E+08	-1,39E+08	-2,50E+08	176,3	423,9	860,8	1,20
	2-MSS	3,54E+08	-1,76E+08	-3,48E+08	197,8	298,5	571,1	1,62
	3-LSS	1,77E+08	1,13E+08	1,20E+08	-1380,9	-2,4	593,7	1,50
	4-LSS	4,22E+08	1,88E+08	-9,38E+07	-525,2	210,8	576,0	1,50
	5-SSU	3,61E+08	-1,49E+08	6,97E+07	442,6	153,6	-102,1	1,20
	6-SUU	4,56E+08	3,70E+07	-1,36E+08	-177,7	447,1	-105,4	1,20

Table B.22: Estimated bias of tailored model configurations

Object	Window						
Chang'e 2	1-SSS	Ground station	703	C94	G68	K95	Q65
		Bias - RA [arcsec]	-12,2	-7,6	8,4	9,8	-9,8
		Bias - Decl. [arcsec]	-6,7	-7,6	10,4	14,5	-1,1
	2-MSS	Ground station	703	C94	G68	K95	Q65
		Bias - RA [arcsec]	2,2	2,3	-7,4	-9,0	0,1
		Bias - Decl. [arcsec]	0,0	-1,4	9,2	14,3	5,2
	7-MUS	Ground station	703	C95	G96		
		Bias - RA [arcsec]	13,8	15,8	17,7		
		Bias - Decl. [arcsec]	-0,5	-4,3	2,3		
Chang'e 3	2-MSS	Ground station	T05	T08			
		Bias - RA [arcsec]	-9,2	-11,1			
		Bias - Decl. [arcsec]	-1,4	-1,5			
	3-LSS	Ground station	I41	T05	T08		
		Bias - RA [arcsec]	-34,6	-29,1	-28,6		
		Bias - Decl. [arcsec]	-13,9	-2,1	-1,6		
	4-LSS	Ground station	703	I41	K88	T05	T08
		Bias - RA [arcsec]	-29,4	-34,3	-14,7	-29,1	-28,6
		Bias - Decl. [arcsec]	-2,5	-14,1	9,8	-2,1	-1,6

B.6.5 Residuals of tailored model framework

The total residual behaviour of the tailored model configurations over time can be seen in Figure B.5 and Figure B.6. The plots are also shown in Figure 5 and Figure 6 in the paper. In-depth comparison between different use cases is performed in subsection 4.3 of the paper.

As mentioned in subsection B.5.3, the PDFs of the generic model framework for various use cases indicate that it is a poor fit and improvements can still be made. Comparing the PDFs of the tailored model framework for each use case in Figure B.7 and Figure B.8, to those in Figure B.3 and Figure B.4, clear improvements can be seen. The residual distributions for both right ascension and declination have become much more Gaussian. Indicating that the tailored model configurations did not only improve overall performance, but also is a more realistic fit on the observations. Though, analyzing the PDFs of for example stable estimation window 2 of the Chang'e 3 booster (which has been tailored), it is seen that the tailored model framework still leaves room for improvement.

Furthermore, analyzing windows 1, 5 and 6 of the Chang'e 3 upper stage, it is seen that some model tailoring is really required to improve the fit for those windows. Though their out-of-sample performance is competitive (as seen in Table B.20), the fit is poor.

In conclusion, tailoring of the generic model framework for each use case individually is vital for improving out-of-sample performance. But elaborate tweaking of different settings (estimated parameters, observation weights, a priori covariance etc.) is required to actually achieve this. The tailoring step is essential, but it is not the solution to all problems and needs to be performed with care.

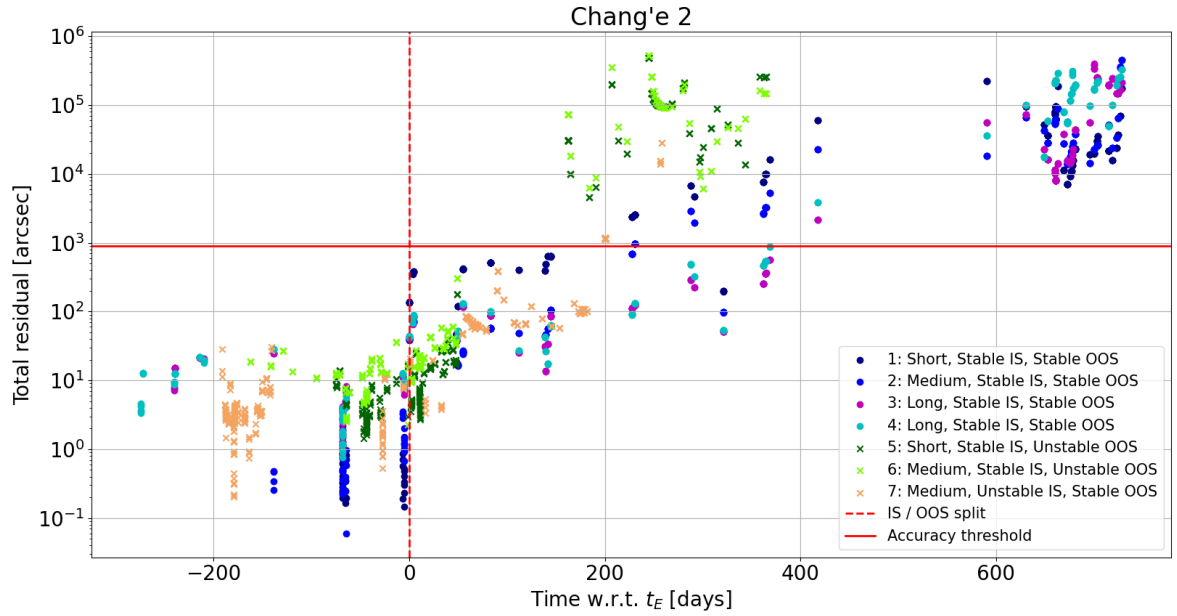


Figure B.5: Total residual of tailored propagations over time per Chang'e 2 use case, left of dashed line are residuals in training set, right are residuals in test set

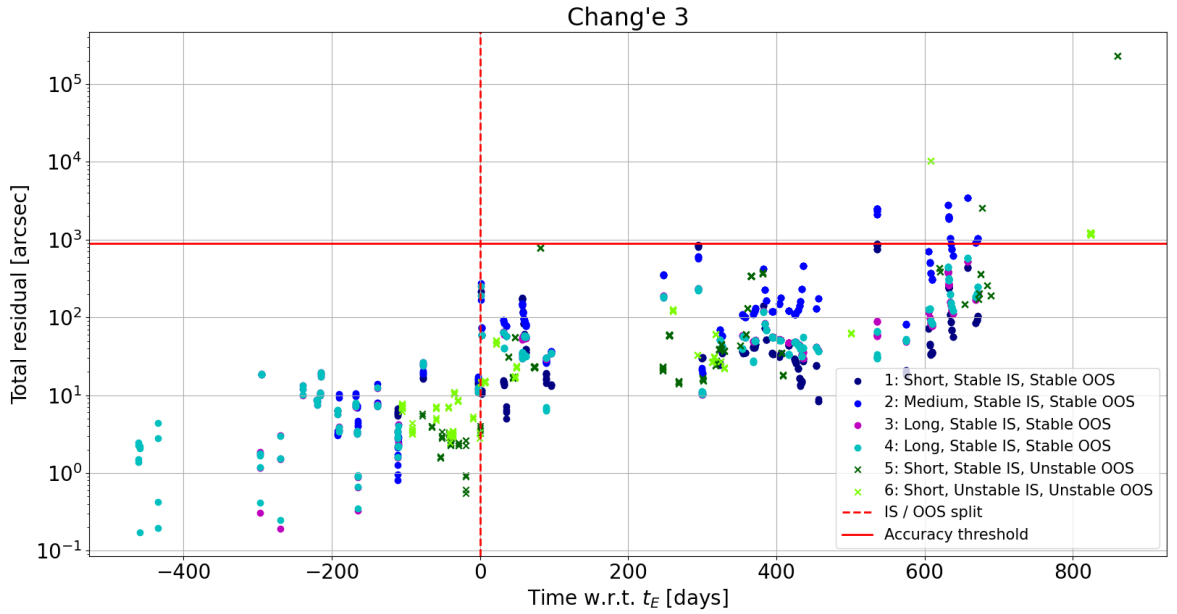


Figure B.6: Total residual of tailored propagations over time per Chang'e 3 use case, left of dashed line are residuals in training set, right are residuals in test set

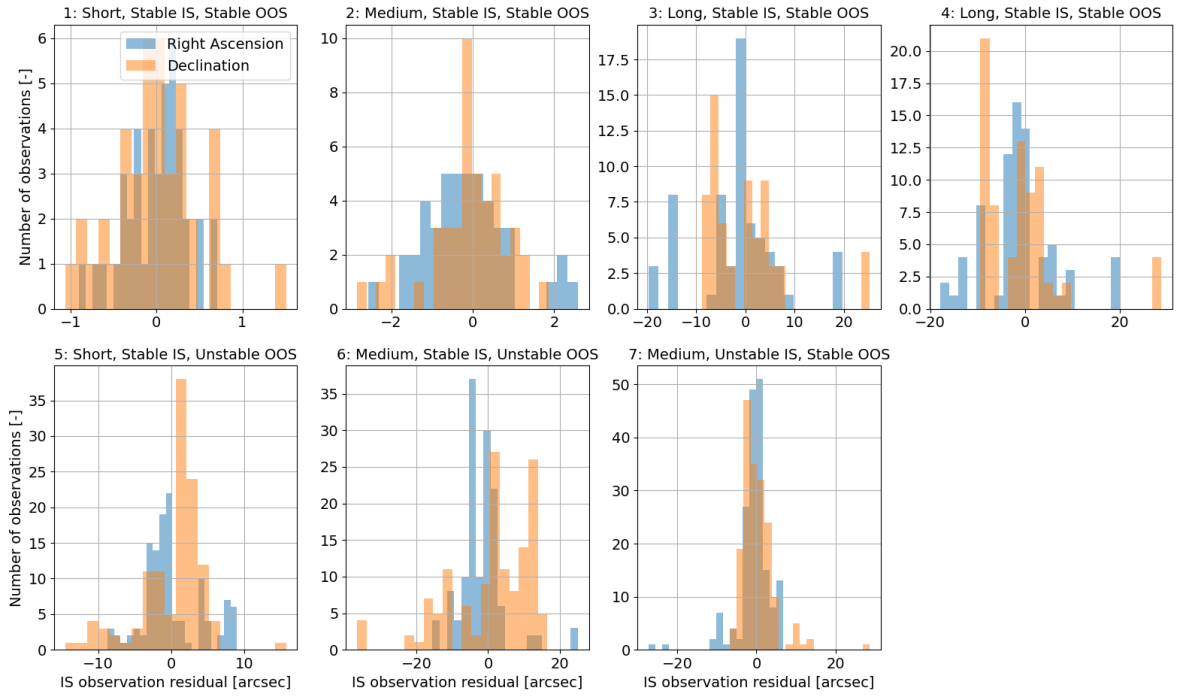


Figure B.7: PDFs of in-sample propagation residuals of tailored model per Chang'e 2 use case

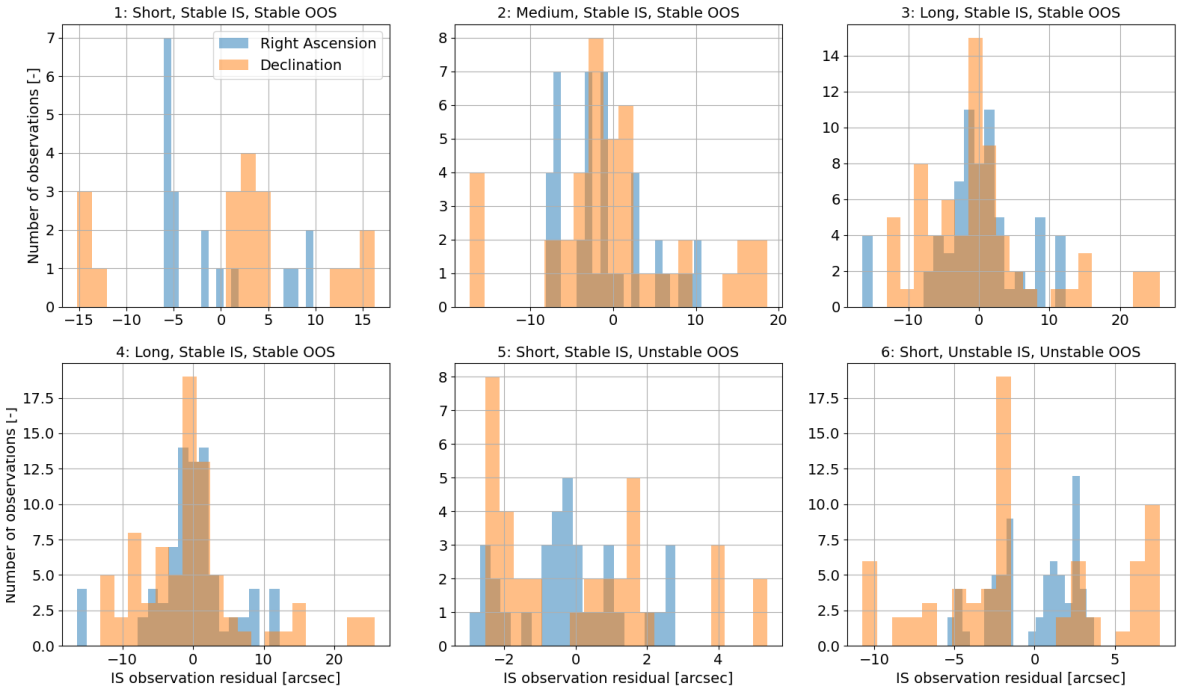


Figure B.8: PDFs of in-sample propagation residuals of tailored model per Chang'e 3 use case

C

Uncertainty propagation

This chapter provides additional material for the uncertainty analyses performed in subsection 4.5, based on the theory described in subsection 3.5 of the paper. First, many uncertainty propagators have been considered during the literature phase of this research which will be presented in section C.1. Afterwards, analyses on formal error and k-ratio are performed in section C.2, which have lead to the average k-ratios described in Table 10 of the paper. The Monte Carlo propagation results, including linearity tests and MC size verification, are presented for each use case in section C.3.

C.1 Uncertainty propagators

During the literature phase of this research, Monte Carlo (MC) simulation using the covariance matrix has been found to be suitable for the cislunar space debris problem. But several other uncertainty propagators have been considered. These can be categorized in linearized and non-linear uncertainty propagators. An overview can be found in Table C.1, including advantages and disadvantages of each method.

The ease of use and robust properties of the Monte Carlo approach have been found to be useful for the chaotic problem at hand, using similar reasoning to preliminary design choices made in section B.1. Furthermore, since only 2 objects are analyzed with in total 13 estimation windows, the computational inefficiency is found to be less relevant.

As suggested in subsection 4.5, one could also decide to apply non-linear uncertainty propagators like Polynomial Chaos Expansion (PCE) instead of the current Monte Carlo approach. As shown in Table C.1, PCE has already been applied in cislunar context [Chow et al., 2022][Giordano et al., 2023]. This is likely much more efficient and analysis on uncertainty propagation performance is more straightforward.

Table C.1: Overview and characteristics of uncertainty propagators

Uncertainty propagator	Advantages	Disadvantages	Applications
Linear - Local linearization	- Easy to implement - Computationally efficient	Poor accuracy for: - Highly non-linear systems - Large initial, non-Gaussian uncertainty - Long-term uncertainty propagation	- Autonomous mission design[1]
Linear - CCA	- Easy to implement - Computationally efficient - Good at quantifying effect of specific parameter uncertainties	Poor accuracy for: - Problems where realistic parameter variance is unknown	- GEO orbit determination incl. SRP and time bias[2]
Monte Carlo	- Provides a mean for non-linear and non-Guassian uncertainty propagation - Can achieve high accuracy results - Easy to implement	- Computationally expensive	- Space debris re-entry[3] - Space debris collision risk[4] - Cislunar SSA[5]
Non-linear - UT	- Computationally efficient - Existing dynamic solvers can be used	- Non-linear propagation only for first 2 moments - Insufficient highly non-Gaussian uncertainty - Poor long-term performance	- Space debris collision risk[6]
Non-linear - GMM	- Just the first two moments of the PDF need to be propagated	- Curse of dimensionality, can be computationally expensive	- Collision risk of GEO objects[7] - Uncertainty propagation with SRP and lunisolar perturbations[8] - Cislunar SSA[9]
Non-linear - PCE	- More efficient than Monte Carlo - Existing dynamic solvers can be used - Accurate for higher order PDF moments - Easy to implement	- Poor efficiency for high dimensional input uncertainty, curse of dimensionality - Application can become complex	- Uncertainty propagation of sun synchronous / Molniya orbits[10] - Satellite collision risk[11] - Cislunar transfer orbits[12]

[1][Geller, 2006]
[2][Cano et al., 2022]
[3][Virgili et al., 2021]
[4][Virgili et al., 2021][Hoogendoorn et al., 2018]
[5][Frueh, Howell, K. J. DeMars, et al., 2021]
[6][Virgili et al., 2021]
[7][K. J. DeMars, Cheng, et al., 2014]
[8][K. J. DeMars, Jah, et al., 2009]
[9][Chow et al., 2022]
[10][Jones, Doostan, and Born, 2013]
[11][Jones and Doostan, 2013][Jones, Parrish, et al., 2015]
[12][Giordano et al., 2023]

C.2 Formal errors and true-to-formal error ratio

As explained in subsection 3.5, a realistic mean and covariance at t_E is required to initialize the Monte Carlo simulation. The estimated covariance, retrieved from the Weighted Least-Squares algorithm and propagated to t_E, is usually an underestimation of the true covariance. This is further confirmed when comparing the formal errors of the estimation over time, to the actual residual or true error over time. The formal error tends to be an order magnitude lower than the true error. This is caused by imperfections in the dynamical model and unmodeled observation noise, which causes overconfidence in the fit. To compensate the estimated 'formal' covariance matrix for this difference, it is multiplied by the square of the average ratio k between formal and true error. This is shown in Equation 2.16 and Equation 2.17 of the paper.

In the following sections, an in-depth analysis will be performed on formal error and k-ratio over time. It will become more apparent why the k-ratio is averaged over the first three months out-of-sample, but also describe certain challenges with this approach.

C.2.1 Formal errors per use case

In Figure C.1 and Figure C.2, the formal errors over time in terms of right ascension and declination are plotted. Comparing to the total residuals in Figure 5 and Figure 6 of the paper, it becomes clear that the formal errors follow a similar increasing trend as the total residuals do. The total residual calculated from the right ascension and declination residual as shown in Equation 2.19. Visually, one can already tell that the formal errors, though being plotted in right ascension and declination, are smaller than the total residuals.

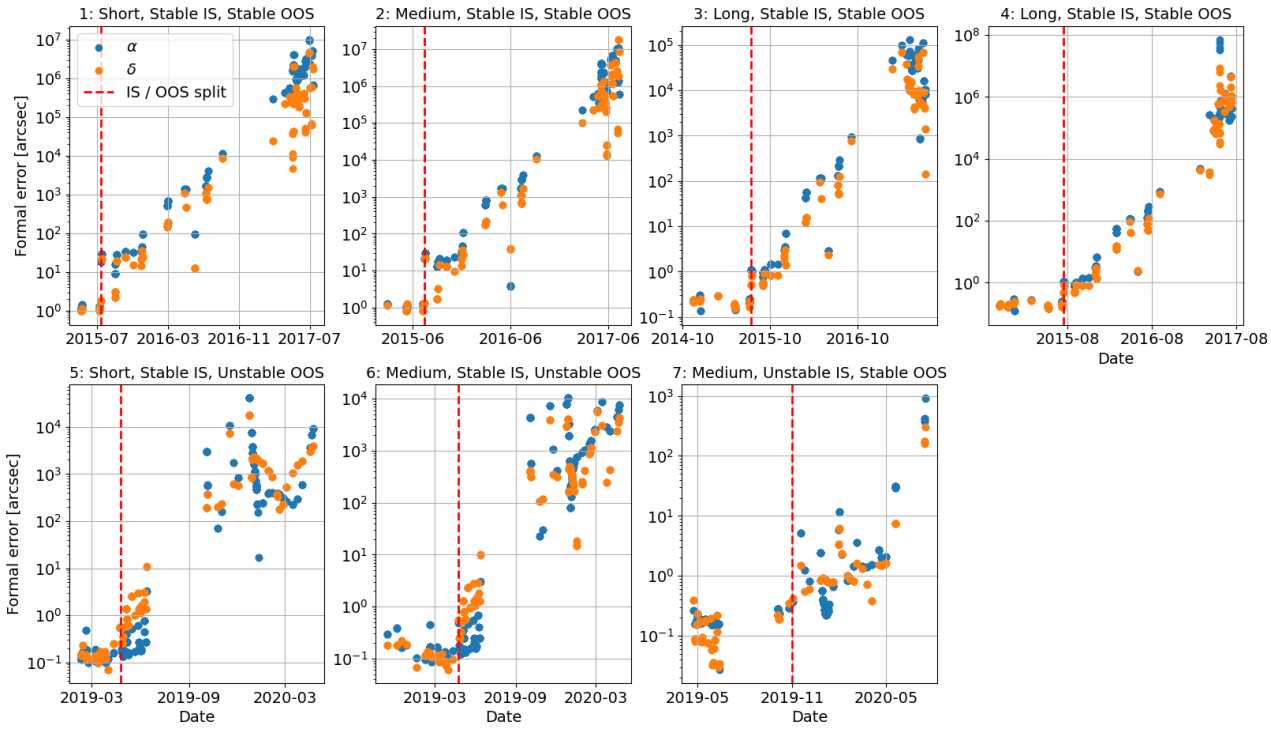


Figure C.1: Formal error of right ascension and declination over time for Chang'e 2 use cases

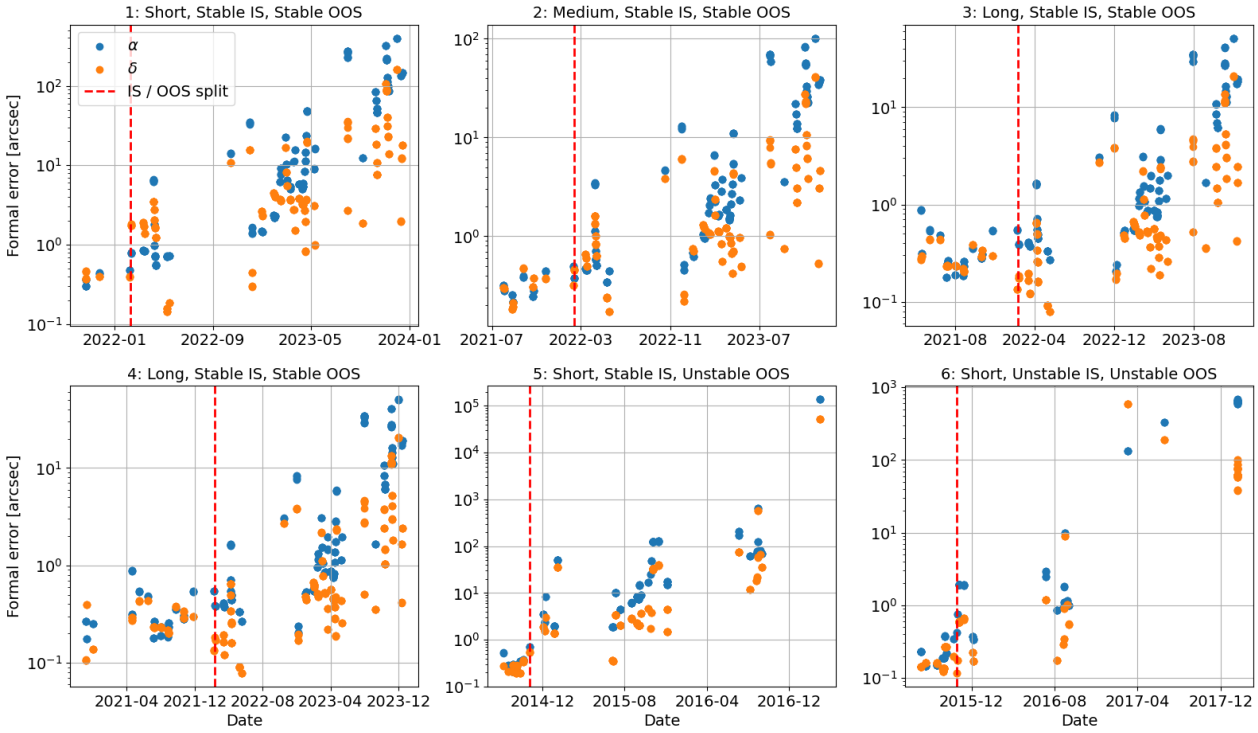


Figure C.2: Formal error of right ascension and declination over time for Chang'e 3 use cases

C.2.2 True-to-formal error ratio per use case

Dividing the residuals of the tailored model framework by the previously shown formal errors, the true-to-formal error ratio k, can be calculated. The k-ratio is plotted over time in Figure C.3 and Figure C.4.

A couple of things are noted. The true-to-formal error ratio is not constant, and for most use cases decreases over time. After a while, the ratio becomes smaller than 1, meaning formal errors have actually become larger than true errors. This behaviour is especially observed for stable use cases 1-4 of both objects, behaviour of the true-to-formal error ratio is much less predictable for unstable in-sample or out-of-sample use cases.

For this research, it is decided that the average k-ratio will be calculated for the first three months using Equation 2.16 in the paper. This k-ratio will be used for Monte Carlo initialization. From the plots, it becomes clear that this approach is more suitable for stable windows 1-4, as the k-ratio is more constant in this period than for the other unstable windows. Still, since all estimation windows have not yet crossed the accuracy threshold of 900 arcsec (described in subsection 3.6) in the first three months, this is assumed to be the best approximation of the actual true-to-formal error.

The median k-ratio in terms of right ascension (α) and declination (δ), as well as the average overall k-ratio, are shown in Table C.2.

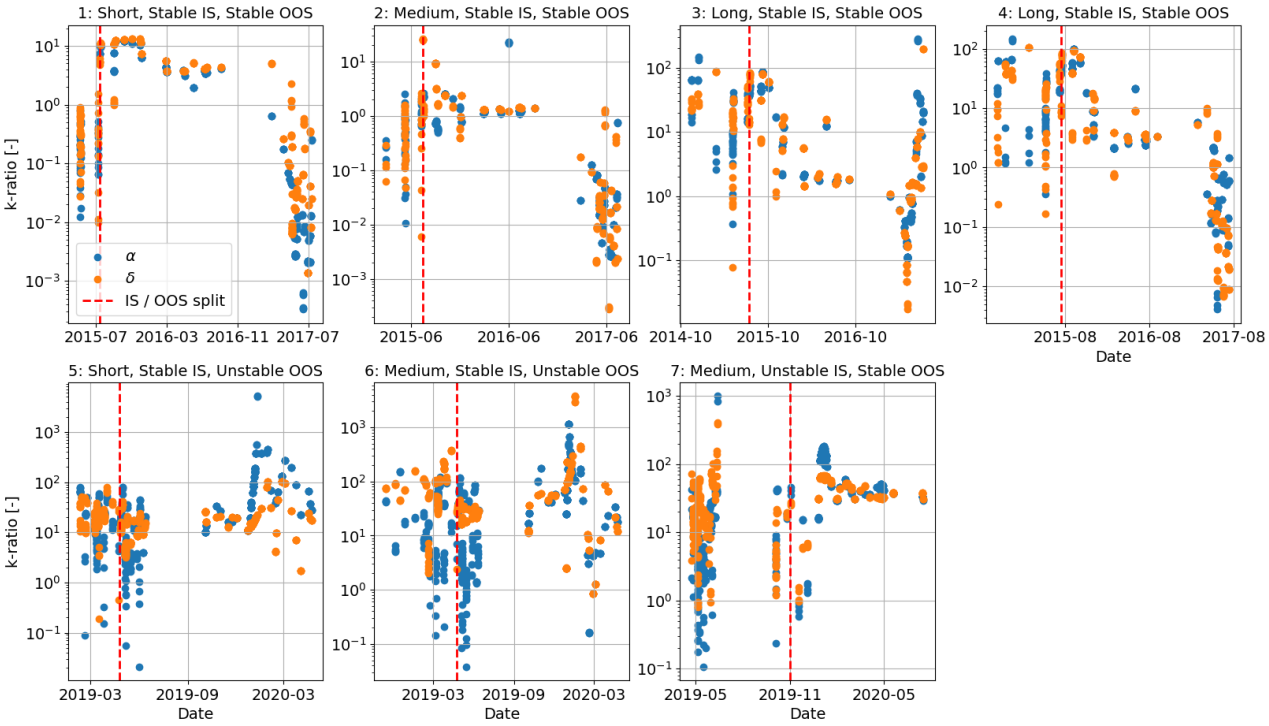


Figure C.3: k-ratio of right ascension and declination over time for Chang'e 2 use cases

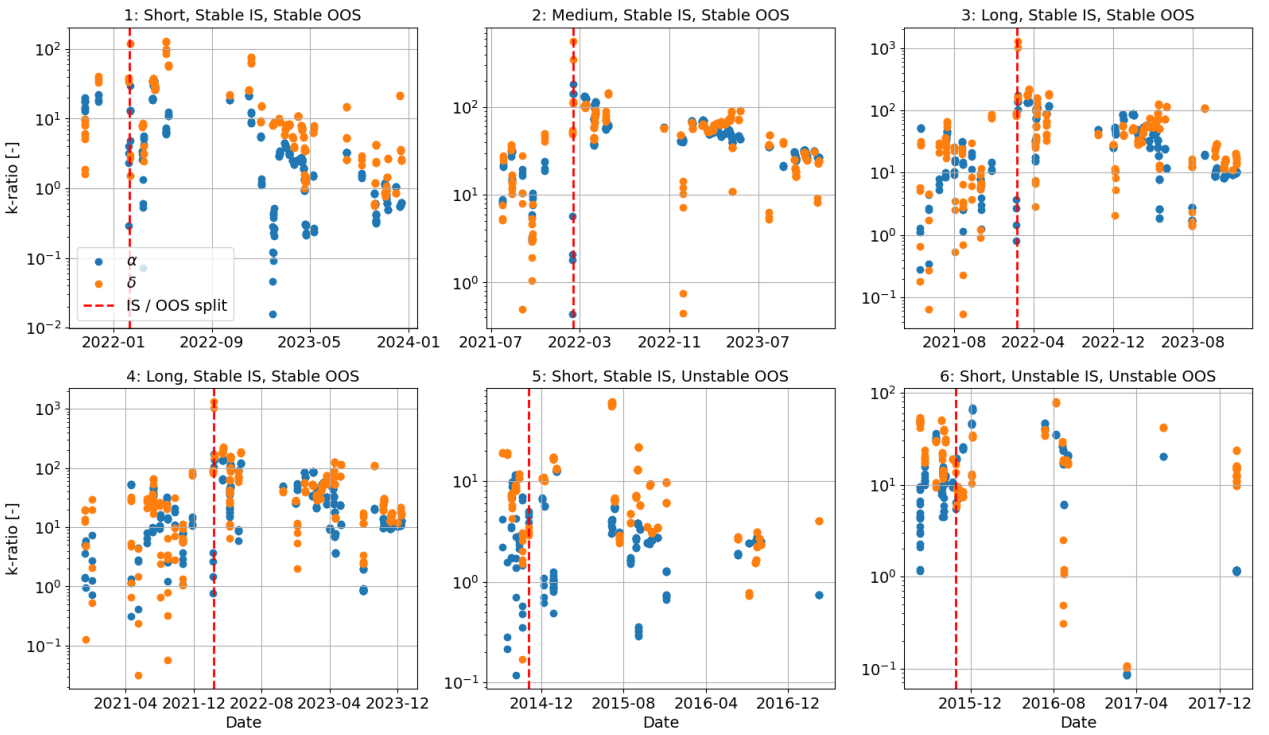


Figure C.4: k-ratio of right ascension and declination over time for Chang'e 3 use cases

Table C.2: Median k ratio (α, δ) and average overall k ratio for each use case for first 3 months out-of-sample

Object	Window	k-ratio α (med.) [-]	k-ratio δ (med.) [-]	Overall k-ratio (avg.)
Chang'e 2	1-SSS	9,7	11,0	10,3
	2-MSS	1,0	2,4	1,7
	3-LSS	48,7	60,1	54,4
	4-LSS	57,0	72,0	64,5
	5-SSU	11,4	9,5	10,4
	6-MSU	5,9	23,1	14,5
	7-MUS	133,1	62,8	98,0
Chang'e 3	1-SSS	19,3	33,1	26,2
	2-MSS	107,5	80,9	94,2
	3-LSS	104,5	98,1	101,3
	4-LSS	88,6	98,1	93,3
	5-SSU	1,1	13,3	7,2
	6-SUU	7,4	8,3	7,9

C.3 Monte Carlo simulation results

Using the average overall k-ratios shown in Table C.2 and Equation 2.17, an approximation of the true covariance at a certain time is made. Using the mean values at that time (t_E or t_0) and the approximate covariance, 200 different initial states and parameter sets can be generated using a Gaussian Random Number generator[13]. These 200 randomly generated states can then be propagated to analyze uncertainty distribution over time. These analyses are described in subsection 4.5 of the paper, additional information, plots and validation will be provided below.

C.3.1 Uncertainty propagation results

The main conclusions drawn on uncertainty behaviour over time, are based on uncertainty propagations starting on the final time of the estimation windows t_E. This is the split between in-sample and out-of-sample, often called 'IS / OOS split' in plots. If orbit estimation is performed well and in-sample orbit is sufficiently linear, a good approximation of the mean and covariance at this time should be available and covariance propagation to t_E should be valid. This will be validated by comparing to MC propagations from t_0 in subsection C.3.2.

The residuals of the 200 MC propagations from t_E against the residual of the tailored model framework are shown in Figure C.5 and Figure C.5. The difference in total position between the 200 MC propagations and the tailored model propagation can be found in Figure C.7 and Figure C.8. Several analyses have been performed and conclusions have been drawn on effect of uncertainty over time in subsection 4.5 of the paper.

[13] `https://numpy.org/doc/stable/reference/random/generated/numpy.random.normal.html`

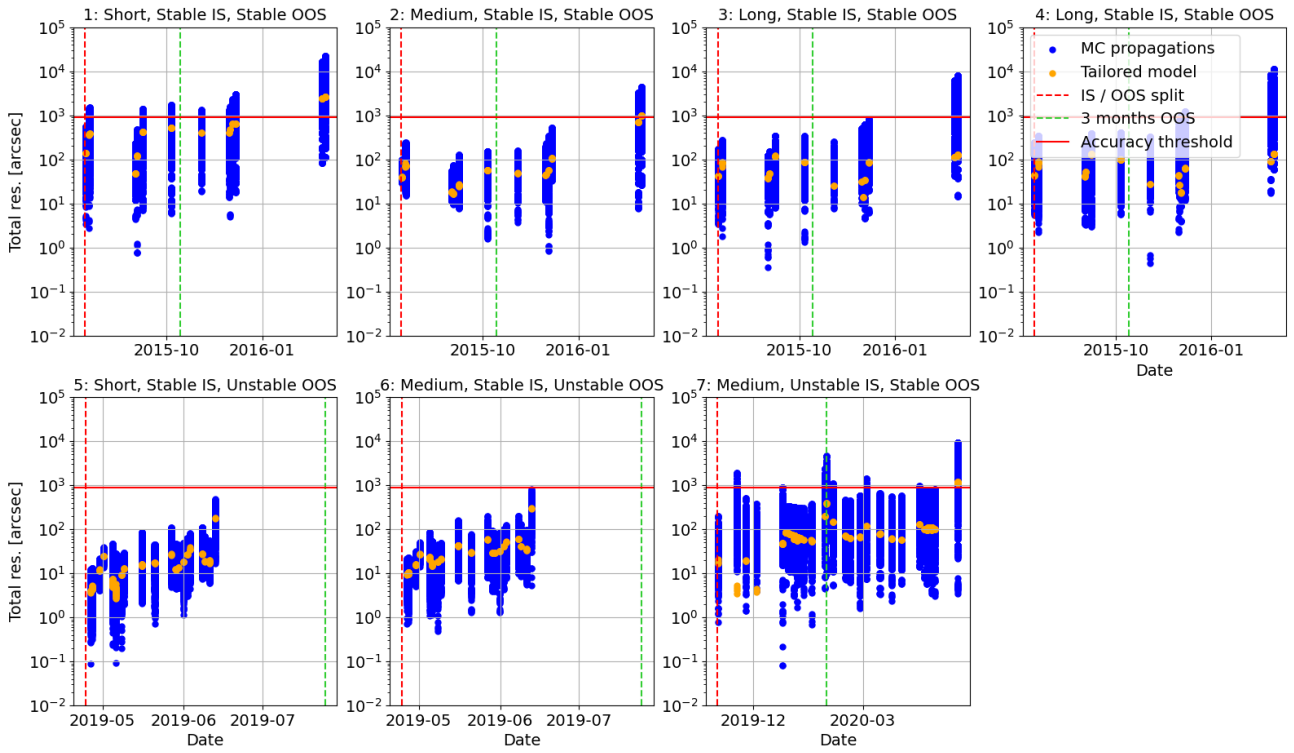


Figure C.5: Residual of MC propagations from t_E vs tailored model propagation for Chang'e 2 use cases

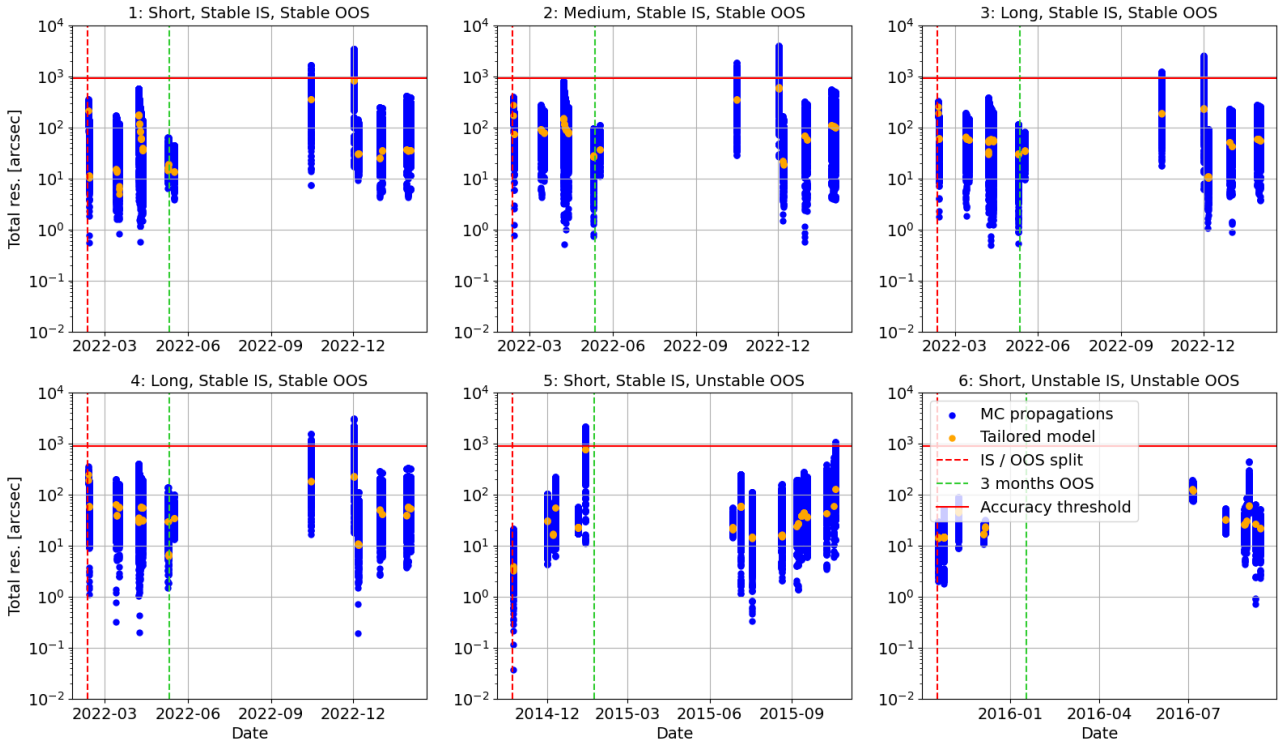


Figure C.6: Residual of MC propagations from t_E vs tailored model propagation for Chang'e 3 use cases

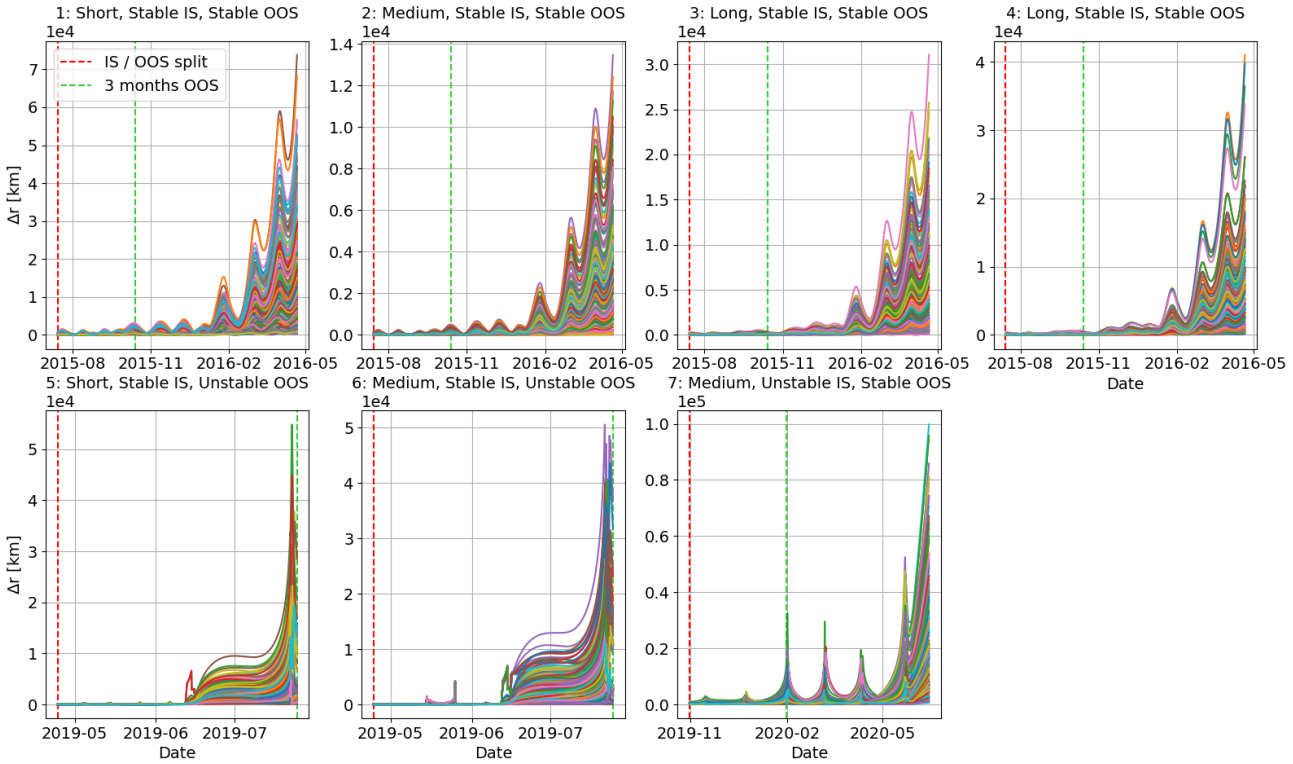


Figure C.7: Position difference between MC and tailored model propagations for Chang'e 2 use cases

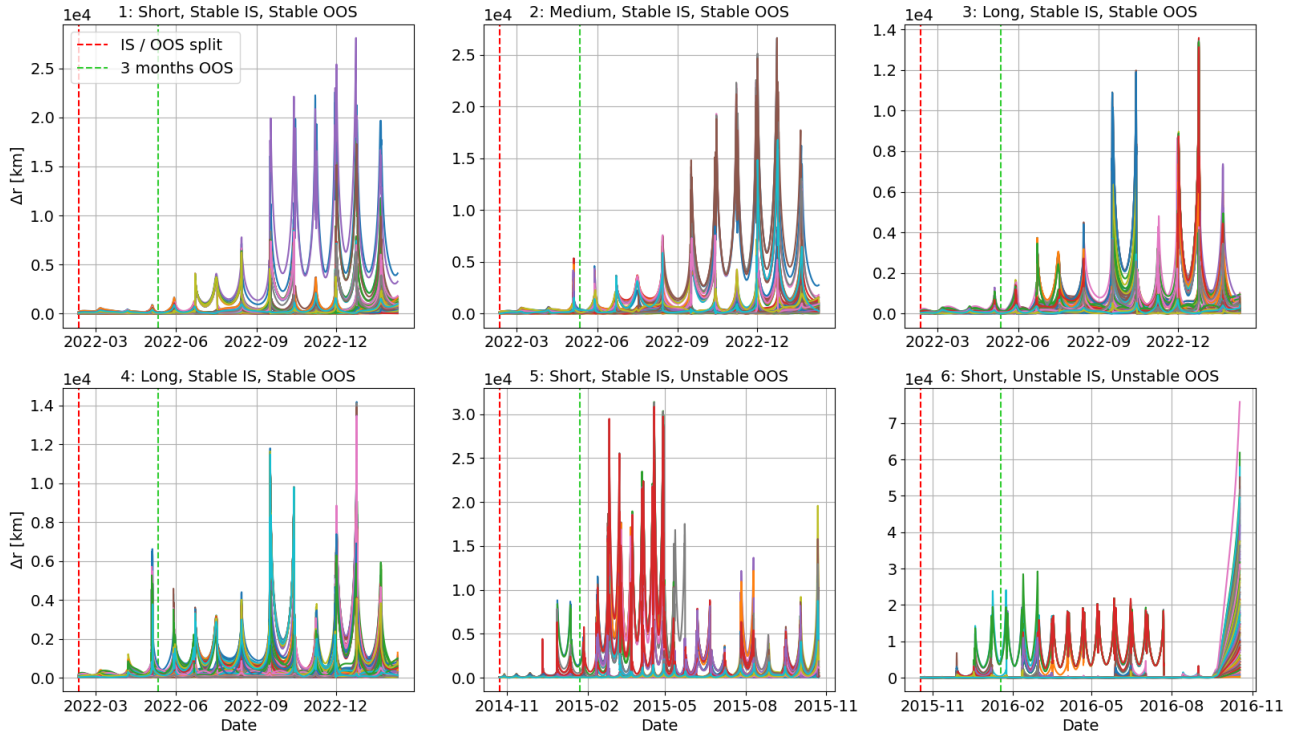


Figure C.8: Position difference between MC and tailored model propagations for Chang'e 3 use cases

C.3.2 Analysis on in-sample (non-)linearity

As explained before, Monte Carlo propagation from t_E assumes linear uncertainty propagation in-sample for covariance propagation. This can be validated by performing the same Monte Carlo propagations from t_0, the initial time in-sample. Comparing the Monte Carlo propagations from t_0 and t_E, one can confirm sufficient in-sample linearity if the residual spread does not differ much between the different initial times.

In subsection 4.5 of the paper, it is mentioned that the largest difference between MC propagation from t_E and t_0 is seen in (unstable) estimation window 7 of the Chang'e 2 booster. Which makes sense, as this use case includes a close Moon approach in-sample. This causes errors when performing linear covariance propagation

in-sample. Comparing the residual plots for the other use cases in Figure C.9 and Figure C.10, with the ones from t_E in Figure C.5 and Figure C.6, one finds that this difference is much less significant and the covariance propagation approach seems valid.

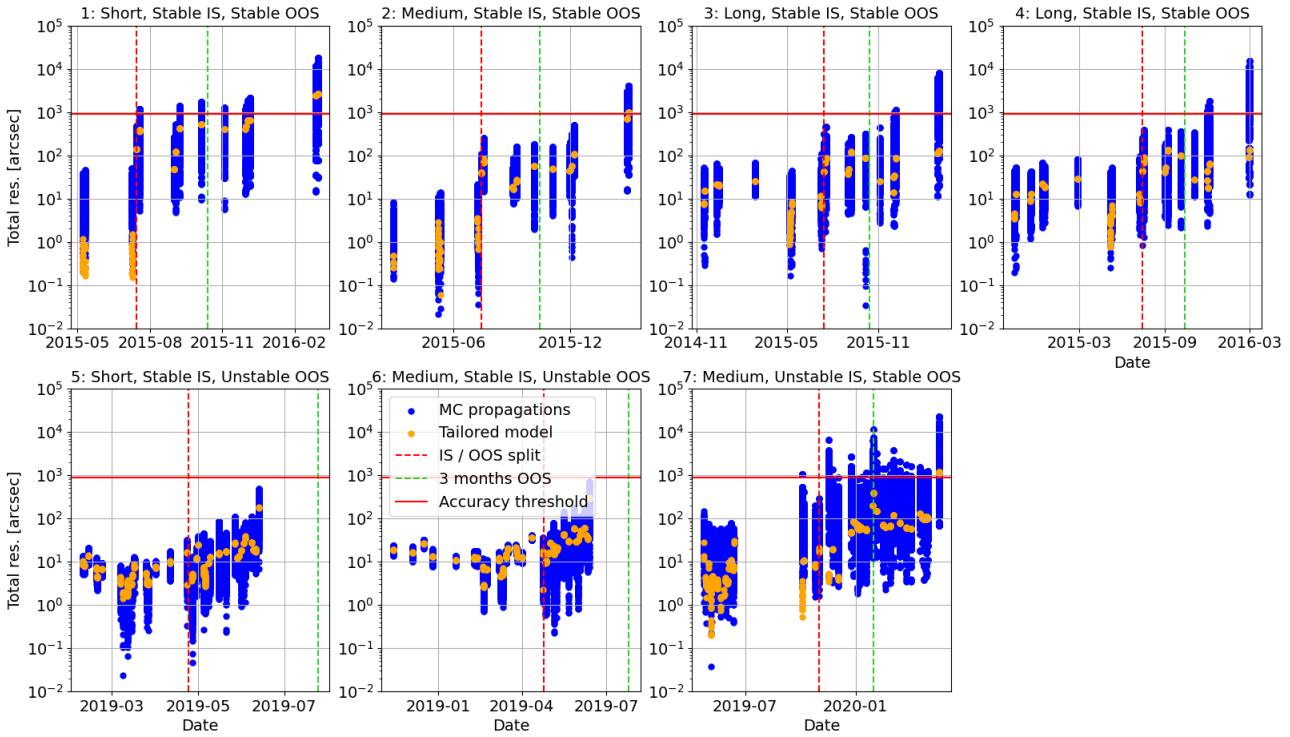


Figure C.9: Residual of MC propagations from t_0 vs tailored model propagation for Chang'e 2 use cases

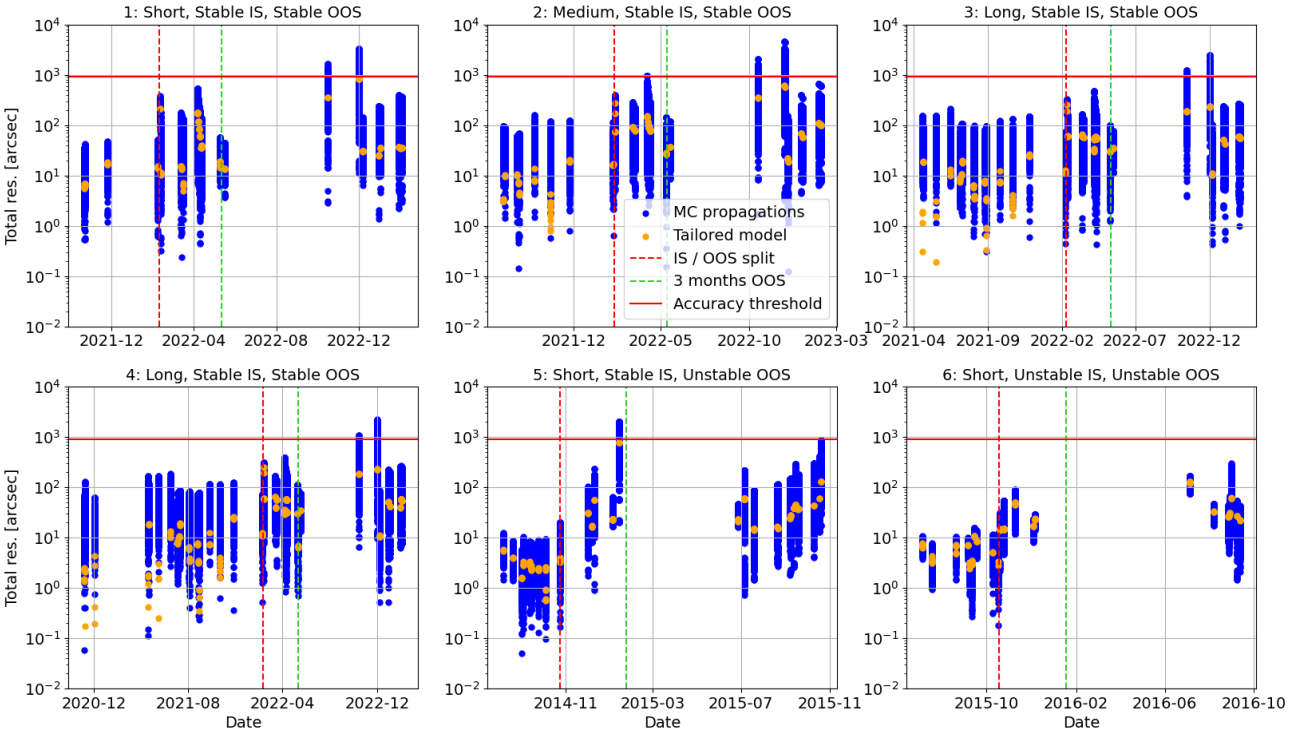


Figure C.10: Residual of MC propagations from t_0 vs tailored model propagation for Chang'e 3 use cases

C.3.3 Kolmogorov-Smirnov testing

As explained in subsection 3.5 of the paper, it needs to be analyzed how long uncertainty propagation from t_E is linear for various use cases. This is achieved using the Kolmogorov-Smirnov (KS) test. The KS test compares a sample distribution to a theoretical Gaussian distribution with sample mean and variance, and tests the null hypothesis (H0) that the sample distribution at a specific time is Gaussian. A significance level of 0.05 is used to test this hypothesis. Meaning if the p-value statistic of the KS test is smaller than 0.05, H0 is rejected and the distribution is concluded to be non-Gaussian. Once the distributions become non-Gaussian, the linearity assumption is said to be violated. Overall results of KS testing are shown in subsection 4.5 of the paper, this section provides additional material, plots and verification.

The overall findings of the KS tests, also shown in the paper, are presented in Table C.3. It is found that the linearity assumption of uncertainty propagation holds for 162-262 days and 55-110 days for the stable windows 1-4 of the Chang'e 2 and 3 booster respectively. For the unstable windows (window 5, 6 and 7), the linearity assumption only holds for 1-3 weeks, after which the sample distribution has already become non-Gaussian.

As explained in subsection 4.5, a modified KS test is applied that does not apply a zero mean in its theoretical Gaussian distribution but uses the sample mean. This is because the KS test is sensitive to non-zero means, quickly concluding a distribution to be non-Gaussian. But the non-zero mean is often caused by the limited sample size of 200, where the tested distribution actually is Gaussian. This is confirmed when using zero mean in the KS test and testing normality of the initial distribution (which is generated to be Gaussian), sometimes leading to non-Gaussian conclusions by the KS test. Using the sample mean in KS testing (sample = difference between MC and tailored propagated state), is said to be valid if the mean is sufficiently zero. This is tested using the standard error of the mean (σ_μ), which says something about how far the sample mean is from the mean of the actual distribution and is calculated as follows from the standard deviation (σ) and sample size (N) [Lee et al., 2015]:

$$\sigma_\mu = \frac{\sigma}{\sqrt{N}} \tag{C.1}$$

As long as the standard error of the mean ('non-zero threshold') is larger than the mean itself, the mean is said to be sufficiently zero. The mean of the sample is plotted against the non-zero threshold for the $\boldsymbol{x}$-distribution in Figure C.11 and Figure C.12, for the $\boldsymbol{y}$-distribution in Figure C.13 and Figure C.14, and for the $\boldsymbol{z}$-distribution in Figure C.15 and Figure C.16. It is found that in general the mean does not exceed the non-zero threshold for the periods shown in Table C.3. This means that the conclusions on (non-)linear uncertainty propagation still hold. Similar to the Chang'e 2 booster allowing linear uncertainty propagation for a longer period, the sample means also seems to stay below the non-zero threshold much longer compared to the Chang'e 3 booster. In future work, the MC sample size should be increased significantly, so the modified KS test is not necessary.

Finally, the p-values of the MC propagations at each propagation step which have lead to the results in Table C.3, are also plotted in this section. The p-values of the $\boldsymbol{x}$-distributions over time are shown in Figure C.17 and Figure C.18 for the Chang'e 2 and 3 booster respectively, for the $\boldsymbol{y}$-distributions these are found in Figure C.19 and Figure C.20 and for the $\boldsymbol{z}$-distributions in Figure C.21 and Figure C.22. Analyzing the plots, it is found that p-value for a specific state distribution is highly dependent on where the object is positioned. But for unstable orbits, the distributions quickly become non-Gaussian, while for stable orbits the distributions fluctuate in the Gaussian regime for a longer time. It can be seen that the $\boldsymbol{x}$-, $\boldsymbol{y}$- or $\boldsymbol{z}$-distribution of the propagated MC samples at a time after than the H0 rejection time, can suddenly become Gaussian again. But this still means that the linearity assumptions have already been violated, and uncertainty propagation is no longer linear.

Table C.3: Kolmogorov-Smirnov test results (of different estimation windows) for initial distribution ($\boldsymbol{x_0}$, $\boldsymbol{y_0}$ and $\boldsymbol{z_0}$), and first moment of null hypothesis rejection (H0 rej.) indicating non-Gaussian $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$-distributions

Object	Window	x0 distribution	y0 distribution	z0 distribution	x - H0 rej.	y - H0 rej.	z - H0 rej.
CH2	1-SSS	Gaus. | p=0,75	Gaus. | p=0,98	Gaus. | p=1	162,1 days	239,3 days	242,6 days
	2-MSS	Gaus. | p=0,76	Gaus. | p=0,91	Gaus. | p=0,6	261,8 days	255,5 days	257,8 days
	3-LSS	Gaus. | p=0,54	Gaus. | p=0,87	Gaus. | p=0,7	233,3 days	239,3 days	257,5 days
	4-LSS	Gaus. | p=0,77	Gaus. | p=0,91	Gaus. | p=0,72	233,4 days	239,4 days	257,7 days
	5-SSU	Gaus. | p=0,82	Gaus. | p=0,85	Gaus. | p=0,61	8,9 days	8,9 days	8,9 days
	6-MSU	Gaus. | p=0,8	Gaus. | p=0,74	Gaus. | p=0,72	8,7 days	8,9 days	8,9 days
	7-MUS	Gaus. | p=0,34	Gaus. | p=0,18	Gaus. | p=0,15	15,4 days	15,4 days	15,4 days
CH3	1-SSS	Gaus. | p=0,95	Gaus. | p=0,85	Gaus. | p=0,97	56,6 days	56,6 days	56,9 days
	2-MSS	Gaus. | p=0,89	Gaus. | p=0,53	Gaus. | p=0,78	84,1 days	84,1 days	84,1 days
	3-LSS	Gaus. | p=0,99	Gaus. | p=0,83	Gaus. | p=0,94	108,5 days	108,5 days	108,5 days
	4-LSS	Gaus. | p=0,54	Gaus. | p=0,51	Gaus. | p=0,98	56,4 days	56,4 days	56,9 days
	5-SSU	Gaus. | p=0,88	Gaus. | p=0,99	Gaus. | p=0,87	19,6 days	5,3 days	5,3 days
	6-SUU	Gaus. | p=0,89	Gaus. | p=0,9	Gaus. | p=0,95	4,1 days	4,2 days	4,2 days

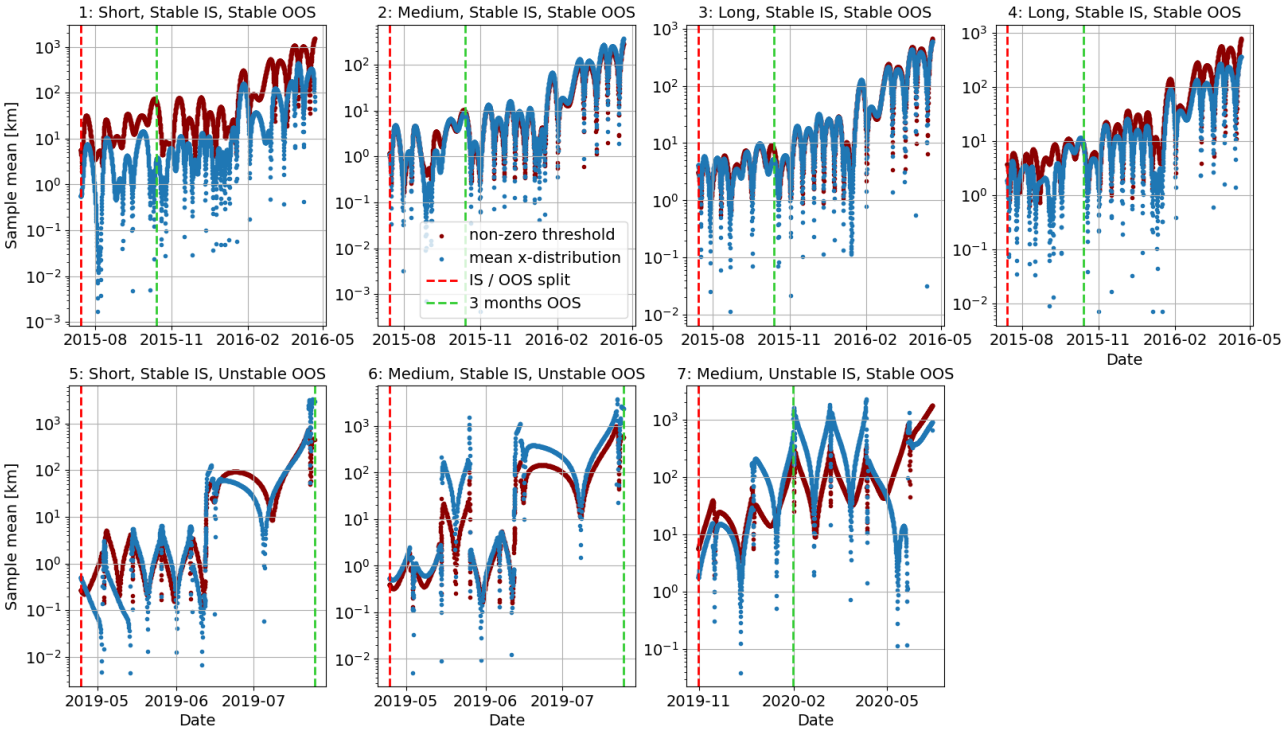


Figure C.11: Sample mean of x-distribution of MC samples over time for Chang'e 2 use cases

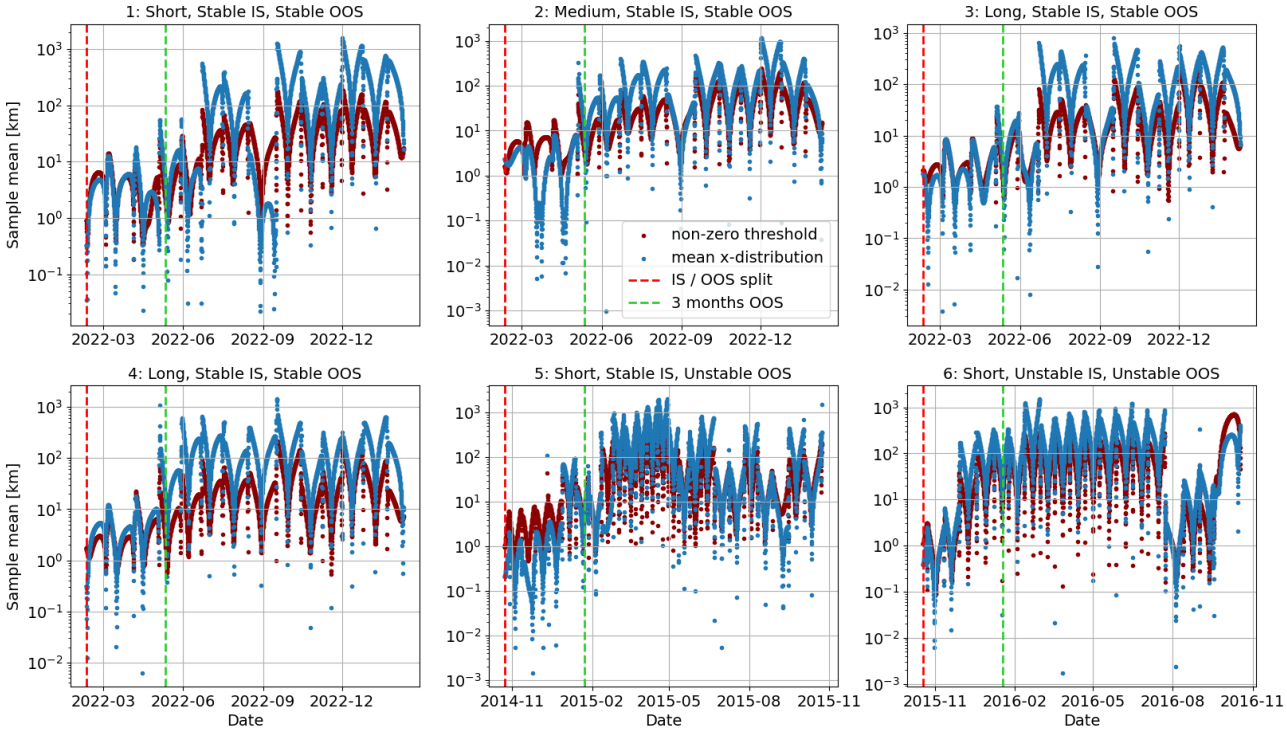


Figure C.12: Sample mean of x-distribution of MC samples over time for Chang'e 3 use cases

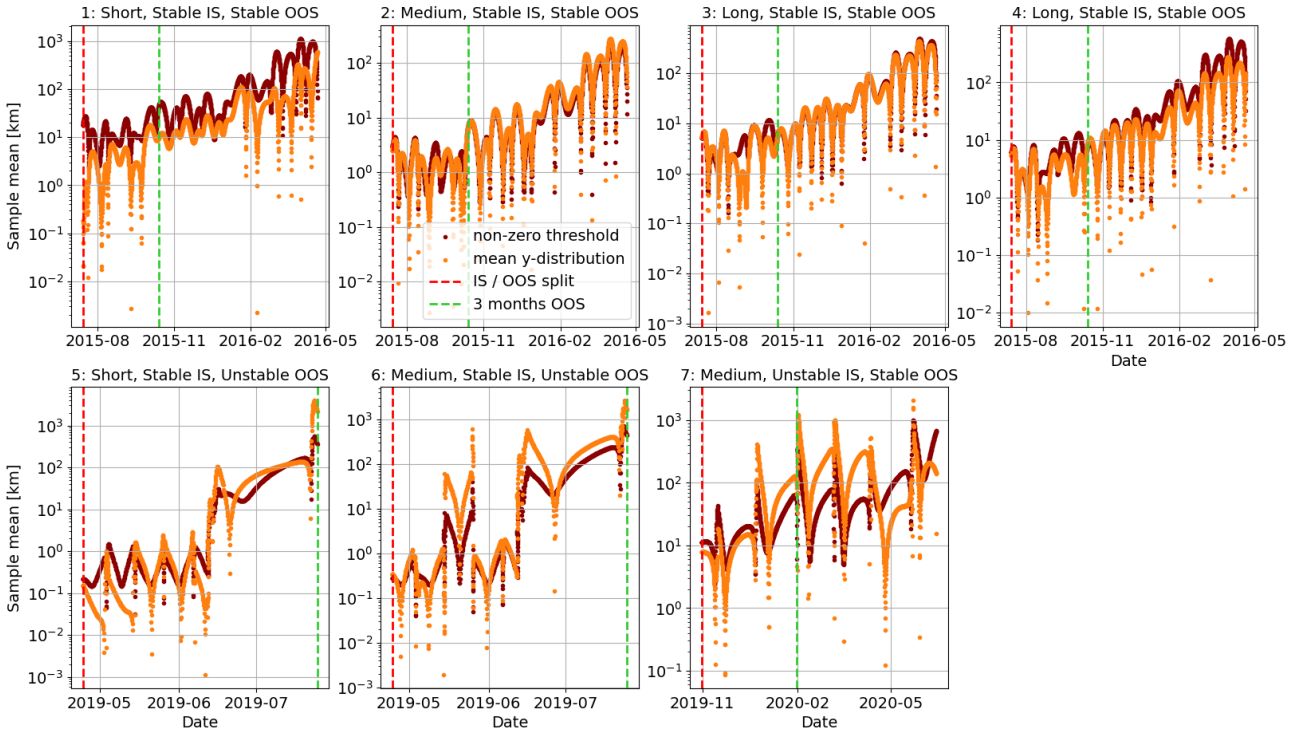


Figure C.13: Sample mean of y-distribution of MC samples over time for Chang'e 2 use cases

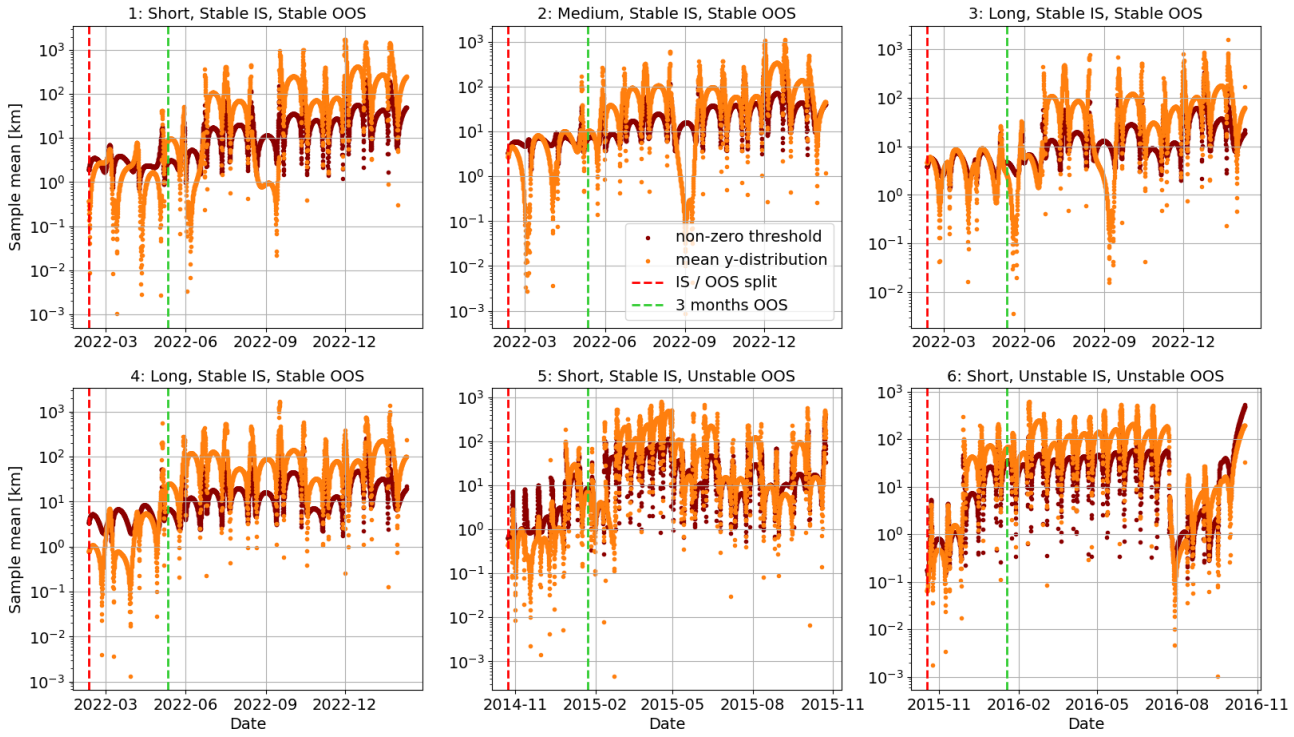


Figure C.14: Sample mean of y-distribution of MC samples over time for Chang'e 3 use cases

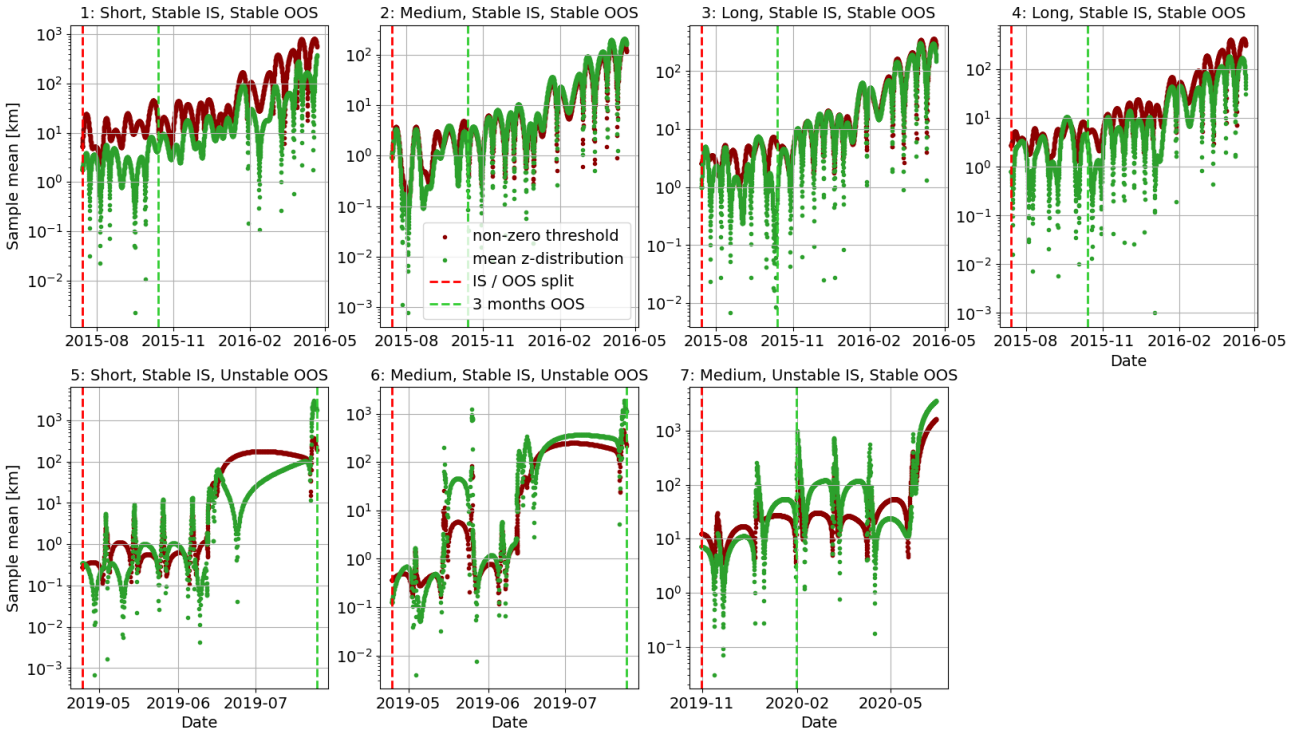


Figure C.15: Sample mean of z-distribution of MC samples over time for Chang'e 2 use cases

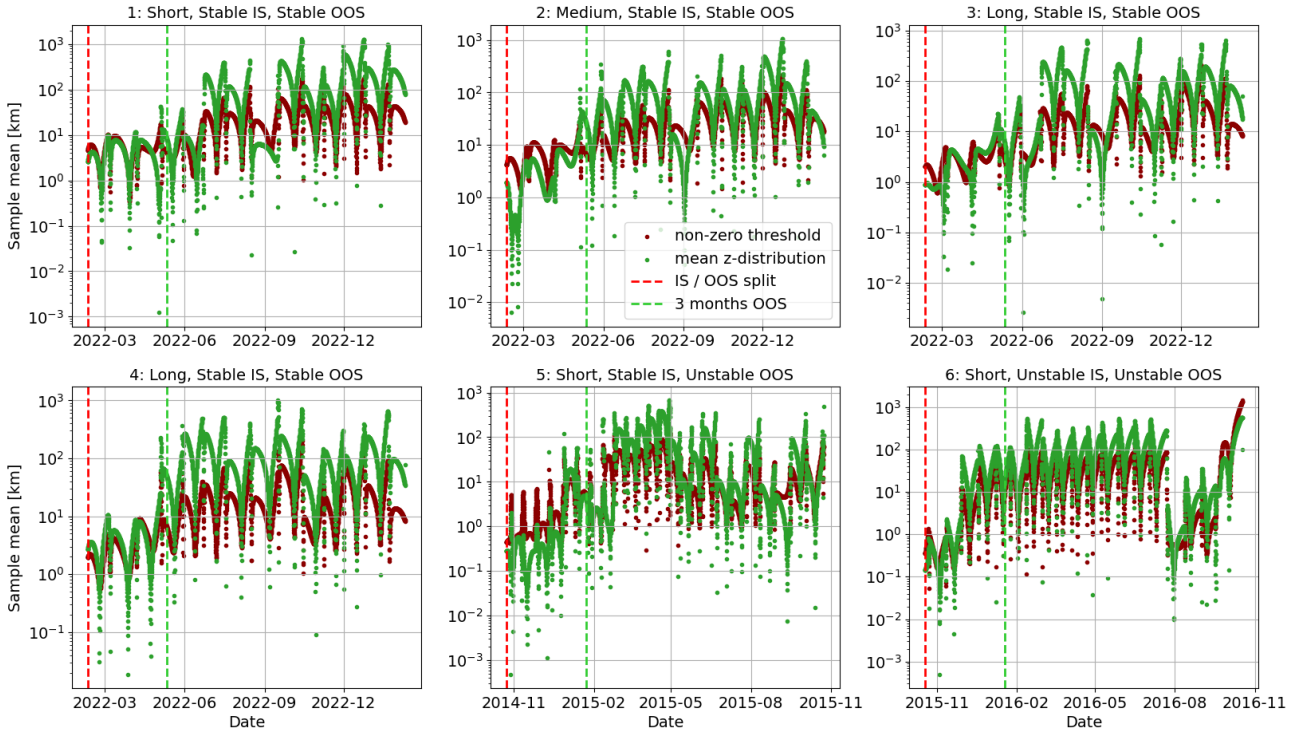


Figure C.16: Sample mean of z-distribution of MC samples over time for Chang'e 3 use cases

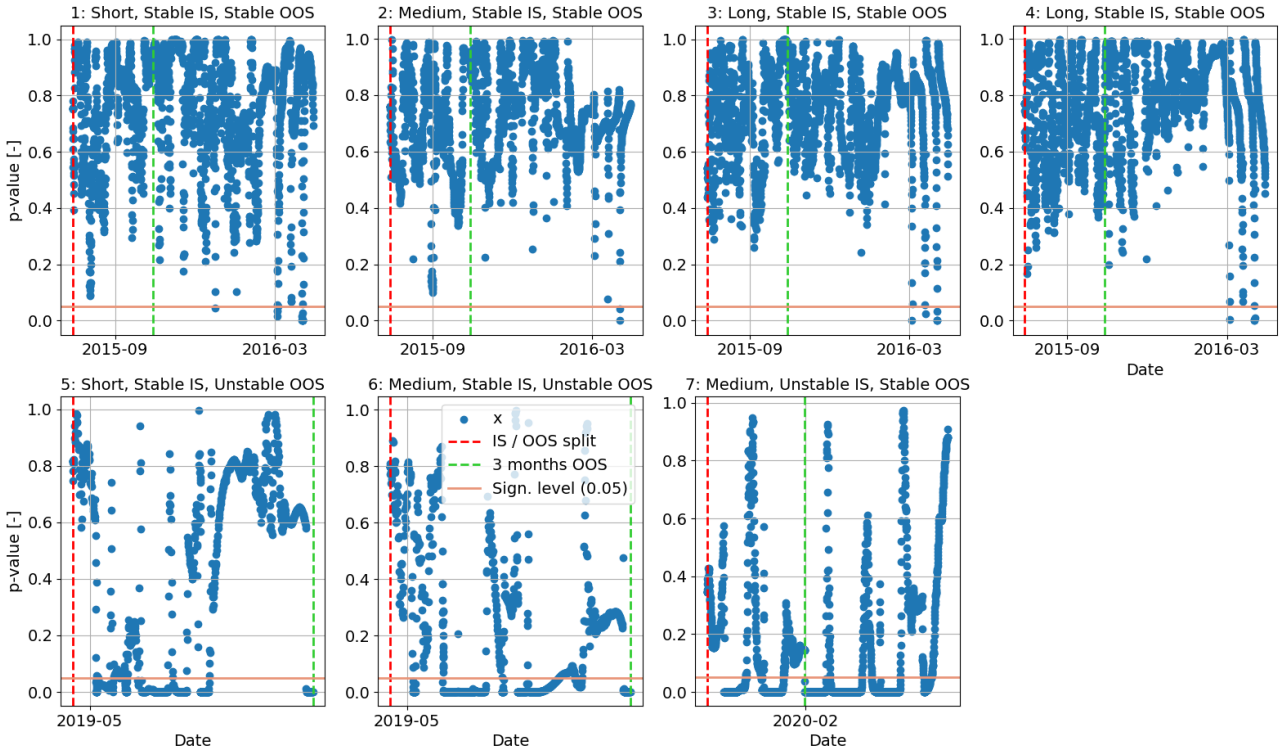


Figure C.17: p-value of x-distribution of MC samples over time for Chang'e 2 use cases

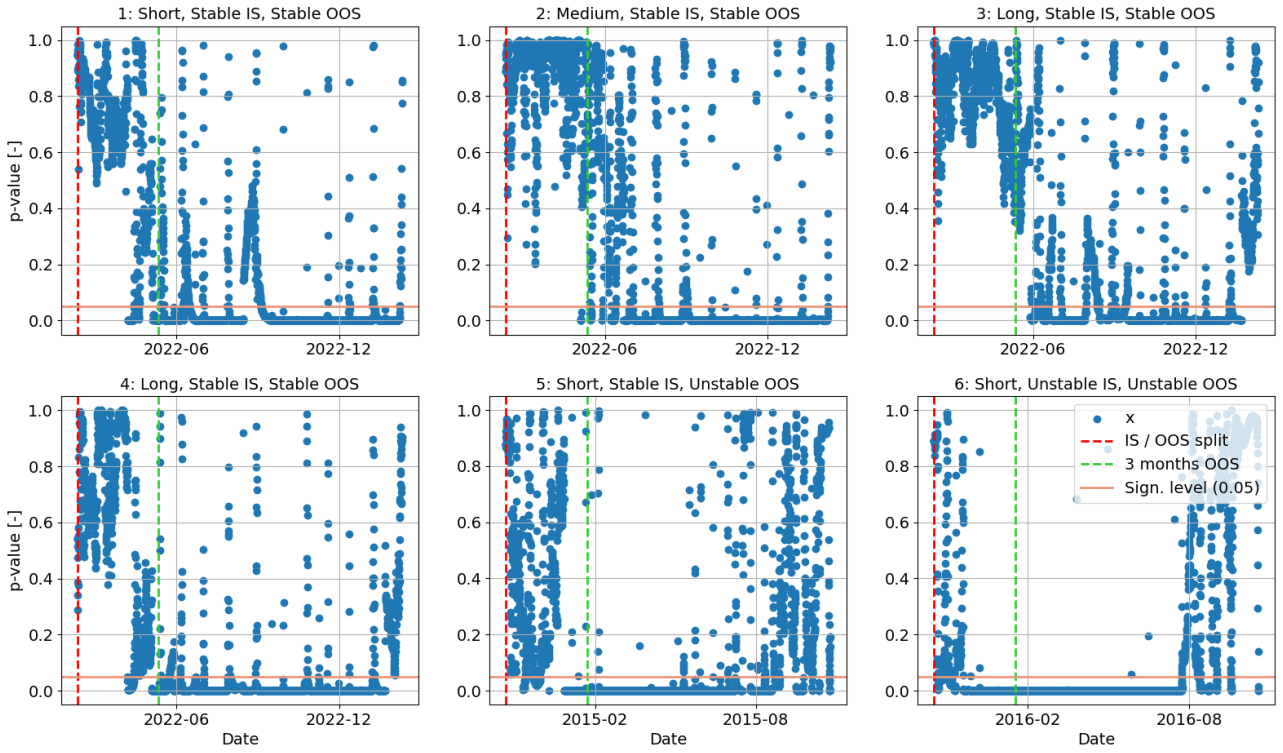


Figure C.18: p-value of x-distribution of MC samples over time for Chang'e 3 use cases

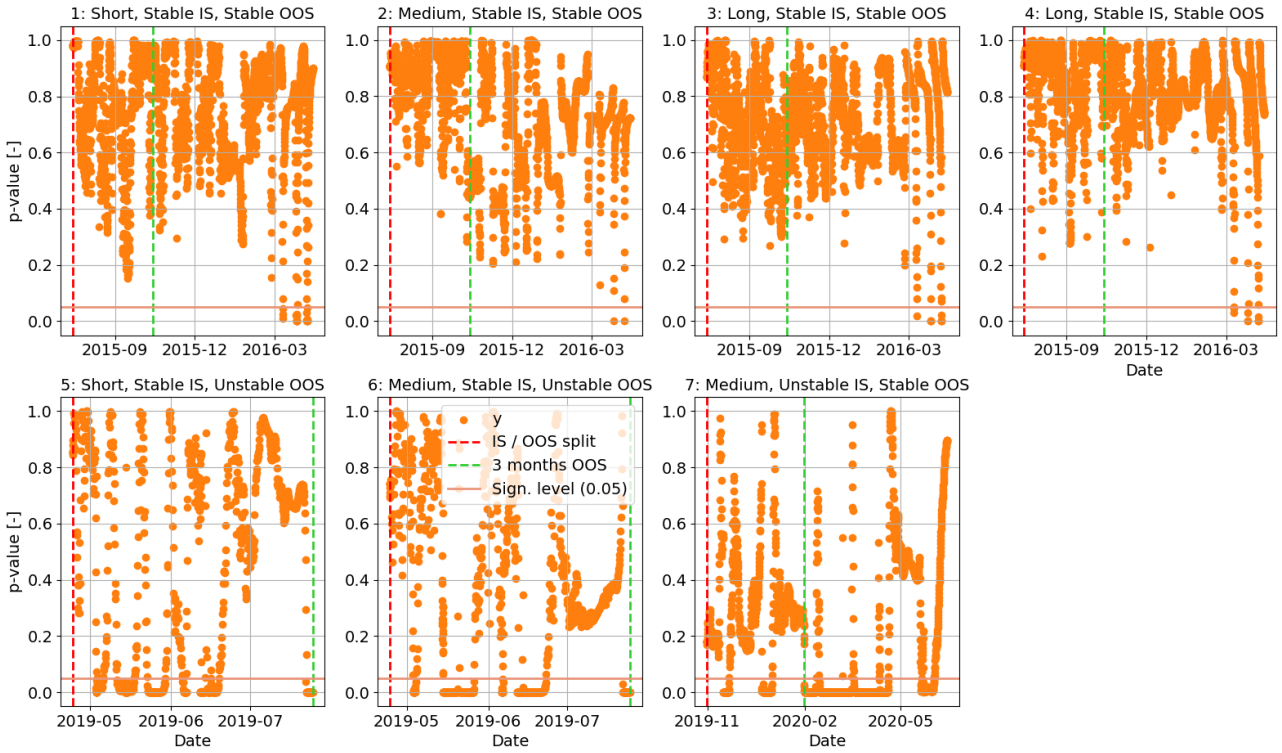


Figure C.19: p-value of y-distribution of MC samples over time for Chang'e 2 use cases

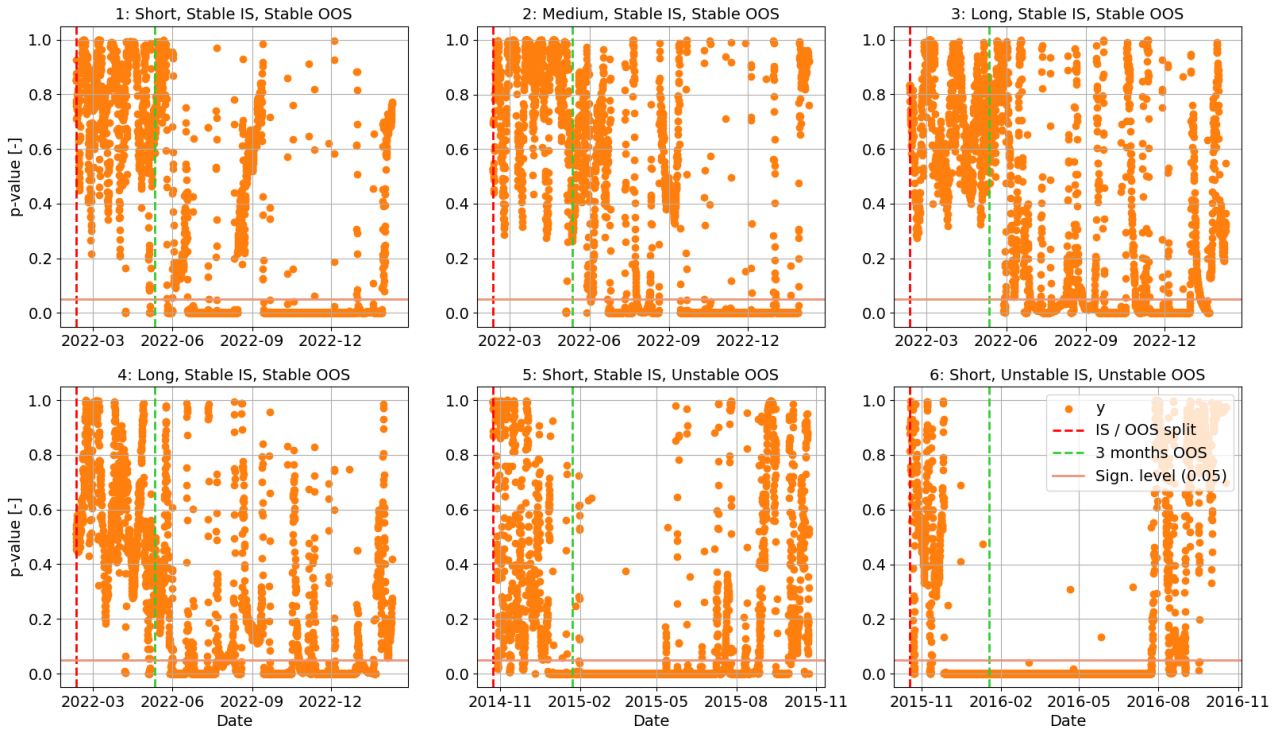


Figure C.20: p-value of y-distribution of MC samples over time for Chang'e 3 use cases

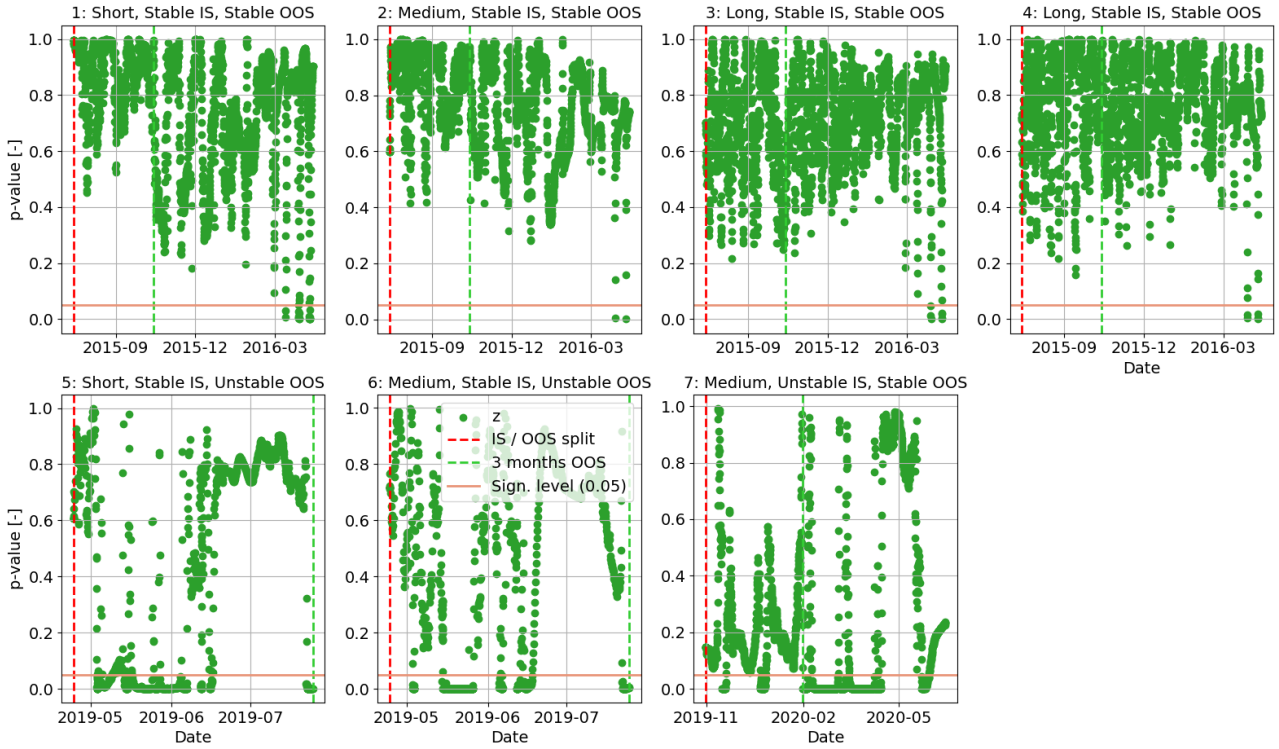


Figure C.21: p-value of z-distribution of MC samples over time for Chang'e 2 use cases

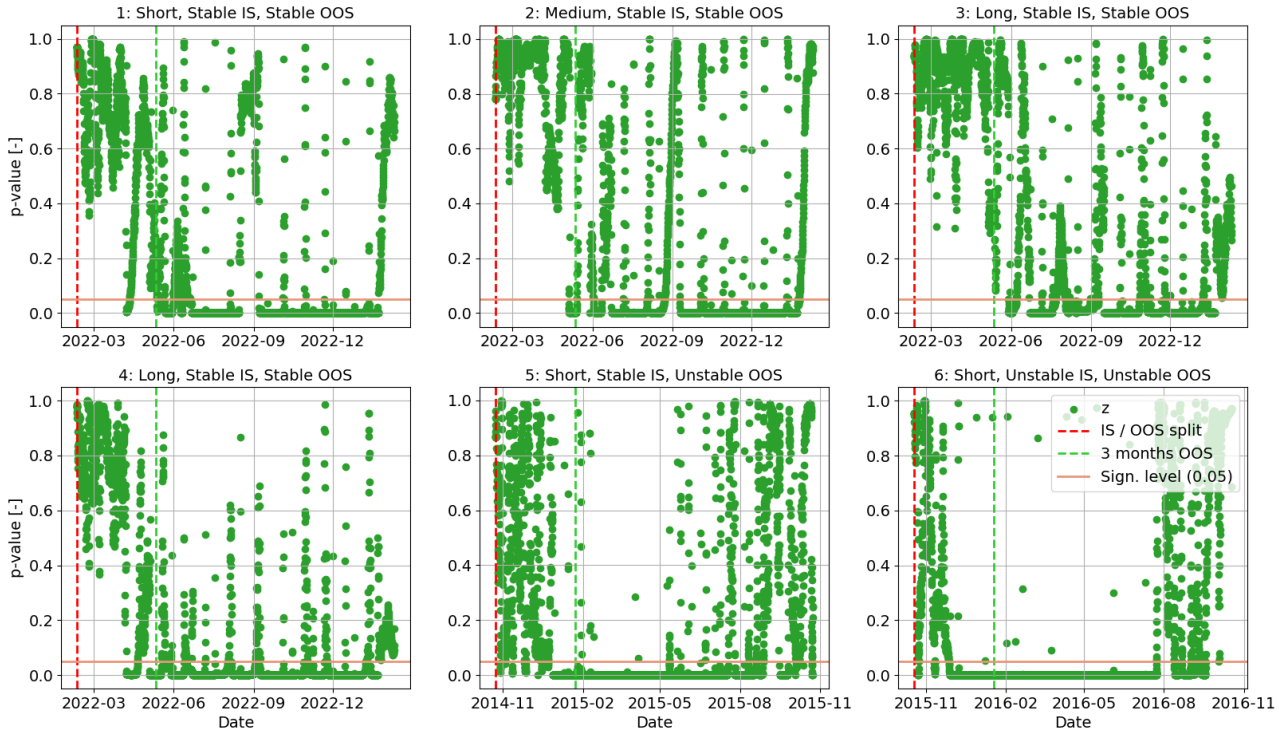


Figure C.22: p-value of z-distribution of MC samples over time for Chang'e 3 use cases

C.3.4 Analysis on Monte Carlo sample size

Finally, as explained in subsection C.3.3, a sample size 200 is insufficient to consistently provide an initial sample that follows a Gaussian distribution. A much larger sample size (N=1000-10000) is likely required to solve this, which is currently found to be computationally out-of-scope.

Still, it is important to verify the conclusions on effect of uncertainty over time based on 200 MC propagations, described in subsection 4.5. This is done by performing a more expensive computation of 400 Monte Carlo propagations, and comparing plots to see whether residual distributions are similar and the conclusions still hold. Size 400 Monte Carlo propagation residuals can be found for the Chang'e 2 and 3 booster in Figure C.23

and Figure C.24 respectively. Comparing to the 200 MC propagations in Figure C.5 and Figure C.6, it is found that the distributions are highly similar for all use cases. The conclusions on uncertainty effect over time for various use cases stay the same. Thus, it can be said that a size 200 Monte Carlo simulation is sufficient for preliminary analysis on uncertainty over time.

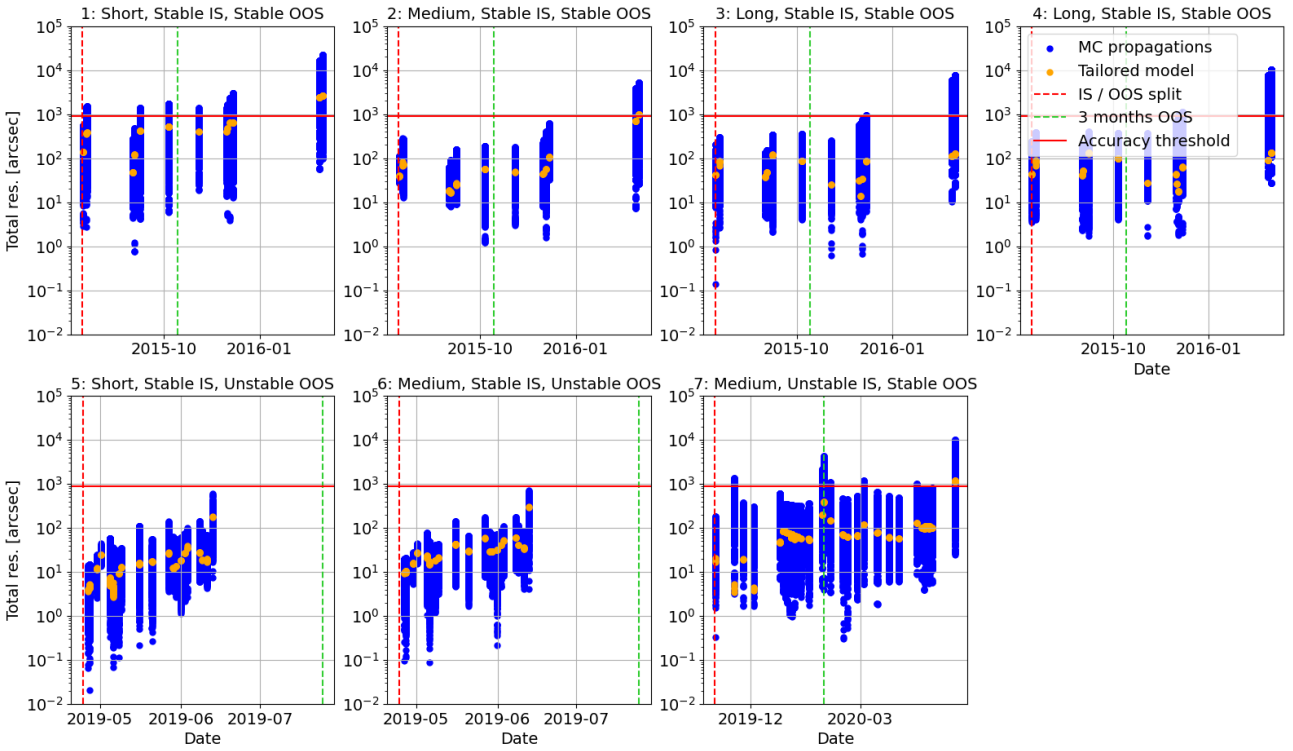


Figure C.23: Total residual of 400 Monte Carlo propagations over time for Chang'e 2 use cases

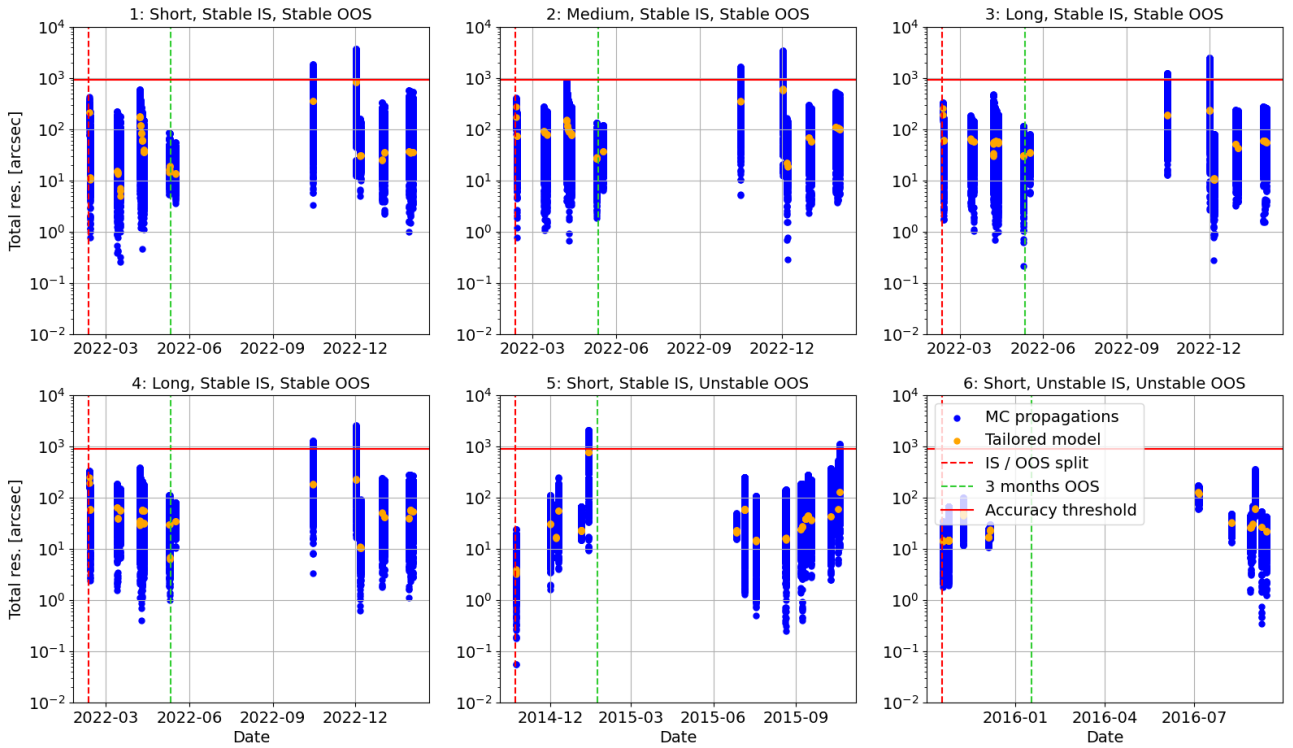


Figure C.24: Total residual of 400 Monte Carlo propagations over time for Chang'e 3 use cases

D

Verification & Validation

Verification and validation has been an integral part of the generic model framework development and tailoring process, described in the paper and Appendix B. In each model framework iteration, the effect of design choices on propagation accuracy is continuously checked. The accuracy is found by comparing the propagated orbits to actual observations, this is an important validation step. Furthermore, the proper functioning of the Monte Carlo approach over time has also been verified and validated extensively. This is done by comparing the residuals of MC propagations from t_E to the ones of MC propagations from t_0 (subsection C.3.2), testing whether distributions are Gaussian using Kolmogorov-Smirnov tests (subsection C.3.3) and partly verifying the Monte Carlo size of 200 samples using a sample size of 400 (subsection C.3.4).

But in the process of building the actual algorithm corresponding to this model framework using Tudat, many preliminary verification and validation steps have been performed. This has been done to ensure that all individual building blocks of the model framework, shown in Figure 2 of the paper, function properly. The preliminary verification and validation process will be elaborated upon in section D.1.

Additionally, though the model framework results are already validated based on residuals with respect to optical observations, a final verification step is performed in section D.2. Here, the orbit propagations for all use cases will be compared to the estimated ephemeris retrieved from Find_Orb estimations. The Find_Orb estimations have been used as an initial guess in the orbit determination algorithm of this model framework. This not only serves as an extra verification step for the output of this model framework, but will also allow investigation on the accuracy of the Find_Orb estimations.

D.1 Preliminary V&V process

During preliminary verification and validation of the model framework, three important building blocks are tested extensively. Orbit propagation, orbit estimation and data sources. An overview of the preliminary V&V process is shown in Figure D.1 and will be elaborated upon below.

Starting with verification of the models and input used for orbit propagation in Tudat. The main building blocks used in orbit propagation are the dynamical models, integrator and propagator. The proper functioning of these components for cislunar space is tested by propagating a cislunar object with a well-known ephemeris (which is not the case for the Chang'e 2 and 3 upper stages) and comparing the propagation to the ephemeris. To achieve this, orbit propagation is performed for the Spektr-R spacecraft [Zakhvatkin et al., 2020][Zhamkov et al., 2016], for which accurate ephemeris can be retrieved from JPL Horizons[1]. A highly accurate model is created which is similar to the baseline model described in Table B.5. Comparing the baseline propagation to the JPL ephemeris, as well as performing similar tests as in Table B.6 and seeing the effect on ephemeris difference, the various building blocks can be verified. This process has shown that the dynamical models, integrator and propagator function properly. The object parameters (mass, area, C_r etc.) of the Chang'e 2 and 3 boosters, which are important for orbit propagation, are validated using the material of the Long March user manual [Cen, 2011]. The Chang'e 2 and 3 upper stages are boosters of a Long March 3C and 3B rocket respectively.

The orbit determination algorithm, or estimator, has also been preliminarily verified. This is done using two objects for which optical observations are available, SWARM-B and Eros. For SWARM-B, the estimations on angular observations are propagated and compared to Cartesian ephemeris retrieved from GNSS data of ESA. Orbit determination on Eros is especially performed to validate the proper conversion of Minor Planet Center (MPC) data in this algorithm. A complete set of observations is stored by MPC for Eros, as this is a Near-Earth Asteroid (NEA). Tudat has a class called BatchMPC, which allows for direct downloading and conversion of MPC data to angular observations. But for the Chang'e 2 and 3 upper stages, this observation set is stored in the

[1] https://ssd.jpl.nasa.gov/horizons/app.html/

MPC format by Project Pluto (not by MPC) and conversion to angular observations has to be done by writing our own conversion Python code. By comparing the orbit determination results for Eros using BatchMPC to loading in the Eros MPC formatted data by hand and performing estimation, another check is performed for the proper functioning of the estimator. Both tests performed for SWARM-B and Eros showed similar or the same orbit determination and propagation results, thus verifying the orbit determination algorithm.

Finally, some validation is performed on the reference frames, time frames and data conversions for the various data sources. The main data sources are various Cartesian ephemerides used for initial state guesses, angular observations retrieved from Project Pluto in MPC format (converted by our own algorithm) and ground station locations used in orbit determination. To verify the Find_Orb and TLE-derived (which also originate from Find_Orb) Cartesian ephemeris, each source is ensured to be in the J2000 frame and TDB time frame. For Find_Orb and TLE ephemeris, this required conversion from UTC to TDB time frame. JPL Horizons has two months of ephemeris for the Chang'e 3 booster (April and May 2021). Comparing the converted Find_Orb and TLE ephemeris for the Chang'e 3 booster to the JPL Horizons ephemeris (in J2000, TDB frame) in these 2 months, the accuracy can be verified. It is found that Cartesian ephemeris of Find_Orb estimations are most accurate, while in the TLE-derived ephemeris there seems to be an issue with the time conversion. Thus, for initial state guesses it is decided to use the direct Find_Orb estimations. For the conversion of optical observations from in MPC format (J2000, UTC) to usable angular data (right ascension, declination, time) in the right format for orbit determination (J2000, TDB), a check is done using Eros and BatchMPC. The observations for Eros are loaded in using BatchMPC and using the self made conversion algorithm, after which the two are compared. This way the conversion of Project Pluto stored MPC data on the Chang'e 2 and 3 upper stages has been verified. Lastly, the orbit determination algorithm requires ground station locations as described in subsection 3.3 of the paper. These are retrieved manually from the MPC database[2] and verified by comparing to the ground station locations found for Eros using BatchMPC.

D.2 Verification using Find_Orb estimations

As mentioned previously, the orbit estimations performed by Find_Orb on the Chang'e 2 and 3 upper stages, have been found to be the most accurate a priori approximations of the Cartesian location of the objects (before this thesis). Whether this model framework generates more accurate orbit propagations than Find_Orb, is something to investigate in future studies. But the two orbits can be compared for a final verification step of the model framework results.

In Figure D.2 and Figure D.3, the total position difference between the propagated orbits of the tailored model framework and the orbits retrieved from Find_Orb estimations are shown. All propagated orbits of the tailored framework are sufficiently accurate, thus have a residual smaller than 900 arcseconds. Comparing both orbits for various use cases, it can be seen that generally the orbits only differ up to 1000 km and are usually much closer. This is another verification for the correctness of the orbit determination and propagation of the tailored model framework.

But there are two major outliers. For unstable estimation windows 5 and 6 of the Chang'e 3 upper stage, which are in 2014-2016, the position difference diverges to over 100.000 km. Since the orbit propagation of the tailored model framework is known to be sufficiently accurate in this period, this is an indication that Find_Orb has trouble performing accurate orbit estimation reliably in this period. This has also been discussed in Appendix A. The process of finding converging initial state guesses (from Find_Orb estimations) between 2014-2016 for the Chang'e 3 booster has taken many iterations with different initial times. Thus indicating that accuracy of the Find_Orb estimations in this period is poor.

In conclusion, comparison with Find_Orb estimations has given extra verification on the correctness of the estimated and propagated orbits of the tailored model framework. It has also provided evidence that Find_Orb estimations (and probably also the TLEs) for the Chang'e 3 upper stage in 2014-2016 are inaccurate and should be revised.

[2] https://www.minorplanetcenter.net/iau/lists/ObsCodesF.html

V&V of model framework for orbit propagation		V&V of model framework for orbit estimation		V&V of space debris data sources	
What	*How*	*What*	*How*	*What*	*How*
Dynamical models Integrator Propagator	**1.** Propagate Cislunar object with well-known ephemeris (Spektr-R through JPL Horizons) **2.** Create benchmark model and compare to ephemeris to validate overall propagation framework **3.** Validate specific models by modelling different dynamics, integrator / propagator configurations	Estimator (Angular observations)	**1.** Validate by estimating on SWARM-B observations, compare to Cartesian estimation and propagation of SWARM-B GNSS data. **2.** Validate by estimating on Eros MPC data and comparing to the BatchMPC example provided by Tudat Space	**Cartesian ephemeris / initial state:** - TLEs - Find_Orb - JPL	**1.** Validate reference frames - TLEs / Find_Orb = J2000 - JPL = J2000 **2.** Validate time frames: - TLEs / Find_Orb = UTC - JPL = TDB **3.** Validate conversion: - Convert TLE / Find_Orb (UTC to TDB) and compare with JPL
Input parameters (Chang'e objects)	**1.** Check input parameters for different Chang'e boosters with literature			**Angular observations:** - MPC - Project Pluto	**1.** Validate reference frames - MPC = J2000 - Project Pluto = J2000 **2.** Validate time frames: - MPC = UT/UTC - Project Pluto = UT/UTC **3.** Validate conversion: - Convert time frames (UTC to TDB) and compare with importing Eros using BatchMPC
				Ground station location	**1.** Convert ground station location **2.** Compare ground station locations with BatchMPC import

Figure D.1: Preliminary V&V process overview

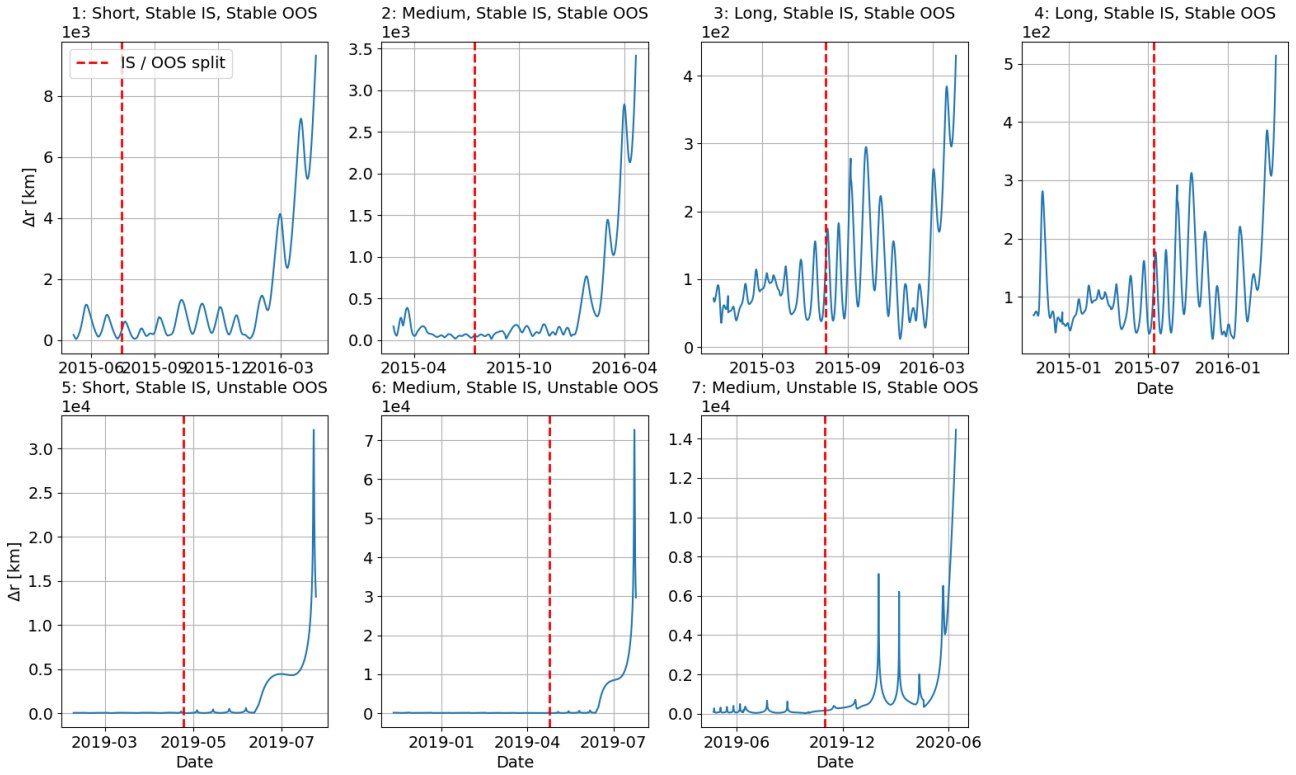


Figure D.2: Total position difference between propagation of tailored model framework and Find_Orb estimations for Chang'e 2 use cases

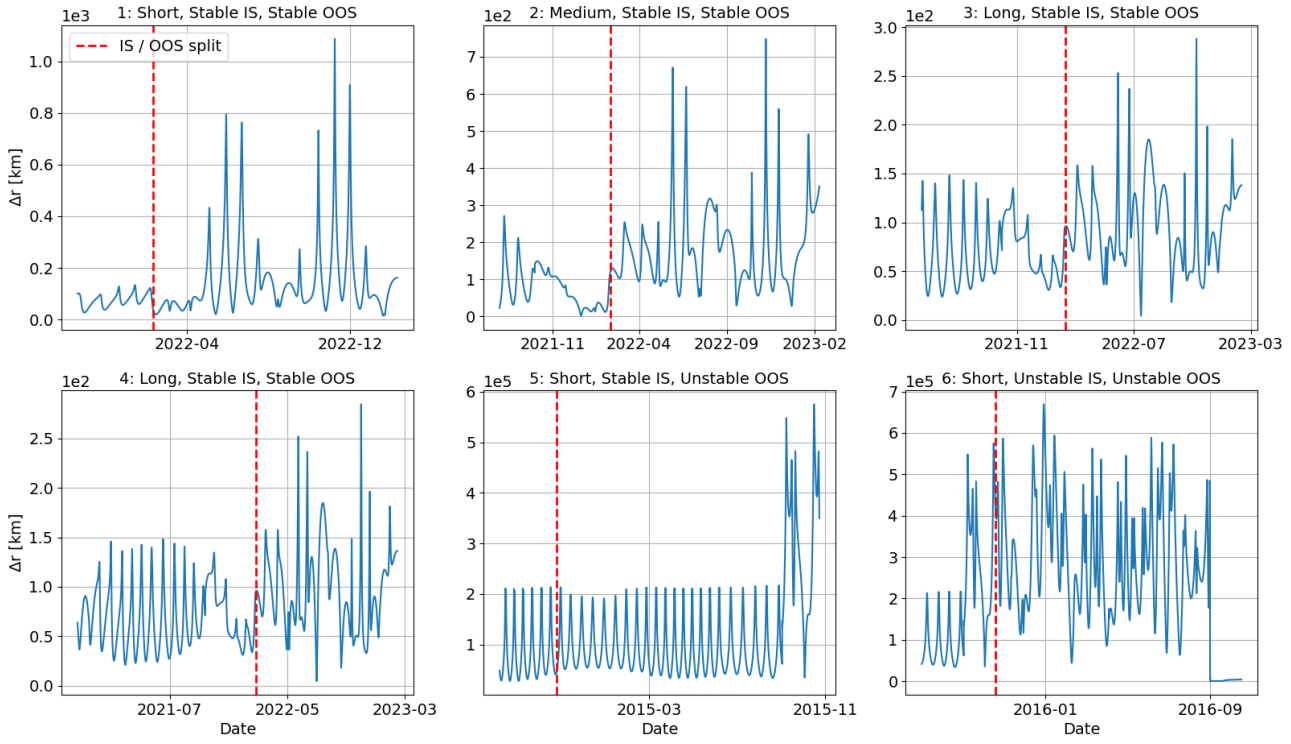


Figure D.3: Total position difference between propagation of tailored model framework and Find_Orb estimations for Chang'e 3 use cases