Combination of Swarm Kinematic Orbits

by

Mattijs Berendsen

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Thursday June 23rd, 2025, at 09:00 AM.

Student number:	4687272	
Project duration:	July 2024 – June 2025	
Thesis committee:	Dr. ir. João de Teixeira da Encarnação,	TU Delft, supervisor
	Ir. Kevin J. Cowan MBA,	TU Delft, chair
	Dr. ir. Ernst J.O. Schrama,	TU Delft, external examiner
	Dr. ir. Jose A.A. van den IJssel	TU Delft, additional

TU Delft

Contents

Preface

This thesis marks the culmination of nearly eight years of study here in Delft, from my Bachelor's through to my Master's. My time at TU Delft has had its share of ups and downs; while it has not always been easy, it has certainly never been dull. The opportunity to study here, and to learn from and engage with dedicated professors, has been a privilege.

In that light, I would like to express my sincere gratitude to my supervisor, João, who dedicated considerable time to our many meetings. His rigorous feedback has been as extensive as it has been precise, and his willingness to accommodate my circumstances—particularly during a period of medical difficulties—was deeply appreciated.

I also extend warm thanks to Jose van den IJssel for generously replying to my many emails and listening to me talk for hours. Her assistance has been invaluable. More broadly, I am grateful to all the professors who have taught me throughout my time at TU Delft; their guidance and support have helped shape the person I am today.

As I said, I faced some health challenges during this thesis; particularly relating to mental well-being. I consider myself incredibly fortunate to have been supported throughout by my wonderful partner, Ari – who found herself reading my thesis on a beach in Italy –, by my family, and by my many friends. To them, I express my deepest thanks.

Mattijs Berendsen,
Delft, June 2025

List of Acronyms

AC	Analysis Centre
AIUB	Astronomical Institute of the University of Bern
CMA	Celestial Mechanics Approach
CHAMP	CHAllenging Minisatellite Payload
COST-G	International Combination Service for Time-variable Gravity Fields
CV	Coefficient of Variation
DAA	Decorrelated Acceleration Approach
DE	Differential Evolution
ECEF	Earth-Centred Earth-fixed
EGSIEM	European Gravity Service for Improved Emergency Management
ESA	European Space Agency
GHOST	GPS High precision Orbit determination Software Tools
GLS	Generalised Least Squares
GOCE	Gravity Field and Steady-State Ocean Circulation Explorer
GPS	Global Positioning Service
GRACE	Gravity Recovery and Climate Experiment
GRACE-FO	Gravity Recovery and Climate Experiment Follow On
STD	Standard Deviation
GNSS	Global Navigation Satellite System
IERS	International Earth Rotation and Reference Systems Service
IFG	Institute of Geodesy Graz
IGFS	International Gravity Field Service
IEBA	Improved Energy Balance Approach
ILRS	International Laser Ranging Service

ITRF	International Terrestrial Reference Frame
KO	Kinematic Orbits
LEO	Low Earth Orbit
LOS	Line of Sight
MIT	Massachusetts Institute of Technology
MJD	Modified Julian Date
NCEI	National Centers for Environmental Information
NGA	National Geospatial-Intelligence Agency
OLS	Ordinary Least Squares
PDF	Probability Density Function
POD	Precise Orbit Determination
PyGMO	Python Parallel Global Multiobjective Optimizer
RDO	Reduced dynamic orbit
RMSE	Root Mean Squared Error
RMS	Root Mean Square
SAA	Short-Arc Approach
SLR	Satellite Laser Ranging
TUD	Technical University of Delft
Tudat	TU Delft Astrodynamics Toolbox
UTC	Coordinated Universal Time
VCE	Variance Component Estimation

Abstract

The Swarm satellite mission, in operation since November of 2013, has been collecting high-accuracy GPS tracking data for over a decade. Kinematic orbit solutions derived from this data have been used as input to gravity field estimation techniques for some time, with research initiatives such as the 'Multi-approach gravity field models from Swarm GPS data' dedicated to the production of highly accurate time-varying models. The production of these models is of great value to the geodetic community, as they provide an alternative and independent source of gravity field information which can be used to bridge gaps in the service of dedicated gravity field missions such as the Gravity Recovery and Climate Experiment, and provide additional temporal and spatial coverage of Earth's gravity field. The determination of this time-varying component is of great scientific value, as these models capture the complex motion of mass about the Earth occurring over short periods of time; such as the melting of polar icecaps and drought. The focus of this thesis is the improvement of the Swarm kinematic orbits used as input for such models.

We begin with an introduction to the research, including background of the topic, the scope of the research, the current state of the art and the gap in knowledge to be addressed culminating in the presentation of research questions. Input kinematic orbits from the Technical University of Delft, the Astronomical Institute of the University of Bern and the Institute of Geodesy Graz are retrieved. We apply input consolidation using interpolation to bring each input to consistent 1-second frequency sampling. We then apply a 0.3m outlier-filtering using the residuals of each input with respect to a reduced dynamic reference orbit. The consolidated and pre-screened input orbits are then combined using the following techniques: arithmetic mean weighting, inverse-variance weighting, iterative variance component estimation, weighting with the residuals with respect to a Reduced dynamic orbit (RDO) reference, and optimisation using an RDO reference.

The resulting combined orbits are then assessed using Satellite Laser Ranging (SLR) residuals retrieved using the 'GNSS High Precision Orbit Determination Software Tools' of the German Aerospace Center. An additional analysis of the uncertainty of these orbits is performed using the projection of the orbit uncertainty along the line-of-sight vector from the satellite laser ranging station to the satellite. Results demonstrate that the combined orbit solutions outperform the input orbits across a range of metrics, with the best-performing combined orbits produced using the arithmetic mean method achieving an SLR residual root mean-square error of 2 cm in 2022, and 4 cm in 2023, improving upon the next-best input orbit by 20 cm of error. This improvement is achieved without the loss of any volume of data. The combined orbits are additionally less sensitive to poor tracking data, and are therefore more robust than the inputs. These combined orbits show promise as use in gravity field estimation from Swarm kinematic orbits, as an improvement in the input orbits corresponds to an improvement in the gravity field model.

Finally, we explore the advantage of using SLR residuals to re-scale the input orbit noise such that each input is statistically consistent. We find that this technique further improves the quality of the combined orbits created using the inverse variance and variance component estimation techniques.

Chapter 1

Introduction

Gravity field models are essential tools in Earth sciences, providing critical insights into various geophysical processes such as ocean circulation and climate change. Of particular interest in light of the prevalent risk presented by climate change is the time-varying component of the gravity field; the component of the gravity field which meaningfully changes within a 5-year period. This provides us insight into the movement of water through the Earth's hydrosphere, and hence insight into global groundwater supply, drought and aquifers, as well as ice loss in the Antarctic and Greenland.

Gravity field models are typically estimated using data from dedicated satellite missions, with significant contributions from missions such as CHAllenging Minisatellite Payload (CHAMP), Gravity Recovery and Climate Experiment (GRACE), and Gravity Field and Steady-State Ocean Circulation Explorer (GOCE). The advent of high-accuracy Global Navigation Satellite System (GNSS) technology has enabled the estimation of gravity field models from tracking data, with the only requirement being an accurate dual-frequency GNSS receiver. This has the potential to significantly improve the spatial and, most importantly, temporal resolution of estimated gravity field models, enabling us to better monitor short-term geophysical and hydrological processes, which has in recent history been the sole purview of the GRACE-FO mission, launched in 2018 and active since.

It is of great benefit to have additional sources of high-accuracy gravity field models as an alternative to those derived from the GRACE mission. This is especially pertinent given the 10-month gap in GRACE data between the decommission of GRACE and the launch of its follow-on. Not only that, but we additionally know that the combination of several gravity field solutions produces more accurate models [65], and therefore that the collection of any additional gravity field information contributes to the improvement of our understanding of Earth's gravitational field. For this, we take advantage of the European Space Agency (ESA) Swarm mission, launched in 2013 and consisting of three satellites in Low Earth Orbit (LEO), which was specifically designed to measure Earth's magnetic field, but also provides valuable data for gravity field modelling through its accurate onboard GNSS receivers. Therefore, kinematic orbit solutions derived from Swarm GNSS data can be used to estimate gravity field models and provide us an alternative and independent source for the time-varying field, for which it is most important to have frequent data collection.

Improving models derived from Swarm data, then, is of great value to the geodetic community. Current practice is for independent analysis centres to produce unique kinematic orbit solutions from the raw GNSS data. They then apply their – typically unique – gravity field estimation techniques to solve for the gravity field coefficients. The resulting collection of models is then combined by research groups partaking in initiatives such as the International Combination Service for Time-variable

Gravity Fields (COST-G) to produce a final combined gravity field model, typically provided on a monthly basis.

What has not yet been investigated is the application of combination techniques to the kinematic orbits themselves prior to gravity field estimation. By improving the quality of the inputs to gravity field estimation, the models themselves are improved. In this thesis, we aim to improve the quality of the Swarm kinematic orbits by combining a set of input orbits provided by the Institute of Geodesy Graz (IFG), Astronomical Institute of the University of Bern (AIUB) and the Technical University of Delft (TUD) such that the combined orbit is closer to the underlying truth. In service of this, several combination techniques are proposed and assessed in order to determine if this novel approach has the potential to result in superior quality kinematic orbits for the purpose of gravity field estimation.

This thesis opens with the necessary background and context in chapter 2, outlining the state of the art in Swarm kinematic and RDO solutions, and formulating the central research questions. The combination and evaluation methodology is then detailed in chapter 4. Verification and sensitivity analyses of the developed approach are presented in chapter 5. The performance of the resulting combined orbits for 2022 and 2023 is assessed in chapter 6, and the findings are summarised in chapter 7.

Chapter 2

Background

The focus of this thesis is the improvement of Swarm Kinematic Orbits (KO)'s. To support the formulation of research questions, required background knowledge is first provided in this chapter. It begins by describing the Swarm mission and the nature of the orbit data it produces. The fundamental concepts of kinematic and RDO's are discussed, followed by an explanation of how these orbits are evaluated using SLR. Then, the current state of Swarm orbit solutions is reviewed, and finally, the open-source tools used in this thesis are introduced.

2.1 The Swarm mission

The Swarm mission, launched by ESA in 2013, consists of a constellation of three identical satellites – Swarm A, B and C – designed to study the Earth's magnetic field. Swarm A and C fly side-by-side with an initial altitude of 462 km at 87.35 degree inclination, while Swarm B had a higher initial altitude of 511 km and an inclination of 87.75 degrees [22]. Swarm's main goal is to provide high-precision and high-resolution measurements of the geomagnetic field, its temporal variations, and its interaction with the magnetosphere. Each satellite is equipped with a vector field magnetometer, an electric field instrument, and accelerometers, which together enable detailed mapping of the magnetic field generated by the Earth's core, mantle, crust, oceans, ionosphere, and magnetosphere. The Swarm satellites carry highly accurate GNSS receivers [67].

2.2 Orbit solutions

The field of Precise Orbit Determination (POD) deals with the determination of satellite orbits from tracking data. The precise tracking data of the Swarm satellites is disseminated from ESA servers. Several analysis centres produce their own orbit products from this data. Of key importance to this thesis are Swarm KO and RDO solutions. In this section, we describe the theory behind these orbits.

2.2.1 Kinematic orbits

Kinematic orbits are position time-series of a satellite's trajectory, derived through orbit determination techniques using only the geometry of the context retrieved from GNSS tracking data; that is, there are no force models or dynamics used in the kinematic orbit determination process [33].

Observation Model and Estimation Framework

The GNSS tracking data used in kinematic orbit determination consists primarily of pseudo-range (code) and carrier-phase measurements on two frequencies, f_1 and f_2, typically referred to as L1 and L2. These observations can be combined to reduce atmospheric and hardware biases. A common choice in POD is the use of the ionosphere-free linear combination, which effectively eliminates the first-order ionospheric delay – the effect of free electrons in the ionosphere on the propagation speed of the GNSS signal, introducing a frequency-dependent delay – and provides improved robustness for orbit determination [64].

The basic observation equations for code (R) and carrier-phase (Φ) in their ionosphere-free forms are expressed as:

$$R_{\mathrm{IF}} = \rho + c\,\delta t + d_{\mathrm{IF}} + \epsilon_{\mathrm{R}}, \tag{2.1}$$

$$\lambda_{\mathrm{IF}}\Phi_{\mathrm{IF}} = \rho + c\,\delta t + \lambda_{\mathrm{IF}} n + d_{\mathrm{IF}} + \epsilon_{\Phi}. \tag{2.2}$$

Here, R_{IF} and Φ_{IF} are the ionosphere-free combinations of the raw code and phase measurements, respectively. ρ denotes the geometric range between receiver and satellite, $c\,\delta t$ is the receiver clock offset, λ_{IF} is the wavelength of the ionosphere-free combination, n is the integer ambiguity, and d_{IF} represents the combined hardware delays from the satellite and receiver. ϵ_{R} and ϵ_{Φ} model the measurement noise for the code and phase observations. The integer ambiguity refers to the unknown number of full carrier wavelengths between the satellite and receiver at the start of tracking, as this information is not directly observable from the signal. Integer ambiguity resolution algorithms are applied to estimate this value and enable precise range retrieval from the carrier-phase measurements.

In practice, the phase observations are weighted more heavily than code observations due to their much lower noise levels (sub-centimetre compared to metre-level). However, code observations are essential to help constrain the clock and ambiguity parameters [64].

The general linearised observation model [70] can be written in least-squares form as:

$$y = A\hat{x} + \epsilon. \tag{2.3}$$

With y the observation vector, A the design matrix containing partial derivatives of observations with respect to the unknown parameters $\hat{x}$ (which include the satellite position components and receiver clock bias), and ϵ the residual vector.

The parameters $\hat{x}$ are obtained by minimising the cost function:

$$J = \epsilon^{\mathrm{T}} P_{yy}^{-1} \epsilon, \tag{2.4}$$

where P_{yy} is the covariance matrix of the observation noise. The weighted least-squares solution is then given by:

$$\hat{x} = \left(A^{\mathrm{T}} P_{yy}^{-1} A\right)^{-1} A^{\mathrm{T}} P_{yy}^{-1} y. \tag{2.5}$$

The covariance matrix of the estimated parameters is then:

$$P_{xx} = \left(A^{\mathrm{T}} P_{yy}^{-1} A\right)^{-1}. \tag{2.6}$$

This matrix expresses the formal precision of the estimated positions and clock parameters. These formal errors are highly sensitive to the satellite–receiver geometry and the weighting of the GNSS observations [33]. The resulting Standard Deviation (STD)'s and correlations (for example between along-track and cross-track directions) provide insight into the quality of the kinematic solution. Importantly, the covariance information primarily reflects the internal uncertainty distribution of the solution — that is, how precisely the parameters are estimated relative to one another under the assumed model and measurement noise. However, it does not necessarily reflect the absolute accuracy of the orbit with respect to the true trajectory, as it neglects modelling errors, measurement biases, and other external sources of error [33, 34].

2.2.2 Reduced-dynamic orbits

RDO determination aims to combine the strengths of kinematic and dynamic methods by estimating satellite positions using force models while preserving the inclusion of empirical accelerations. Unlike purely kinematic solutions, RDO's incorporate physical models—such as gravitational, atmospheric drag, and solar radiation pressure forces—to constrain the trajectory over time. However, they remain less rigid than fully dynamic orbits due to their treatment of unmodelled accelerations [33].

The observation model remains fundamentally the same as in the kinematic case, with GNSS phase and code measurements providing the observational data. However, instead of estimating the position at each epoch independently, a continuous trajectory is estimated by integrating the satellite's equation of motion, supplemented by stochastic parameters such as piecewise-constant accelerations or process noise.

This approach introduces a smoothness constraint, which increases robustness to poor geometry or measurement outliers, while still allowing for some model mismatches. The orbit is fit over longer arcs (typically several minutes to hours), and additional parameters such as accelerations $\vec{a}(t)$ are estimated to absorb unmodelled forces. This results in a solution that is both more accurate and more stable than kinematic orbits, particularly in the along-track direction where GNSS geometry is weakest [64].

2.2.3 Evaluation of orbit quality using SLR

Kinematic and RDO's are primarily evaluated through the use of SLR observations. SLR involves measuring the distance between a ground station and a satellite through the reflection of a laser off a retro-reflector on the surface of it [48]. This provides exceptionally accurate measurements of the range. These range observations are filtered for quality, corrected for known effects such as atmospheric refraction, and then used to produce normal points. Normal points are the reported range values, time-binned into 5-second windows, with an associated mean range. These points are timestamped at the midpoint of each window.

SLR stations are sparse, and therefore SLR observations are not available the vast majority of the time. In addition, there is a large discrepancy between the quality of observations from different stations, and a reduction in the quality of observations with decreasing elevation angle as the signal experiences a thicker atmosphere. Therefore, it is necessary to make a selection of specific high-performing stations and to optionally define an elevation cut-off angle [44], such as by selecting stations with a reported Root Mean Square (RMS) of the RDO SLR residuals below a certain threshold.

In this thesis, SLR normal point data is retrieved and processed using the GPS High precision Orbit determination Software Tools (GHOST) software toolkit further discussed in subsection 2.4.2, which provides station positions through the International Laser Ranging Service (ILRS) [51] and filtered observations corrected for known effects, including atmospheric refraction. The observations are time-binned into 5-second intervals, yielding normal points — reported mean ranges timestamped at the midpoint of each bin [68]. These are used to evaluate orbit accuracy by computing residuals with respect to the predicted satellite–station range. The computation of this is described in section 4.6.

2.3 State of the art Swarm orbits

The current state of the art of KO's derived from Swarm GNSS data are produced by AIUB, the TUD, and IFG. The full methods by which each institute derives their KO's is beyond the scope of this thesis. For a comprehensive description of the different KO's including their background models used, refer to Appendix G. These KO's are provided as time-series of positions along with their covariance data. KO data is available for each of the Swarm satellites; Swarm A, B and C.

State of the art RDO's are produced by the TUD and disseminated by ESA, as well as by IFG. These are provided as position and velocity time series for each of the Swarm satellites. In this thesis, we make use of the TUD RDO's, as the retrieval and processing of them is most convenient, and because the TUD RDO's performed better in preliminary testing; this is further detailed in Appendix E.

Several publications discuss the quality of these orbits. In general, Swarm orbit solutions have centimetre-level accuracy. A paper assessing the quality of the Swarm KO's produced by the TUD for the year of 2013–2014 reported SLR RMS errors of 3.25, 2.74 and 3.11 cm for Swarm A, B and C [30]. RMS errors of 2 cm were reported for the RDO's [29]. Another paper addressing four years of Swarm precise orbit determination from 2013 to 2018 reported a median daily RMS of the SLR residuals below 2.5 cm for their KO's and 0.5–1 cm for their RDO's [44]. In general, Swarm RDO's are 50–70% better than their KO counterparts, as evidenced by the referenced literature [29, 44].

This is a general indication of the performance of these orbits; an issue with directly comparing the performance of them — especially at different points in time — is their dependence on the quality of the tracking data which is sensitive to the satellite geometry, as well as other environmental factors such as solar activity. Additionally, the errors reported by the SLR validation are affected significantly by the particular station selection. For the assessment of orbits produced in this thesis, a direct comparison must be made using a consistent station selection.

2.4 Open-source software

Open source software is used in this thesis to synthesise orbits for verification tests, interpolate input orbits and validate the combined orbits.

2.4.1 Tudat - TU Delft Astrodynamics Toolbox

The TU Delft produces its own astrodynamics toolbox, TU Delft Astrodynamics Toolbox (Tudat) [16]. In this thesis, this toolkit is taken advantage of to propagate synthetic orbits for the purpose of testing. It is also used for interpolation functionality.

2.4.2 GHOST - GNSS High Precision Orbit Determination Software Tool

In order to assess the performance of the combined orbits, we require SLR validation. To perform this, we use the GHOST toolkit, which contains the SLRRES function used to retrieve the SLR observations of the satellite and the station positions, enabling the retrieval of orbit residuals along the station–satellite Line of Sight (LOS) vector [68].

Chapter 3

Research Proposal

This chapter outlines the motivation for this thesis, summarises the rationale for improving Swarm kinematic orbit quality and combining existing solutions. It contextualises this work within existing efforts in gravity field modelling, introduces known combination strategies from the geodetic community, and presents the central research we wish to address.

3.1 Why improvement in these orbits is desired

Researchers have identified a need for additional sources of gravitational field data supplementary to dedicated missions. The highly-accurate tracking data of Swarm may be used to estimate gravity field models, and since Swarm coverage has been continuous since its launch it is able to cover gaps in data from dedicated gravity field missions such as GRACE. Launched in 2002, on October 27, 2017, the first GRACE mission finally came to an end having exceeded its planned nominal lifetime by 10 years. The follow-on mission, Gravity Recovery and Climate Experiment Follow On (GRACE-FO), launched in May 2018 and began operations shortly after. Between these two missions lies a gap of around a half year wherein no high-quality satellite to satellite ranging time-varying gravity field data is available [29]. This highlights a problem with reliance on gravity field missions; coverage is not guaranteed, missions can fail and budgets for follow-on missions can run dry. The addition of quality gravity field models from Swarm orbit data is therefore desired.

Swarm kinematic orbits are used to estimate these supplementary gravity field models. These orbits are used as inputs in gravity field estimation techniques due to them not containing any a-priori gravity field models which would bias the resulting field [66]. Therefore, the improvement of Swarm kinematic orbits directly contributes to the improvement of gravity field models estimated from Swarm mission data [30]. Methods by which these gravity field models are estimated are provided in section F.5 for the curious reader.

3.2 COST-G and the Multi-Approach Swarm Gravity Field Initiative

COST-G, the International Gravity Field Service (IGFS) product centre for time-variable gravity fields, provides monthly global gravity field models derived by combining input from several analysis centres. These combinations are based on solutions or systems of normal equations, and aim to improve noise characteristics and overall model quality. The models are derived from multiple data sources, including GRACE, GRACE-FO, GNSS missions such as Swarm, and satellite laser ranging [32].

COST-G supports the goals of the European Gravity Service for Improved Emergency Management (EGSIEM) initiative, seeking to improve the robustness, reliability, and frequency of time-variable gravity field products. The TUD contributes to COST-G through its involvement in the generation and combination of Swarm-based gravity field models [65]. These Swarm models are based on kinematic orbits, and their accuracy is directly affected by the quality of the input orbits.

A specific research initiative within this context is the Multi-Approach Gravity Field Models from Swarm GPS Data initiative [65]. This group combines gravity field models generated from diverse processing strategies and input KO's across multiple analysis centres.

3.3 Rationale for combining kinematic orbits

As detailed in section 2.3, kinematic orbits produced by AIUB, TUD, and IFG differ due to variations in processing strategies, tracking networks, ambiguity resolution, and filtering [19]. These differences result in discrepancies at the centimetre level, both in terms of random error and systematic biases. Moreover, these solutions contain unique data gaps, outliers, and inconsistencies in uncertainty estimation. These variations are not aligned across Analysis Centre (AC)'s, presenting an opportunity for error reduction through combination.

In many areas of space geodesy, the combination of independently derived products has proven effective for improving accuracy. This principle underpins, for example, the realisation of global terrestrial reference frames by the International Earth Rotation and Reference Systems Service (IERS), the merging of Earth orientation parameters, and the combination of gravity field models from the Swarm mission [10, 65]. The most accurate gravity field models are themselves the product of combinations of many different forms of gravity field information, such as the XGM2019e model which is the product of the combination of satellite data from GOCE, GRACE with terrestrial data and topographic information [77]. Applying a similar principle to Swarm kinematic orbits may yield improved orbits with improved accuracy, robustness, and data density.

3.4 Overview of Time Series Combination Strategies in Geodesy

A range of strategies have been proposed in literature – particularly in GNSS and gravity field modelling – for combining independent datasets with differing accuracy, resolution, and noise characteristics. Broadly speaking, combination strategies fall into two categories:

- **Unweighted or equally weighted methods**, which assume that all inputs are of similar quality and simply average the available data.
- **Weighted methods**, which attempt to assign different levels of influence to each input based on its quality ; either through reported uncertainty, empirical variance, or optimisation against

a reference.

Weighted approaches can further be divided by how the weights are determined:

- *A priori*, using errors reported by the ACs.
- *A posteriori*, using residuals, variance estimates, or reference data.
- *Optimisation-based*, fitting weights to minimise an objective function such as RMS difference with a reference RDO.

Challenges arise when input uncertainties are not comparable or are inconsistently scaled, when biases are present, or when data gaps exist. This makes it non-trivial to define a single best approach. Some methods, such as Variance Component Estimation (VCE), attempt to solve this by estimating weights iteratively from internal consistency. Additional risks arise with methods which involve using RDO reference data due to the possibility of biasing the combination to an a-priori gravity field model. In this thesis, these general strategies are adapted and tested for the Swarm orbit combination context. Full descriptions, derivations, and verification and validation procedures are presented in chapter 4.

3.5 Research gap and research questions

While previous efforts have combined gravity field models and normal equations from multiple sources [10, 65], and this principle has been applied to gravity field models estimated from Swarm tracking data, we identify that no study has yet attempted to directly combine the state of the art Swarm kinematic orbits from which gravity field models are estimated. This is despite the known differences in outliers, epoch-specific accuracy and volume of data among these orbits from the IFG, TUD, and AIUB – as discussed in section 2.3 –, as well as the known benefits of combination techniques discussed in section 3.3.

The central hypothesis of this work is that a combined Swarm kinematic orbit product – derived from independent state of the art solutions – will yield a more accurate, complete, and robust time series suitable for gravity field estimation, which depends critically on the quality of the underlying orbits. This leads to the primary research question:

> **What improvement in the quality of Swarm kinematic orbits can be achieved through the combination of unique solutions?**

To support this, three sub-questions are explored, each addressing a methodological challenge:

- **Sub-question 1: To what extent do combination methods which utilise a reference RDO risk introducing bias or over-fitting?**
  The use of RDO's as reference in a combination method may bias the combined solution towards it, and therefore to an a-priori gravity field model. This is an issue, as improvement in these orbits is sought for its downstream effect on the quality of Swarm gravity field models, which we specifically do not want to bias towards an a-priori field. We therefore evaluate how sensitive combination methods leveraging an RDO reference are to this risk.

- **Sub-question 2: Can the uncertainty information of independent kinematic orbits effectively be propagated into a combined solution?**
  Each analysis centre provides per-epoch uncertainty estimates. However, different analysis centres compute their covariances differently, and scaling is inconsistent between them. Therefore, this inconsistency may be propagated into the combined solution. This question investigates whether the per epoch uncertainties can be used to construct realistic uncertainty estimates for the combined solution, which is useful for downstream gravity field estimation. For this purpose, it is key that the uncertainty is distributed realistically, as it is used as a relative measure of the accuracy of the input positions. It is propagated through the models, reflecting in the uncertainties reported for the spherical harmonic coefficients. Generally, it is most important that the uncertainty is internally consistent; that the distribution is correct such that when the orbit performs worse this is reflected in the uncertainty information, even if the magnitude does not map on to reality. The magnitude of the uncertainty is less important for individual gravity field solutions, however combined gravity fields such as those produced by COST-G use this uncertainty in the weighting of the combination [32]. In this case, it becomes important to have realistic and consistent uncertainty values in the orbits.
- **Sub-question 3: Can we utilise independent sources of validation to re-scale the uncertainty of the input orbits in order to improve the combination?**
  Given the inconsistencies observed in provided uncertainty information, this question explores the use of external validation tools such as SLR residuals to calibrate and rescale uncertainty information. We seek to determine whether this makes the propagated uncertainty more realistic, and whether it improves the performance of combination methods which depend upon the input uncertainty.

These questions motivate the design and evaluation of multiple combination strategies – including statistical and optimisation-based approaches – which are presented in detail in chapter 4. To frame this research, we define a reference case to which we compare the output of our combination methods. This is the current Swarm state of the art kinematic orbits and RDO detailed in section 2.3. An advantage of the combined orbit over the input kinematic orbits is considered a success, while we use the RDO as a reference as it is known to be more accurate and robust.

3.6 Research requirements

In order to address the research questions outlined above, a number of requirements must be satisfied across data availability, methodological capability, and validation strategy:

1. Input and reference orbits

To test combination strategies, we use three state of the art Swarm kinematic orbit solutions provided by the IFG, TUD and AIUB as input. For combination methods leveraging a reference orbit, we use a state of the art Swarm RDO produced by the TUD and disseminated by ESA. These must be:

- Covering overlapping time periods for meaningful comparison and combination,

- Including per-epoch formal uncertainty or covariance information,
- Of comparable accuracy.

The availability of these products and their uncertainty information is a prerequisite for combination.

2. Consistent pre-processing and epoch alignment

Differences in data sampling, data gaps, and inconsistent time tagging must be resolved. This requires the following:

- Interpolation, truncation, or epoch matching strategies to ensure a common set of epochs for comparison and combination,
- Robust handling of missing or corrupted data to avoid including outliers and poor data in the combined solution.

This step ensures consistency and compatibility between input orbit solutions by enforcing the requirement that we combine orbits one a per-epoch basis.

3. Implementation of multiple combination strategies

To investigate the potential benefits and limitations of different approaches, the following families of methods must be implemented:

- **Unweighted methods:** Simple arithmetic mean to act as a baseline,
- **Weighted methods:** Including a-priori (uncertainty-based), a-posteriori (residual-based), and optimisation-based techniques (e.g., RMS minimisation) and VCE.

Implementation should include propagation of uncertainties into the combined solution where possible. Furthermore, combination should be designed to retain as much data as possible. As such, methods are required to make use of the maximum possible data per-epoch; they should use the union of each input dataset, rather than the intersection.

4. Validation of orbits

To evaluate the quality of the combined orbits, access to external validation sources is necessary. In this thesis, we use SLR observations which are retrieved using the GHOST software. This data must be available across the chosen case study period. We make use of the RMS of the SLR residuals as a measure of orbit accuracy. In addition, we compare the number of epochs in the combined solutions with the input data and the reference RDO to assess data retention.

5. Evaluation mechanisms

To properly interpret the output of each combination method, the following evaluation mechanisms are required:

- Sensitivity analyses to bias, noise, and incomplete data inputs,
- Residual analysis with respect to the RDO,
- Reduced Chi-squared metrics and RMS-to-STD ratios to assess uncertainty realism.

These tools are essential to answering the sub-questions related to over-fitting and uncertainty propagation.

6. Reproducibility and automation

Given the scale of the orbit datasets and the number of strategies under consideration, the research also requires:

- A reproducible and automated pipeline for data acquisition, pre-processing, combination, and validation,
- A framework to support large-scale testing across time periods and satellites,
- The ability to apply combination methods to each Swarm satellite.

This ensures that the methodology is extensible to future missions and datasets, and that multiple test cases can be analysed. The methodology presented in chapter 4 is defined with these requirements in consideration.

Chapter 4

Methodology

In this chapter, we introduce the approach to answering the research questions proposed in section 3.5. The methods by which KO's are combined in this thesis are defined. We discuss the verification test set-ups, sensitivity analyses to be performed and validation methodology using SLR observations used to assist with the answering of our research questions.

In terms of performance, there are two key metrics; the accuracy of the orbits and volume of data. Accuracy refers to how close the combined solution is to reality, while volume refers to the number of epochs it contains. It is clear that accuracy is important, however volume is equally so as the more quality data points provided to gravity field inversion methods, the better their solutions become. The goal of each combination method is to maximise both metrics. Several combination methods will be introduced including stand-alone methods dependent only on the KO data and methods using the available RDO's.

We begin with the retrieval and pre-processing of the data used to answer our research questions Sections 4.2 and 4.3 introduce the combination methods which, in conjunction with the validation scheme defined in section 4.6, are used to answer our primary research question. The method by which sub-question 1 will be answered is introduced in section 4.4. The approach to answering sub-questions 2 and 3 is given by the uncertainty propagation introduced along with the methods in sections 4.2 and 4.3, with the validation scheme provided in section 4.5.

4.1 Retrieval of data and pre-processing

Before any combination methods are applied, it is necessary to pre-process data. We can separate the data used into two types; input and validation data. The KO's and RDO's are considered input data, while SLR data is considered validation data.

Kinematic orbits from IFG, TUD and AIUB are retrieved from the TUD university servers. RDO's provided in sp3c format produced by the TUD are retrieved from the ESA dissemination servers [21]. In this thesis, we will henceforth refer to the RDO's as the ESA RDO's, to distinguish them clearly from the TUD kinematic orbits. The sp3k format is an extension of the sp3c format [28] developed by the Multi-approach gravity fields from Swarm GNSS data team [19]. The relevant change to the standard sp3c format is the increase of precision to sub-millimetre of the position and variance information. This increased precision is valuable as kinematic (and especially reduced-dynamic) orbits continue to improve and their uncertainties reduce. These orbit files contain the following information relevant to the combination:

- The trajectory data in Cartesian coordinates (x,y and z) in kilometres. The KO data is in the IGS20 reference frame [52]; it is an Earth-centred, Earth-fixed reference frame. The epochs are given in Global Positioning Service (GPS) time; therefore currently 18 seconds ahead of Coordinated Universal Time (UTC). The ESA RDO's is in IGS20 reference frame and GPS time.
- STD and covariance information for the KO's. This includes the STD of each parameter (σ_{xx}, σ_{yy}, σ_{zz}) given in mm, and the covariances of each parameter in the form of correlation coefficients multiplied by 10,000,000.
- The epoch information. This is used to synchronise different sets of data.

4.1.1 Pre-processing

The formats of the provided data are not consistent in terms of the time stamps, and the units of data must be consolidated. That process is described in this section.

Consolidating data format

All files are retrieved from their respective servers automatically and parsed in order to extract the desired information. The trajectory data is then converted from kilometres to meters, the STD's converted from millimetres to meters, the correlation factors converted to covariances in m^2, and the epoch time converted to UTC from GPS by accounting for leap-seconds using the Astropy library [1]; this is done for improved compatibility with the astrodynamics libraries available in Python required for coordinate transformations. The correlation factors are converted to covariances by the following:

$$\sigma_{xy} = \rho_{x,y} \cdot (\sigma_{xx}\sigma_{yy}). \tag{4.1}$$

The variable $\rho_{x,y}$ is the correlation factor between the x and y component. The end result is all orbit data in the IGS20 reference frame.

Consolidating epoch time stamps

While the provided KO data each have a 1 second sampling, it was observed to not always be exact. Similarly, the RDO data does not have exact 10 second sampling. This presents problems; KO's are to be combined at epochs at which several data sources are available, and therefore slight deviations in time must be consolidated to ensure compatibility between them. Any methods which depend upon RDO data rely upon having RDO data at each epoch, therefore it needs to be resampled to a 1 second frequency. We choose to resample the RDO's to 1 second instead of downsampling the KO's to 10 second sampling as our goal is to produce as much quality data as possible. In order to solve these issues, interpolation is applied to resample the input data to whole seconds. Interpolation is performed using a Lagrange interpolator from the Tudat package [16]. The Lagrange interpolating polynomial is given by:

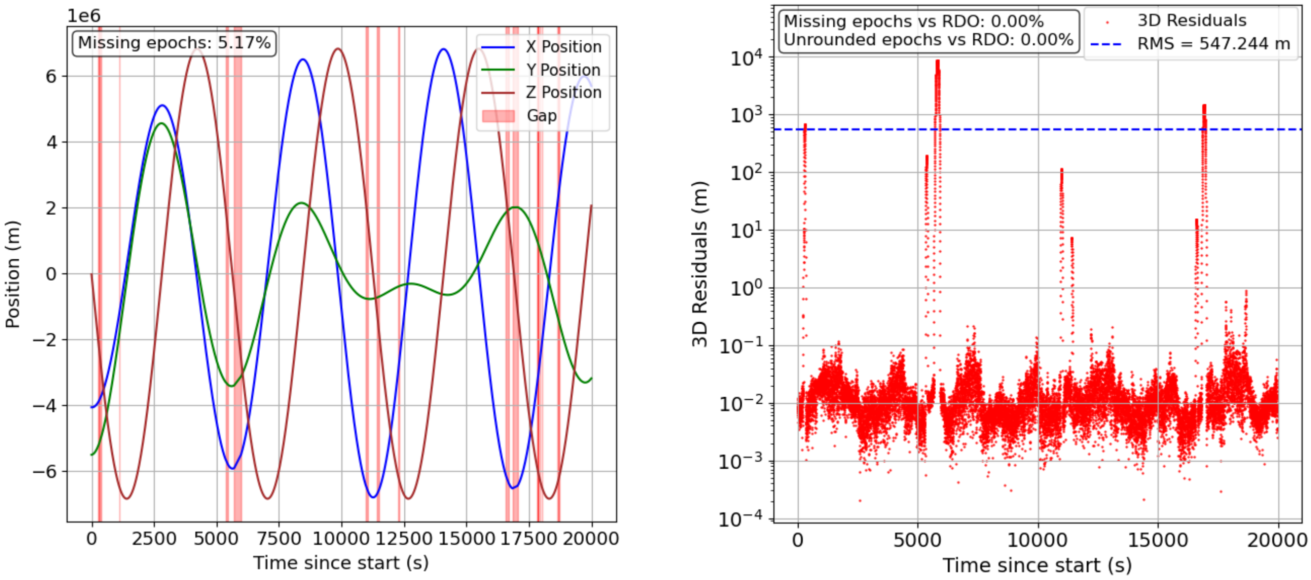
$$P_n(t) = \sum_{i=0}^{n} y_i \, \ell_i(t), \tag{4.2}$$

where the Lagrange basis polynomials are defined as:

$$\ell_i(t) = \prod_{\substack{j=0 \\ j \neq i}}^{n} \frac{t - t_j}{t_i - t_j}, \tag{4.3}$$

with $t_0, t_1, \ldots, t_n$ being the interpolation nodes and $y_i = f(t_i)$ the known function values at these nodes.

This polynomial provides a unique fit through all $n+1$ data points such that $P_n(t_i) = y_i$ for all i. In this work, a polynomial of order $n = 10$ is used, meaning that 5 data points either side of a central point are used to construct the interpolating polynomial.



(a)

(b)

Figure 4.1: Presentation of gaps in AIUB input orbit for a typical day (1st January 2023) and the resulting interpolation error which results if they are unaccounted for.

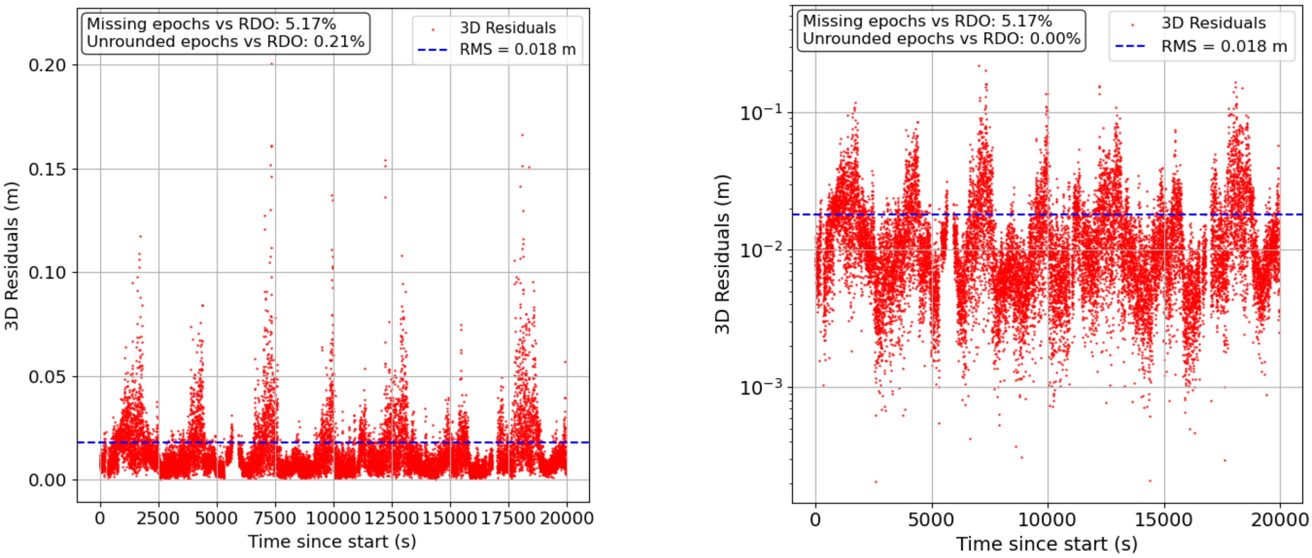
Straightforward application of this interpolation scheme does not work for this data, as it tends to have gaps of missing data ranging from a few seconds to hours; especially in the KO's, which can be seen in Figure 4.1a where an example of AIUB input data is presented showing the frequency of gaps on a normal day. In total, in the first month of January 2023, the percentage of missing epochs for each dataset is shown in Table 4.1. Gap sizes of many minutes are observed. While the interpolator handles consistent data with minimal gaps – that is, gaps of no more than 3-5 missing epochs –, they fail when presented with much larger gaps of minutes, as the polynomial is essentially fit to two independent arcs of the orbit at vastly different positions in space.

Table 4.1: Percentage of missing epochs in kinematic orbit data for January 2023.

Analysis Centre	**Missing Epochs (%)**
IFG	0.0%
AIUB	7.9%
TUD	13.2%

By applying interpolation without taking these gaps into account, we erroneously create many data points which are very inaccurate. In order to determine the impact of this, the RDO data is used as reference for comparison as it has consistent epochs and consistent frequency of 10 seconds, meaning that interpolation can safely be applied without gap-handling. When no gap-handling is used alongside interpolation, extremely high residuals in excess of 1km are observed as in plot b of Figure 4.1b where gaps were present.

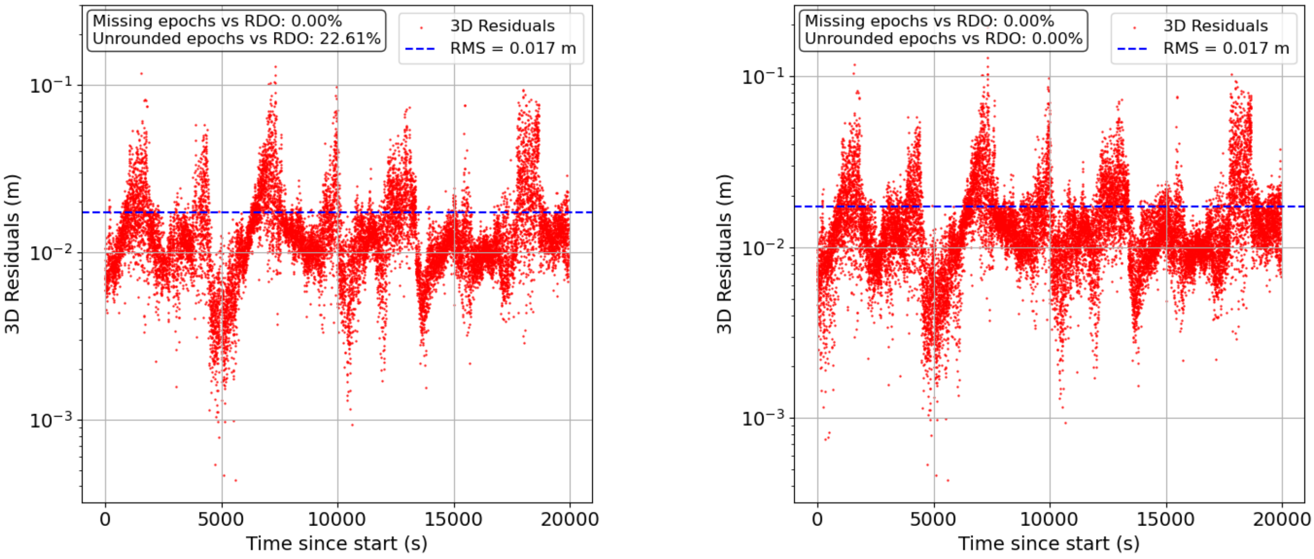
To avoid this, we segment the data into continuous chunks separated by gaps exceeding 5 seconds, selected as it does not reduce number of epochs as shown in figures 4.2 and 4.3. We then apply interpolation to each chunk independently before concatenating them to retrieve the full interpolated data, with gaps left in place. A limitation of this method is that near the boundaries of gaps there are insufficient points from which to construct the full polynomial, leading to less accuracy. To account for this, following the interpolation 5 points are removed from each boundary to avoid including this data. In figures 4.2 and 4.3 the residuals of non-interpolated and interpolated KO's from the AIUB and TUD with respect to the RDO data is presented, including the number of epochs which do not overlap due to either missing data or unconsolidated epochs times.



(a) AIUB input data, without interpolation

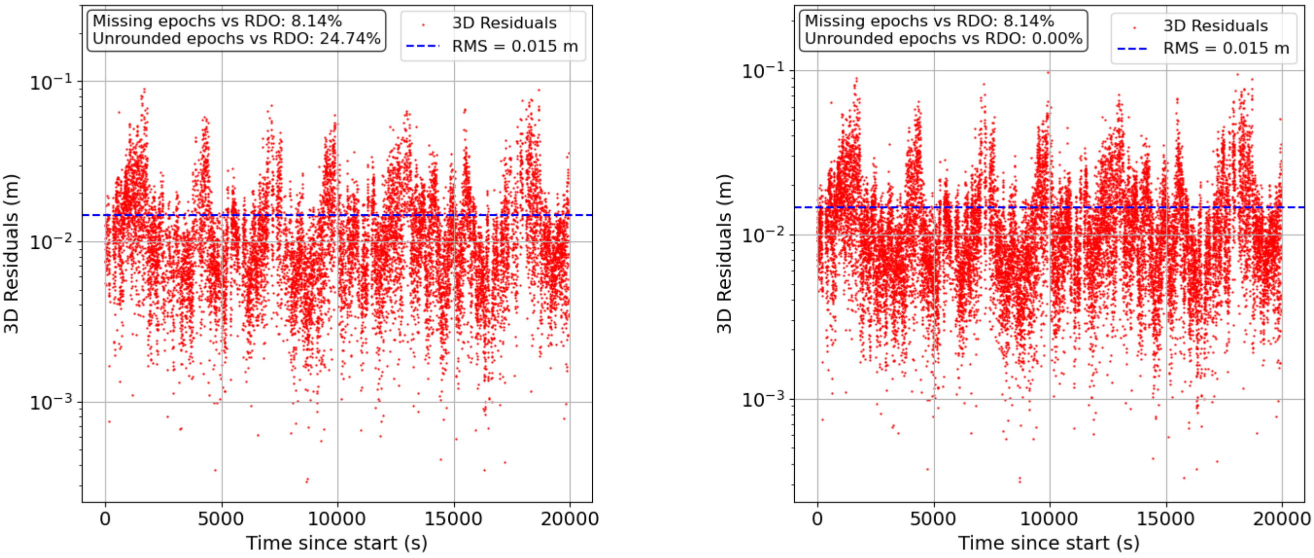
(b) AIUB input data, with interpolation

Figure 4.2: KO residuals for AIUB before and after interpolation. Demonstrates the effect of epoch interpolation on alignment with RDO data.



(a) IFG input data, without interpolation

(b) IFG input data, with interpolation



(c) TUD input data, without interpolation

(d) TUD input data, with interpolation

Figure 4.3: KO residuals for IFG and TUD before and after interpolation. Epoch consolidation improves alignment with RDO data while preserving orbit fidelity.

Figures 4.2 and 4.3 show that interpolation is correctly consolidating the timescale as the number of unrounded epochs compared to the RDO reduces to 0%. It is also retaining epochs and is not introducing error into the input orbits, demonstrated by the consistent RMS and the same number of missing epochs before and after its application.

Screening of kinematic orbits with RDO data

This step is designed to remove outliers from the input data. Due to the combined solution taking advantage of multiple sets of input data, filtering can be applied without excessive loss of volume of data; this takes advantage of the fact that different sets of input data are produced using different methods which lead to different sources of error, and therefore result in different outliers. In order to filter outliers, the magnitude of the position difference with respect to the RDO data is used. At time t and position state $\vec{x}$, we get the following:

$$e_t = \left\| \vec{x}_t^{\,(\mathrm{RDO})} - \vec{x}_t^{\,(\mathrm{KO})} \right\| \tag{4.4}$$

A maximum is defined, and input orbit data which has greater residuals than this is discarded; the specific selection of this screening threshold is addressed in section 5.5. This leverages the fact that RDO's are more robust than KO's [76] due to the dynamic model acting as a filter for outliers.

4.2 Independent combination methods

In this section the independent combination methods are introduced. These are defined as combination methods which only involve information contained within the input KO data. The advantage of this is that these methods are only reliant on KO's being available. Additionally, they are computationally inexpensive. Expectations of the performance and behaviour of the methods is discussed, to be tested in chapter 5.

4.2.1 Arithmetic mean

The most straightforward approach is to take the average of the inputs. At the level of the KO's, this means taking the arithmetic mean of each parameter of the positions at each epoch using:

$$\vec{r}_{i,t}^{\,(\mathrm{mean})} = \frac{\sum_{i=1}^{N_{\mathrm{sol}}} \vec{r}_{i,t}}{N_{\mathrm{sol}}}. \tag{4.5}$$

The superscript t represents each epoch. N_{sol} is the number of input KO's, $\vec{r}^{\,t}_{\mathrm{mean}}$ is the position of the satellite, while i refers to a particular input orbit. This approach is computationally inexpensive. Conversely, by not including available stochastic information, we ignore the estimated variance of each solution in the combination, potentially leading the combined solution to be biased towards inputs with greater errors. The expectation is that this solution produces a combined solution superior to the set of input solutions only if the input solutions are approximately randomly distributed about the 'true' position of the satellite. Essentially, as the distribution of the set of input orbit solutions approaches a random distribution about the underlying 'truth', the mean of the set of solutions approaches the 'true' value. Of course, in reality each input solution can be perceived as having a bias (or offset) with respect to the true value, as well as different variance.

In practice this means that the fewer the number of input solutions, the less confident we can be that the combined solution is superior when using this method. In addition, this method blindly trusts the input data by assuming equally valid inputs; therefore, it is particularly sensitive to biases in input data.

Uncertainty propagation is applied under the assumption that the input solutions are uncorrelated. The combined STD for each component $j \in \{x, y, z\}$ is computed as:

$$\sigma_{j,t}^{(\mathrm{mean})} = \sqrt{\frac{1}{N_{\mathrm{sol}}^2} \sum_{i=1}^{N_{\mathrm{sol}}} \left(\sigma_{i,j,t}\right)^2}. \tag{4.6}$$

The variable $\sigma_{i,j,t}$ is the reported STD of component j from solution i at epoch t.

The corresponding off-diagonal covariance terms for each pair $(j,k) \in \{(x,y),(x,z),(y,z)\}$ are computed as:

$$\sigma_{j,k,t}^{(\text{mean})} = \frac{1}{N_{\text{sol}}^2} \sum_{i=1}^{N_{\text{sol}}} \sigma_{i,j,k,t}. \tag{4.7}$$

This formulation propagates the variance and covariance assuming equal confidence in all input solutions.

4.2.2 Inverse-variance weighting

A weighted combination can be performed by taking advantage of the known variance information. In practice, different observations of a parameter using different methods or instruments have different variance associated with them. In this case, combining the observations of the parameter by taking the mean is not an optimal estimator, as each observation is not equally accurate. Instead, a combination method weighting the input observations by the inverse of their variance becomes the optimal estimator, as it accounts for the known difference in the qualities of the different observations [71].

Our observations are different estimations of the same parameters, all based on (approximately) the same input GPS data. Weights are then computed as the inverse of the variance:

$$P_{i,t} = \begin{bmatrix} \frac{1}{(\sigma_{x,i,t})^2} & 0 & 0 \\ 0 & \frac{1}{(\sigma_{y,i,t})^2} & 0 \\ 0 & 0 & \frac{1}{(\sigma_{z,i,t})^2} \end{bmatrix}. \tag{4.8}$$

The weighted combination is then performed using:

$$\vec{r}_{i,t}^{\,(\text{inv})} = \left(\sum_{i=1}^{N_{\text{sol}}} P_{i,t} \right)^{-1} \sum_{i=1}^{N_{\text{sol}}} P_{i,t} \vec{r}_{i,t}. \tag{4.9}$$

In this formulation, $\vec{r}_{i,t}^{\,(\text{inv})}$ is the combined position at epoch t, N_{sol} is the number of KO solutions, $\vec{r}_{i,t}$ is the position from solution i, and $P_{i,t}$ is the weight matrix for solution i at epoch t.

This method is dependent on the variance reported by the KO's, which, as discussed in subsection 2.2.1, is not necessarily consistent across AC's, particularly in magnitude. Additionally, due to the presence of unknown biases in each input orbit, there is no guarantee that an input solution with lower estimated variance is actually closer to the truth.

Uncertainty propagation: The propagated uncertainty of the combined solution is computed directly from the input variances and covariances. For each component $j \in \{x,y,z\}$, the combined STD is given by:

$$\sigma_{j,t}^{(\text{inv})} = \sqrt{\left(\sum_{i=1}^{N_{\text{sol}}} \frac{1}{(\sigma_{j,i,t})^2} \right)^{-1}}. \tag{4.10}$$

This equation assumes uncorrelated input errors between solutions. For the covariance terms between coordinates j and k, the following expression is used:

$$\sigma_{jk,t}^{(\mathrm{inv})} = \sum_{i=1}^{N_{\mathrm{sol}}} w_{j,i,t} \cdot w_{k,i,t} \cdot \sigma_{jk,i,t}. \tag{4.11}$$

Here, $w_{j,i,t}$ is the normalised weight for component j of input i at epoch t, and $\sigma_{jk,i,t}$ is the reported covariance between components j and k at epoch t for input i. This formulation enables full covariance matrix propagation.

4.2.3 Simplified variance component estimation

VCE is a statistical method used to optimally combine different sets of observations. Weighting purely using the variance information provided with each input solution has a flaw; we assume that the variance information is consistent between AC's. In reality, this is not the case [2]. VCE accounts for this by enabling the scaling of the various information between inputs. Internal relations of variance within each input solution are assumed fixed. Then, the variance of each input solution is estimated with respect to the weighted combined solution, leading to a set of variance components corresponding to each input. These newly estimated variance components are used to recompute the combined solution, after which the variance components can be estimated once more. The solution is iterated upon until changes in the variance components fall below a threshold. The result of this process is a single parameter – a variance component – for each parameter for each input solution. To start, we redefine the weighted combination as:

$$\vec{r_t}^{(\mathrm{vce})} = \frac{\sum_{i=1}^{N_{\mathrm{sol}}} \hat{\sigma}_{i,j}^2 Q_{i,t} \vec{r}_{i,t}}{\sum_{i=1}^{N_{\mathrm{sol}}} \hat{\sigma}_{i,j}^2 Q_{i,t}}. \tag{4.12}$$

This is almost identical to Equation 4.9 aside from the weight matrix $P_{i,t}$, which has been split into two terms: $\hat{\sigma}_{i,j}^2$ and $Q_{i,t}$. The former is the variance component, and the latter the internal variance. The variable $\vec{r_t}^{(\mathrm{vce})}$ is the combined solution at epoch t. As discussed, it can be seen that the variance component acts to scale the variance of each solution, causing it to contribute differently to the final combination while maintaining the internal structure across epochs defined by the known variance information.

Now, the method used to estimate the variance components themselves — $\hat{\sigma}_{i,j}^2$ — is introduced. The first approach to VCE was derived by Helmert, which sought to estimate weights between uncorrelated sets of observations. Later additions by Förstner, Kubik, and Ebner extended the approach to consider observations with known correlation [2]. In this thesis, the approach is limited to the estimation of variance components alone, as we assume no correlation between the sets of input data.

Due to the reliance of more complex methods on computationally expensive matrix computations, within the scope of this thesis we limit ourselves to one of Helmert's simplified estimators. This approximates the variance components of each input dataset through the use of its residual with respect to the combined solution. It was decided to use an algorithm already applied to gravity field combination [34] with some modifications to include the a-priori variance information in the combination.

VCE is performed independently for each state component $j \in \{x, y, z\}$, leading to three variance components for each input orbit. An initial guess of the weighted combination is made by assuming initial variance components are equal to one. With this initial combined orbit, the variance components of a parameter are updated as follows:

$$\hat{\sigma}^2_{i,j,k} = \frac{\vec{v}^\top_{i,j} \vec{\sigma}^2_{i,j} \vec{v}_{i,j}}{N_{\mathrm{obs},i} \left(1 - \frac{\hat{\sigma}^{-2}_{i,j,k-1}}{\sum_{i=1}^{N_{\mathrm{sol}}} \hat{\sigma}^{-2}_{i,j,k-1}}\right)}. \tag{4.13}$$

Here, k refers to the iteration index, i to the analysis centre, j to the state component, $\vec{\sigma}^2_{i,j}$ to the variance of the state component at each epoch, and $N_{\mathrm{obs},i}$ to the number of valid observations. A residual vector $\vec{v}_{i,j}$ is computed as:

$$\vec{v}_{i,j,k} = \vec{r}_{c,j,k} - \vec{r}_{i,j}. \tag{4.14}$$

This is evaluated for each coordinate j, where $\vec{r}_{c,j,k}$ is the combined estimate of component j at iteration k, and $\vec{r}_{i,j}$ is the corresponding vector from analysis centre i. These residuals represent the difference between the input KO's and the combined solution at this iteration depth. Convergence is typically achieved within 5–10 iterations, based on the convergence criterion:

$$\|\vec{\sigma}_k - \vec{\sigma}_{k-1}\| < \mathrm{tol}, \tag{4.15}$$

where $\vec{\sigma}$ is the vector containing the variance components of each coordinate. The tolerance is set to $\mathrm{tol} = 1 \times 10^{-5}$, ensuring that changes in the magnitude of the vector of variance components between iterations are negligible compared to the scale of the STD's being estimated (typically on the order of mm).

Once the variance components have been estimated, the uncertainty of the combined solution can be propagated using a VCE-weighted formula. The STD for each coordinate is computed as:

$$\sigma^{(\mathrm{vce})}_{j,t} = \sqrt{\left(\sum_{i=1}^{N_{\mathrm{sol}}} \frac{1}{\hat{\sigma}^2_{i,j} \cdot (\sigma_{i,j,t})^2}\right)^{-1}}. \tag{4.16}$$

For each covariance pair $(j, k) \in \{(x, y), (x, z), (y, z)\}$, the uncertainty is computed using weighted averaging:

$$\sigma^{(\mathrm{vce})}_{jk,t} = \frac{\sum_{i=1}^{N_{\mathrm{sol}}} \sigma_{i,jk,t} \cdot w_{i,jk,t}}{\sum_{i=1}^{N_{\mathrm{sol}}} w_{i,jk,t}}. \tag{4.17}$$

Here, the composite weights $w_{i,jk,t}$ incorporate both the estimated variance components and the uncertainties:

$$w_{i,jk,t} = \frac{1}{\hat{\sigma}^2_{i,j}(\sigma_{i,j,t})^2 + \hat{\sigma}^2_{i,k}(\sigma_{i,k,t})^2}. \tag{4.18}$$

This method retains the internal structure of the covariance while adjusting for inter-centre inconsistencies in scale, offering a more physically realistic estimate of uncertainty.

4.3 Combination using reference data

Combining estimations of parameters without the use of external information is limited by the quality of the input data. In this section, we introduce methods which leverage external RDO data as reference.

4.3.1 Weights derived from residuals with respect to RDO data

RDO data is closer to reality than KO data [44]. Therefore, we can estimate the true variance of our input orbits using their residuals with respect to it. Theoretically, if a set of data is much closer to the truth than another, then the difference between the two mainly consists of the random noise and bias inherent to the less accurate data. In our case, the difference in accuracy is approximately 50-70% as described in section 2.3. We have RDO data available at each epoch; therefore, the weighted combined solution is defined as:

$$\vec{r}_t^{\,(\mathrm{res})} = \frac{\sum_{i=1}^{N_{\mathrm{sol}}} \left(\vec{v}_{i,t}\right)^{-2} \vec{r}_{i,t}}{\sum_{i=1}^{N_{\mathrm{sol}}} \left(\vec{v}_{i,t}\right)^{-2}}. \tag{4.19}$$

In this expression, $\vec{v}_{i,t}$ is the residual vector of the input orbit from analysis centre i with respect to the reference orbit at epoch t, and $\vec{r}_t^{\,(\mathrm{res})}$ is the combined solution at that epoch. This method assumes that errors in the reference data are negligible. In practice, this is not true, and therefore the method may conflate errors from both the input orbits and the RDO. Additionally, there is a risk of over-fitting to the RDO, for which there is no control in this method. Fundamentally, this method down-weights data which is further from the RDO, and up-weights data which is closer. In order to propagate the uncertainty of the combined orbit, we use the residual-derived weights to combine the covariance information of each input. For each coordinate $j \in \{x, y, z\}$, the weights are defined as:

$$w_{i,j,t} = \frac{1}{\left(v_{i,j,t}\right)^2 + \epsilon}. \tag{4.20}$$

Here, $v_{i,j,t}$ is the residual of coordinate j at epoch t for analysis centre i, and ϵ is a small stabilisation constant to prevent division by zero. Using these weights, the STD of the combined solution for each coordinate is given by:

$$\sigma_{j,t}^{(\mathrm{res})} = \sqrt{\sum_{i=1}^{N_{\mathrm{sol}}} \left(w_{i,j,t}\right)^2 \cdot \left(\sigma_{i,j,t}\right)^2}. \tag{4.21}$$

For the covariance terms between coordinate pairs $(j, k) \in \{(x, y), (x, z), (y, z)\}$, the combined value is computed as:

$$\sigma_{j,k,t}^{(\mathrm{res})} = \frac{\sum_{i=1}^{N_{\mathrm{sol}}} w_{i,j,t} \cdot w_{i,k,t} \cdot \sigma_{i,jk,t}}{\sum_{i=1}^{N_{\mathrm{sol}}} w_{i,j,t} \cdot w_{i,k,t}}. \tag{4.22}$$

This approach allows the uncertainty structure of the input orbits to influence the uncertainty of the combined orbit, but is highly sensitive to the accuracy of the reference data. Because the weights are derived from differences with the RDO, any error or bias in the reference data will directly affect the combined solution and its uncertainty estimate.

4.3.2 Fitting to RDO Data - optimisation

There is an assumption in the geodetic community that gravity field models ought not be estimated from orbit solutions which themselves depend on a-priori gravity field information [66]. This is a reasonable assumption as the a-priori gravity field model information implicitly present within the RDO data through the dynamic force model used to estimate them, which is likely to bias the estimated solution towards it. Therefore, we wish to explore a method by which we can exploit RDO's to inform the combination without excessively biasing the combination.

To address this, we can define a set of piece-wise constant weights across arcs, defined as a segment of the overall trajectory with a given length of time (t_{arc}). If we define the arc length well, we may avoid over-fitting. Additionally, the structure of the input orbits is retained; epochs are not weighted independently, therefore the 'shape' of each input orbit is preserved within each arc. In turn, this should help to preserve the signal of the gravity field present implicitly within the orbit data. Uncertainty propagation is not implemented in these optimisation methods due to time constraints.

Fitting methods and algorithms

We apply determine these piece-wise constant weights with optimisation algorithms from Python Parallel Global Multiobjective Optimizer (PyGMO) [9], an open-source library containing a number of state-of-the-art optimisation algorithms which can handle single and multi-objective, (un)constrained and bounded optimization problems.

Combination

The most straightforward approach is to define weights for each parameter x,y and z for each analysis centre. The weighted combination is then simply the weighted sum of each parameter as:

$$\vec{r_t}^{(\text{opt})} = \sum_{i=1}^{N} \begin{bmatrix} w_{x,i} & 0 & 0 \\ 0 & w_{y,i} & 0 \\ 0 & 0 & w_{z,i} \end{bmatrix} \vec{r}_{i,t}. \tag{4.23}$$

The variable t refers to the epoch, w to the weights, and i to the input KO's. This leads to 3 weights for each input KO. While this works in principle, due to the large magnitudes of each parameter x, y and z coupled with the very small deviation of each input solution from the reference solution (Root Mean Squared Error (RMSE) between the input solutions and the RDO is typically in the order of 10mm), very small changes in the values of the weights lead to very large changes in the RMSE, leading to issues with rounding-errors due to limited floating-point resolution of the packages we make use of in Python. When weights are constrained to sum to one for each parameter, optimisation algorithms in the PyGMO suite struggle to make adjustments to the weights to effectively explore the design space; they behave over-constrained and consistently fail to optimise the weights. This is a poor definition of the problem for this application. Therefore, rather than optimising weights on the orbit position states themselves, we redefine the weighted solution through the introduction of the vector $\vec{\delta}$ as:

$$\vec{r}_t^{\,(\mathrm{opt})} = \sum_{i=1}^{N} \begin{bmatrix} w_{x,i} & 0 & 0 \\ 0 & w_{y,i} & 0 \\ 0 & 0 & w_{z,i} \end{bmatrix} \begin{bmatrix} \delta_{x,i} \\ \delta_{y,i} \\ \delta_{z,i} \end{bmatrix} + \frac{\vec{r}_{i,t}}{N}. \tag{4.24}$$

The variable t refers to the epochs, and N to the number of input orbits. Each weight serves to shift the arithmetic mean of the input solutions. Rather than applying the weights directly to the orbit positions, they are multiplied by a scalar value $\vec{\delta}$ and added to the arithmetic mean contribution for each analysis centre. For small values of $\vec{\delta}$, this approach ensures that the design space is limited to a region about the arithmetic mean of the input solutions in which superior combinations are likely to exist.

The value of $\vec{\delta}$ plays a key role in shaping the optimisation landscape. It is determined automatically based on the RMSE of the input solutions with respect to a reference orbit. Specifically, for each parameter (x, y, z), $\vec{\delta}$ is set to one-tenth of the smallest RMSE among the input solutions. This ensures that a weight of 1 corresponds to a shift equal to 10% of the lowest RMSE, while the bounds of the weights are fixed at ±10. As a result, the maximum possible shift for any input is equal to the lowest RMSE, keeping the design space both bounded and interpretable.

This choice of $\vec{\delta}$ is justified by the fact that the RMSE provides a physically meaningful scale characterising the deviation of the input orbits from a trusted reference. By tying perturbations to a fraction of this scale, the optimiser is encouraged to search within a realistic and data-driven region of the solution space. This approach is especially useful because it allows fine-grained control when inputs are already of high quality, while still enabling broader adjustments when the input solutions vary more significantly.

This formulation offers two main advantages: the optimisation weights remain within a consistent numerical range across different input datasets, simplifying tuning and convergence, and the scale of $\vec{\delta}$ adapts dynamically to the overall performance of the inputs. However, it is worth noting that this method can be sensitive to large disparities in input quality — if one solution is significantly worse than the others, the resulting small value of $\vec{\delta}$ may restrict the optimiser's ability to explore sufficiently diverse combinations.

Objective function

We define the objective function used to compute the fitness of a set of weights. In this case, we wish to find combination weights which minimise the residuals of the combined orbit with respect to the RDO reference. We therefore use the RMSE between the combined arc and the RDO as the objective function. The objective function is the 3D RMSE, defined as:

$$R_{\mathrm{SLR}} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left((x_{i,\mathrm{RDO}} - x_{i,\mathrm{comb}})^2 + (y_{i,\mathrm{RDO}} - y_{i,\mathrm{comb}})^2 + (z_{i,\mathrm{RDO}} - z_{i,\mathrm{comb}})^2 \right)}. \tag{4.25}$$

The variable R_{SLR} is the 3D RMSE. This is a measure of the 3D position difference between the combined solution and the RDO data. Choosing this as an objective function means that we are optimising to minimise the difference between our reference orbit and our combined solution.

Algorithms

PyGMO offers a wide array of global and local, single and multi-objective, (un)constrained optimisation algorithms. Single-objective optimisation algorithms are considered. Within this category, the selection has been limited to a handful of better-performing algorithms described as such by PyGMO [8]. The selection to be tested includes four global genetic algorithm optimisers and one local optimiser. Global optimisers are selected as they explore the design space about the arithmetic mean considerably more. A local optimiser is considered as it is less computationally expensive, although it is expected to fall into local minima.

Self-adaptive differential evolution Self-adaptive differential evolution is an adapted form of the widely used Differential Evolution (DE) method of global, heuristic optimisation. DE is a genetic algorithm which fundamentally works by direct search of the design space for objective function minimum. A direct search is done by varying parameter vectors in a population and computing the resulting objective function value; or fitness. How variations to the parameter vectors are done is one of the key methods by which algorithms differ as it greatly influences the performance. In addition, the selection criteria – how it is determined whether a variation (or evolution) is accepted into the population or discarded – influences the convergence rate of the algorithm as well as the likelihood of it converging to local rather than global minima; often times, evolutions are accepted if they improve the fitness of the individual, which results in faster convergence but also a greater likelihood of trapping in a local minimum [62]. Several adaptations exist to avoid poor convergence, such as changes to the population size and relaxation of the evolution criteria to allow vectors to occasionally move away from a minima; facilitating their 'escape' from a local minimum.

DE works by basing changes on the population vectors on the weighted difference between two randomly selected vectors within the population; that is, adaptations to the vectors are generated without external input once an initial population is defined through random selection from the design space. This operation is called 'mutation'. A uniform population distribution is the convention. Once a new vector is created through mutation, it is then further mixed with an additional vector known as the 'target vector' in a process referred to as 'crossover' to produce the 'trial vector'. The performance of this trial vector is then assessed by evaluating the objective function, and if it performs better than the target vector, it replaces it in the population. This final process of replacement is known as selection. Per evolution, each vector in the population acts as the target vector once, ensuring that each entry in the population is assessed per evolution [62].

Self-adaptive DE improves upon this through the introduction of parameter-self adaptation; namely, the self-adaptation of the crossover probability and the weight coefficient applied to the difference during mutation. There are a number of methods by which this is done; in this thesis DE1220 is used, a variant proposed by the PyGMO team. In essence, control parameters of the algorithm such as the aforementioned crossover rate are encoded into the individuals themselves and undergo genetic operators. Superior control parameters lead to superior individuals which then go on to propagate their encoded control parameters. In this way, the control parameters self-adapt as they evolve alongside the parameters. This self-adaptive DE algorithm provides superior convergence than comparable algorithms [12], and requires less arbitrary tuning as a result of the self-determination of control parameters.

GACO This evolutionary algorithm is based upon the real-world behaviour of ant colonies. Ants explore an area about their nest to find food, at first randomly. If an ant is successful, it returns to the nest while laying a pheromone trail down. Other ants will then follow this trail, assuming that it is more likely to lead to food than a randomly selected path. The pheromones are not permanent; they dissipate over time ensuring that only trails which continue to be fruitful remain. Shorter paths are preferred, as their travel time is shorter resulting in the pheromones being updated more frequently. Over time, the shortest past to a nearby supply of food will be identified [57].

This behaviour is mimicked by modelling the movement of ants as artificial ants moving on a graph representing a problem. An optimal 'path' is considered to be the minimum value of an objective function. Artificial ants are defined as decision vectors, and are part of a solution archive (population); the number of the ant in the population is denoted l. Ants are ranked in the solution archive by their performance, with the first ant being the best performing. The construction of a new ant begins by selecting a mean μ_l^i for each dimension of the decision vector i. This mean is taken as one of the values of the decision variable from one of the ants. It is selected randomly, but with the superior ants having a greater likelihood of being chosen; meaning μ_1^i is the most likely to be selected. Next, a random number is sampled from a Probability Density Function (PDF). This process is done for each dimension of the decision vector, leading to the creation of a new ant.

The PDF used is a weighted sum of several one-dimensional Gaussian functions. These functions are defined by the means of each ant ($\mu_{i,l}$) and STD's computed by first computing the maximum and minimum difference between an ant and every other ant in the solution archive, and then taking the difference between these two values and dividing by the number of generations you are on (the number of evolutions your population has undergone). Weights are linearly distributed from the best to worst solution; i.e., the contribution of each Gaussian distribution is scaled by its rank, with lower rank corresponding to higher scaling.

The various included Gaussian distributions are analogous to the different paths the ant could go down. The ranking system relates to the strength of the pheromones. When a new ant is produced, its fitness is evaluated and it is compared against every ant in the population, from the first rank onwards until it reaches a worse-performing ant at some position k, at which point it takes the place of this ant and shifts solution archive an index down below. The worst performing ant is consequentially discarded, maintaining the constant size of the solution archive.

This method is suited towards highly non-convex non-linear problems, as the method by which new ants are created avoids falling into local optimums. This is because the Gaussian distributions contained within the PDF from which the ants are created effectively represent explorative paths into different optimums, ensuring that the population remains diverse.

XNES Exponential Natural Evolution Strategies (XNES) is a member of the family of Natural Evolution Strategies (NES), a class of black-box optimisation algorithms that iteratively update a parameterised search distribution by following the natural gradient of expected fitness [25]. Unlike classic genetic algorithms, NES operates on the parameters of the search distribution itself rather than directly on candidate solutions. In XNES, the search distribution is a multivariate normal distribution, defined by a mean vector and a covariance matrix.

XNES introduces an efficient and numerically stable parametrisation of the covariance matrix through the use of its exponential mapping, ensuring that the covariance matrix remains symmetric and positive-definite during updates. Specifically, the covariance matrix is represented as an exponential of a symmetric matrix, which allows updates to be performed in the log-space of the covariance matrix. This approach avoids numerical issues commonly encountered when updating the covariance matrix directly and allows the algorithm to perform natural gradient ascent efficiently in the space of the distribution parameters.

At each iteration, a population of candidate solutions is sampled from the current distribution. Their fitness values are used to estimate the natural gradient, which is then used to update both the mean and the covariance matrix of the search distribution. This update mechanism allows XNES to automatically adapt the shape, orientation, and scale of the search distribution to the local structure of the objective function.

XNES is particularly effective in high-dimensional, non-linear, and non-convex optimisation problems, as it can adaptively scale the search along directions of low and high sensitivity, effectively handling ill-conditioned problems. By maintaining and updating a full covariance matrix, XNES can exploit correlations between parameters, providing a powerful method for complex optimisation landscapes [25].

CMAES The Covariance Matrix Adaptation Evolution Strategy (CMAES) is one of the most widely used derivative-free, population-based optimisation algorithms for continuous search spaces [27]. It is particularly well-suited for high-dimensional, ill-conditioned, and non-separable problems. CMAES works by iteratively sampling candidate solutions from a multivariate normal distribution and updating its parameters—mean, covariance matrix, and step size—based on the performance of these samples.

The key innovation of CMAES lies in the adaptation of the covariance matrix, which allows the algorithm to learn the shape of the objective function landscape during the search process. The covariance matrix encodes information about the dependencies between variables, enabling the algorithm to perform efficient search in any direction, even in highly anisotropic problem spaces. This is achieved by updating the covariance matrix using evolution paths, which accumulate information about successful steps over multiple generations. These evolution paths enable the algorithm to make informed updates to both the covariance matrix and the global step size, improving the efficiency of the search.

CMAES is designed to be invariant under linear transformations of the search space, making it highly robust to scaling and rotation of the problem. The update mechanism balances exploration and exploitation dynamically, allowing the algorithm to efficiently converge to the global optimum even in rugged fitness landscapes with multiple local minima. However, the computational cost of maintaining and updating the full covariance matrix grows quadratically with the number of variables, which can limit its application to very high-dimensional problems without using modified variants or approximations.

Due to its robustness and versatility, CMAES is considered a state-of-the-art algorithm for challenging optimisation problems and is often used as a benchmark against which other algorithms are compared [27].

Nelder-Mead This is a derivative-free local optimisation method that uses a simplex-based approach to iteratively search for a minimum of an objective function in an n-dimensional space. The algorithm begins with an initial simplex. Each vertex represents a candidate solution, and the objective function is evaluated at all vertices. These evaluations guide the algorithm in reshaping the simplex to explore the search space and approach the optimum iteratively [46].

At each iteration, the vertices are ordered based on the objective function values. The algorithm computes the centroid of the n best vertices (excluding the worst) and applies geometric transformations to adaptively adjust the simplex. The first transformation is reflection, where the worst vertex is reflected across the centroid to explore a new candidate solution on the opposite side. If the reflected point is significantly better than the current best vertex, the algorithm performs an expansion by moving further along the reflection direction to seek even greater improvements. If the reflected point does not provide a sufficiently good solution, the algorithm attempts a contraction by moving the worst vertex closer to the centroid, either from the outside or inside the simplex. When none of these transformations yield satisfactory progress, the simplex is shrunk towards the best vertex, reducing its size to focus the search in a smaller region.

These transformations allow the algorithm to balance exploration of the search space with exploitation of promising regions. The simplex evolves iteratively, adapting its shape and size to the problem's landscape. The process continues until convergence criteria are met, such as the simplex becoming sufficiently small or changes in the objective function value becoming negligible [46].

Algorithm comparison and selection rationale

The selected algorithms represent a diverse set of evolutionary strategies and local optimisation methods, each with unique mechanisms for navigating complex design spaces, and which are state-of-the-art algorithms [8]. Self-adaptive Differential Evolution (DE1220) and GACO are global, population-based optimisers that rely on population diversity and adaptive mechanisms to prevent premature convergence to local minima. DE1220 leverages self-adaptation of control parameters to improve convergence robustness, while GACO employs an archive-based pheromone model to balance exploration and exploitation effectively.

XNES and CMAES both maintain and adapt a full covariance matrix, enabling them to capture correlations between parameters and adjust the search distribution accordingly. This makes them particularly well-suited for problems where the parameter sensitivities differ significantly. CMAES is especially robust for non-separable, ill-conditioned problems due to its adaptive step size and evolution paths, while XNES offers efficient and stable covariance matrix updates via its exponential parametrisation, reducing numerical instabilities and enabling reliable adaptation in high-dimensional spaces.

Finally, the Nelder-Mead simplex method provides a simple local optimisation approach which, despite its limitations in non-convex problems, offers a computationally inexpensive option for fine-tuning weights around promising solutions found by global methods.

Given the non-linear, possibly non-convex nature of the orbit combination problem, and the challenges posed by the limited scale of deviations between the input and reference orbits, global optimisers such as DE1220, GACO, XNES, and CMAES are expected to provide the necessary robustness to navigate the design space effectively. However, due to the computational cost associated with these methods, the combination of a global optimiser followed by a local optimiser (Nelder-Mead) is considered a pragmatic strategy. This hybrid approach leverages the global optimisers' capability to explore the design space and avoid local minima, while using the local optimiser to refine the solution and ensure convergence to an optimal set of weights within the local region of the best-found solution.

4.4 Test setup to assess over-fitting to the reduced-dynamic reference orbit

In order to answer sub-question 1, we design a validation test to assess the risk of over-fitting to the reference RDO when it is used in the combination process targeting the CMAES, Nelder-Mead, and residual-weighted methods. These methods rely on the RDO as a reference and may inadvertently fit too closely to it, risking the introduction of implicit bias from the a-priori gravity field and force models embedded in the RDO.

To test for over-fitting, we compute the residuals between each input or combined orbit and the RDO over a long-term dataset. These residuals are calculated at each epoch t as the position difference between the tested orbit $\vec{r}^{(\text{test})}(t)$ and the RDO orbit $\vec{r}^{(\text{RDO})}(t)$:

$$e(t) = \left\| \vec{r}^{(\text{test})}(t) - \vec{r}^{(\text{RDO})}(t) \right\| . \tag{4.26}$$

We then compute the RMS of these residuals over all N epochs:

$$\text{RMS}_{\text{RDO}} = \sqrt{\frac{1}{|N|} \sum_{t \in N} e(t)^2}. \tag{4.27}$$

This metric quantifies how closely each orbit fits to the RDO. A method exhibiting significantly lower residual RMS with respect to the RDO compared to others may be over-fitting to it, especially if its performance in independent validation using SLR residuals does not show a similar improvement.

To help contextualise these findings, we also compare each method's RMS against the STD of the residuals. The combination methods that do not use the RDO (arithmetic mean, inverse-variance, VCE) as well as the input orbits serve as a reference group to assess what level of RMS is expected in the absence of over-fitting.

4.5 Propagation of uncertainty verification setup

To assess the accuracy and reliability of the covariances reported by the combined orbits in order to answer sub-questions 2 and 3, a controlled verification of this methodology is performed using synthetic data in section 5.3. This approach allows us to verify that the uncertainties reported by the orbit combination methods reflect the actual variability observed in the orbit residuals. Distinction is made between the empirical and reported covariance; the former is the values computed from the experimental output – the residuals of the combined orbits with respect to the truth –, and the latter is the covariance information propagated through the combination. The verification procedure is as follows:

- **Truth orbit generation:** A reference trajectory is generated using a Keplerian orbit model. True positions are computed at regular intervals over a 24-hour arc, sampled at 1-minute epochs.
- **Synthetic input orbit generation:** For each test run, three input orbits are created by adding known noise to the reference trajectory. The per-epoch noise level is varied indepen-

dently for each coordinate and input orbit, with STD's drawn uniformly from the interval $[0.5, 2.5]$ m and covariances from $[0, 0.1]$ m^2.

- **Missing data simulation:** Each input orbit omits 0.5% of its epochs at random to simulate asynchronous data availability across sources as is the case with real input data, which contains between 0 and 13.2% missing data as shown in Table 4.1.
- **Combination methods:** The synthetic input orbits are combined using the four methods which propagate uncertainty information detailed in sections 4.2 and 4.3.1: the arithmetic mean, inverse-variance weighting, residual-weighted combination using differences with the truth orbit and variance component estimation (VCE).
- **Statistical sampling:** The entire procedure is repeated for 50 independent noise realisations. For each run, all metrics are computed per method and per coordinate, allowing statistical analysis over multiple realisations.
- **Validation metrics:** The following metrics are computed:
  - The mean reported and empirical STD for each coordinate. The empirical STD is calculated from the residuals $r(t)$ between the combined orbit and the ground truth:

    $$\sigma^{(\mathrm{emp})} = \sqrt{\frac{1}{|N|} \sum_{t \in N} \left(r(t) - \bar{r}\right)^2}. \tag{4.28}$$

  - The mean reported and empirical covariances for the xy, xz, and yz coordinate pairs. The empirical covariance is computed as:

    $$\sigma_{xy}^{(\mathrm{emp})} = \frac{1}{|N|} \sum_{t \in N} \left(r_x(t) - \bar{r}_x\right) \left(r_y(t) - \bar{r}_y\right), \tag{4.29}$$

    where $\sigma_{xy}^{(\mathrm{rep})}$ is the reported covariance value from the combined orbit solution.
  - The ratio of the empirical to reported covariance. This metric indicates whether the reported covariance is consistent with the observed orbit errors. A ratio close to one suggests agreement. It is computed for each off-diagonal term (xy, xz, yz), and averaged across 50 runs to avoid the influence of randomness:

    $$\mathrm{R}_{xy} = \frac{\sigma_{xy}^{(\mathrm{emp})}}{\sigma_{xy}^{(\mathrm{rep})}}. \tag{4.30}$$

  - The root-mean-square (RMS) of the residuals, providing a combined error magnitude:

    $$R_{(\mathrm{SLR})} = \sqrt{\frac{1}{|N|} \sum_{t \in N} \|\mathbf{r}(t)\|^2}$$

  - The reduced chi-square statistic, computed as:

    $$\chi_\nu^2 = \frac{1}{|N| \cdot D} \sum_{t \in N} \sum_{k=1}^{D} \left(\frac{r_k(t)}{\sigma_k(t)}\right)^2, \tag{4.31}$$

    where $r_k(t)$ is the residual in coordinate $k \in \{x, y, z\}$ at time t, and $\sigma_k(t)$ is the reported STD.

- The Coefficient of Variation (CV) of the daily RMS/STD ratio, used to assess the **temporal stability** of the uncertainty information. It normalises the variability of the ratio across time by its mean, enabling comparison across orbits with different uncertainty magnitudes:

  $$\mathrm{CV} = \frac{\sigma_{\mathrm{RMS/STD}}}{\mu_{\mathrm{RMS/STD}}}, \tag{4.32}$$

  where $\mu_{\mathrm{RMS/STD}}$ and $\sigma_{\mathrm{RMS/STD}}$ are the mean and STD of the daily RMS-to-STD ratios, respectively. Lower CV values indicate more consistent calibration between the orbit uncertainty and the SLR residuals throughout the validation period.
- The estimated variance components (VCE weights) per coordinate and input orbit, which quantify the relative confidence assigned to each input based on its residual statistics.

4.6 SLR observations

To validate the combined orbit solutions, we compare the computed satellite–station range against the observed range from SLR normal point data. These observations are retrieved using the GHOST software.

We make use of these timestamped range measurements to compute the residuals of our orbit solutions. For a given epoch t, $\vec{r}_{(\mathrm{sat})}(t)$ is the satellite position in a terrestrial reference frame (International Terrestrial Reference Frame (ITRF)) and $\vec{r}_{(\mathrm{sta})}(t)$ is the corresponding position of the SLR ground station. The computed range based on the orbit solution is:

$$\rho_{(\mathrm{model})}(t) = \left\| \vec{r}_{(\mathrm{sat})}(t) - \vec{r}_{(\mathrm{sta})}(t) \right\|. \tag{4.33}$$

The observed range $\rho_{(\mathrm{obs})}(t)$ is provided by the SLR normal point data. The SLR residual is then computed as:

$$\Delta\rho(t) = \rho_{(\mathrm{obs})}(t) - \rho_{(\mathrm{model})}(t). \tag{4.34}$$

This residual quantifies the discrepancy between the satellite's computed orbit and the true observed range from the ground station. The lower this value, the better the orbit solution comports with reality according to the independent SLR validation. The key metric to establish performance in a given time frame is the RMS of SLR residuals denoted $R_{(\mathrm{SLR})}$, computed as:

$$R_{(\mathrm{SLR})} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \left(\Delta\rho(t_i)\right)^2}. \tag{4.35}$$

The variable N is the number of valid SLR observations within the considered interval, and $\Delta\rho(t_i)$ is the residual at epoch t_i. This provides a measure of orbit accuracy, with lower values indicating better agreement between the predicted orbit and the independent SLR measurements.

4.6.1 Use of SLR residuals to validate uncertainty propagation

The line of sight vector between the SLR stations from which we receive observations and the Swarm satellites may be computed. With this, we then project the STD data reported by our combined orbits onto it, enabling us to compare these to the SLR residuals, thereby allowing us to gauge how physically realistic it is.

We start by projecting the orbit STD onto the station-satellite LOS. At each epoch, the station position vector $\vec{r}_{\text{(station)}}$ and the satellite position vector $\vec{r}_{\text{(sat)}}$, both in the ITRF frame, are used to compute the LOS unit vector $\vec{u}_{\text{(LOS)}}$:

$$\vec{u}_{\text{(LOS)}} = \frac{\vec{r}_{\text{(sat)}} - \vec{r}_{\text{(station)}}}{||\vec{r}_{\text{(sat)}} - \vec{r}_{\text{(station)}}||}. \tag{4.36}$$

The orbit STD in Earth-Centred Earth-fixed (ECEF) frame is diagonal, represented by its STD's σ_x, σ_y, and σ_z along the X, Y, and Z axes respectively. The projected orbit STD along the LOS direction is then computed as:

$$\sigma_{\text{(LOS)}} = \sqrt{\vec{u}^2_{\text{(LOS)},x}\ \sigma^2_x + \vec{u}^2_{\text{(LOS)},y}\sigma^2_y + \vec{u}^2_{\text{(LOS)},z}\ \sigma^2_z}\ . \tag{4.37}$$

Where $\vec{u}_{\text{(LOS)},x}$, $\vec{u}_{\text{(LOS)},y}$, and $\vec{u}_{\text{(LOS)},z}$ are the components of the LOS unit vector.

For the validation, the following key metrics are considered:

- **Residual RMS:** Root-mean-square of the SLR residuals. This reflects the overall magnitude of the orbit errors, including both bias and random noise.
- **Mean of the LOS projected orbit STD:** This represents the reported uncertainty in the combined orbit.
- **Ratio** RMS^2 / STD^2**:** The ratio between the squared SLR residual RMS and the squared mean orbit STD LOS. A value close to 1 implies that the reported orbit uncertainties reflect the true magnitude of the residuals, including any bias. This metric is sensitive to both random noise and systematic offsets.
- **Reduced Chi-square:** Calculated as

  $$\chi^2 = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\vec{e}_i}{\sigma_{i,\text{(LOS)}}}\right)^2, \tag{4.38}$$

  where N is the number of residuals, $\vec{e^i}$ is the residual at epoch i, and $\sigma_{i,\text{(LOS)}}$ is the reported line-of-sight STD at that epoch. This metric evaluates whether the scatter of the residuals is consistent with the reported uncertainty, assuming zero-mean errors. Unlike the RMS/STD ratio, it is not sensitive to bias in the residuals. Including both metrics allows a more complete assessment of the uncertainty modelling, distinguishing between mismatches in spread and mismatches due to bias.

Chapter 5

Verification

This chapter presents the verification of the implemented software and methodology, ensuring correctness, absence of bugs, and compliance with expectations outlined in chapter 4. Verification of the methodology is performed using synthetic data with known truth, while validation of orbits is conducted using SLR observations. Where applicable, sensitivity analyses are conducted to evaluate the chosen parameters.

5.1 Methodology Verification Using Synthetic Data

The methods and their corresponding plot labels are contained within Table 5.1.

Table 5.1: Summary of tested combination methods with internal references to relevant sections.

Plot Label	**Full Method Name**	**Reference**
Corrupted arcs average	-	-
Mean	Arithmetic Mean	subsection 4.2.1
Inverse variance	Inverse Variance Weighting	subsection 4.2.2
VCE	Variance Component Estimation	subsection 4.2.3
Residual weighted	Residual weighted	subsection 4.3.1
de1220	DE1220	section 4.3.2
gaco	GACO	
cmaes	CMA-ES	
xnes	XNES	
Nelder mead	Nelder-Mead	
de1220 LO	DE1220 + Nelder-Mead	section 4.3.2
gaco LO	GACO + Nelder-Mead	
cmaes LO	CMA-ES + Nelder-Mead	
xnes LO	XNES + Nelder-Mead	

5.1.1 Approach

To rigorously test the combination methods, synthetic datasets were created by corrupting known KO or fully synthetic Keplerian orbits with noise sampled from known distributions. This allows quantitative assessment of the methods' ability to recover the uncorrupted orbits under controlled scenarios.

Synthetic orbits are generated by adding or subtracting noise with defined STD σ and optional bias β:

$$\vec{x}_{\text{corrupt}} = \vec{x} \pm \mathbf{v}, \;\; \mathbf{v} \sim \mathcal{N}(\beta, \sigma^2 I) \tag{5.1}$$

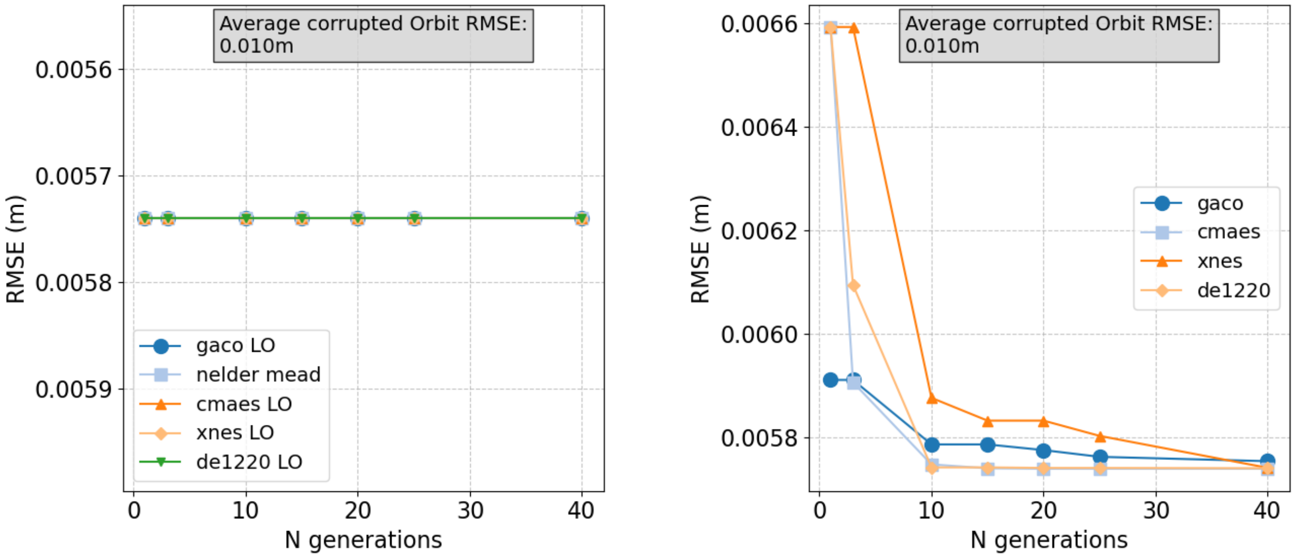
The corrupted orbits, along with their defined variance information used as inputs to the combination methods listed in Table 5.1.

5.1.2 Sensitivity of the optimisers to generation and populations size

Before the optimisation methods can be broadly tested, a sensitivity analysis of the relevant parameters – generation and population size – is conducted to fix them. The former determines how many evolutions or iterations the population experiences, and the latter the size of the population to evolve or be sampled from. Three synthetic input orbits are used, each constructed using noise with increasing STD's; 0.0035 m, 0.006 m and 0.0075 m.

Varying generation size:

Generation sizes of 1, 3, 10, 15, 25, and 40 are tested. The results are presented in Figure 5.1.



(a) Locally optimised methods.

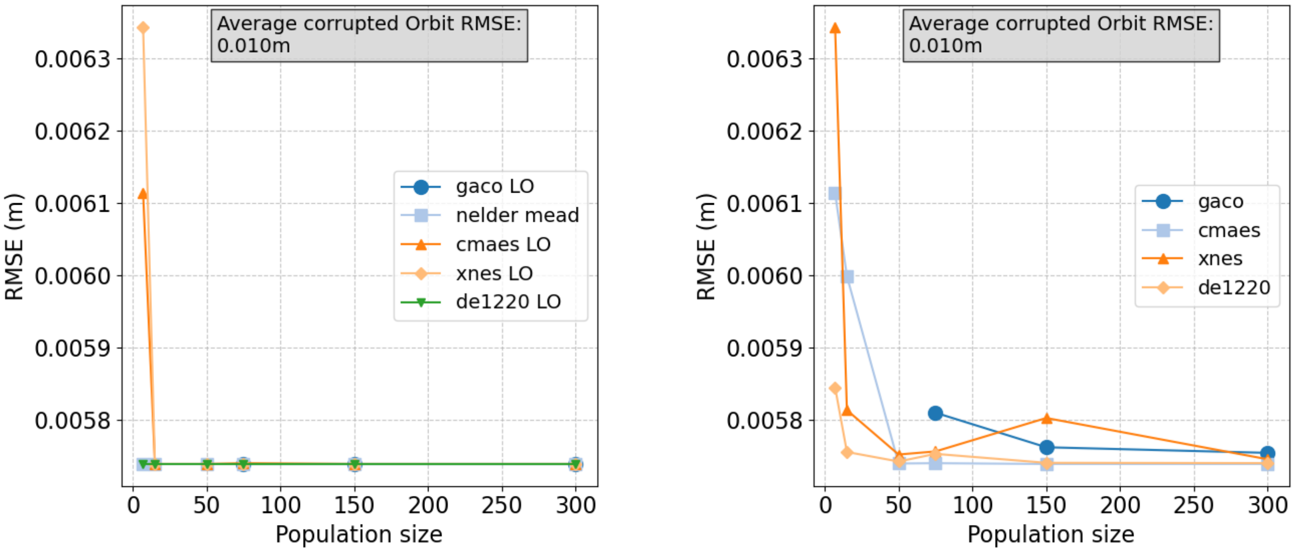
(b) Globally optimised methods

Figure 5.1: Impact of number of generations on RMS of the combined orbit with a random seed of 42.

Changing the number of generations has no effect on the locally optimised solutions, indicating that the initial conditions are sufficiently close to a local minimum such that the local optimiser consistently achieves the same performance. Global optimiser performance improves steadily with population increase, flattening out at 10 for all methods aside from xNES. This is expected behaviour, as more generations provide more opportunities for the evolutionary algorithms to find superior solutions, and indeed they converge towards the same minima discovered by the combined and locally-optimised solutions with increasing generations.

Varying population size:

Population sizes of 7, 15, 50, 75, 150, and 300 are tested. As stated in section 4.3.2, the Gaco optimiser requires a minimum population size of 75, hence why its results begin there. The results are presented in Figure 5.2.



(a) Local and combined optimisers

(b) Global optimisers

Figure 5.2: Change in optimiser performance with increasing population size with a random seed of 42.

The purely local optimiser, Nelder-Mead, is not sensitive to population size as the initial guess of equal weights – corresponding to the arithmetic mean – is a good initial guess for our synthetic data. The combined optimisers require a very small population to tend towards the same minima as Nelder-Mead. The low population results in a sparse random sampling of the design space, which leads to poor initial conditions resulting into falling into local minima which are inferior to that found by Nelder-Mead and higher-population optimisers. With a larger population size, this effect is removed and they again perform equally well. The results of the global optimisers demonstrate that the population size greatly affects their performance. However, it is not necessarily the case that increasing the population size leads to a smooth improvement in the performance of these algorithms. xNES in particular shows erratic behaviour with changing population size, with error first increasing before decreasing with population size. DE1220 similarly shows this behaviour. CMAES and GACO show smooth improvement with population size, flattening at 75 and 150 respectively, with CMAES outperforming the competition overall. Population size is affected by the choice of random seed, as the population is sampled randomly within the design space. The result is that an increased population could lead to a random selection of worse initial conditions, resulting in poorer solutions.

Optimisation parameters

The trends observed in this section remain consistent when changing the random seed of the Pygmo optimiser. The additional plots can be observed in demonstrating this can be viewed in Appendix C. Based on these results, it is decided to use a population size of 150 and 25 generations, as this strikes a balance between robustness and computational efficiency, and increasing either past this point provides diminished returns.

5.1.3 Symmetric synthetic orbit pairs

With the configuration of the optimisers fixed, we can proceed to other tests. First, we produce mirrored synthetic orbits. Given an uncorrupted orbit $\mathbf{x}_{\text{true}}$ and a noise vector ϵ sampled from a distribution with zero mean ($E[\epsilon] = 0$), we generate a pair of synthetic orbits:

$$\begin{cases} \mathbf{x}_{+} = \mathbf{x}_{\text{true}} + \epsilon \\ \mathbf{x}_{-} = \mathbf{x}_{\text{true}} - \epsilon \end{cases} . \tag{5.2}$$

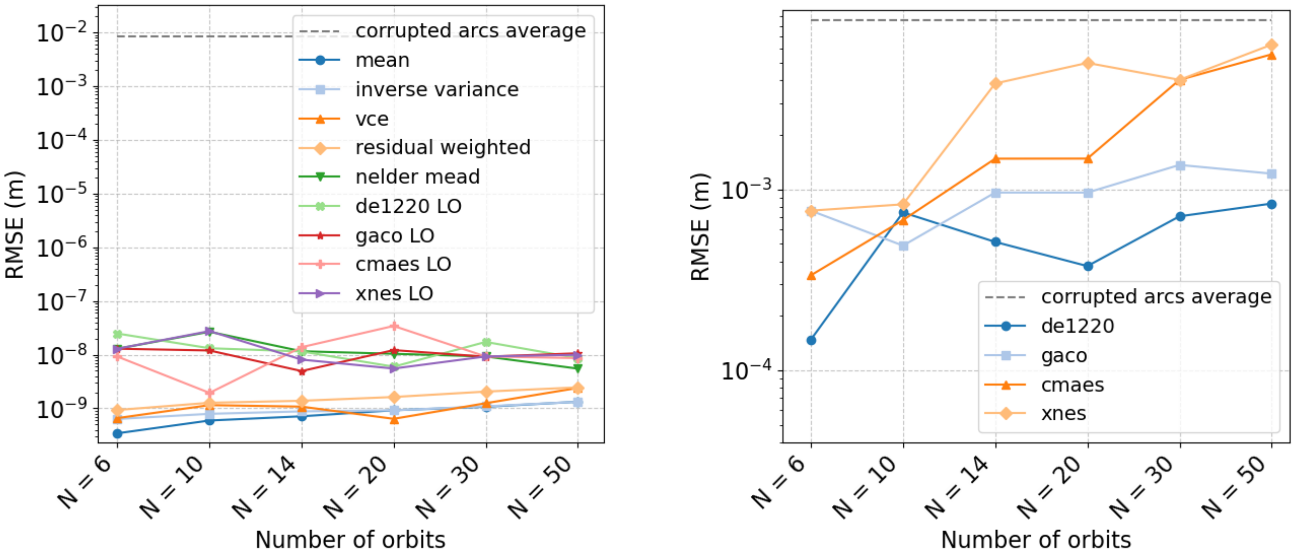
This pair of orbits is symmetric about the truth ($\mathbf{x}_{\text{true}}$), satisfying the relation:

$$\mathbf{x}_{\text{true}} = \frac{1}{2}(\mathbf{x}_{+} + \mathbf{x}_{-}). \tag{5.3}$$

The objective of this is to create a set of input orbits which we know can be combined to reproduce the uncorrupted orbit exactly. This remains true if we introduce additional pairs of corrupted orbits, regardless of what the STD of the corrupting noise is.

The arithmetic mean of the sets of pairs is the optimum combination, as each pair is averaged to the uncorrupted orbit,therefore this method of combination should perfectly reproduce the uncorrupted orbit. Each pair of orbits has equal STD, but is mirrored, which should mean that the inverse-variance weighting method produces equal-weights for each member of the pair. Therefore, as you introduce more pairs, each pair simply cancels itself out in the combination, leading to a perfect reconstruction. Variance component estimation selects variance components based on the differing variances of each input with respect to the combined orbit. We expect this method to cancel the corrupted pairs out, as they will result in equal variance components. The optimisation methods use the uncorrupted orbit as reference data. Whether or not the optimisation algorithms reproduce the uncorrupted orbit depends on the initial guess and number of generations.

The STD is fixed at 0.005 m. Two random seeds are tested – 42 and 20 – to ensure that results are not a fluke and can be reproduced. 6, 10, 14, 20, 30 and 50 total input orbits are tested.



(a) Independent, local optimisers and residual-weighted

(b) Global optimisers

Figure 5.3: Results for symmetric synthetic orbit pairs.

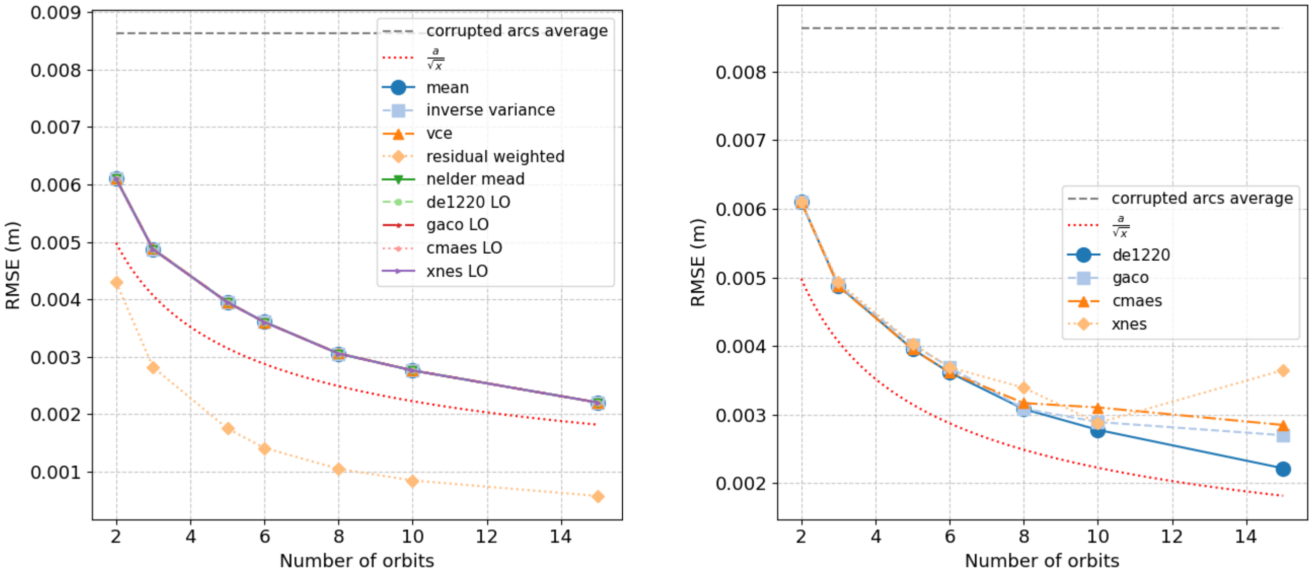
As expected, the arithmetic mean, inverse variance, and VCE methods perfectly reconstructed the uncorrupted orbit, with RMS errors accounted for by rounding error due to limited machine epsilon ($\sim 10^{-10}$ m). This confirms their theoretical behaviour under symmetric noise conditions. Locally optimised methods remained stable and converged close to the truth. This demonstrates that while global optimisers enable a broader search of the design space, they fail as the number of input orbits increase due to the growth of the parameter space.

The global optimisers perform worse. While they produce combined orbits with orders of magnitude lower RMS than the average of the synthetic data, they fail to reach the global minimum. As the number of input orbit pairs increase, these methods perform worse. This is expected, as the number of parameters the optimisers solve for scales directly with the number of orbits, and solving the weights becomes considerably more difficult. It is likely that the local optimiser alone well-optimises a local minimum, and the global optimiser successfully discovers a superior local minimum which is then improved more by the local optimiser.

In addition, the variance components computed by variance component estimation were equal pairs corresponding to the corrupted orbit pairs as expected.

5.1.4 Combining non-mirrored corrupted orbits

In this test, n corrupted orbits are created by adding noise with fixed STD σ and mean μ of 0. As before, no bias is introduced. We test increasing values of n in order to observe whether the combined orbits converge towards the uncorrupted orbit as n tends towards infinity. It is expect that this occurs for the arithmetic mean, inverse variance and VCE methods; due to the formulation of the optimisation problem, the number of parameters to be estimated increases proportionally to the number of orbits, meaning it becomes more challenging to solve. In practice, the number of available KO's available is highly unlikely to exceed 5, so this is not envisioned to cause issues. Results are presented in Figure 5.4.



(a) Non-global corruption, seed 42.

(b) Global corruption, seed 42.

Figure 5.4: Impact of increasing the number of non-mirrored corrupted orbits. An $\frac{a}{\sqrt{x}}$ line is included as a trend-line.

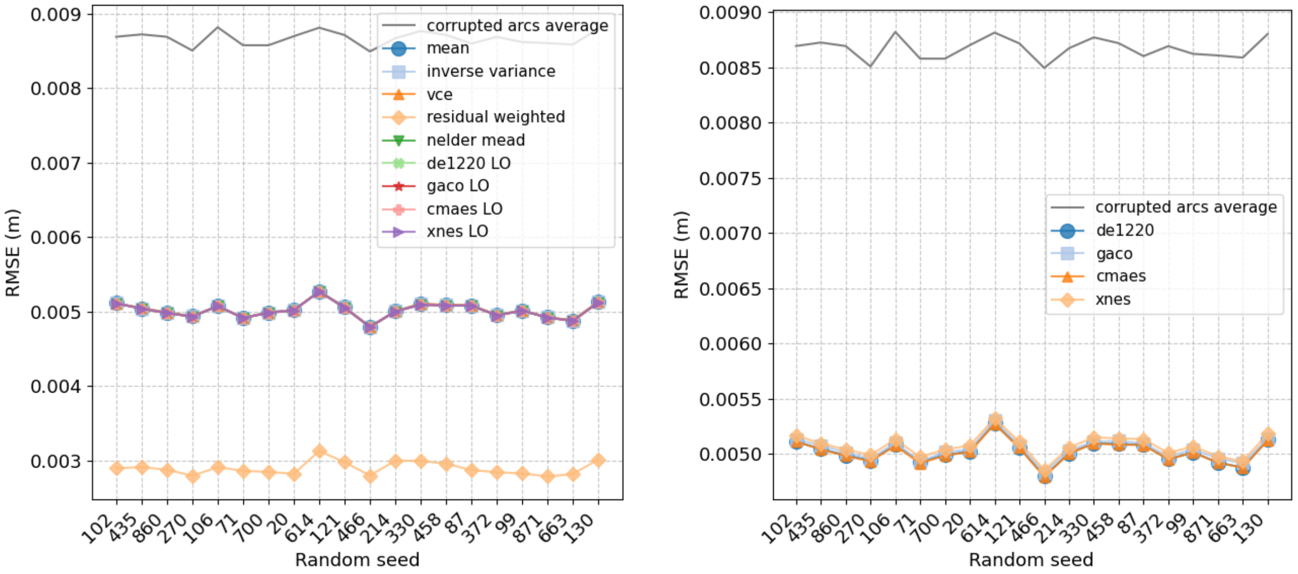
A convergence towards the truth following a $\frac{a}{\sqrt{x}}$ trend is observed for all methods. This confirms that independent (arithmetic mean, inverse variance, VCE), the residual-weighted and locally-optimised methods converge towards the truth as number of input orbits increases, consistent with statistical expectations. Residual-weighted in particular tends towards reproducing the original orbit, demonstrating that determining weights from the residuals with respect to the truth is more effective than using the variance information. The global optimisers other than DE1220 struggled increasingly with growing input orbits due to the rising number of parameters, with divergence from the trend observed from 5 input orbits on, once more highlighting that these methods are not appropriate for high input orbit combinations.

5.1.5 Sensitivity to Random Seeds

Here we evaluate each method's robustness to variations in the random seed used. The random seed used affects several sections of the methodology; the noise used to produce the synthetic input orbits and the initial population sampling and any randomisation used for mutation by the Pygmo optimisers. It is important to assess the sensitivity of these to the random seed used by testing a number of them and observing how results are affected.

Sensitivity to the random seed used to generate noise for synthetic orbits

By ensuring that the random seed is constant aside from that used to sample the noise distribution for producing the synthetic input orbits, we can observe the sensitivity of the methods to the particular sampling of the noise. We expect to observe no meaningful difference between the different seeds, as the magnitude of the noise remains constant, and it is only the distribution which changes as a result of the choice of a different seed. In practice, a positive result for this test suggests that our methods are stable and behave predictably with variations in inputs. For this test, the three synthetic orbits input are regenerated with varying random seeds. The results are presented in Figure 5.5.



(a) Independent and locally-optimised methods, as well as the residual-weighted method.

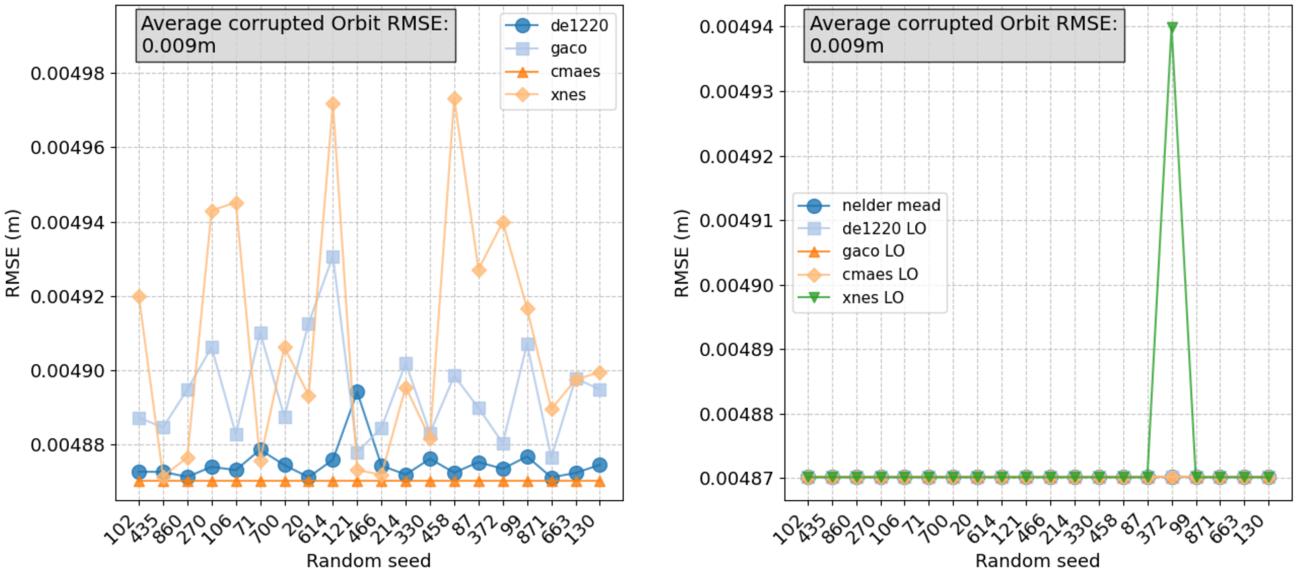
(b) Global optimiser methods.

Figure 5.5: Random seed tests for the noise instantiation used to produce the synthetic input orbits.

The average RMSE of the synthetic, corrupted orbits is provided in gray to show how changing the random seed affects the input orbits. The independent and residual-weighted methods produced consistent results regardless of the seed, as expected. The local optimisers similarly exhibited stable performance. In general, the globally optimised methods show no sensitivity to the random seed. The performance of each method tends to follow the behaviour of the initial set of synthetic inputs, indicating – as expected – that the combination improves with increasing quality of the inputs, regardless of method applied.

Sensitivity of the optimisers to the random seed

The optimisers introduced in section 4.3.2 are sensitive to the random seed selected in two key ways; as input to the methods through the initially randomly sampled population, and internal to the methods depending on how randomisation is integrated into them through, for instance, random mutation in the evolutionary algorithms. In order to test this, the same set of input synthetic orbits is used for each seed, limiting randomisation to that inherent to the pygmo optimiser. The results are presented in Figure 5.6.



(a) Global optimisers

(b) Global optimisers with a final local-optimisation stage and Nelder-mead

Figure 5.6: Random seed tests for Pygmo optimisers.

Variation in performance is observed for global and locally-optimised methods depending on the random seed selected. The locally optimised methods demonstrate no sensitivity to random seed outside of xNES, which is explained by its noticeably poorer performance as demonstrated in Figure 5.6a. Deviation is observed between the varying global optimisers, with CMAES and DE1220 performing the best. This difference is not described by the randomised instantiation of the population, as it does not affect every method equally; therefore, it is a product of the algorithms themselves. xNes and GACO perform noticeably worse, demonstrated by their greater fluctuations in RMS. This behaviour is consistent with the theoretical characteristics of xNES, which relies on natural gradient updates based on the curvature of the objective function. Because the natural gradient is sensitive to the local geometry of the fitness space, even subtle changes induced by different noise realisations can alter the direction of search and result in convergence towards a poorer quality local minima [25].

GACO, while performing better than xNES overall, also shows slightly higher variability and a tendency to converge to suboptimal solutions more frequently. This can be attributed to its pheromone reinforcement mechanism, which favours early good solutions. If the input noise leads to misleading early fitness evaluations, GACO may prematurely hone in on a suboptimal section of the search space. Without mechanisms for reset or exploration encouragement, such behaviour accounts for GACO's increased sensitivity to seed-dependent initial conditions [57].

In accordance with the results of subsection 5.1.2, this test indicates that the choice of population and generation size contributes greatly to the sensitivity of the converged result to the random seed. This random behaviour is observed in all but CMAES, which is the only algorithm shown in figures 5.1 and 5.2 to flatten out with generation and population size.

Summary

The results indicate that none of the methods are sensitive to the random seed used to instantiate the noise of the synthetic orbits. The locally-optimised methods show no sensitivity to the random seed of the pygmo optimisers other than xNES. Of the global optimisers, only CMAES is stable with respect to random seed. Among the global optimisers, CMAES stands out as particularly robust, displaying highly consistent performance.

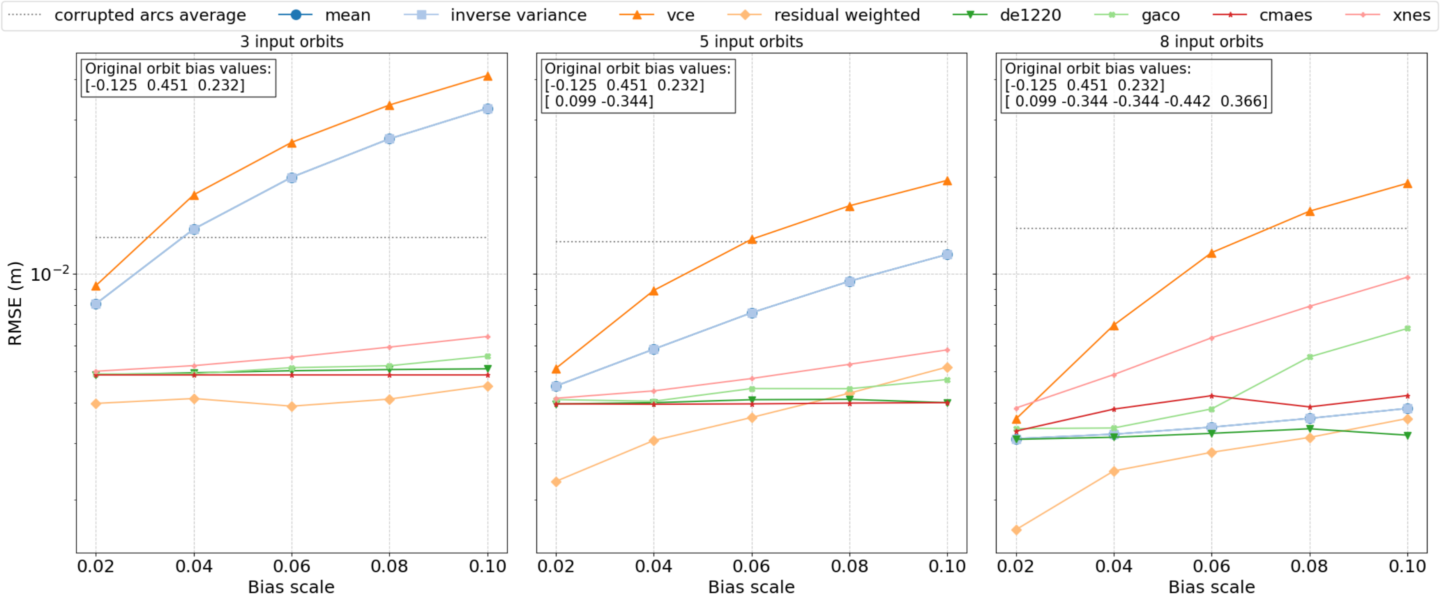
5.1.6 Changing the scale and length of the input biased orbit

Random noise is not the only deviation from truth that can affect an orbit solution. Bias is also important to consider. It describes the offset of the solution such that it can no longer be described as a random sampling of the truth with some STD. In practice, this is what we expect the optimised methods to account for better than the methods dependent on variance information, which only account for variance. The optimisation methods are able to shift the orbits by some constant, which is independent of the STD of the input orbits. The STD of the synthetic input orbits is fixed at 0.005 m for these tests.

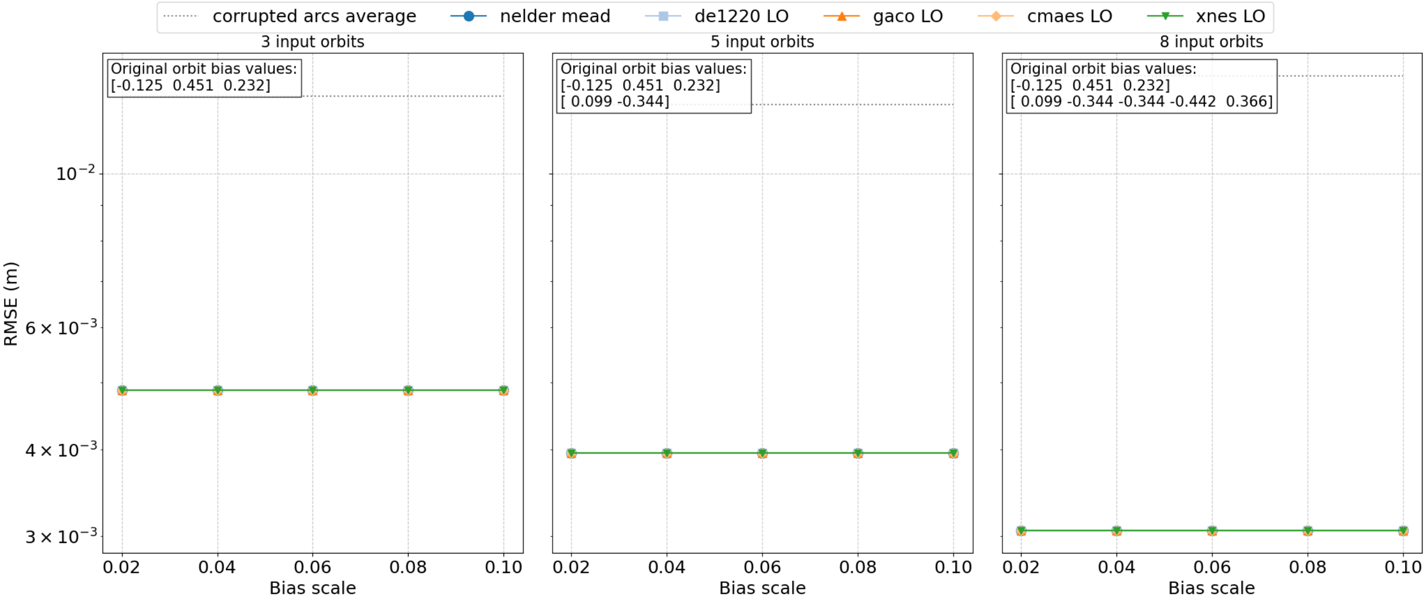
The change in scale (magnitude) of the bias applied to the input orbits is provided in Figure 5.7, with 3-orbit, 5-orbit and 8-orbit combinations included in order to show how increasing the number of biased orbits affects the results.

Independent methods fail to correct biases, as expected. Due to the constant STD of the input orbits, the mean and the inverse-variance perform exactly the same, and demonstrate that they cannot account for bias in the input orbits at all. VCE is a worse-performing inverse-variance method, as again it cannot account for bias, but worsens the combination due to the fact that it tends to fit towards the combined solution of the input orbits, which in this case are all biased. These methods performed better as the number of input orbits was increased, as the bias is sampled between ±0.5 m, meaning that as the number of input orbits increases, the mean of them approaches the truth.

The residual-weighted method performed excellently, as it derives per-epoch weights directly from the residuals with respect to the truth. As such, it down-weights orbits which are more biased with respect to the truth, and up-weights orbits which are closer, ultimately resulting in a much closer fit and a better combination.



(a) Globally optimised, independent, and residual-weighted methods.



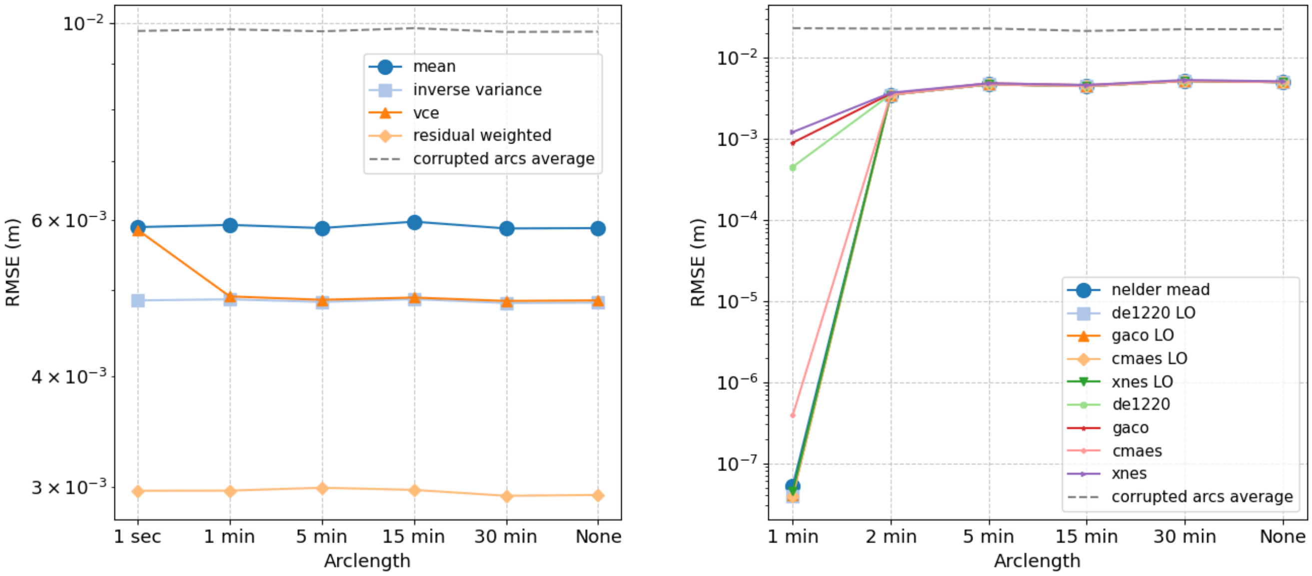
(b) Locally optimised methods.

Figure 5.7: Impact of increasing input bias magnitude. From left to right in each panel, the number of biased input orbits increases from 3 to 8, showing how the methods respond to progressively stronger bias contamination with increasing number of input orbits. The un-scaled (scale=1) input bias values are provided, as well as the average of the corrupted arcs.

The global optimisers performed consistently well, showing close to flat lines with increasing bias scale whether local, global or a combination thereof. This demonstrates that piece-wise weighting is most effective for accounting for systematic bias in the input orbits. Of the global optimisers, CMAES and DE1220 perform the best with changing bias length and scale, while xNES and GACO begin to perform noticeably worse as the number of input orbits increase. The locally-optimised methods prove the most robust to both increasing number of biased inputs, and increasing scale thereof. While none of the methods fully eliminated bias and returned the truth, all of the locally-optimised methods, CMAES and DE1220 and the residual-weighted method produced combined solutions with RMS values 2-3 orders of magnitude lower than the average of the synthetic input orbits.

5.1.7 Changing input arc length

The input orbits may be segmented into arc lengths of a given duration, which affects the methods involving the application of piece-wise weights, or piece-wise variance components. In practice, this means that all optimisation methods and VCE are affected by the arc-length. Therefore, a test is performed to see if there is an optimal arclength, and to determine the effect of this parameter on the methods. For the independent and residual-weighted methods, a 1-second sampling of a day of trajectory data is used, and arc-lengths from single-epoch (1 second) to None (the full day) are tested. For the optimisers, this input data is resampled to 1-minute frequency, as we wish to test the single-epoch performance, but the optimisers take considerable time to run, and so we compromise by resampling to reduce the number of epochs to be solved for while still observing a full orbit. Results are presented in Figure 5.8.



(a) Non-optimised methods.

(b) Optimised methods.

Figure 5.8: Effect of input arc length on combination performance.

Increasing arc length has no meaningful impact on the arithmetic mean, inverse-variance and residual-weighted methods aside from some slight numerical-noise induced fluctuations as expected, as these methods are not sensitive to this variable. On the other hand, VCE breaks down when a single-epoch arc length is used, as the method requires having multiple epochs to well-estimate the variance components as described in subsection 4.2.3. Past a single-epoch arc length, no difference is observed when using different arc lengths, indicating that the variance components tend to be stable regardless of the length of the arc used to determine them. This is to be expected, as the input orbits have consistent noise across epochs, which means that VCE should estimate close to identical variance components regardless of the segmentation. Optimised methods tend to reproduce the reference orbit when a single-epoch arc length is used, with all of the locally-optimised methods and CMAES reporting this. This demonstrates that piece-wise weights are required, as we specifically wish to avoid fitting exactly to the reference data. Additionally, this demonstrates that the optimiser design space is well-defined, as the optimisers can reproduce the reference orbit when a single-epoch arc length is used, shown by the RMSE below e^{-7}m in Figure 5.8.

5.2 Selection of methods from verification results

With the results of these tests, we narrow down the selection of methods to produce results with. Each of the independent methods perform well, and moreover require very little computation time to run. The residual-weighted method outperforms on several tests; the limitation of this method is potential over-fitting to the RDO, which we are interested in testing. Of the optimisation methods, CMAES and Nelder-Mead stand out as the best-performing. The former performed best of the global optimisers across multiple tests, the latter performed very well and requires less computation time as it is a simpler algorithm [46]. These methods similarly have the potential to suffer from over-fitting to the reduced-dynamic reference data. For each method, an arc length of one day is used.

5.3 Verification of the propagation of uncertainty

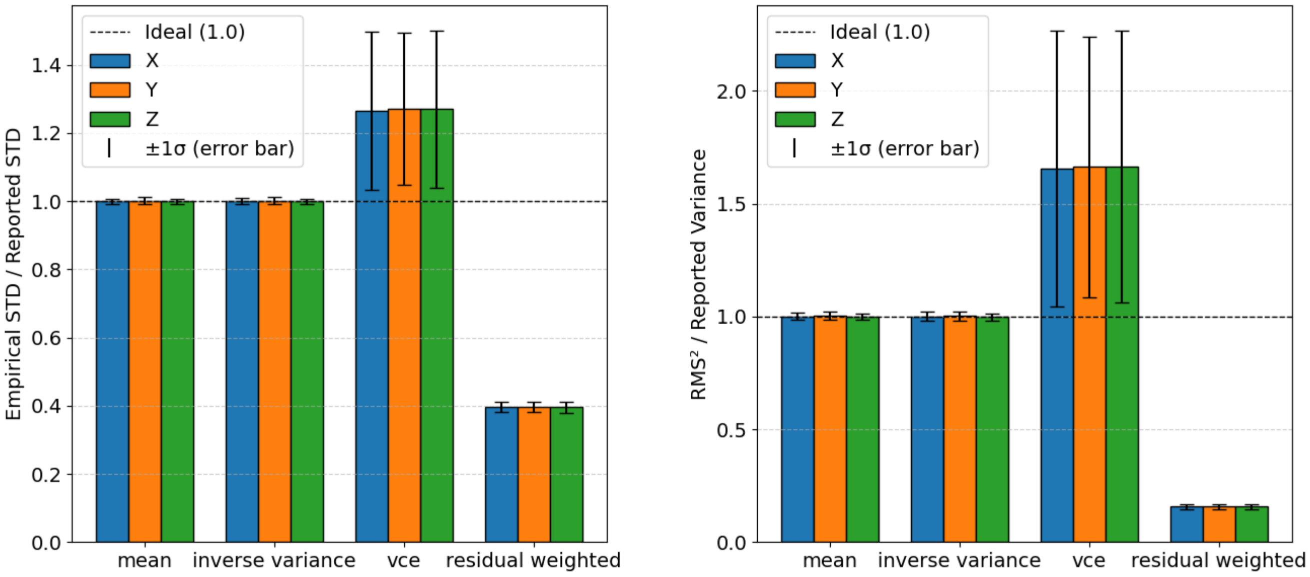
In sections 4.2 and 4.3.1, the combination of covariance information to propagate uncertainty from the input orbits to the combined solution is described for four methods; the arithmetic mean, inverse-variance, VCE and residual-weighted. In this section, a verification of these combined covariance values is performed using synthetic data with the method described in section 4.5.

The results of the 50-run synthetic verification are presented and analysed to assess whether the reported uncertainties from the combined orbits reflect the true error distribution, and whether the variance component estimation method produces statistically meaningful weights under varying noise conditions.

5.3.1 Comparison of propagated uncertainty with true residuals

To assess the statistical validity of the reported uncertainties, we compare them to averaged empirical metrics computed from the residuals across runs.

Figure 5.9a shows the ratio between the empirical STD and the reported STD for each method and coordinate. A ratio of 1.0 indicates that the reported STD correctly captures the random scatter in the residuals. Both the arithmetic mean and inverse-variance methods achieve nearly ideal ratios across all coordinates. VCE exhibits a tendency to underestimate the true uncertainty, as evidenced by ratios of roughly 1.3, while the residual-weighted method significantly overestimates the variability demonstrated by its very low ratio close to 0.4.



(a) Empirical STD / Reported STD

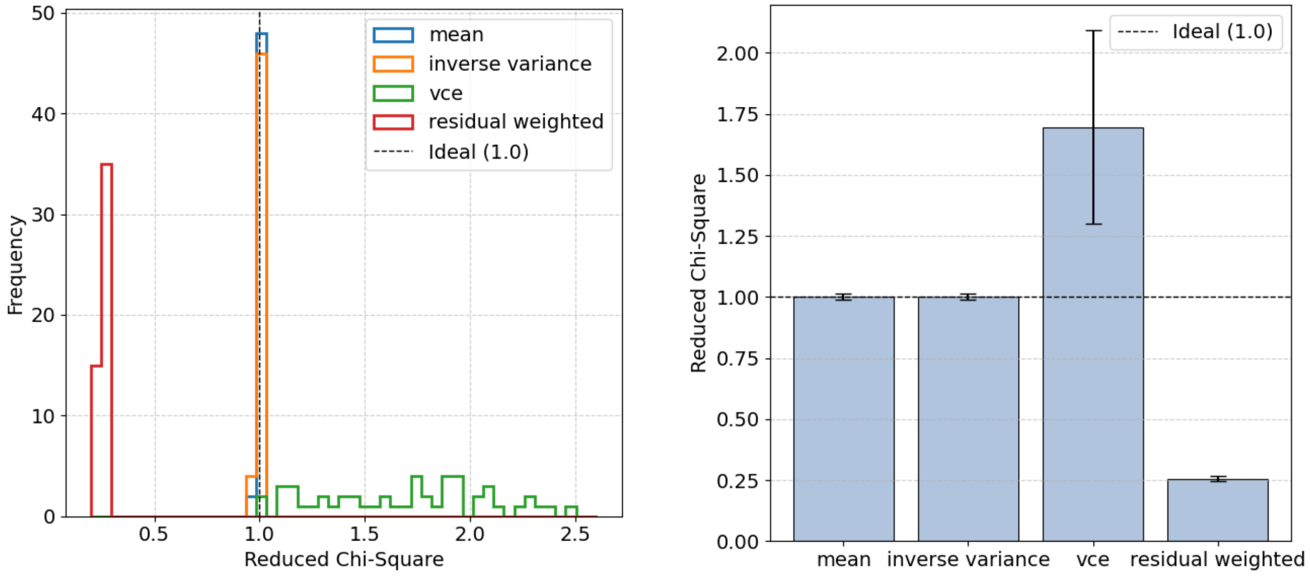
(b) RMS^2 / Reported Variance

Figure 5.9: Comparison of empirical error metrics to reported orbit uncertainty. Each bar represents the mean ratio over 50 runs, grouped by method and coordinate. Error bars denote $\pm 1\sigma$ STD.

Figure 5.9b presents a similar comparison using RMS^2 rather than STD, thus representing both STD and the mean. Results are consistent, though the underestimating of the uncertainty by VCE and the overestimation of it by residual-weighting is exacerbated. This is expected, as VCE does not compensate for bias in any way, therefore the further deviation is explained by the introduction of the inclusion of bias, which simply degrades the results. The mean and inverse-variance methods are unaffected by this change, as bias across runs cancels out as they are statistically unbiased estimators [73]. The residual-weighted method, which explicitly uses differences with the reference orbit, tends to interpret bias as high variance and overestimates uncertainty accordingly.

5.3.2 Reduced chi-square analysis

To verify the global consistency of the reported uncertainty with the actual error distribution, we compute the reduced chi-square per method and per run. Figure 5.10 shows the distribution across all 50 runs and the mean values.



(a) Histogram of reduced chi-square values for each combination method over 50 runs. The vertical dashed line indicates the ideal value of 1.0.

(b) Bar plot of the mean reduced chi-square values for each combination method over 50 runs. The horizontal dashed line indicates the ideal value of 1.0, and the error bars indicate the $\pm 1\sigma$.

Figure 5.10: Comparison of the reduced chi-square of the orbit solutions across different methods, demonstrating the quality of the uncertainty propagation with respect to the reported residuals.

The arithmetic mean and inverse-variance methods cluster tightly around 1.0, as expected. The VCE method exhibits a much wider distribution with underestimation of the uncertainty; it is overconfident. The spread indicates that we cannot compensate for the overconfidence of this method by up-weighting the noise by some estimated constant, which suggests that the noise information produced by VCE is less reliable. The residual-weighted method yields very low and consistent chi-square values, indicating over-reported uncertainties and under-confidence. This aligns with expectations, as residual-weighting in this case uses the residuals to the ground truth to weight the orbit combination, which means the produced orbit is more accurate than the combined uncertainty of the input orbits predicts.

5.3.3 Covariance assessment

A single set of test results is used to demonstrate how the covariance information maps on to the residuals. The STD's are chosen at random from between 1 m and 2.5 m for each parameter, and the covariances selected between $-0.1m^2$ and $0.1m^2$. Random seeds 100-150 are used for the noise realisation. Time-varying deviations are added on top of this base by adding a sample of normally distributed noise with a magnitude of 0.01 m at each epoch. Results are presented for the xy, xz and yz covariances in figures 5.11a, 5.11b, 5.12. The covariances and the variances of each pair of parameters both computed from the residuals (empirical) and retrieved from the combined orbit data are provided numerically in Table 5.3 and Table 5.2, supporting the plotted residuals.

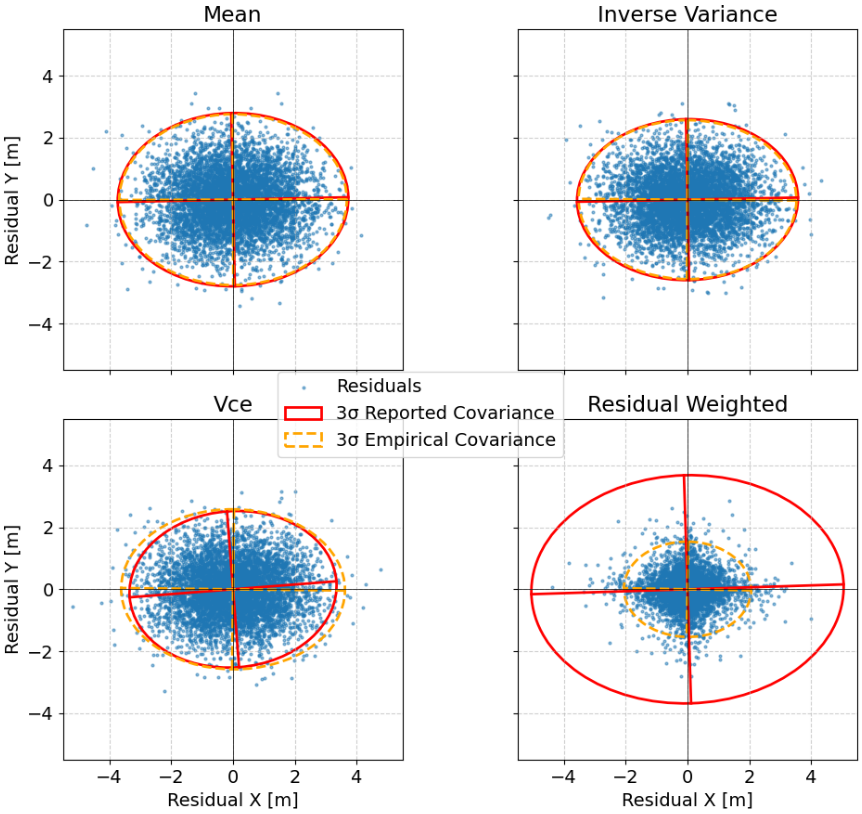
Table 5.2: Reported vs empirical 1σ variances for each method in the XY, XZ, and YZ planes. Each pair of columns shows reported and empirical values for σ_1^2 and σ_2^2, followed by the percentage difference.

Plane	**Method**	σ_1^2 Rep.	Emp.	$\Delta\%$	σ_2^2 Rep.	Emp.	$\Delta\%$
XY	Mean	1.5432	1.5016	+2.77%	0.8700	0.8458	+2.86%
	Inverse Variance	1.4166	1.3770	+2.88%	0.7508	0.7322	+2.55%
	VCE	1.2445	1.4541	-14.42%	0.7109	0.7415	-4.12%
	Residual Weighted	2.8313	0.4698	+502.58%	1.5095	0.2629	+474.08%
XZ	Mean	1.5432	1.5016	+2.77%	1.2307	1.2532	-1.80%
	Inverse Variance	1.4166	1.3770	+2.88%	0.8495	0.8412	+0.98%
	VCE	1.2445	1.4541	-14.42%	0.4746	0.9676	-50.95%
	Residual Weighted	2.8313	0.4698	+502.58%	1.8451	0.3014	+512.36%
YZ	Mean	0.8700	0.8458	+2.86%	1.2307	1.2532	-1.80%
	Inverse Variance	0.7508	0.7322	+2.55%	0.8495	0.8412	+0.98%
	VCE	0.7109	0.7415	-4.12%	0.4746	0.9676	-50.95%
	Residual Weighted	1.5095	0.2629	+474.08%	1.8451	0.3014	+512.36%

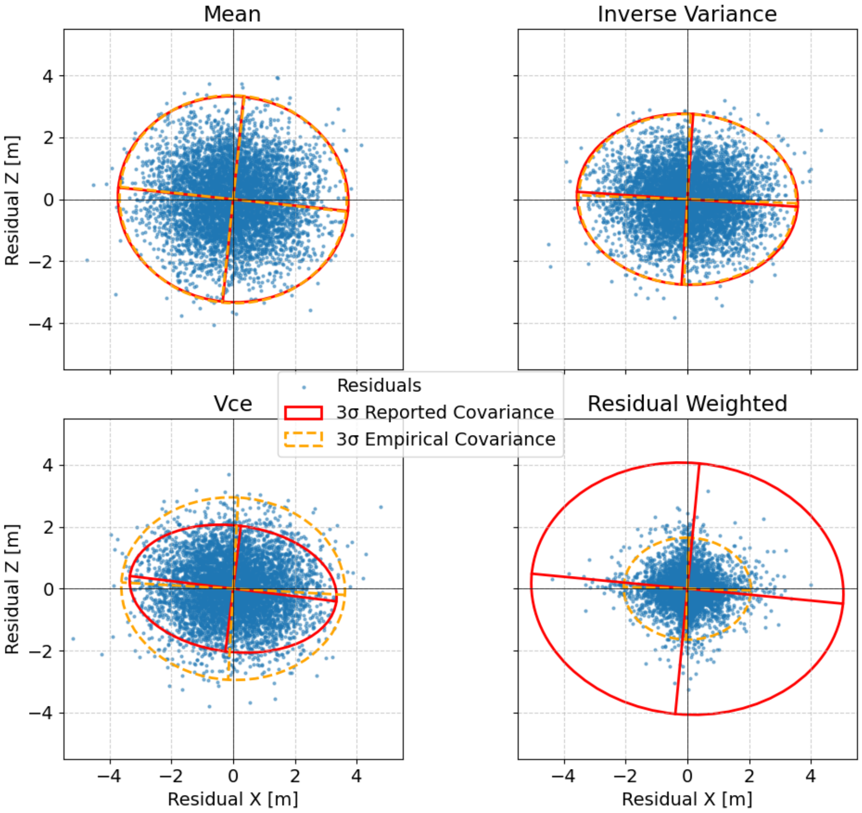
Table 5.3: Reported vs empirical 1σ covariances for each method in the XY, XZ, and YZ planes. Each pair shows the reported and empirical covariance.

Plane	**Method**	Cov Rep.	Emp
XY	Mean	0.0140	0.0020
	Inverse Variance	0.0126	-0.0044
	VCE	0.0416	-0.0081
	Residual Weighted	0.0416	-0.0036
XZ	Mean	-0.0322	-0.0275
	Inverse Variance	-0.0386	-0.0198
	VCE	-0.0959	-0.0271
	Residual Weighted	-0.0959	-0.0046
YZ	Mean	0.0316	0.0642
	Inverse Variance	0.0248	0.0485
	VCE	0.0940	0.0369
	Residual Weighted	0.0940	0.0072

The arithmetic mean and inverse-variance methods demonstrate excellent performance with covariance combination, typically closely following the empirical covariance. Inverse-variance slightly outperforms the arithmetic mean, remaining more consistent. VCE and residual-weighting report variance and covariance information less consistent with the empirical results. As identified in subsection 5.3.1, VCE underestimates the variance, while residual-weighting overestimates it. When it comes to covariance, however, both methods tend produce combined covariances which are correlated, despite the residuals not displaying the same correlation.



(a) XY plane



(b) XZ plane

Figure 5.11: Residuals and covariance ellipses in the XY and XZ planes. The solid red ellipse corresponds to reported 3σ uncertainty, and the dashed orange ellipse is derived from the residuals. The semi major and minor axes of the ellipses are plotted to ensure that correlation is visible.

In the XY plane, the mean, inverse variance and VCE methods show excellent agreement between reported and empirical covariance, in both scale and orientation, with VCE only slightly under-estimating σ_x. Conversely, the residual-weighted method overestimates uncertainty, with the residuals considerably smaller than the 3σ ellipse. In the XZ plane, the mean and inverse variance methods again yield close alignment between reported and observed uncertainty. VCE shows moderate underestimation of the Z variance and slight underestimation of the X variance (consistent with the XY plane). There is some negative correlation in the covariance information which is not observed in the residuals. Residual-weighted again overestimates variance.

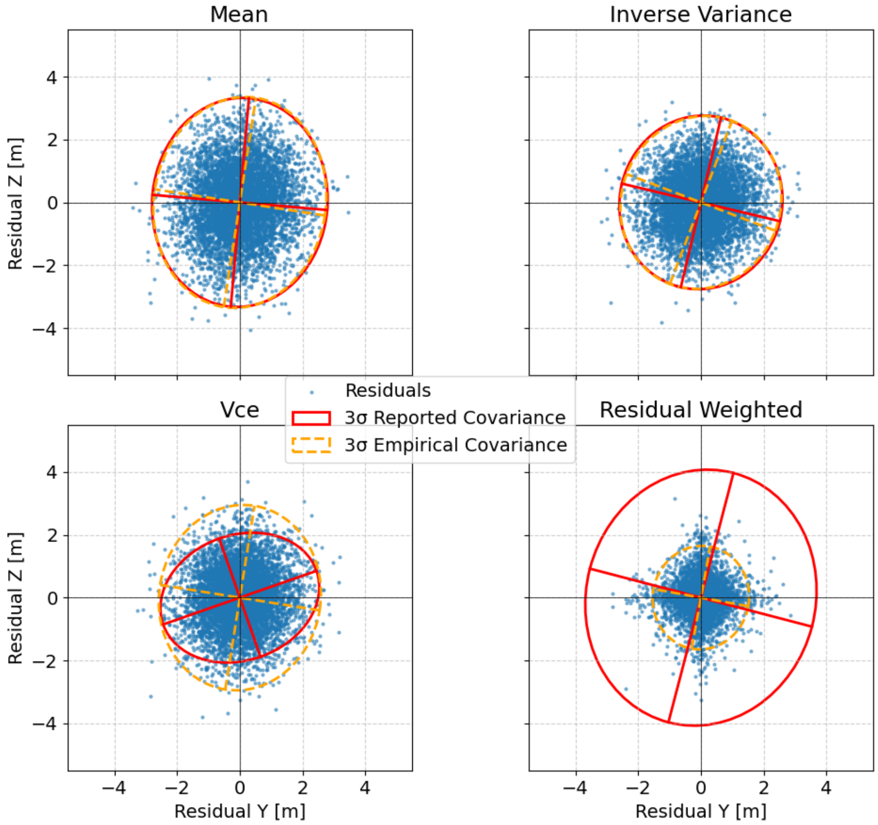


Figure 5.12: Residuals and covariance ellipses in the YZ plane. The solid red ellipse corresponds to reported 3σ uncertainty, and the dashed orange ellipse is derived from the residuals. The semi major and minor axes of the ellipses are plotted to ensure that correlation is visible.

In the YZ plane, the mean and inverse variance continue to demonstrate conformity with the data. VCE reports a similar STD in Y, but underestimates it in Z. The diagonal orientation of the ellipse suggests a positive YZ correlation we do not observe in the residuals themselves. Residual-weighted is again very conservative with the uncertainty information, and shows a slight positive correlation in the covariance. In all three planes there is a clustering of the residuals along each axis for the residual-weighted method. This occurs because the weights are computed per-parameter in this method using the residuals with respect to a reference orbit as per the methodology described in subsection 4.3.1. In this case the reference orbit is ground truth, meaning that any deviation away from the axis (representing zero residuals) has been down-weighted exactly by the inverse of this residual. The result is clustering in an x-shape along the axis, as off-axis deviation is down-weighted. Additionally, there is no distinction between each plane; they behave similarly. This is expected for a full orbit.

Induced correlation by VCE and residual-weighting:

The results demonstrate that correlation is sometimes erroneously estimated in combined orbits when it is not present in the residuals. In order to assess if there is a mistake in either method causing correlation to be induced spuriously, we first assess a combination of input orbits with no covariance. The results for the XY plane are presented in Figure 5.13.

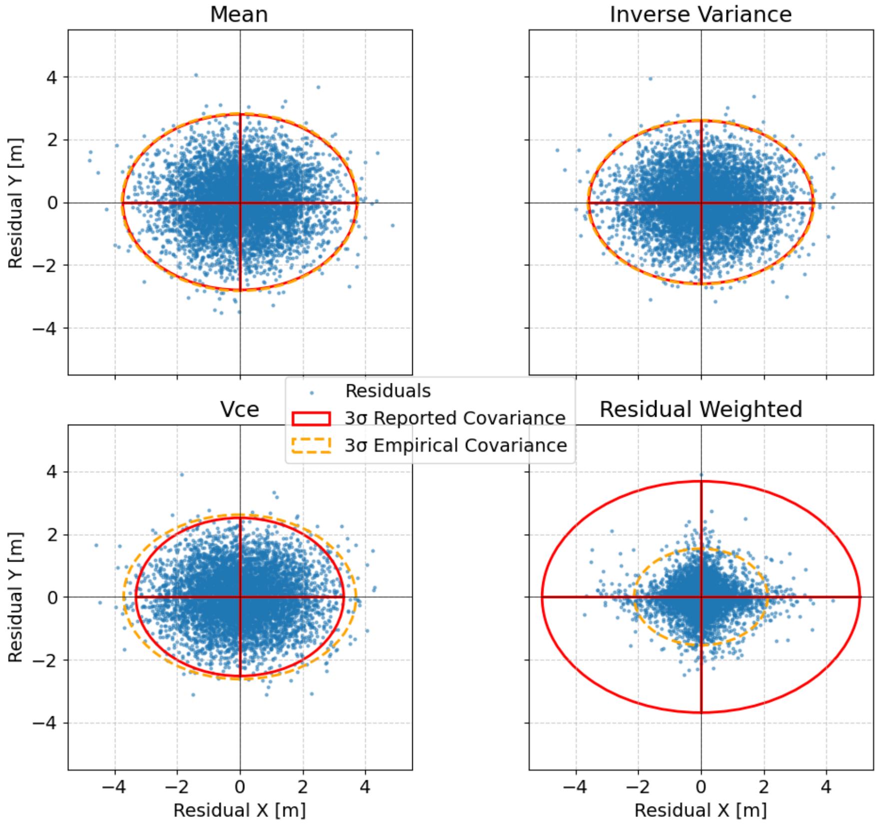


Figure 5.13: Residuals and covariance ellipses in the XY plane. The solid red ellipse corresponds to reported 3σ uncertainty, and the dashed orange ellipse is derived from the residuals. The semi major and minor axes of the ellipses are plotted to ensure that correlation is visible.

Results remain consistent across the other planes. No covariance is introduced into the combined orbit when there is none present in the input data. To test whether correlation is correctly preserved when it is intentionally injected, a follow-up experiment includes input orbits with constant and equal STD's and non-zero covariance only in the XY plane. The results, shown in Figure 5.14, demonstrate that all methods correctly preserve the structure of the input covariance. The empirical and reported ellipses align closely, and no correlation appears in the uncorrelated XZ and YZ planes.

Finally, to test the hypothesis that artificial correlation arises from mixing independently dominated parameters, we construct a test where each coordinate (x, y, z) is dominated by a different orbit by assigning very high noise to all parameters but x in the first orbit, y in the second and z in the third. All inputs remain uncorrelated. The result for the XY plane is shown in Figure 5.15, and demonstrates that no correlation is induced — all methods respect the uncorrelated structure even under coordinate-wise orbit dominance.

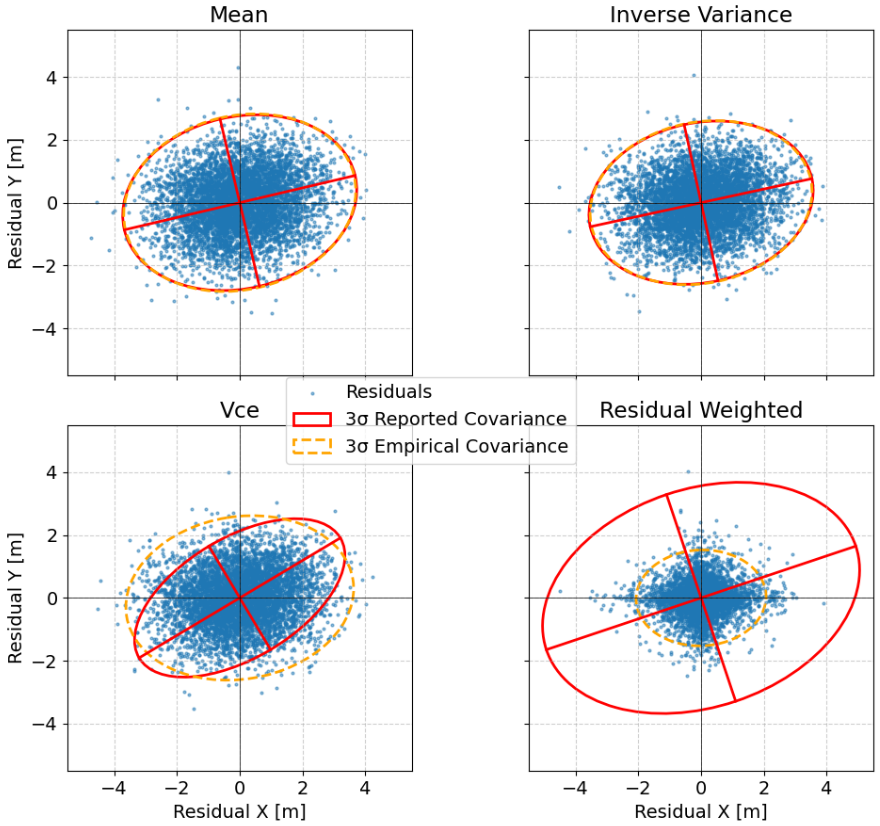


Figure 5.14: Residuals and covariance ellipses in the XY plane for input orbits with correlation injected only in the XY direction. Reported and empirical covariances match well, confirming the methods preserve true input correlation. The semi major and minor axes of the ellipses are plotted to ensure that correlation is visible.

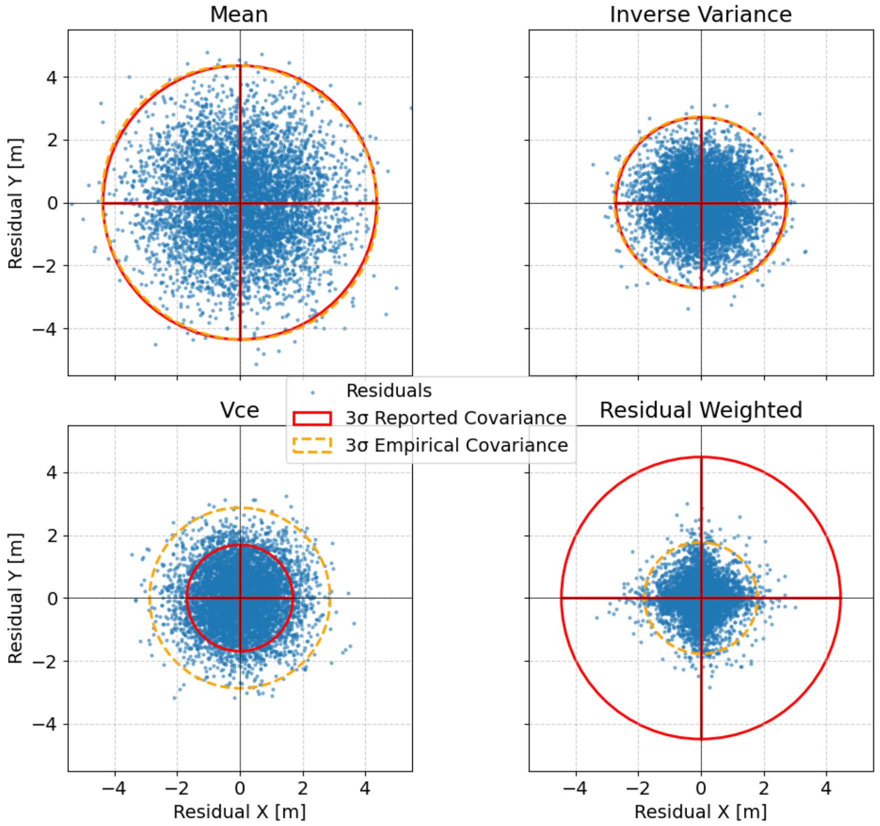


Figure 5.15: Residuals and covariance ellipses in the XY plane when each coordinate is dominated by a different input orbit, all uncorrelated. No artificial correlation is observed in the combined solutions. The semi major and minor axes of the ellipses are plotted to ensure that correlation is visible.

Under controlled conditions, all methods propagate uncertainty predictably. Therefore, the incorrect correlation introduced by residual-weighting and VCE instead must be a product of the interaction of covariance and variance in more complex conditions. Both methods tend towards 'selecting' the best x, y and z parameters across each input orbit, independent to some extent (entirely in the case of the residual-weighted method) of the variance and covariance data associated with it. If, for example, a particular x value has much less error with respect to the combination or the truth, then both methods will tend to fit towards it, entirely neglecting the covariance values of that parameter. The result is that the covariance of that parameter is excessively propagated into the combined orbit, leading to erroneous and larger covariance than is realistic.

Overall, the mean and inverse-variance demonstrated excellent consistency with empirical covariance estimated from the residuals. Provided the input orbit covariance is physically realistic, these methods are demonstrated to propagate the uncertainty well. Residual weighting consistently overestimates the variance information. Both Residual-weighting and VCE struggles to combine more complex covariance information, erroneously injecting correlation into the covariance where none is physically present.

5.3.4 Conclusion of the propagation of uncertainty verification

For the four implemented methods, uncertainty propagation reports values in reasonable agreement with the reported RMS of the combined solutions. The arithmetic mean and inverse-variance report excellent propagation of uncertainty, with uncertainties being directly comparable to the residuals. VCE under-estimates the uncertainty and residual-weighting consistently reports overestimated uncertainty as expected. Both VCE and residual-weighting have the potential to inject spurious correlation into the combined orbit which is not physically present as a result of their weighting scheme.

5.4 Verification of the estimated variance components

Following the uncertainty tests, the effect of the random noise of the input orbits on the variance components estimated by VCE is explored. The estimated variance components (VCE weights) are expected to increase with the input orbit noise, thereby down-weighting noisier inputs in accordance with the methodology described in subsection 4.2.3. We reuse the test setup introduced in section 4.5 to observe resulting variance components across 50 random input orbit instantiations. Figure 5.16 shows the estimated variance components for all coordinates against the corresponding input STD. A strong positive correlation is observed in all directions, confirming that VCE is correctly adapting to the noise levels and behaving as expected by down-weighting noisier inputs above what inverse-variance alone would do.

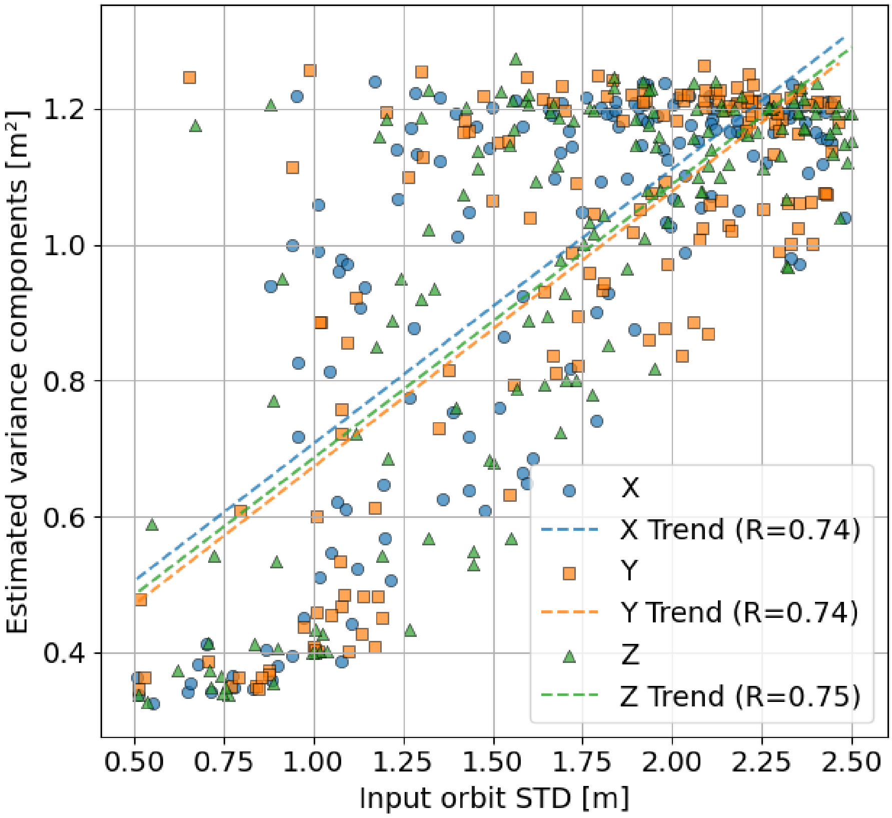


Figure 5.16: Estimated VCE variance component versus input STD, over all coordinates and runs. Dashed lines show fitted trends with correlation coefficients.

For completeness, Figure 5.17 shows the same data separated by coordinate and input orbit, where it can be observed that the results do not vary per-parameter, as expected. The observed trends are consistent across orbits. Saturation occurs at higher noise levels as a result of VCE down-weighting orbits in the combination which have significantly greater noise with respect to the combined estimate to the extent that they effectively no longer contribute to the combined orbit, resulting in this plateau effect towards the top right corner. This test has been conducted for a different noise realisation using seeds 1000-1050, and the results presented in subsection C.1.3 remain consistent with the same trend-lines observed.

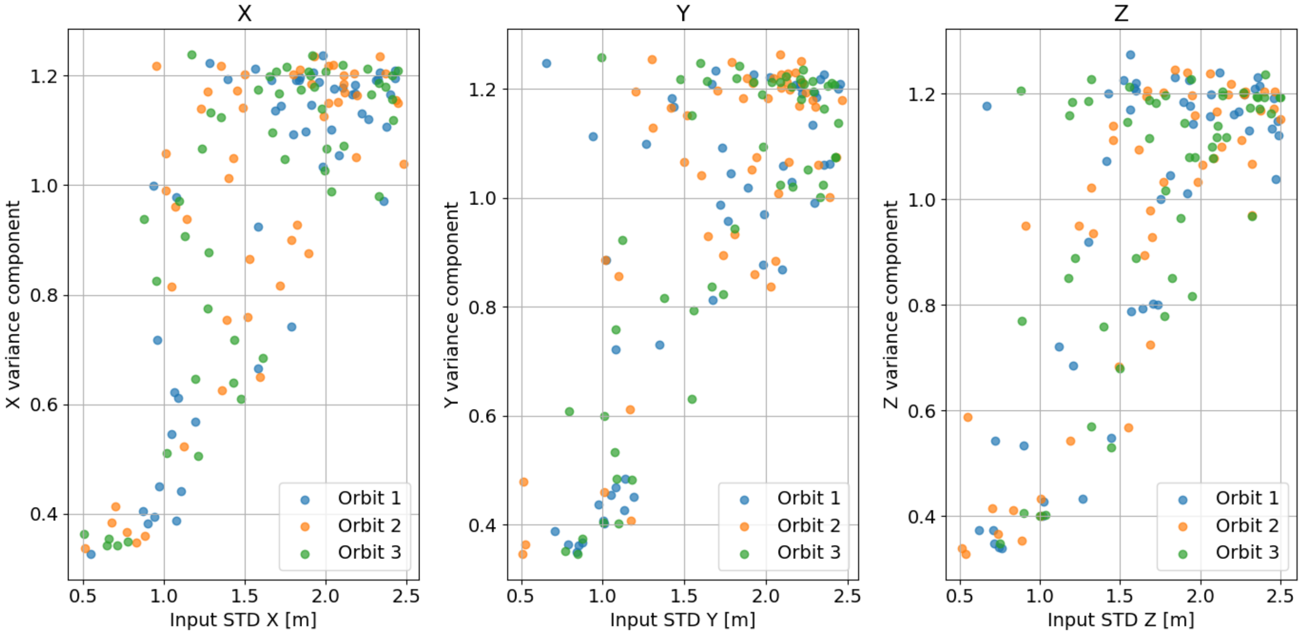


Figure 5.17: Estimated variance components vs input STD, shown per coordinate and input orbit.

In order to demonstrate that when equal-noise input orbits are used approximately constant variance components of 1 are estimated, we test this for 50 runs using a constant STD of 2.5 m, and a constant covariance of 0.1 m. The results are presented in Figure 5.18.

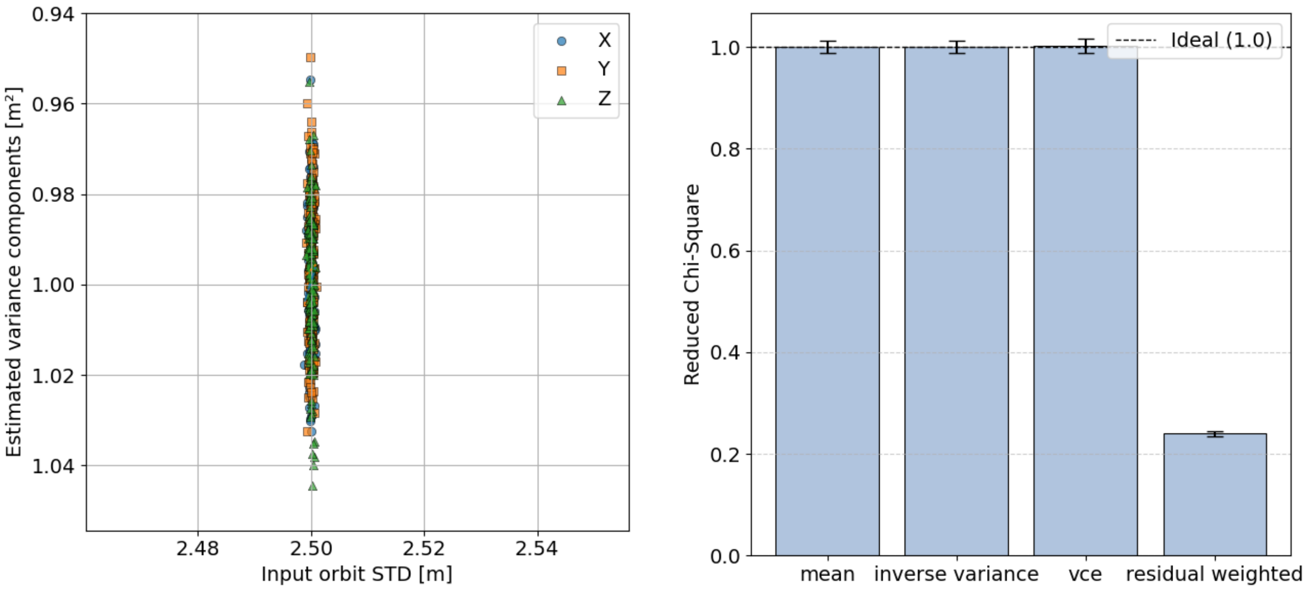


Figure 5.18: The 50-run constant noise estimated variance components, and the resulting reduced chi-square of the orbit solutions across different methods.

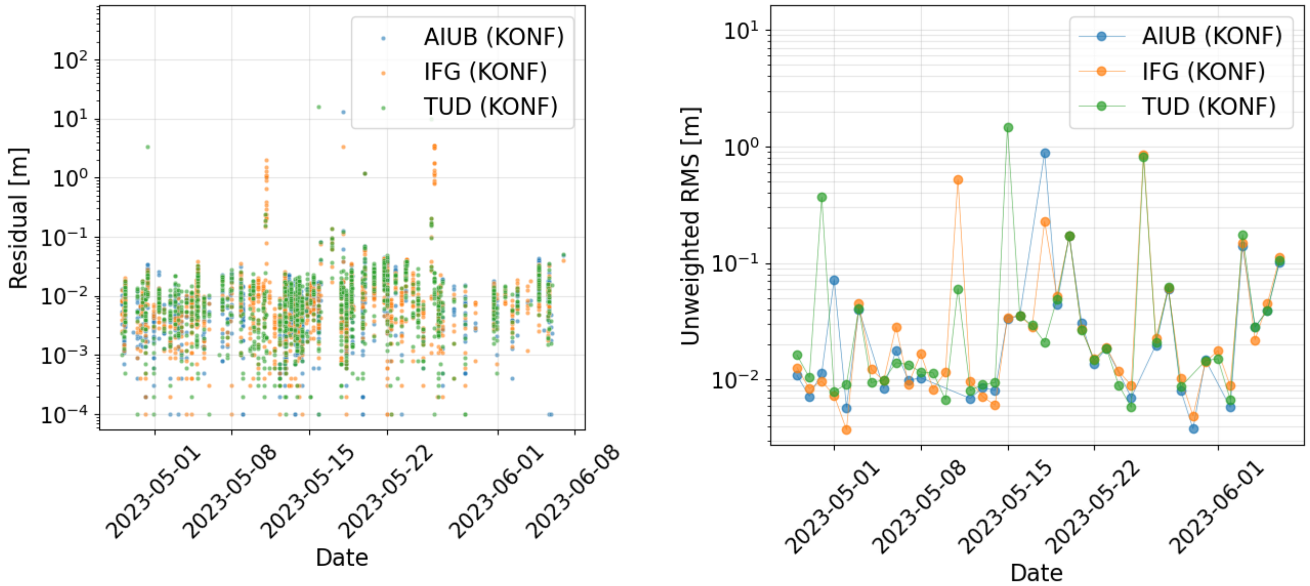
The estimated variance components cluster around 1 for the input orbits with STD 2.5 m. there is some slight deviation which is explained by the random noise realisations leading the variance components for different parameters to be estimated slightly differently each run.

The mean reduced χ^2 is again plotted for this test in Figure 5.18, showing that for a constant noise realisation VCE achieves a ratio of one – a perfect uncertainty representation – when with different-noise inputs it did not. This supports our understanding that the method tends to excessively up-weight inputs it determines are better using the estimated combination, which results in this no longer being an unbiased estimator, but perfectly combines the uncertainty when it is constant due to this excessive up-weighting not occurring.

5.5 Sensitivity analysis of the input-screening threshold

The application of input-screening is described in section 4.1.1, however the actual threshold set is important as it has both positive and negative effects. The screening threshold defines the maximum allowable 3D position difference between any input orbit data point and the RDO at each epoch where this data overlaps. The purpose of this is to screen out outliers in the input data, taking advantage of the fact that RDO's are more stable and consistent due to the inclusion of a dynamic model [76]. A stricter filtering by setting a lower threshold has the effect of more aggressively removing points which differ greatly from the RDO, which in principle should result in the quality of the combined orbit improving in terms of RMS. However, this comes at the expense the data points; screening removes data, and if it is too aggressive a large percentage of epochs is lost. The goal, then, is to balance the improvement in RMS of the input orbits against the epoch loss due to screening.

In order to do so, a particularly noisy period of actual data is selected in 2023 from 2023-04-28 to 2023-06-07. The RMS of SLR residuals is used to compare the performance of the orbits, as well as the number of epochs of data. The unfiltered input orbit (denoted by 'KONF' - Kinematic Orbit, No Filter) SLR residuals are shown for this period in Figure 5.19a, and the daily RMS is provided in Figure 5.19b.



(a) SLR residuals.

(b) Daily RMS of the SLR residuals.

Figure 5.19: SLR observation residuals and daily RMS of each unfiltered input solution for the selected period in 2023 from 2023-04-28 to 2023-06-07.

Significant residuals in the order of meters are present on several days for each input orbit, reflecting the exact type of input data we are aiming to clean using the pre-screening process.

This is a two-objective optimisation problem; we wish to minimise the RMS of the SLR residuals while minimising the epoch loss. As such, we test a set of threshold values and plot the RMS against the epoch loss. The superior solutions form the Pareto-optimal front; the set of solutions for which no other solution improves one objective without simultaneously worsening another. This allows us to objectively identify and compare orbit combination strategies that best balance precision and data availability.

To assess epoch loss properly, we include the results of a combination method, otherwise we cannot observe the total epoch loss of the combined solutions, which is required in order to properly measure the effect of screening. To constrain this test as much as possible to the effect of screening rather than the combination methods themselves, we use the arithmetic-mean method presented in subsection 4.2.1 for comparison. The inclusion of this method represents the total possible number of epochs arising from the union of all epochs present across the set of input orbits, which we are interesting in maximising. The resulting scatter plot including the Pareto front is presented in Figure 5.20.

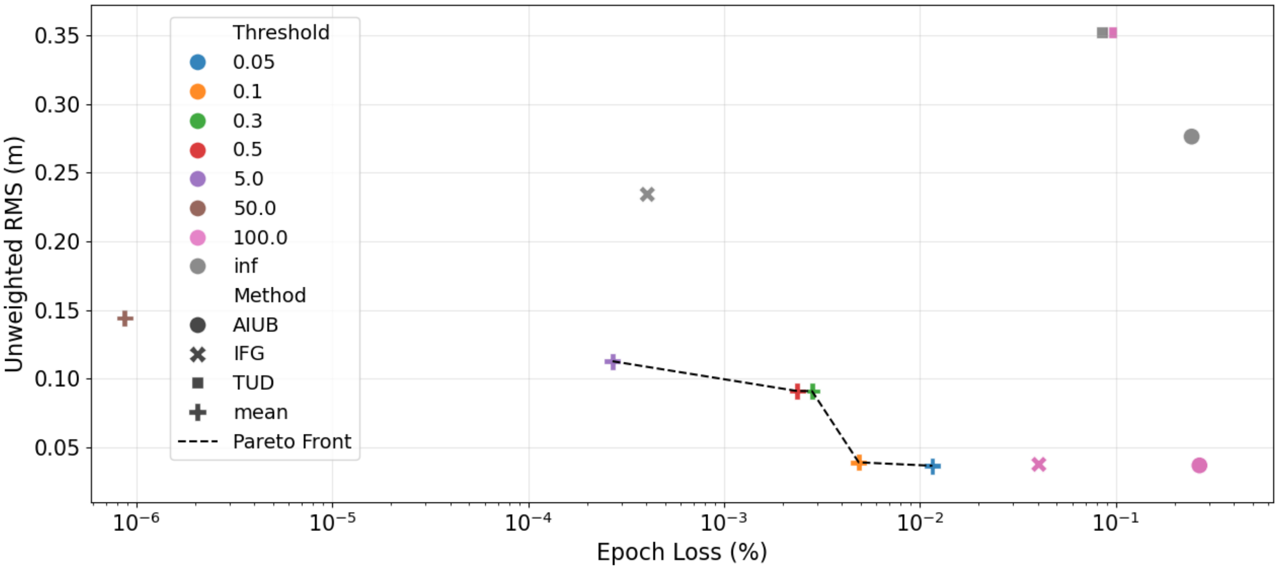


Figure 5.20: Pareto front including all input orbits tested with various filtering thresholds. Linear-log scale. The threshold ceilings are denoted by the colour of the markers, while their shape refers to the method or analysis centre which produced the orbit. An infinite threshold refers to no pre-screening.

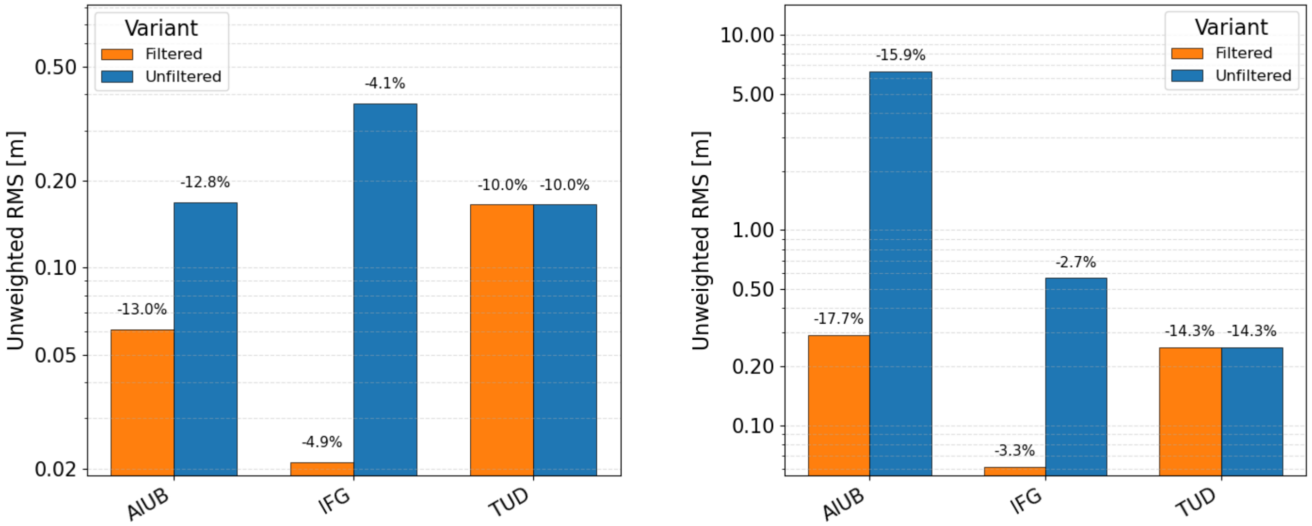
A large discrepancy is observed between the optimality of the input orbits, with IFG generally outperforming the other analysis centres. Regardless, none of the input solutions approach the Pareto front. The Pareto front is dominated by the Arithmetic mean method, as expected as the union of all input epochs is used meaning it consistently outperforms the input orbits in terms of epoch loss, regardless of filtering threshold used. With no filtering, the arithmetic mean experiences no epoch loss in accordance with theory. Epoch loss increases while RMS decreases with decreasing filtering threshold, as expected. No filtering leads to a noticeably less accurate combined solution with an RMS of 14 cm, but provides the maximum data volume. When a threshold of 5 m is used, the mean outperforms every input orbit both in epoch loss and RMS, with the closest input competitor – unfiltered IFG – having similar epoch loss but more than double the RMS. Thresholds of 0.1 and 0.05 demonstrate the very quick increase in epoch loss as the threshold becomes too restrictive. Values of 0.5 and 0.3 m produce similar results, striking a good balance between RMS and epoch loss. Additionally, this test indicates that the input orbits indeed demonstrate outliers at different epochs due to the different processing strategies used to produce them, which is exactly the behaviour we exploit in our combination methods. The exact values of each point on the Pareto front are provided in Table 5.4. When performing the same test with any of the other combination methods, the Pareto front remains dominated by the combined solution; that is, there is never a situation in which one of the input orbits lies on it. These results are available in subsection C.1.4.

Table 5.4: Pareto Optimal Solutions.

Method	**Threshold** [m]	**Unweighted RMS [m]**	**Epoch Loss** [%]
mean	0.05	0.036444	0.011586
mean	0.10	0.038927	0.004876
mean	0.30	0.090867	0.002816
mean	0.50	0.090913	0.002364
mean	5.00	0.112590	0.000270

Based on these results, a pre-screening threshold of 0.3 m is selected for the inputs used for the combination methods. Although it performs very similarly to 0.5 m, the slightly lower value is chosen to be more conservative, giving some additional priority to improving RMS over epoch loss. While values of 0.1 and 0.05 m provide significant further RMS improvements, these thresholds are considered to be too aggressive and risk discarding too much data, as observed in the doubling of epoch loss from 0.3-0.1 and the quadrupling of it from 0.3-0.05. Epoch loss of 0.0028% is observed, which represents approximately 0.001% fewer epochs than IFG's unfiltered input orbit. However, this negligible loss of epochs leads to an improvement in RMS of more than 60% comparatively, with an RMS of 0.1 m observed, demonstrating the benefit of pre-screening.

In order to demonstrate the change in residual RMS and number of epochs available in the input orbits over a long period of time when applying this screening threshold, the RMS of the SLR residuals over 2022 and 2023 are presented in Figure 5.21 for each input solution, with and without screening.



(a) 2022.

(b) 2023.

Figure 5.21: RMS of the SLR residuals over 2022 and 2023 for each input orbit with and without pre-screening. A split axis is used in the right plot due to outliers present in the unfiltered AIUB data.

The decrease in RMS observed demonstrates that performance is improved with pre-screening, while the percentage loss in number of epochs remains minimal in the order of 0.5%. The TUD displays no change in performance with or without screening. This is because the TUD already employs pre-screening with the RDO in its processing strategy, using a threshold of 1 m, and are more conservative with outlier detection [19].

Chapter 6

Combination of Swarm Kinematic Orbits

In order to answer our primary research question – what, if any, improvement is there to be made to the performance of kinematic orbit solutions through their combination – we produce combined orbits over defined periods and assess their quality. With the various components of the combination methods verified and validated in chapter 5, the combination of the kinematic input orbits provided by IFG, AIUB and the TUD is conducted in accordance with the methodology detailed in chapter 4. Two case-study years are selected on which to apply the selected combination methods. 2022 and 2023 are chosen, the former as a more stable year, the latter as a more active year due to the steady increase in solar activity since the beginning of the 25th solar cycle, as observed in Figure 6.1 [72].

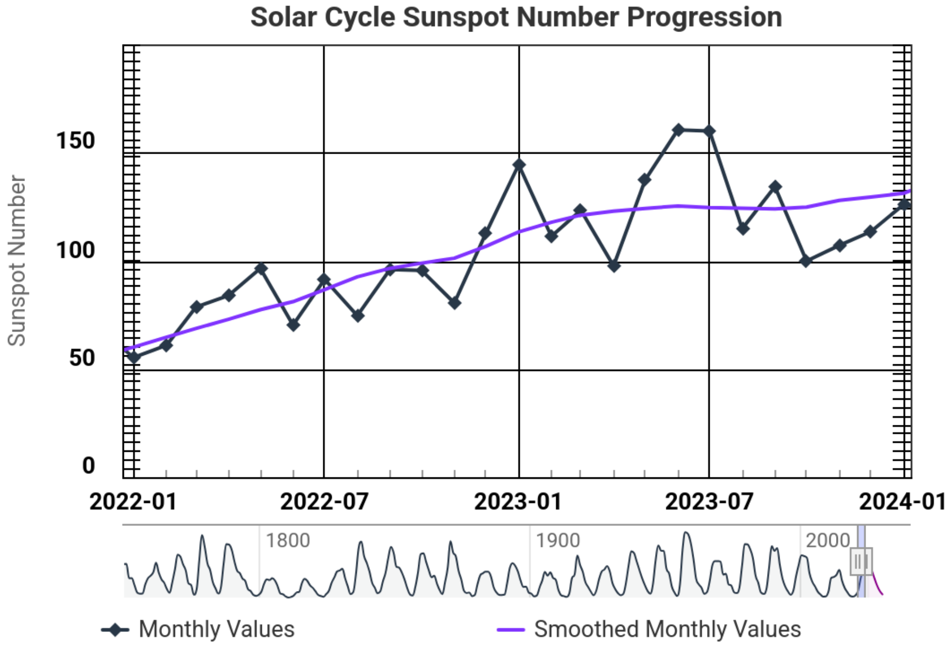


Figure 6.1: Solar activity from 2022-2023 demonstrating that activity is increasing. Image credit: National Oceanic and Atmospheric Administration.

For the purpose of assessing the combination methods, only Swarm A orbits are used. Additional results for 2023 are available for Swarm B and C in Appendix D.

Input orbits are consolidated to consistent epoch times using interpolation and then pre-screened using their residuals with respect to the more accurate and reliable TUD RDO's disseminated by ESA. The screening threshold is fixed at 0.3 m of 3D residual.

The final selected methods for generating results for the case study years are as follows:

- Arithmetic mean
- Inverse variance weighting
- Variance component estimation
- Residual-weighting
- Optimisation using Nelder-Mead
- Optimisation using CMAES

Selected based upon by their performance in the tests conducted in chapter 5. The metrics used to assess the performance of the combined orbits compared to the input orbits is the RMS, mean and distribution of the SLR residuals, and the percentage of lost epochs of data with respect to the maximum possible epochs available.

In both 2022 and 2023, coverage by the input orbits is not complete. The TUD orbits miss no days, however AIUB misses a significant number of days totalling 44, and IFG misses a total of 4 days. This is visualised in Figure 6.2.

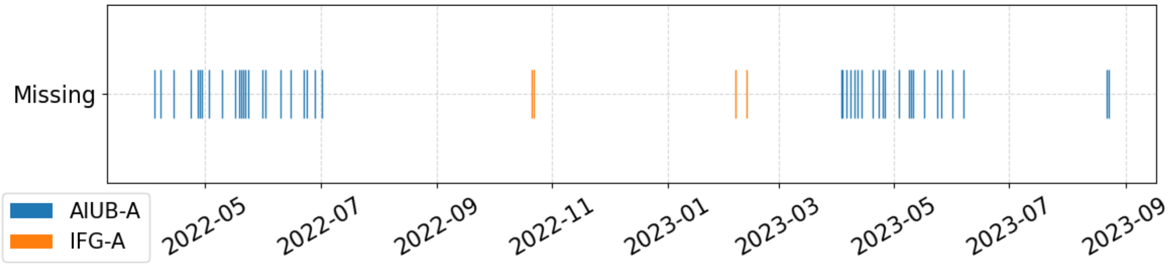


Figure 6.2: Missing days for IFG and AIUB across 2022 and 2023. vertical bars represent where data is missing.

In addition, the weekly RMS of the input orbits for 2022-2023 are presented in Figure 6.3. There is significant variation week to week, and a general upward trend in the RMS of the residuals in accordance with the increasing solar activity. In some weeks, RMS values ar far in excess of meters, highlighting the value of the combination and especially the pre-screening detailed in section 4.1.1.

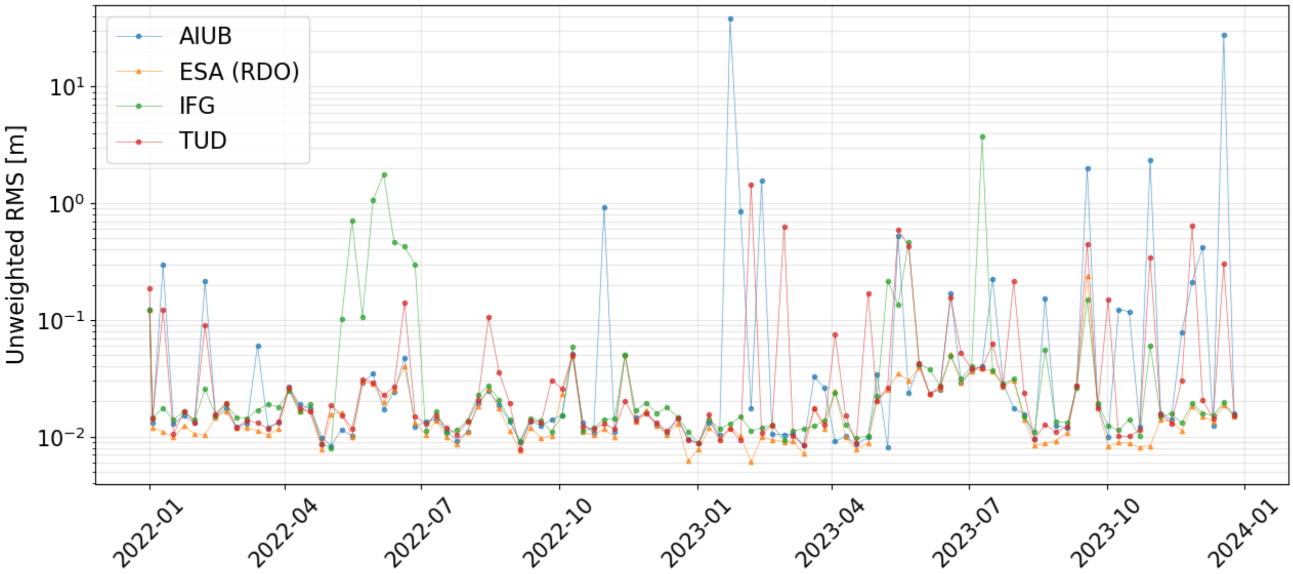


Figure 6.3: The weekly RMS of the residuals for each input orbit and the reference RDO for 2022 and 2023.

6.1 SLR observation sensitivity analysis

The SLR observation residuals are used to assess the performance of orbits in accordance with the method introduced in section 4.6. However, as SLR observations are taken at multiple ground stations – each with their own unique sets of uncertainties, errors and varying performance – it is essential to make a selection of high-performance stations and to filter out observations at which there is an issue with the station or the satellite itself. Otherwise, the RMS of the residuals can be polluted by a small number of erroneous data points. To make a selection of SLR stations and determine which – if any – filtering is required, we perform a sensitivity analysis.

6.1.1 Normal point outlier filtering and station quality control using RDO

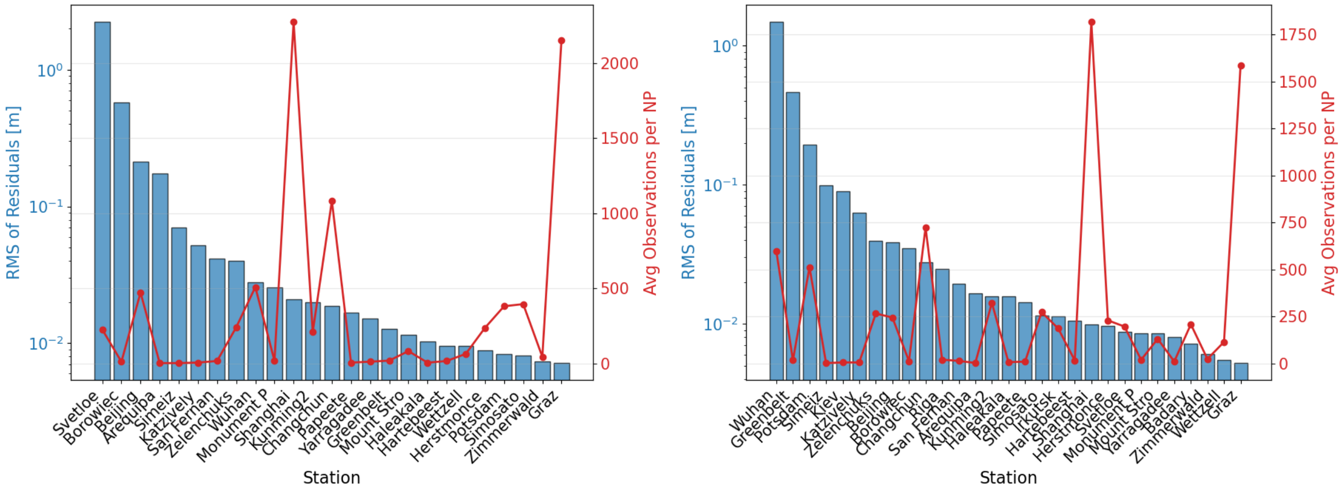
To ensure the reliability of the SLR residuals used for orbit validation, a filtering procedure is applied to remove problematic data, both at the observation and station level. This process leverages the ESA RDO as a high-accuracy independent reference. The RDO solution is known for its high precision and robustness, and is free from outliers or dynamic inconsistencies [76]. It therefore serves as an effective benchmark for assessing the quality of SLR stations. For each normal point observation, the residual with respect to the ESA orbit is computed as:

$$\Delta\rho_{(\mathrm{ESA})}(t) = \rho_{(\mathrm{obs})}(t) - \left\| \mathbf{r}_{(\mathrm{ESA})}(t) - \mathbf{r}_{(\mathrm{sta})}(t) \right\| \tag{6.1}$$

where $\mathbf{r}_{(\mathrm{ESA})}(t)$ is the satellite position from the ESA RDO at epoch t, and $\mathbf{r}_{(\mathrm{sta})}(t)$ is the geocentric position of the SLR station in the same terrestrial reference frame. To eliminate the influence of gross errors and unstable station performance, a two-stage filtering process is applied; epoch rejection by outliers and station quality screening.

Epoch rejection

Data at any epoch where the ESA residual exceeds 10 metres in absolute value is rejected entirely, across all methods and stations. This ensures that all observations coinciding with known outlier events – such as manoeuvres or corrupted tracking data – are excluded from the normal point dataset. A threshold of 10 m is selected as we are certain that residuals above this for the RDO are a product of the station performance rather than the orbit itself, which is known to be robust and consistent [76]. If this is not done, SLR results are dominated by the few extremely bad residuals. To demonstrate the effect of neglecting this poor station data filtering, the per-station RMS of the residuals are presented in Figure 6.4 for 2022 and 2023, plotted alongside the average number of points used to construct a normal point by each station.



(a) 2022.

(b) 2023.

Figure 6.4: SLR residual RMS per station in 2022 and 2023 with respect to the ESA RDO orbit. The twin axis shows the average number of data-points used to construct each normal point (NP).

In both years there are station outliers; Svetloe, Borowiec, Beijing, Arequipa, Wuhan, Potsdam and Greenbelt each demonstrate RMSE above one meter. The increased RMS does not correlate with the observation number, suggesting that it is not very dependent on the number of observations used to construct the normal point and is rather the stations themselves performing worse. After this screening, remaining outliers in the other orbits are products of their processing strategies, not a result of poor station performance. The number of outliers removed for each orbit for each year is provided in Table 6.1, along with the orbits which retain unfiltered outliers.

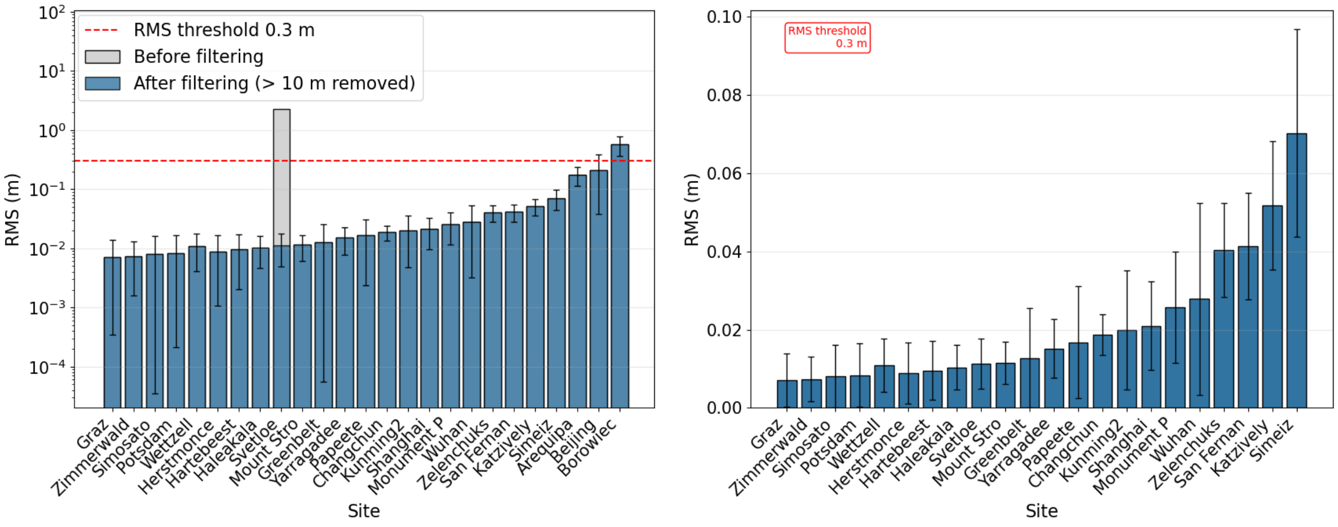
Table 6.1: Summary of SLR outlier filtering. The second column shows the number of epochs removed using RDO residuals $\geq 10\,\text{m}$. The third column includes the total the number of remaining epochs with residuals $\geq 10\,\text{m}$ from non-RDO methods. The fourth column lists the orbits with non-filtered residuals $\geq 10\,\text{m}$ remaining.

Year	**Epochs Removed**	**Remaining High Residuals**	Orbit solutions with $\geq$ 10m residuals remaining
2022	2	2	TUD
2023	1	95	AIUB, TUD, IFG

The outliers inherent in the combined solutions align with those present in the RDO, which are erroneous outliers; essentially, these results indicate that our combined solutions have no outliers, or outliers are at the very least of significantly lower magnitude when compared to those present in the input orbits.

Station benchmarking and selection

For the remaining ESA RDO residuals, the RMS is computed separately for each station. Stations for which the sum of the RMS and the STD of the residuals exceeds a threshold of 0.3 metres for the RDO data are considered unreliable and excluded from further analysis. This threshold is selected as once we have removed outlier observations for the RDO, we know we are left with quality observations. The threshold serves as a method to select a set of higher-performing stations, and a conservative value is chosen above which a station's performance is determined to be poorer.

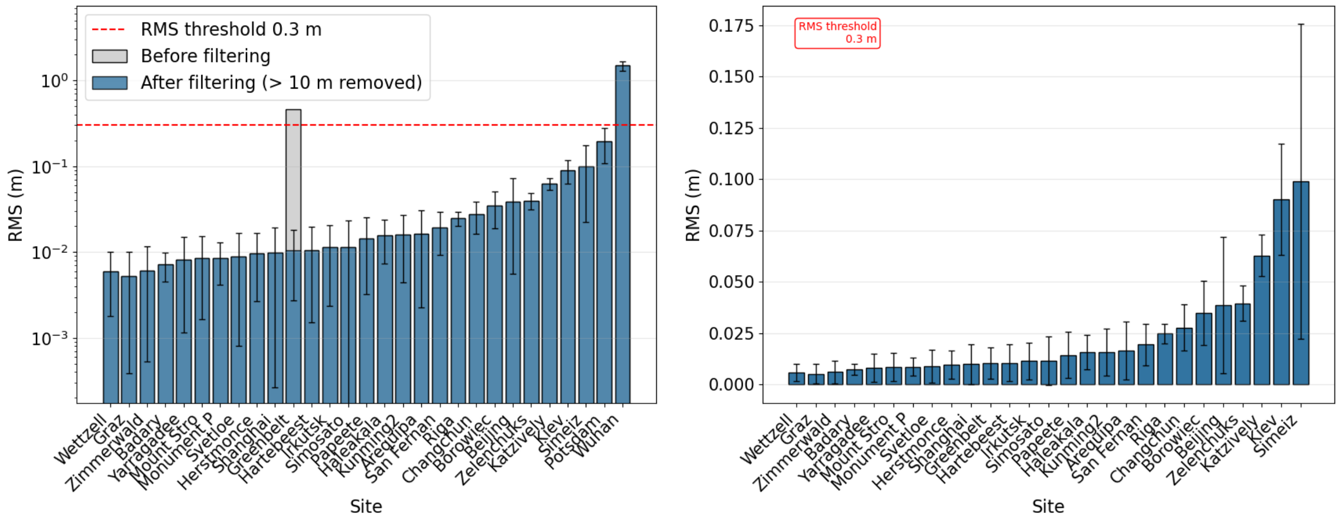


(a) Before filtering.

(b) After filtering.

Figure 6.5: SLR residual RMS per station in 2022 with respect to ESA RDO orbit. The 0.3 m threshold is indicated.

The effectiveness of this selection is visualised in Figure 6.5 for 2022 and Figure 6.6 for 2023, where the RMS per station is shown before and after outlier filtering. Once these two steps are applied, the selected stations show residual RMS' in the range of millimetres to tens of centimetres, rather than the RMS values in the order of metres observed earlier.



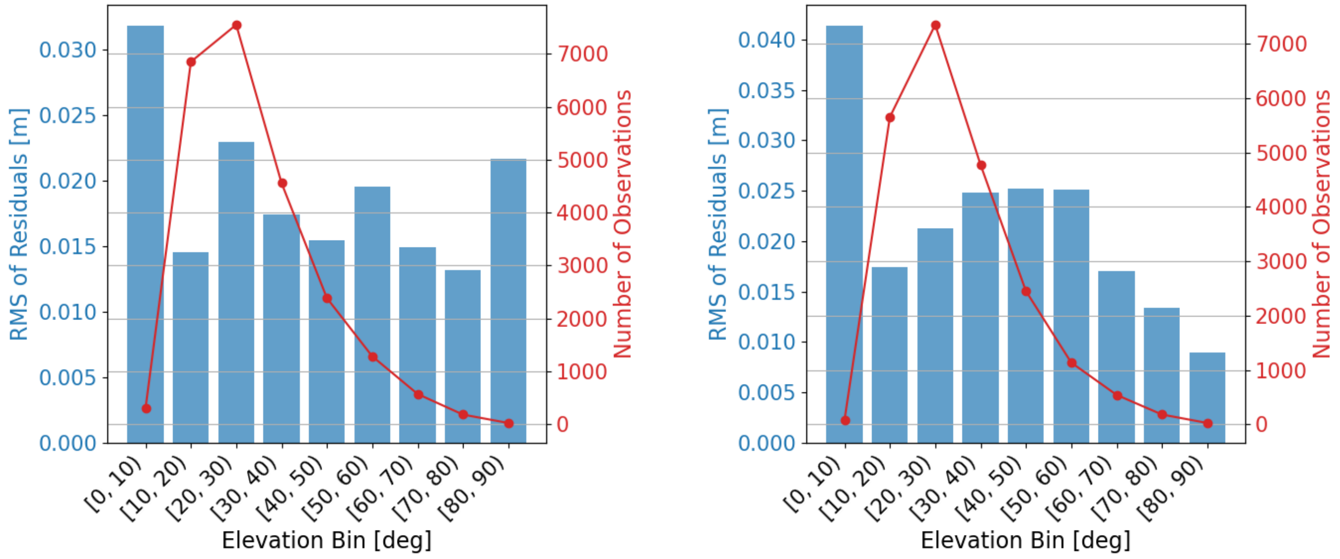
(a) Before filtering.

(b) After filtering.

Figure 6.6: SLR residual RMS per station in 2023 with respect to ESA RDO orbit. The 0.3 m threshold is indicated.

Elevation cut-off angle

Finally, we perform a filtering of observations based on the elevation angle they were taken at. Lower elevation angles introduce systematic errors into the SLR residuals, as the laser must pass through considerably more atmosphere leading to greater atmospheric refraction [63]. Typically, this cut-off angle is set to between 10 and 20 degree depending on application [18]. To determine where to set the cut-off, an analysis of the sensitivity of the residuals to the elevation angle is performed. The residual observations are binned between elevation angle ranges, and the RMS of the binned residuals is computed, allowing us to observe how the RMS changes with elevation angle.



(a) 2022.

(b) 2023.

Figure 6.7: Overall SLR residual RMS of the ESA RDO for 2022 and 2023 computed between fixed elevation angle range bins of 10 degrees. The number of observations between each bound is provided on a second y-axis.

The results are presented in Figure 6.7. The number of observations associated with each bin is included to indicate the frequency of observations at different elevation angles. Residual quality significantly reduces for observations taken at elevation angles below 10 degrees, with an RMS almost twice that of the next greatest. Therefore, we filter out observations taken below an elevation angle of 10°. In total, once we apply filtering, we are left with an average of 23377 SLR observations per orbit, compared to a pre-filtered total of 31100.

6.2 Final combined orbit results

The remaining observations can be used to make an objective comparison of the orbits. We begin by addressing our primary research question. The performance of the combined orbits for the full years of 2022 and 2023 are presented in figures 6.8 and 6.9 as the RMS of the residuals across the full years for each method. The input orbits are included for the purpose of comparison. The epoch loss as a percentage with respect to the maximum possible epochs is annotated atop the bars.

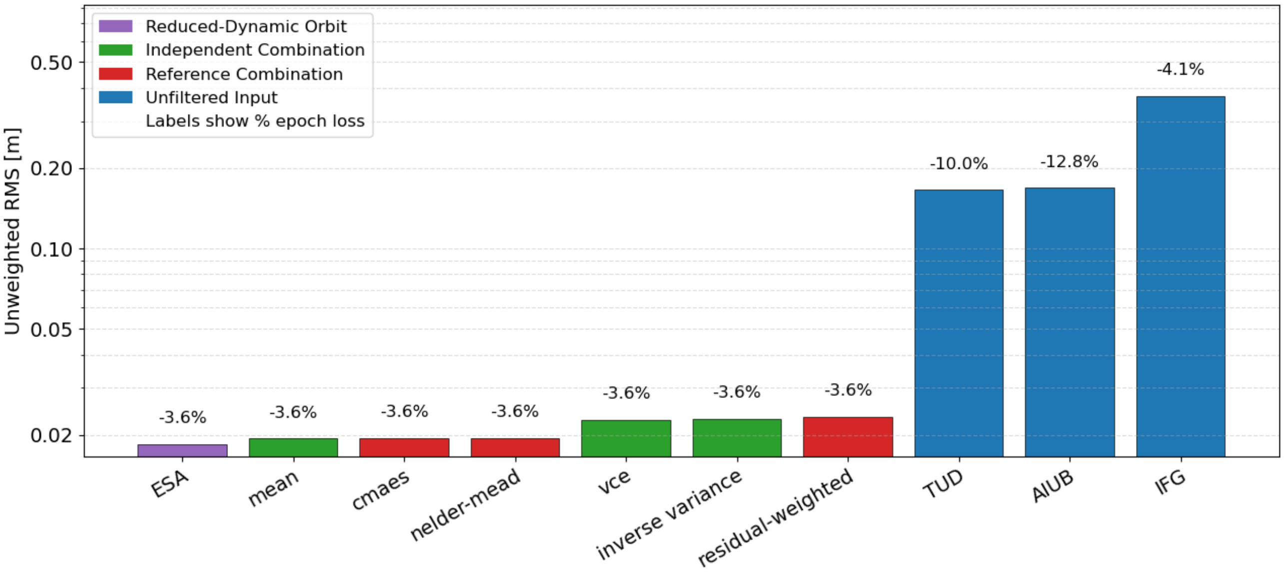


Figure 6.8: RMS of the SLR residuals for each orbit for the year of 2022. % epoch loss is annotated atop the bars.

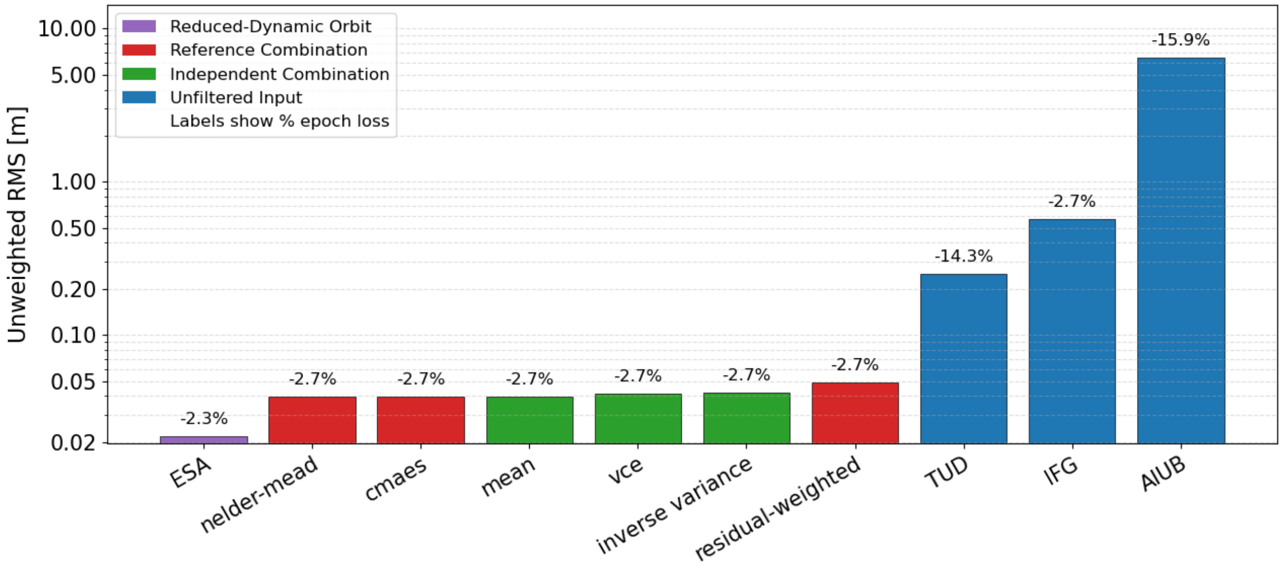


Figure 6.9: RMS of the SLR residuals for each orbit for the year of 2023. % epoch loss is annotated atop the bars.

The numerical results are provided in Table 6.2, including the percentage epoch loss, residual RMS, and the mean and STD of the residuals.

Table 6.2: Epoch loss and residual statistics for tested orbit solutions in 2022 and 2023. Total possible epochs: 31,512,665 (2022); 31,361,315 (2023). ESA RDO is shown as a reference only. The green cells indicate best performance among the tested methods. All metrics refer to residuals (e.g. 'Mean' refers to 'Mean residual').

Orbit Solution	2022				2023			
	Epoch Loss [%]	RMS (m)	Mean (m)	STD (m)	Epoch Loss [%]	RMS (m)	Mean (m)	STD (m)
RDO (ESA)	3.6	0.01849	0.00103	0.01846	2.3	0.02175	0.00053	0.02174
IFG	4.1	0.37272	0.01083	0.37257	2.7	0.56937	0.00583	0.56935
TUD	10.0	0.16579	0.00095	0.16579	14.3	0.25057	0.00233	0.25056
AIUB	12.8	0.16872	0.00195	0.16871	15.9	6.50626	0.31537	6.49877
Arithmetic mean	3.6	0.01936	0.00060	0.01935	2.7	0.03941	-0.00001	0.03942
Inverse variance	3.6	0.02288	0.00079	0.02287	2.7	0.04207	0.00046	0.04207
VCE	3.6	0.02276	0.00066	0.02275	2.7	0.04122	0.00048	0.04122
Residual-weighted	3.6	0.02331	0.00090	0.02330	2.7	0.04917	0.00049	0.04917
Nelder-Mead	3.6	0.01948	0.00067	0.01947	2.7	0.03937	-0.00006	0.03937
CMAES	3.6	0.01948	0.00068	0.01947	2.7	0.03937	-0.00006	0.03937

All of the combined solutions outperform each input orbit in both years, both in terms of epoch loss and residual RMS. In 2022, the combined orbits retain 96.4% of epochs, gaining 0.5% more than IFG, 6.4% more than TUD, and 9.2% more than AIUB. In 2023, they retain 97.3% of epochs; matching IFG, and gaining 11.6% and 13.2% relative to TUD and AIUB, respectively.

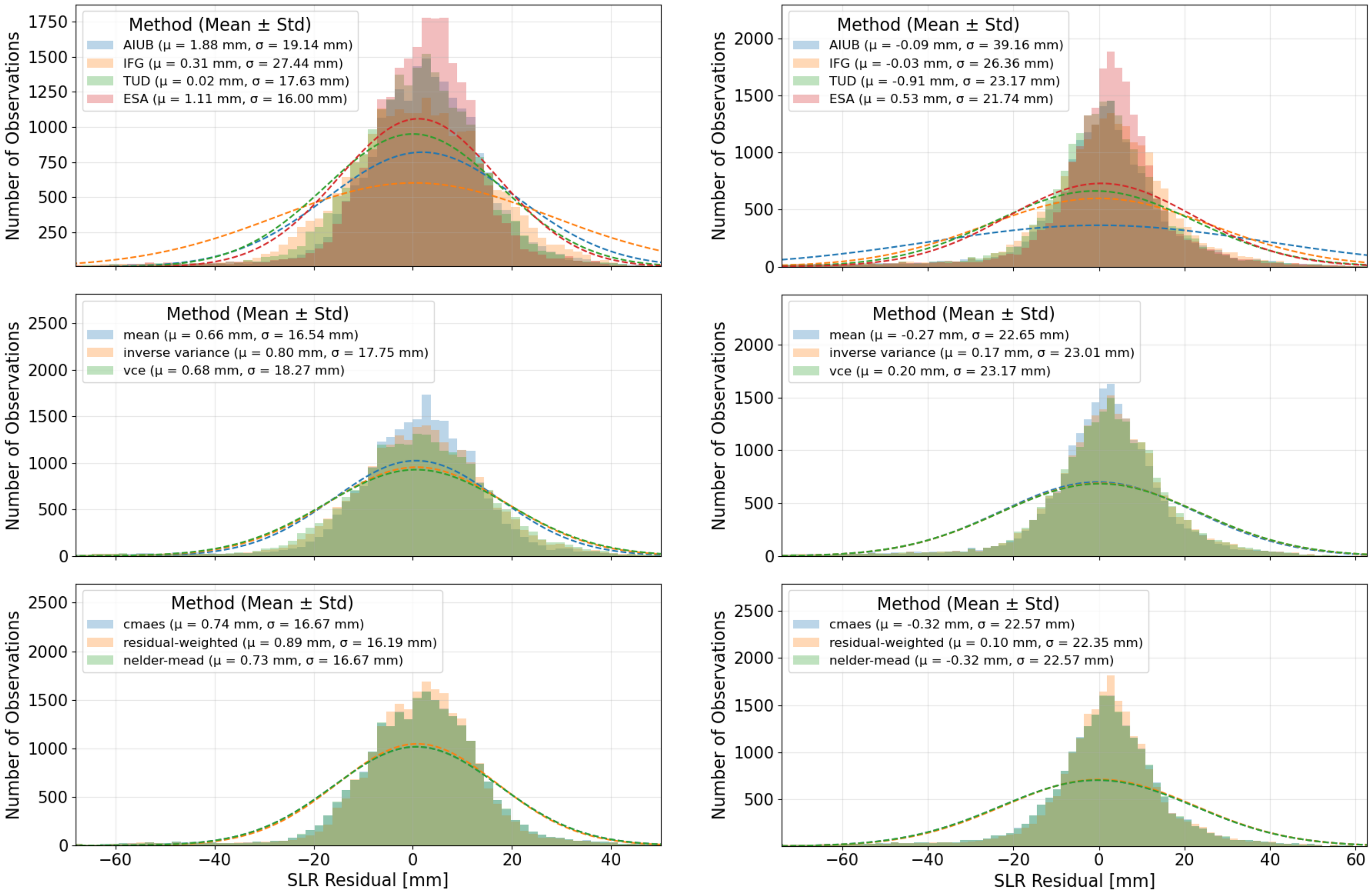
The average percentage improvement in RMS across the combined methods exceeds 90% compared to the input orbits in both years. For example, in 2022, input RMS values range from 0.16579 m (TUD) to 0.37272 m (IFG), while the combined solutions range from 0.01936 m to 0.02331 m. In 2023, input RMS values span from 0.25057 m (TUD) to 6.50626 m (AIUB), while the combined methods lie between 0.03937 m and 0.04917 m.

Compared to the RDO, the arithmetic mean, Nelder-Mead, and CMAES methods are only 5% worse in 2022 (RMS: 0.01849 m vs 0.01936 m), while the inverse-variance and VCE methods are 23.7% worse (RMS: 0.02288 m and 0.02276 m, respectively). In 2023, the best combined orbit (Nelder-Mead and CMAES, both at 0.03937 m) is approximately 81% worse than the RDO (0.02175 m), and the arithmetic mean is 81.2% worse (0.03941 m). The residual-weighted method performs the worst among the combined solutions, with an RMS of 0.04917 m — 126% higher than the RDO in 2023. Despite this, the RMS of the best combined orbits in 2023 is still 93.1% lower than that of the best-performing input (IFG, 0.56937 m), indicating that the combination process remains highly effective even under poorer tracking conditions.

The differences in RMS performance among the combined methods are modest: 0.00395 m (20.4%) in 2022 and 0.00980 m (24.9%) in 2023. The arithmetic mean performs exceptionally well in both years, yielding the lowest RMS in 2022 (0.01936 m) and the second-lowest in 2023 (0.03941 m). It also maintains the lowest STD in both years (0.01935 m in 2022; 0.03942 m in 2023). In contrast, the inverse-variance method yields RMS values of 0.02288 m (18.2% worse than the mean) in 2022 and 0.04207 m (6.7% worse) in 2023. The VCE method performs similarly, with RMS values of 0.02276 m (17.6% worse) and 0.04122 m (4.6% worse), respectively. These methods consistently underperform the arithmetic mean, likely due to imperfections in the scaling of the input covariance information, which leads to sub-optimal weighting. A detailed investigation of this issue is presented in subsection 6.2.2.

There is limited benefit in applying the optimiser-based combinations under current data conditions. In 2023, the best-performing solutions – Nelder-Mead and CMAES, both at 0.03937 m – improve over the arithmetic mean by only 0.00004 m (0.04 mm). This marginal improvement is not consistent; in 2022, the arithmetic mean outperforms all optimiser-based methods. However, this trend aligns with the low bias of the input orbits. The optimisers are designed to shift the centre of the combination to compensate for systematic biases; when bias is minimal, little improvement is observed. This becomes apparent in 2023, where the AIUB input exhibits a substantial radial bias of 0.31 m, and the optimisers yield a small improvement over the mean. This suggests that optimiser-based approaches may offer tangible benefits when significant input bias is present.

In addition to improved accuracy and data retention, the combined solutions also exhibit lower STD's than the inputs, reflecting reduced noise. For example, the arithmetic mean maintains the lowest STD across all methods in both years (0.01935 m in 2022; 0.03942 m in 2023). To properly observe differences in spread, histograms of the residuals are visualised in Figure 6.10, with a 0.5 m ceiling applied to suppress outliers and focus on the core distribution. Gaussian curves are overlaid to highlight shape and symmetry.



(a) 2022.

(b) 2023.

Figure 6.10: Histogram of the SLR residuals, and the mean and STD of the residual distribution representing the bias of the orbit solution. Gaussian curves are fit overtop. Outlier filtering of 0.5m is applied as the focus is on the distribution, which is otherwise dominated by excessively large residuals.

The RDO displays the narrowest distribution in both years. The combined orbits show consistently tighter residual spreads than the input solutions, suggesting that the combination process effectively suppresses measurement noise and outliers. Among the combined methods, the residual-weighted solution yields the narrowest distribution, providing visual evidence that it closely fits to the RDO. In contrast, the inverse-variance and VCE methods show noticeably broader spreads, with STD's approximately 8% higher than other combined orbits in 2022 and 1.5% higher in 2023. This is consistent with their tendency to up-weight inputs with low reported uncertainty, even when those inputs exhibit greater actual residuals, leading to overconfidence. This behaviour is explored further in subsection 6.2.2.

Systematic bias, as indicated by the mean residual, is small across all orbits; always less than 1 mm for the combined solutions. In 2022, all methods exhibit positive mean residuals, consistent with a small radial bias, while the TUD input has a near-zero mean (0.00095 m), suggesting minimal systemic error. In 2023, all orbits again exhibit non-zero means, with AIUB and IFG closest to zero. These findings align with values reported in the literature [63], with any minor differences explained by the station selection and elevation angle threshold used in this analysis.

The RDO displays the narrowest distribution in both years, with STD's of 0.01600 m in 2022 and 0.02174 m in 2023. The combined orbits have consistently tighter residual spreads than the input solutions, suggesting that combination effectively suppresses measurement noise and outliers. Among the combined methods, the residual-weighted solution yields the narrowest distribution in 2023, with a STD of 0.02235 m – very close to the RDO's 0.02174 m – and the lowest among the combined solutions that year. In 2022, it achieves an STD of 0.01619 m, just above the RDO's 0.01600 m.

In contrast, the inverse-variance and VCE methods show noticeably broader spreads than the other combined solutions. In 2022, their STD's are 0.01775 m and 0.01827 m, which are 6.5% and 9.6% higher than the next-best combined methods – CMAES and Nelder-Mead – both at 0.01667 m. In 2023, their spreads increase to 0.02301 m (inverse-variance) and 0.02317 m (VCE), approximately 1.9% and 2.7% higher than CMAES and Nelder-Mead (both 0.02257 m). This persistent gap highlights the poorer noise performance of the statistically informed methods, consistent with their tendency to up-weight inputs with underestimated uncertainty, even when those inputs exhibit higher actual residuals. This results in overconfident, but noisier, combinations. This behaviour is explored further in subsection 6.2.2.

Systematic bias, indicated by the mean residual, is small across all orbits and always less than 0.001 m for the combined solutions. In 2022, all methods exhibit positive mean residuals between 0.00066 m (arithmetic mean) and 0.00089 m (residual-weighted), consistent with a small radial bias. The TUD input has a near-zero mean of 0.00002 m, suggesting minimal systemic error. In 2023, the combined orbits again show low mean residuals, ranging from -0.00032 m (CMAES and Nelder-Mead) to 0.00020 m (VCE), while AIUB and IFG inputs have means closest to zero at -0.00009 m and -0.00003 m, respectively. These findings align with values reported in the literature [63], with any minor differences explained by the station selection and elevation angle threshold used in this analysis.

6.2.1 Over-fitting to the reference reduced dynamic orbit

In order to answer our first sub-question regarding the risk of over-fitting to the RDO with the methods which use it as reference – CMAES, Nelder-Mead and residual-weighting – we perform an investigation following the methodology described in section 4.4. The results are presented in Figure 6.11

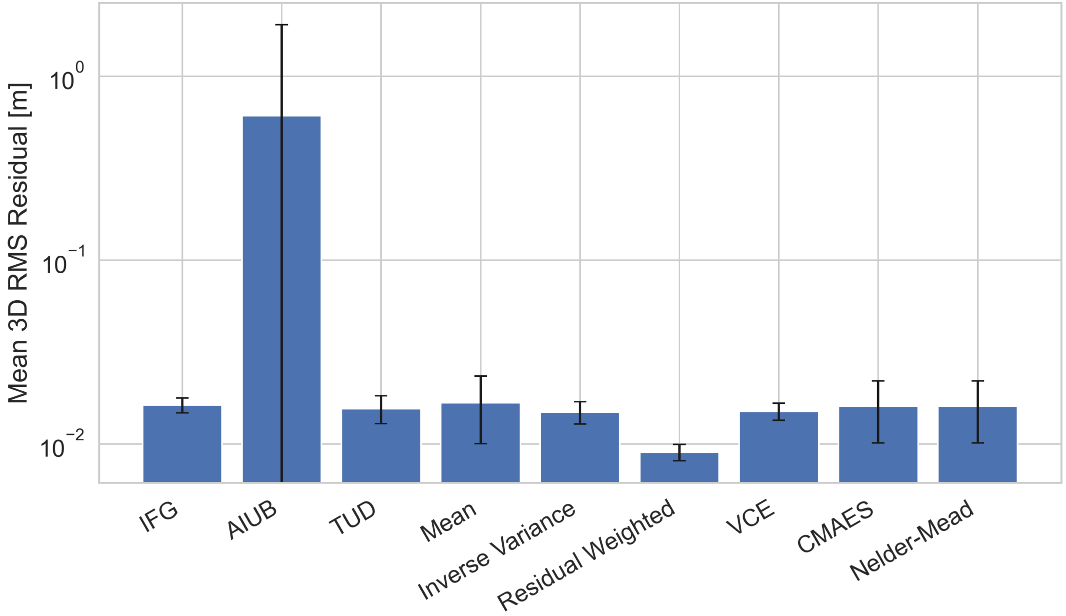


Figure 6.11: RMS of the residuals of each input and combined orbit with respect to the RDO used as reference in the dependent combination methods, as well as the STD of the shown as error-bars. Due to outliers, AIUB shows RMS comparable to the STD of its residuals, hence the error-bar overlaps. Results for the full year of 2023.

We notice that AIUB has a significantly higher RMS with respect to the RDO due to the presence of outliers. There is no sign of over-fitting in the CMAES or Nelder-Mead methods, as these solutions both have a greater RMS with respect to the RDO than all but the AIUB input orbit and the arithmetic mean combined orbit, indicating that the piece-wise constant weights – as well as the narrow search space defined about the mean of the input orbits – serve to avoid this phenomena. We apply no such compensation to the residual-weighted method, which directly uses the residuals of the input orbits with respect to the RDO to inform the combination. As a result there is a strong indication of over-fitting, with a 40% decrease in RMS observed with respect to the next-closest method. In addition, it has the lowest standard deviation, further indicating that the distribution of residuals is closer to the RDO as well. Therefore, we conclude that the residual weighted method presents a risk of over-fitting to the reference orbit.

This test is not conclusive, and in order to thoroughly investigate and quantify the effect – if any – of potential biasing toward the RDO (and therefore implicitly the a-priori gravity field), gravity field models must be estimated and analysed; something left to the recommendations of this thesis.

6.2.2 Variance information of the combined orbits

To evaluate the realism and quality of the variance information provided by the combined orbits and compare it to the individual kinematic orbit solutions and to answer our second sub-question regarding the propagation of uncertainty into the combined orbits, a dedicated assessment is performed using the SLR range residuals. This validation determines whether the reported orbit STD adequately reflect the actual orbit errors as observed via independent SLR measurements. For this, the period of January and February 2023 is examined. The methodology used to produce this variance information validation is presented in subsection 4.6.1. The methods assessed are those for which uncertainty propagation is implemented; the mean, inverse-variance, VCE and residual-weighted.

The results are presented in Table 6.3. The reduced χ^2 is plotted for input and combination methods in Figure 6.12. If the SLR residuals show agreement with the STD projected along the LOS vector determined from the orbit data, we can state that the uncertainty information is physically realistic. We include the mean of the residuals to observe whether or not there is any systematic bias in each orbit solution.

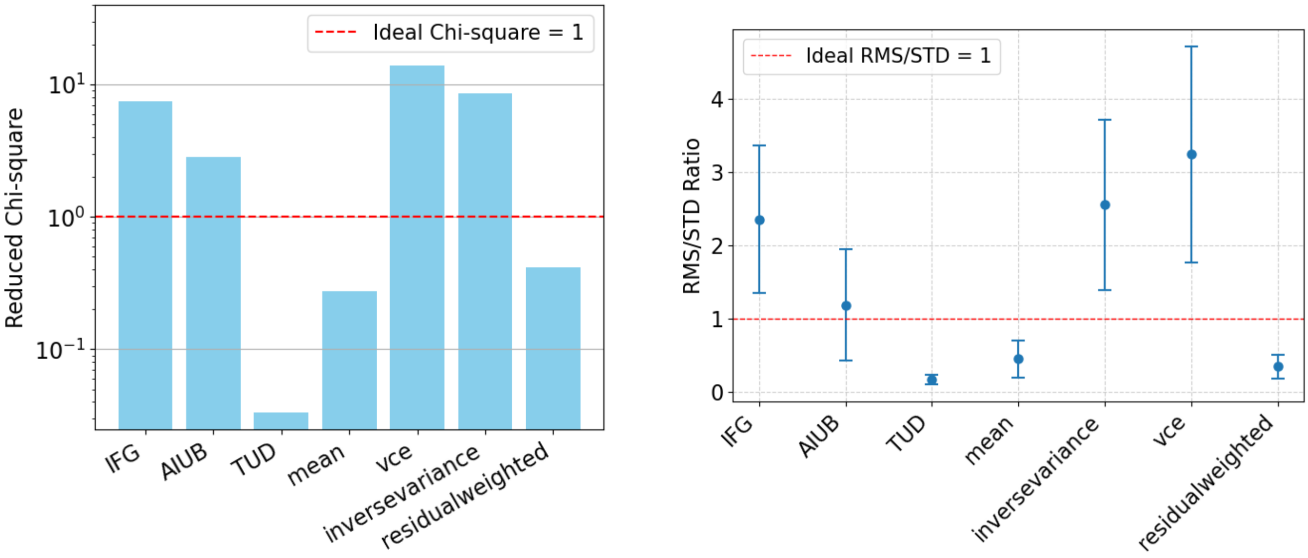
A reduced χ^2 greater than one indicates systematic overconfidence in the uncertainty of the orbit, as the residuals are consistently greater than the STD. Conversely, a value lower than one indicates systematic under-confidence. An ideal ratio of one means that the reported uncertainty maps on to the residuals. The ratio of residual RMS to the STD provides us an additional metric to judge the correctness of the uncertainty information, which implicitly includes bias, whereas the reduced χ^2 does not.

To assess how this calibration behaves over time, we use the coefficient of variation of the daily RMS/STD ratio. A lower CV indicates that the relative calibration between the orbit uncertainties and the SLR residuals remains more consistent throughout the month. However, stability does not imply correctness: an orbit may report highly stable yet physically unrealistic uncertainties. Therefore, the RMS/STD ratio and its CV together allow us to assess both the accuracy and stability of the reported uncertainty information. A description of how these metrics are computed is available in subsection 4.6.1.

Table 6.3: Summary of SLR variance validation metrics for each method, including the residual RMS, the mean of the residuals, the mean LOS standard deviation, ratio of the RMS to the LOS STD, the CV of the daily RMS/STD ratio and the reduced-Chi squared ($\chi^2_{(\text{red})}$).

Method	**SLR RMS (m)**	**Mean Residual (m)**	**Mean LOS STD (m)**	**RMS/STD**	**STD(RMS/STD)**	**CV**	$\chi^2_{(\text{red})}$
IFG	0.0110	0.0042	0.0043	2.55	1.01	0.429	7.51
AIUB	0.0138	0.0037	0.0090	1.54	0.76	0.643	2.82
TUD	0.0099	0.0038	0.0591	0.17	0.06	0.394	0.03
Mean	0.0097	0.0038	0.0205	0.47	0.26	0.567	0.28
VCE	0.0104	0.0040	0.0030	3.49	1.47	0.454	13.97
Inverse-variance	0.0103	0.0041	0.0037	2.74	1.16	0.457	8.54
Residual-weighting	0.0088	0.0038	0.0252	0.35	0.16	0.465	0.41

To complement these metrics, the ratio of residual RMS to mean orbit LOS STD is presented in the rightmost plot of Figure 6.12 along with STD of the daily ratio error-bars indicating temporal variability. The ideal ratio of 1 is indicated by the red dashed line.



(a) Reduced Chi-square ratios.

(b) RMS/STD.

Figure 6.12: Left: Bar plot of the reduced χ^2 of each input combination method. This give an indication of the physical accuracy of the noise in terms of magnitude. Right: Residual RMS to orbit STD ratio for each orbit, including the STD of the daily ratio to show the temporal stability.

Input orbit results

IFG is greatly overconfident; the RMS of the residuals is 2.55 times the mean reported STD, and the reduced χ^2 is 7.51, as shown in Figure 6.12. Together this indicates that the reported uncertainty underestimates the true residual scatter. The CV of the daily RMS/STD ratio is 0.43, indicating moderate fluctuation in the consistency of the uncertainty calibration over time.

This is also apparent in Figure 6.13, where the SLR residuals are plotted over time, including the orbit STD projected along the station-satellite line of sight vector shown as daily-averaged envelopes. The reported uncertainty envelopes are narrower than the spread of the residuals.

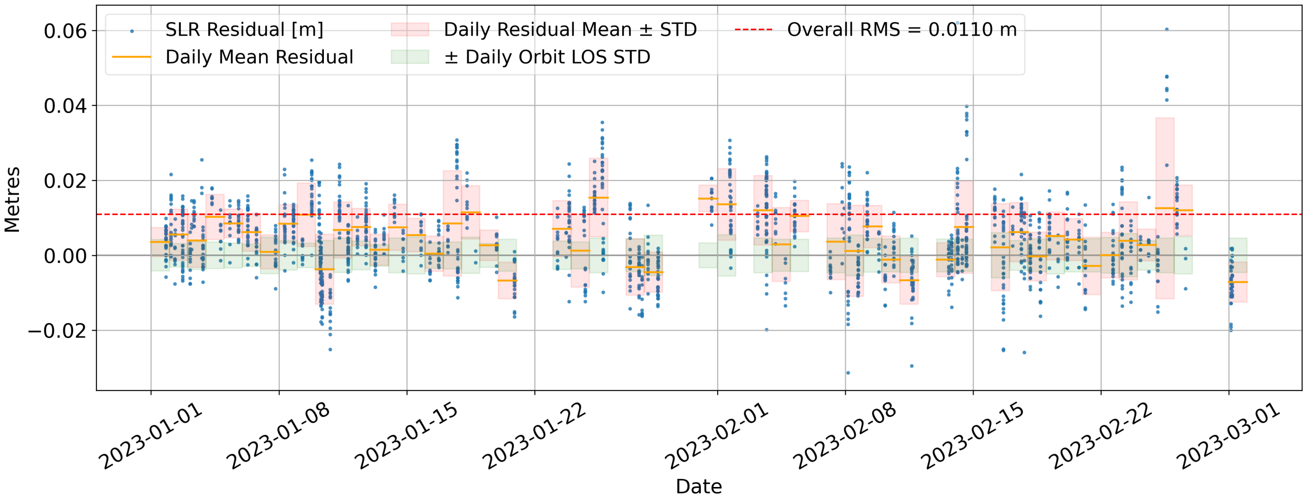


Figure 6.13: SLR residuals, orbit STD projected along the line of sight vector envelope and the residual mean and STD envelope for IFG. The uncertainty envelopes represent daily average uncertainty.

The AIUB input orbit is also overconfident, with a reduced χ^2 of 2.82 and an RMS/STD ratio of 1.54. The CV of the daily RMS/STD is 0.64, the highest of the three input orbits, reflecting greater day-to-day fluctuations in uncertainty calibration despite the lower overconfidence compared to IFG.

This is supported by Figure 6.14, where the reported STD envelope overlaps more frequently with the residual STD, indicating improved calibration, albeit with greater temporal variation.

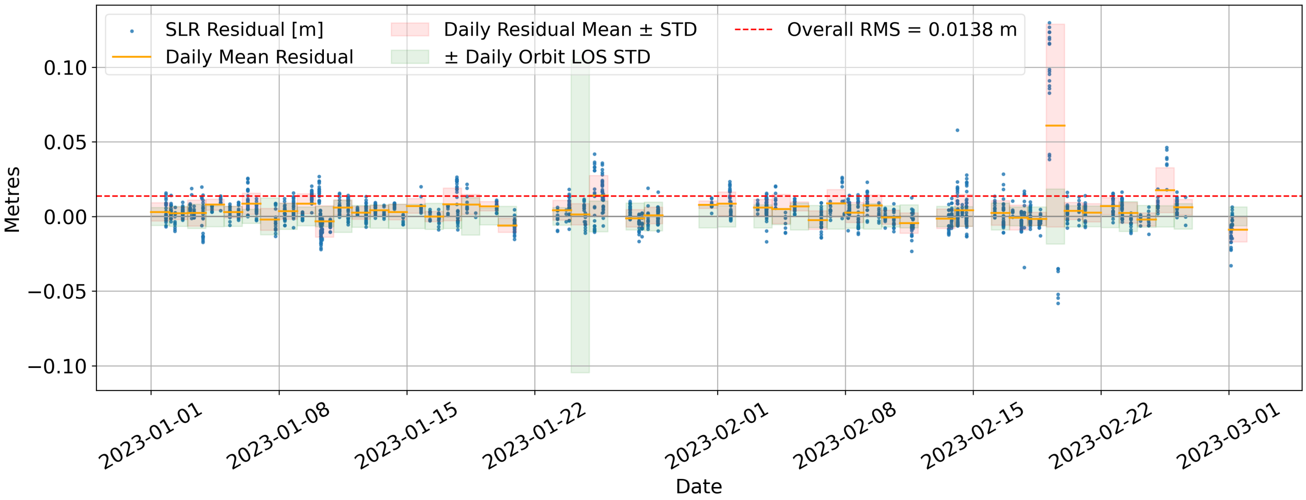


Figure 6.14: SLR residuals, orbit STD projected along the line of sight vector envelope and the residual mean and STD envelope for AIUB. The uncertainty envelopes represent daily average uncertainty.

The TUD input orbit is systematically under-confident. It reports significantly inflated uncertainty compared to the residuals, with a reduced χ^2 ratio of 0.03 and an RMS/STD ratio of only 0.17. Despite this poor calibration, it demonstrates the most stable temporal behaviour, with a CV of 0.39, the lowest among all orbits. This suggests that although the reported uncertainty magnitude is unrealistic, it is more consistent over time.

This is supported by Figure 6.15, where the reported STD envelope almost fully contains the residuals and their spread, with minimal variation throughout the time series.

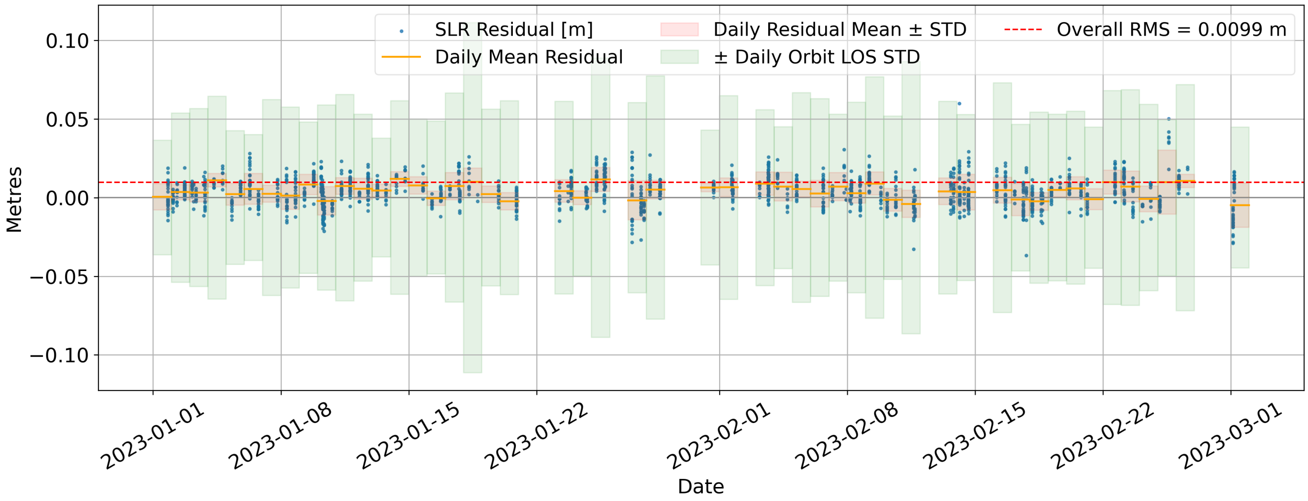


Figure 6.15: SLR residuals, orbit STD projected along the line of sight vector envelope and the residual mean and STD envelope for TUD. The uncertainty envelopes represent daily average uncertainty.

Combined orbit analysis

Now we assess how the uncertainty of the input orbits — demonstrated to be inconsistent — is propagated through the combination.

The arithmetic mean is under-confident, but less so than the TUD, with a reduced χ^2 of 0.28 and an RMS/STD ratio of 0.47. Its CV is 0.57, indicating moderate variation in how well the uncertainty calibration matches the residuals over time. This relatively stable performance reflects the averaging effect of the mean combination: it gives no preference to any set of input uncertainty values, and thus tends to smooth the temporally variable uncertainty of IFG and AIUB using the flatter uncertainty of TUD. Figure 6.16 shows that the reported STD envelope mostly encompasses the residual STD envelope, as expected of under-confidence. On 19 February, a pronounced spike in the mean residual of 0.02 m is observed, aligning with a peak on the same day in the AIUB orbit (Figure 6.14), highlighting the sensitivity of the mean combination to outliers.

The arithmetic mean is under-confident, though less so than the TUD, with a reduced χ^2 of 0.28 and an RMS/STD ratio of 0.47. However, it exhibits relatively poor temporal stability, with a CV of 0.57; the highest among the combination methods. This suggests that the quality of the uncertainty calibration fluctuates more significantly over time than for the other approaches. This instability arises because the mean treats all inputs equally, regardless of the reliability or consistency of their uncertainty information. As such, the temporally varying behaviour of IFG and AIUB is not effectively damped and instead contaminates the combined uncertainty. Figure 6.16 shows that the reported STD envelope mostly encompasses the residual STD envelope, consistent with under-confidence. On 19 February, a pronounced spike in the mean residual of 0.02 m is observed, aligning with the peak on the same day in the AIUB orbit (Figure 6.14), illustrating the method's sensitivity to outliers in the inputs.

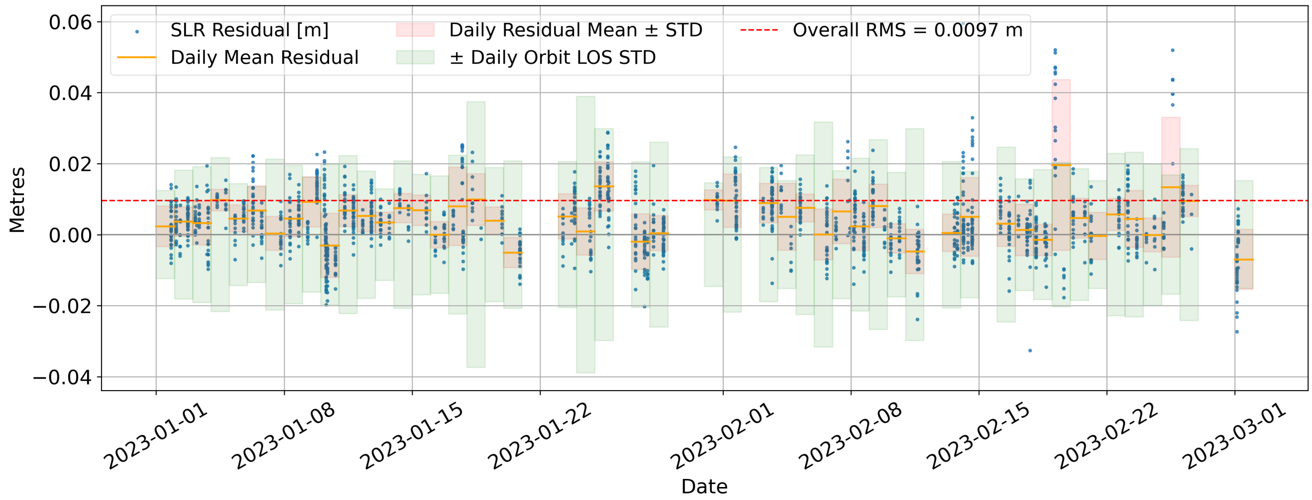


Figure 6.16: SLR residuals, orbit STD projected along the line of sight vector envelope and the residual mean and STD envelope for the Mean combination. The uncertainty envelopes represent daily average uncertainty..

The inverse-variance and VCE methods both exhibit strong overconfidence, with reduced χ^2 values of 8.54 and 13.97, and RMS/STD ratios of 2.74 and 3.49 respectively. Their CVs are 0.46 and 0.45, showing slightly better temporal consistency than the mean but still reflecting fluctuations in calibration. As introduced in section 4.2, these methods propagate uncertainty using the harmonic mean of inverse variances, which heavily penalises high uncertainties. Since IFG and AIUB report unrealistically small uncertainty, the combination — especially VCE — is biased towards them, further reducing the reported uncertainty. Consequently, the output uncertainty is smaller than any of the individual inputs, resulting in extreme overconfidence. This behaviour is shown clearly in Figures 6.17 and 6.18, where the reported STD envelopes are significantly smaller than the spread of the residuals.

Figures 6.17 and 6.18 show this systematic overconfidence, where the reported STD envelope is significantly smaller than the residuals over time.

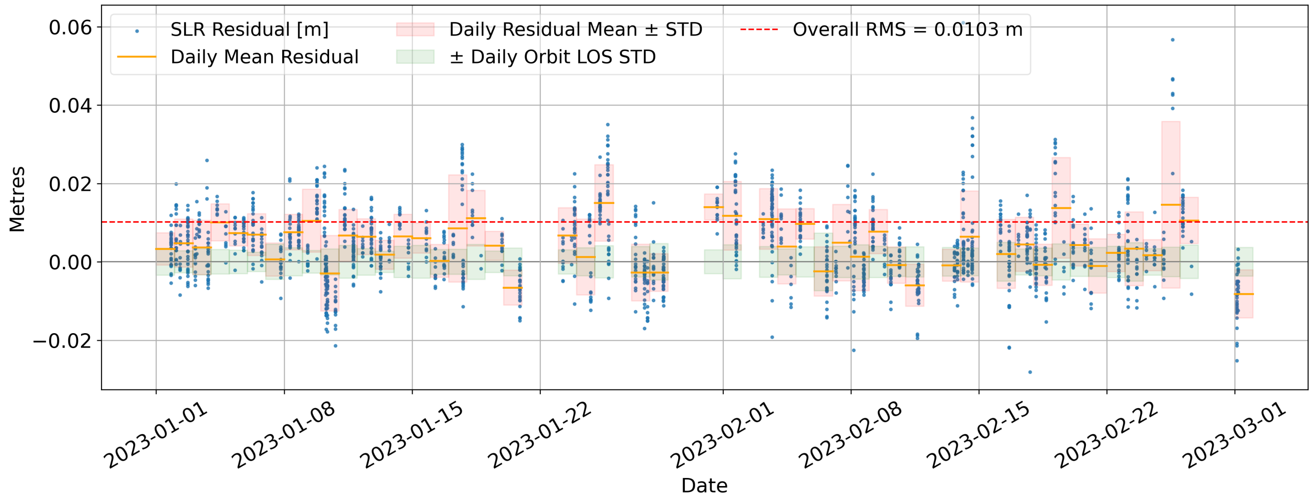


Figure 6.17: SLR residuals, orbit STD projected along the line of sight vector envelope and the residual mean and STD envelope for the inverse-variance combination. The uncertainty envelopes represent daily average uncertainty.

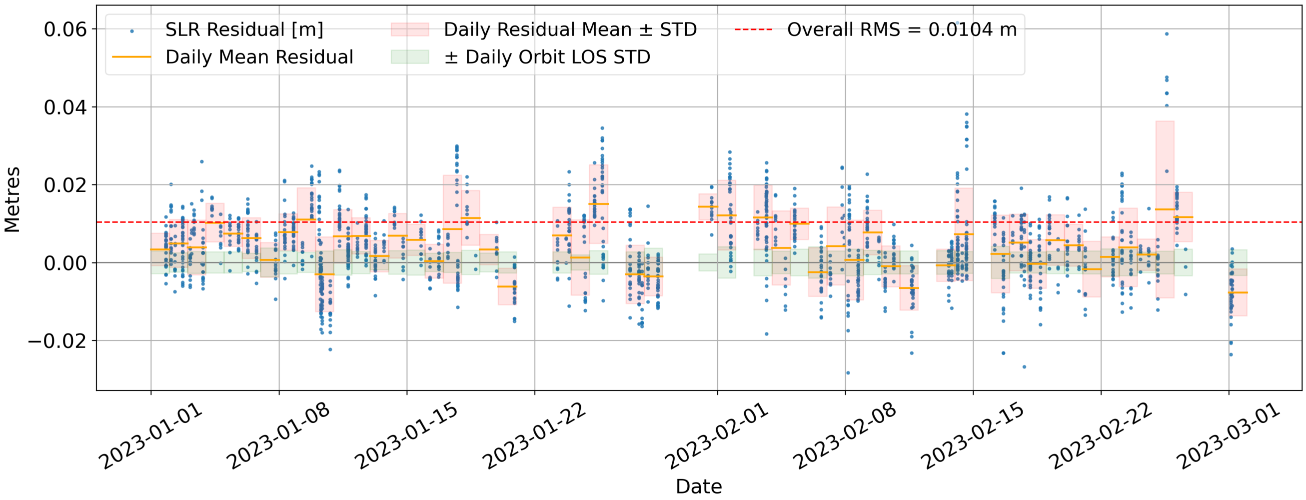


Figure 6.18: SLR residuals, orbit STD projected along the line of sight vector envelope and the residual mean and STD envelope for the VCE combination. The uncertainty envelopes represent daily average uncertainty.

Finally, the residual-weighted method reports a consistent under-confidence, with a reduced χ^2 of 0.41 and an RMS/STD ratio of 0.35. its CV is 0.47, reflecting moderate consistency in the reported uncertainty. Since this method entirely bypasses the input orbit uncertainty — instead using the empirical residuals against a reference RDO — its uncertainty behaviour is more consistent and less biased by poor input calibration. Figure 6.19 shows that the uncertainty envelope generally exceeds the residuals, except during periods of bias.

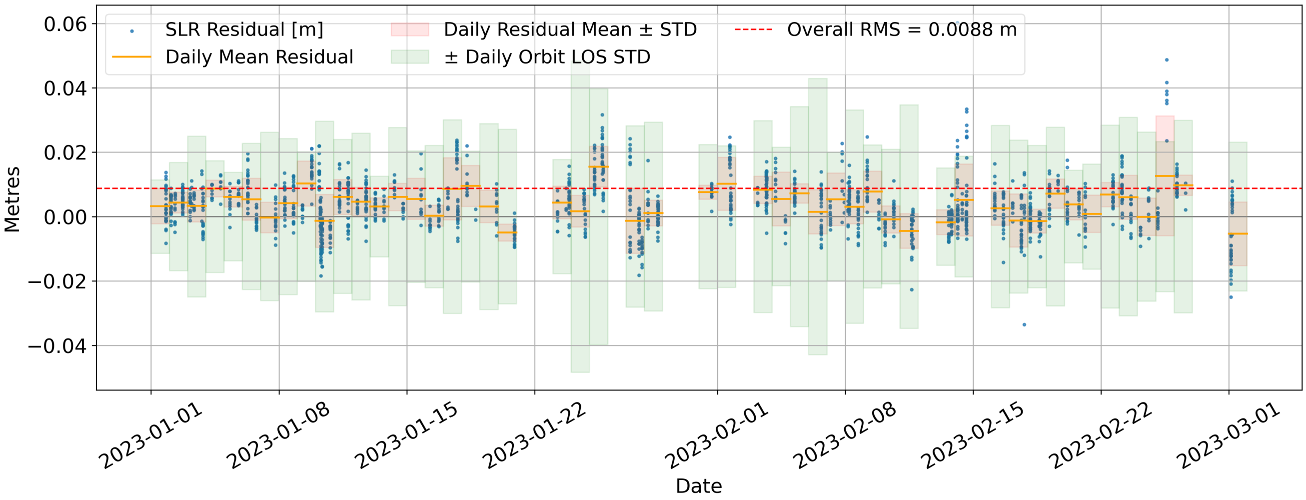


Figure 6.19: SLR residuals, orbit STD projected along the line of sight vector envelope and the residual mean and STD envelope for the residual-weighted combination. The uncertainty envelopes represent daily average uncertainty.

To conclude this validation, the uncertainty in the combined orbits is within acceptable ranges compared to the input orbits, indicating that the combination methods propagate uncertainty correctly. The key observation is that the input orbits do not have particularly well-scaled uncertainty information, which leads to all methods similarly reporting imperfect uncertainty. Furthermore, methods which use the variance information in the weighted combination – inverse-variance and

VCE – suffer from the inconsistent magnitude of the input orbit reported STD's with respect to the residuals, which vary by a factor 10. This inconsistency results in these methods over-fitting to the excessively confident input orbits – IFG and AIUB – leading to worse-performing combinations with significantly overconfident uncertainty. The mean and residual-weighted methods somewhat mitigate the poorly conditioned uncertainty of the input orbits by virtue of them not including this information in the combination. As such, their χ^2 values are closer to 1. However, this should not be viewed as a general benefit of these methods, rather it is a product of the particular set of input orbits. In his case, we have one extremely under-confident input (TUD, with a χ^2 of 0.03) and two overconfident inputs (AIUB, IFG, with χ^2's of 2.82 and 7.51 respectively); methods which combine without using the variance information simply report uncertainties somewhere between the three inputs, which in our case happens to be closer to the ideal χ^2 of 1. These results suggest that the combination methods would benefit greatly from input orbit noise scaling to consolidate the magnitudes of the uncertainty.

6.2.3 Scaling of input uncertainty based on reduced Chi-squared

In order to observe the effect of scaling the input uncertainty to improve its consistency with the observed residuals and answer sub-question 3, we determine scaling factors as the square root of the reduced χ^2 values provided in Table 6.2. These factors correct for discrepancies between the reported uncertainties and the actual orbit performance as assessed via residuals. Specifically, if the observed residuals are larger (or smaller) than expected based on the reported STD's, the reduced χ^2 value will exceed (or fall below) 1, indicating an under- (or over-) estimation of the uncertainty.

To address this, the reported STD's $\sigma_{ij,t}$ for each coordinate j of input orbit i at epoch t are scaled by a factor $s_i = \sqrt{\chi^2_{\nu,i}}$, such that the updated uncertainties better reflect the empirical error characteristics:

$$\sigma_{ij,t}^{(\text{scaled})} = s_i \cdot \sigma_{ij,t} = \sqrt{\chi^2_{\nu,i}} \cdot \sigma_{ij,t}. \tag{6.2}$$

This adjustment ensures that the scaled STD's yield reduced χ^2 values closer to 1, thereby improving the consistency between reported uncertainty and observed residuals.

Under the assumption that the structure of the input covariances is valid, but that their absolute scale may be poorly estimated, we apply the square of the same factor to the covariance terms:

$$\sigma_{jk,t}^{(\text{scaled})} = s_i^2 \cdot \sigma_{jk,t}. \tag{6.3}$$

We examine the effect of applying constant scaling factors over the entire two-month period. These factors, derived for each input solution based on the average residual statistics, are as follows: IFG — 2.74, AIUB — 1.68, and TUD — 0.17. These are used to scale both STD's and covariances prior to orbit combination.

Result of rescaling by reduced χ^2

The resulting χ^2 plot is presented for the re-scaled input orbits as well as the re-processed combined orbits produced using them in Figure 6.20.

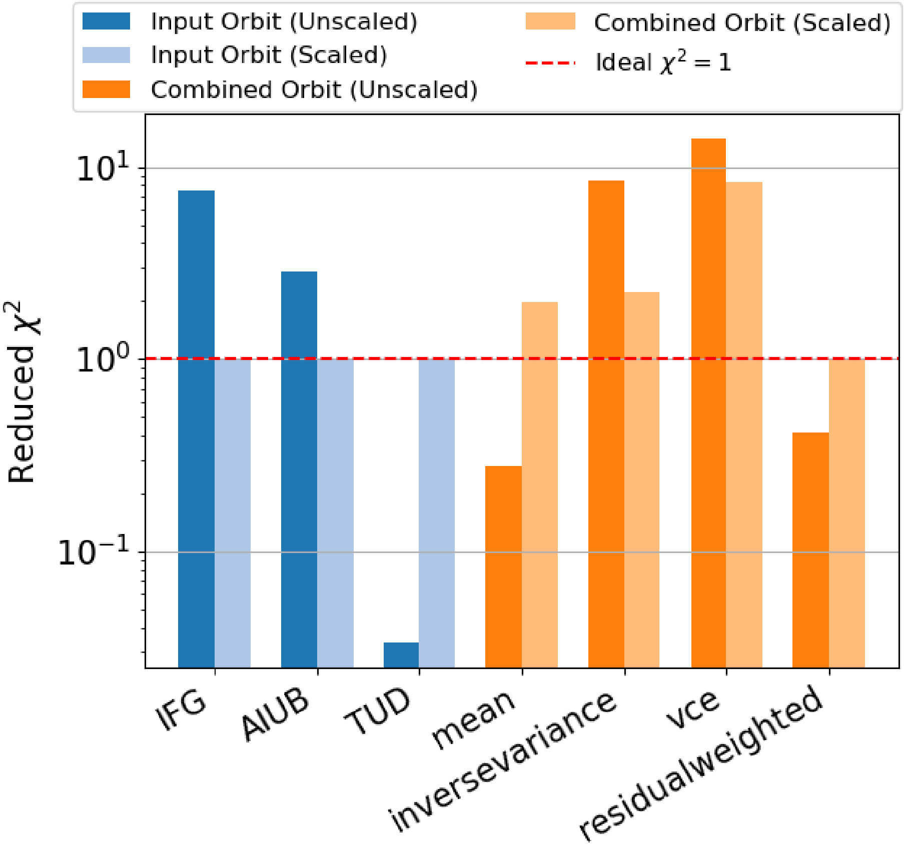


Figure 6.20: Bar plot of the reduced χ^2 of each input combination method. This give an indication of the physical accuracy of the noise in terms of magnitude.

By applying these scaling factors to the inputs, their reduced χ^2 values are exactly one.

By re-scaling the uncertainty of the input orbit's we achieve much better agreement between their reported residuals and the resulting STD projected along the LOS vector, indicating improved statistical consistency. In principle, this should improve propagation of uncertainty, and methods which use the uncertainty information in the combination. Figure 6.20 shows that the reduced χ^2 changes significantly for each combined orbit. Numerical results are provided in Table 6.4.

Table 6.4: Validation metrics for orbits with rescaled input uncertainty. Input orbits (IFG, AIUB, TUD) were individually scaled to achieve $\chi^2 \approx 1$, and combined orbits generated using these scaled inputs, evaluated with reduced χ^2, SLR residual RMS, mean residual, mean line-of-sight STD, the RMS-to-STD ratio, its STD, and CV.

Method	$\chi^2_{(\text{red})}$	**SLR RMS (m)**	**Mean Residual (m)**	**Mean STD LOS (m)**	**RMS/STD**	**STD(RMS/STD)**	**CV**
IFG	1.0000	0.0110	0.0042	0.0118	0.93	1.0085	0.4286
AIUB	1.0000	0.0138	0.0037	0.0151	0.92	0.4533	0.6428
TUD	1.0000	0.0099	0.0038	0.0108	0.91	0.3545	0.3941
Mean	1.9723	0.0097	0.0038	0.0077	1.26	0.5742	0.4668
Invee Variance	2.2504	0.0095	0.0040	0.0065	1.46	0.6062	0.4390
VCE	8.3082	0.0098	0.0037	0.0045	2.19	1.1750	0.4688
Residual Weighted	1.0069	0.0088	0.0038	0.0101	0.88	0.3979	0.4660

Reduced χ^2 increases from 0.28 (under-confident) to 1.972 (overconfident) for the mean. This occurs because rescaling the input orbits increases the reported uncertainties of two of the three solutions and reduces their spread, resulting in a tighter mean uncertainty and consequently an overly confident propagation. its CV falls from 0.57 to 0.47, indicating moderately improved temporal consistency.

A significant improvement is observed in the inverse-variance method, where χ^2 drops from 8.54 to 2.25, bringing the reported uncertainty closer to the observed residuals. However, it still remains overconfident. This is because inverse-variance weighting assumes that input uncertainties are fully reliable once rescaled, yet any residual correlation or un-modelled bias between the input orbits is ignored. Correlations in errors observed across the input orbits are expected, as they each broadly use the same input GNSS data for estimation. In addition, its CV improves from 0.46 to 0.44, showing a minor improvement in temporal consistency.

The VCE method improves markedly as well, with χ^2 reducing from 13.97 to 8.31. However, it continues to significantly underestimate uncertainty. This persistent overconfidence likely arises because if one input aligns well with the combination due to chance or shared bias, it can be overweighted, leading to unrealistically small reported variances. In this case, the CV rises from 0.45 to 0.47; a slight worsening of temporal consistency.

By contrast, the residual-weighted method based on the reduced dynamic reference orbit achieves excellent statistical consistency after rescaling. Its reduced χ^2 improves from 0.41 to 1.01, its RMS-to-STD ratio likewise approaches 1, and its temporal stability remains the same with a CV of 0.47. Unlike the other methods, it does not depend on reported variances, but instead selects inputs based on their actual residuals with respect to the reference orbit. Once input uncertainties are correctly scaled, the residual-weighted method assumes the RDO to be truth and at each epoch determines weights from it which more correctly capture the full range of errors, including systemic bias.

After rescaling the input uncertainties, both the inverse-variance and VCE methods show modest but meaningful improvements in orbit position accuracy, as reflected in their SLR residual RMS values. The inverse-variance RMS decreases from 0.0103,m to 0.0095,m (a 7.8% improvement), while VCE improves from 0.0104m to 0.0098m (5.8%). Notably, the inverse-variance method now slightly outperforms the unscaled mean combination (0.0097m) in terms of accuracy, whereas it was previously worse. VCE, while still more overconfident in uncertainty reporting, also surpasses the unscaled mean in RMS after rescaling. However, both methods continue to be less accurate than the residual-weighted method (0.0088m), which remains the most precise combined solution post-scaling for this particular time period.

To conclude, this technique of noise-rescaling has the potential to improve the performance of the inverse-variance and VCE methods significantly, producing new best-in-class orbit combinations.

Chapter 7

Conclusion and Recommendations

In this final section, the thesis is summarised, the results of chapter 6 are concluded and the research questions presented in section 3.5 are answered. Finally, a brief discussion of the contribution of the work and future recommendations are presented in section 7.2.

7.1 Summary & conclusions

In the introduction of this thesis, we presented the goal of this research; to combine Swarm kinematic orbits such that the resulting solution is superior. In chapter 2, the concept of kinematic orbits – satellite orbit solutions which depend only upon the geometry of the orbit, leveraging tracking data obtained from GNSS constellations – are introduced, as well as key background knowledge required to understand the pursuit of orbit combinations. Our interest in the improvement these orbits arises from the fact that they are used to estimate the time-varying gravity field of Earth as an additional source of high-accuracy data, complementary to dedicated gravity field missions such as GRACE [66]. Improving them results in improvements in the gravity field solutions.

In this thesis, we aim to provide an incremental improvement in the performance of the orbits used for gravity field estimation from Swarm kinematic orbits. In service of this, several combination techniques are devised and presented in chapter 4 fitting into two distinct categories; combination techniques dependent solely on the input orbit data including the position and covariance information, and techniques which leverage additional sources of reference data to further inform the combination. The latter techniques leverage the known greater accuracy and robustness of RDO's [76] – orbit solutions determined using a combination of dynamic models with GNSS tracking data – to inform the combination and 'guide' it closer to the truth.

To build upon the theory and facilitate the investigation, software has been implemented to support the application of orbit-combination to Swarm satellite data, as well as to verify and validate results. In chapter 5 the implementation of the methodology is verified, sensitivity analyses are conducted in order to ensure proper functioning prior to producing results, and a final selection of high-performing combination techniques is made; these being combination by the arithmetic mean, inverse-variance weighting, iterative variance-component estimation, weighting by residuals with respect to a reduced-dynamic reference orbit, optimisation with the Nelder-Mead local optimisation algorithm and CMAES global optimisation algorithm.

Kinematic input orbit data is retrieved from three analysis centres; IFG, AIUB and the TUD, with the latter additionally providing the RDO used as reference. Input orbit data is consolidated to consistent 1-second frequency epochs using Lagrange interpolation in order to ensure compatibility. Consolidated input data is then pre-screened for outliers with a 0.3 m ceiling of 3D orbit residuals with respect to the reduced-dynamic reference orbit. Pre-screening of the input data runs the risk of losing significant epochs of data, however this is mitigated by orbit combination as input orbit solutions tend to have outliers at different epochs due to their unique processing strategies, meaning that overall the combination retains a greater density of data than any input.

With this processed input data, the combination techniques are applied to two case study years – 2022, a lower solar-activity year, and 2023, a higher solar-activity year – for the Swarm A satellite; Additional data for Swarm B and C is present in Appendix D. These years are selected to observe the stability of the methods across periods under different conditions. SLR observations of the Swarm satellites computed using the GHOST software are used to perform a comprehensive assessment of the performance of the combined solutions through the computation of RMS errors from the residuals. The results for these years are presented in chapter 6.

With the foundations set, we conclusively answer our primary research question: **What improvement in the quality of Swarm kinematic orbits can be achieved through the combination of unique solutions?**.

Results across 2022 and 2023 show improvement of the combined orbits with respect to the input orbits in terms of SLR residual RMS in the order of tens of centimetres. In particular, the combination processes successfully filter outliers from the input data, which is especially valuable in noisier years like 2023 where AIUB's input orbit has significant outliers leading to an SLR RMSE in the order of meters. These outliers are successfully removed from the combination by the pre-screening process.

In 2022, the arithmetic mean achieves the best RMS among all methods, at 0.01936 m, which is only 0.87 mm greater than the reduced-dynamic reference orbit provided by ESA (RMS: 0.01849 m). In 2023, the same method is narrowly outperformed by the optimiser-based combinations (Nelder-Mead and CMAES), which yield an RMS of 0.03937 m, only 0.04 mm better than the arithmetic mean. The RMS of the difference with respect to the RDO orbit in 2023 (RMS: 0.02175 m) is 17.66 mm, still representing a substantial gain over the worst input (AIUB, RMS: 6.51 m), and a significant reduction from the best input (IFG, RMS: 0.569 m).

Importantly, these improvements are achieved without loss of temporal coverage. The combination framework retains all epochs where at least one input solution is present. In 2022, this leads to an epoch loss rate of 3.6%, matching the RDO in coverage. In 2023, we observe an epoch loss of 2.7%, equivalent to the best-performing input (IFG), despite filtering significant outliers.

Surprisingly, the inverse variance and VCE methods – techniques which employ the a-priori covariance information provided with the input orbits – perform worse than the others. This shows that the input covariances are inconsistently scaled. The residual-weighting method, which uses residuals with respect to the RDO to assign combination weights, also underperforms. This is consistent with the observation that the combined orbits are not significantly less accurate than the reference orbit itself, meaning the fundamental assumption behind this method — that the reference is far more accurate — does not hold, especially after pre-screening.

We therefore conclude that improvements of up to 0.55 m in the SLR residual RMSE with respect to the best input orbit can be achieved through orbit combination, with reductions from 0.569 m (best individual input, IFG, 2023) to 0.03937 m (best combined solution). In 2022, the best combined orbit differs from the reduced-dynamic reference by just 0.87 mm in SLR RMSE, demonstrating near-equivalence in quality. Even in the noisier year of 2023, the combined SLR RMSE remains within 17.66 mm of the RDO while achieving a 93% reduction relative to the best input solution. These results confirm that orbit combination enables the generation of robust, consistent kinematic orbits that not only outperform all individual inputs but can approach the performance of a reduced-dynamic solution. Among the tested methods, the arithmetic mean offers the most favourable trade-off between accuracy and simplicity. The negligible improvements from optimisation techniques — only 0.04 mm better than the mean in 2023 — do not justify their computational cost under current input conditions. Overall, we recommend the application of orbit combination to Swarm kinematic orbits, as the improvement is meaningful and worthwhile.

Moving on to the sub-research questions defined, we beginning with: **Sub-question 1: To what extent do combination methods which utilise a reference RDO risk introducing bias or over-fitting?**. To assess this, we compute the residuals of the input and combined orbits with respect to the RDO and the resulting RMSE. We compare how each orbit performs in terms of this RMSE and the RMS of the SLR residuals; if a method reports closer agreement with the RDO, but does not show similar improvement in the SLR residuals, we take this as evidence of over-fitting. There are no signs of over-fitting in the optimised methods – Nelder-Mead and CMAES –, supporting the efficacy of the mitigation techniques described in chapter 4. However, there is strong evidence of it occurring with the residual-weighted method, with a 40% decrease in RMSE with respect to the next-closest method. Over-fitting is prevalent for the residual weighted method, but not present at all in the optimised methods.

We may now answer the following question – **Sub-question 2: Can the uncertainty information of independent kinematic orbits effectively be propagated into a combined solution?** – may now be addressed. We perform an uncertainty analysis of the combined orbit solutions which have uncertainty propagation implemented; the arithmetic mean, inverse variance, VCE and residual-weighted techniques. Given that we can determine the satellite-station LOS vector on which the SLR residuals lie using tracking data and the known station positions, we compute the STD of the residuals in LOS direction. We then project the uncertainty reported by the combined orbit onto this vector. The result is two uncertainty values; one empirical, another reported. If the uncertainty is properly scaled in terms of magnitude and correct in its distribution, we should observe similar values for both across time.

The input orbit uncertainty information is not physically realistic in terms of magnitude, with reduced χ^2 values of 7.51, 2.82 and 0.03 for the IFG, AIUB and TUD respectively; the ideal ratio indicating physically realistic uncertainty information is 1, with larger values overconfident and smaller under-confident. As a result, the combination methods propagate this physical unrealism into the combined orbit. This explains the aforementioned worse performance of the inverse-variance and VCE methods, which report worse RMS compared to the other combination methods and overconfident reduced χ^2's of 8.54 and 13.97 respectively as result. Conversely, the arithmetic mean reports under-confidence with a reduced χ^2 of 0.28 due to the under-confidence of the TUD orbit, which leads to this uncertainty information dominating the uncertainty propagation for this method. The residual-weighted method reports uncertainties closer to physical reality while remaining under-confident with a reduced χ^2 of 0.41 as a result of its weights being determined by the RDO rather than the inconsistent input uncertainty information.

Regardless, we find that the reported uncertainties are internally consistent; that is, while the magnitudes are not physically realistic, the uncertainty is correctly reported greater when the observed residuals are greater, and lower when they are lower. This is the most important aspect of the uncertainty propagation, as it is this internal distribution which is used in the gravity field estimation process in order to weight the observations by reported quality. Overall, we conclude that the uncertainty information is correctly propagated through the combination, however the resulting uncertainty remains physically unrealistic in terms of magnitude. Despite this, the distribution of uncertainty over time maps on to the reported residuals, suggesting that the orbit uncertainty will be useful for the purpose of observation weighting in gravity field estimation.

Having identified the inconsistent scaling of the input orbits, we may answer the final sub-question: **Sub-question 3: Can we utilise independent sources of validation to re-scale the uncertainty of the input orbits in order to improve the combination?**. To test this, we use the STD of the residuals and the projected uncertainties on the SLR station-satellite LOS vector of the input orbits to compute reduced χ^2 scaling factors which are then used to rescale their covariance. In this thesis, we examined the effect of constant scaling factors across a two month period from January to February of 2023. The noise-rescaled input orbits report uncertainties in agreement with the SLR residuals.

When we repeat the orbit combination using these rescaled inputs, both the inverse-variance and VCE methods improve sufficiently to outperform the arithmetic mean in terms of SLR residual RMSE. The inverse-variance method becomes the best-performing uncertainty-weighted combination, achieving an RMSE of 0.0095 m compared to 0.0097 m for the arithmetic mean. This reinforces the idea that the input orbits possess internally consistent noise distributions but differ in their reported magnitudes. Uncertainty propagation also improves for both methods: the reduced χ^2 for inverse-variance drops from 8.54 to 2.25, and for VCE from 13.97 to 8.31, indicating a substantial reduction in overconfidence. The RMS-to-STD ratio for inverse-variance also improves from 2.74 to 1.46, approaching consistency with the SLR residuals.

In contrast, the arithmetic mean becomes overconfident after rescaling, with its reduced χ^2 increasing from 0.28 to 1.97 and its RMS-to-STD ratio rising from 0.47 to 1.26, reflecting the uniform scaling of the inputs. Meanwhile, the residual-weighted method achieves the best overall calibration, with a reduced χ^2 of 1.01 and RMS/STD ratio of 0.88, along with the lowest RMSE of 0.0088 m.

Overall, rescaling the input orbit noise using the SLR-derived uncertainties leads to significant improvements both in the realism of reported uncertainties and in the performance of methods that depend on them.

To conclude this thesis and confirm the answer to our research questions, we state that the combination of Swarm kinematic orbits improves their quality in terms of accuracy and data density, and the uncertainty information propagated through the combination is valid and useful for downstream gravity field estimation, and the uncertainty information is improved through through input uncertainty rescaling.

7.2 Recommendations

The output of this research is a set of combined orbits for the years of 2022 and 2023 for the Swarm A, B and C satellites. There are many recommendations for future lines of inquiry. First and foremost, the estimation of a gravity field model from the combined orbits is of great interest so we may determine if the improvements in these orbits results in any improvements in the them. Of special interest is the estimation of gravity field models using the residual-weighted and optimisation (Nelder-Mead, CMAES) methods in order to assess whether they bias the field towards the a-priori gravity field present implicitly in the RDO reference.

Furthermore, the application of noise-rescaling using the reduced Chi-squared ratios computed with SLR residuals should be applied to all input data in totality, as this technique is demonstrated to improve the performance of the inverse variance and VCE combined solutions and the uncertainty propagation; unfortunately, due to time limitations this was not applied for all input data used in this thesis. Additionally, it is recommended to apply this noise-rescaling technique to the output combined orbits too; ultimately, it simply serves to ensure that the reported uncertainty is more in agreement with reality. It is also recommended to explore additional noise-rescaling techniques, perhaps using weekly or even daily scaling factors rather than the single constant factor tested in this thesis.

Finally, the combination methods detailed in this thesis may also be applied to other satellite missions, and additional combination techniques this author missed should be explored and tested.

References

[1] Astropy Collaboration et al. "The Astropy Project: Sustaining and Growing a Community-oriented Open-source Project and the Latest Major Release (v5.0) of the Core Package". In: (Aug. 2022).

[2] Hermann Bähr, Zuheir Altamimi, and Bernhard Heck. "Variance Component Estimation for Combination of Terrestrial Reference Frames". In: ().

[3] G. Beutler et al. *Celestial Mechanics Approach: Theory and Applications*. New York: Springer, 2010.

[4] A. Bezděk et al. "Further Developments in the Decorrelated Acceleration Approach for Gravity Field Estimation". In: *Geophysical Journal International* 205 (2016), pp. 1665–1675. DOI: `10.1093/gji/ggw104`.

[5] A. Bezděk et al. "Gravity field estimation using the Decorrelated Acceleration Approach (DAA)". In: *Advances in Space Research* 53 (2014), pp. 412–429. DOI: `10.1016/j.asr.2013.11.031`.

[6] Aleš Bezděk et al. "Gravity field models from kinematic orbits of CHAMP, GRACE and GOCE satellites". In: *Advances in Space Research* 53.3 (2014), pp. 412–429.

[7] R. Biancale and A. Bode. "Mean Annual and Seasonal Atmospheric Tide Models Based on 3-hourly and 6-hourly ECMWF Surface Pressure Data". In: *Scientific Technical Report STR06/01, GFZ Potsdam* (2006).

[8] Francesco Biscani and Dario Izzo. "A parallel global multiobjective framework for optimization: pagmo". In: *Journal of Open Source Software* 5.53 (2020), p. 2338. DOI: `10.21105/joss.02338`. URL: `https://doi.org/10.21105/joss.02338`.

[9] Francesco Biscani, Dario Izzo, and Marcus Märtens. *esa/pagmo2: pagmo 2.7*. Version 2.7. 2018. DOI: `10.5281/zenodo.1217831`. URL: `https://doi.org/10.5281/zenodo.1217831`.

[10] S Böckmann et al. "International VLBI Service for Geodesy and Astrometry: Earth orientation parameter combination methodology and quality of the combined products". In: *Journal of Geophysical Research: Solid Earth* 115.B4 (2010).

[11] B. R. Bowman et al. "A New Empirical Thermospheric Density Model JB2008 Using New Solar and Geomagnetic Indices". In: *AIAA/AAS Astrodynamics Specialist Conference* 2 (2008), pp. 1976–1997.

[12] Janez Brest et al. "Self-adapting control parameters in differential evolution: A comparative study on numerical benchmark problems". In: *IEEE transactions on evolutionary computation* 10.6 (2006), pp. 646–657.

[13] L. Carrere et al. "FES 2014: A New Tidal Model on the Global Ocean with Enhanced Accuracy in Shallow Seas and in the Arctic Region". In: *Geophysical Research Abstracts* 17 (2015).

[14] R. Dach et al. *Bernese GNSS Software Version 5.2*. Astronomical Institute, University of Bern, Nov. 2015. ISBN: 978-3-906813-05-9. DOI: `10.7892/boris.72297`.

[15] R. Dach et al. *Bernese GNSS Software Version 5.3*. Bern, Switzerland: Astronomical Institute, University of Bern, 2015.

[16] Dominic Dirkx et al. "The open-source astrodynamics Tudatpy software–overview for planetary mission design and science analysis". In: *EPSC2022* EPSC2022-253 (2022).

[17] H. Dobslaw et al. "AOD1B RL06: A Full Revision of Atmospheric and Oceanic Background De-aliasing for Satellite Gravity Missions". In: *Geophysical Journal International* 211.1 (2017), pp. 263–269. DOI: `10.1093/gji/ggx302`.

[18] Howard Donovan. "Reduction of the minimum elevation angle for NASA satellite laser ranging tracking operations". In: *Lanham, april* (2001).

[19] João Encarnação and Pieter Visser. *TN-01: Standards and Background Models*. Apr. 2019. DOI: `10.13140/RG.2.2.12840.32006/1`.

[20] National Centers for Environmental Information. *NOAA National Centers for Environmental Information (NCEI)*. Online. Accessed: 2025-06-13. 2025. URL: `https://www.ncei.noaa.gov/`.

[21] European Space Agency. *Swarm Dissemination Server*. `https://swarm-diss.eo.esa.int/`. Accessed: 2025-06-08. 2025.

[22] European Space Agency (ESA). *Swarm Mission*. `https://earth.esa.int/eogateway/missions/swarm`. Accessed: 2025-06-08. European Space Agency, 2013.

[23] W. M. Folkner et al. *The Planetary and Lunar Ephemerides DE430 and DE431*. 2014.

[24] Filip Gałdyn and Krzysztof Sośnica. "Impact of the combination and replacement of SLR-based low-degree gravity field coefficients in GRACE solutions". In: *Progress in Earth and Planetary Science* 11 (Feb. 2024), p. 7. DOI: `10.1186/s40645-024-00608-z`.

[25] Tobias Glasmachers et al. "Exponential natural evolution strategies". In: *Proceedings of the 12th annual conference on Genetic and evolutionary computation*. 2010, pp. 393–400.

[26] Jiancheng Han et al. "A high-resolution time-variable terrestrial gravity field model of continental North China". In: *Communications Earth & Environment* 5.1 (2024), p. 44.

[27] Nikolaus Hansen. "The CMA evolution strategy: a comparing review". In: *Towards a new evolutionary computation: Advances in the estimation of distribution algorithms* (2006), pp. 75–102.

[28] Steve Hilla. *The Extended Standard Product 3 Orbit Format (SP3-d) 21 February 2016*. 2016.

[29] Jose van den IJssel et al. "Precise science orbits for the Swarm satellite constellation". In: *Advances in Space Research* 56.6 (2015), pp. 1042–1055.

[30] A Jäggi et al. "Swarm kinematic orbits and gravity fields from 18 months of GPS data". In: *Advances in Space Research* 57.1 (2016), pp. 218–233.

[31] A. Jäggi, G. Beutler, and U. Meyer. "AIUB GRACE-only static model, version 3 (AIUB-GRACE03S)". In: *Journal of Geodesy* 85 (2011), pp. 113–133. DOI: `10.1007/s00190-010-0416-1`.

[32] Adrian Jäggi et al. "International Combination Service for Time-Variable Gravity Fields (COST-G)". In: *Beyond 100: The Next Century in Geodesy*. Ed. by Jeffrey T. Freymueller and Laura Sánchez. Cham: Springer International Publishing, 2022, pp. 57–65. ISBN: 978-3-031-09857-4.

[33] Adrian Jäggi et al. "Kinematic and highly reduced-dynamic LEO orbit determination for gravity field estimation". In: *Dynamic Planet: Monitoring and Understanding a Dynamic Planet with Geodetic and Oceanographic Tools IAG Symposium Cairns, Australia 22–26 August, 2005*. Springer. 2007, pp. 354–361.

[34] Yoomin Jean, Ulrich Meyer, and Adrian Jäggi. "Combination of GRACE monthly gravity field solutions from different processing strategies". In: *Journal of Geodesy* 92 (2018), pp. 1313–1328.

[35] Philip Knocke, J Ries, and B Tapley. "Earth radiation pressure effects on satellites". In: *Astrodynamics conference*. 1988, p. 4292.

[36] A. Kvas et al. “GOCO06s – a satellite-only global gravity field model”. In: *Earth System Science Data* 13.1 (2021), pp. 99–118. DOI: `10.5194/essd-13-99-2021`. URL: `https://essd.copernicus.org/articles/13/99/2021/`.

[37] Andreas Kvas et al. “ITSG-Grace2018: Overview and Evaluation of a New GRACE-Only Gravity Field Time Series”. In: *Journal of Geophysical Research: Solid Earth* 124.8 (2019), pp. 9332–9344. DOI: `https://doi.org/10.1029/2019JB017415`. URL: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019JB017415`.

[38] Jet Propulsion Laboratory. *Spherical Harmonics: Representation of the Earth’s Gravity Field*. Online. Accessed: 2025-06-13. 2000. URL: `https://spsweb.fltops.jpl.nasa.gov/portaldataops/mpg/MPG_Docs/Source%20Docs/gravity-SphericalHarmonics.pdf`.

[39] JH Lieske et al. “Expressions for the precession quantities based upon the IAU 1976 system of astronomical constants”. In: *Astronomy and Astrophysics, vol. 58, no. 1-2, June 1977, p. 1-16.* 58 (1977), pp. 1–16.

[40] Florent Lyard et al. “Modelling the global ocean tides: modern insights from FES2004”. In: *Ocean dynamics* 56 (2006), pp. 394–415.

[41] T. Mayer-Gürr. “GOCO05S: A Satellite-Only Global Gravity Field Model”. In: *Geophysical Research Abstracts* 17 (2015).

[42] T. Mayer-Gürr, Annette Eicker, and Karl-Heinz Ilk. “ITG-Grace02s: A GRACE gravity field derived from short arcs of the satellites orbit”. In: *Proceedings of the 1st International Symposium of the International Gravity Field Service Gravity Field of the Earth* (Jan. 2007), pp. 193–198.

[43] T. Mayer-Gürr, E. Kurtenbach, and A. Eicker. “ITG-Grace2010s: The New GRACE Gravity Field Release Computed in Bonn”. In: *Eos Transactions AGU* 91.26 (2010), pp. 221–222.

[44] Oliver Montenbruck et al. “Reduced dynamic and kinematic precise orbit determination for the Swarm mission from 4 years of GPS tracking”. In: *GPS solutions* 22.3 (2018), p. 79.

[45] N/A. *Essentials Of Geophysics*. [Online; accessed 26-August-2024]. 2004. URL: `https://ocw.mit.edu/courses/12-201-essentials-of-geophysics-fall-2004/resources/ch2/`.

[46] John A Nelder and Roger Mead. “A simplex method for function minimization”. In: *The computer journal* 7.4 (1965), pp. 308–313.

[47] Nikolaos K. Pavlis et al. “The development and evaluation of the Earth Gravitational Model 2008 (EGM2008)”. In: *Journal of Geophysical Research: Solid Earth* 117.B4 (2012). DOI: `https://doi.org/10.1029/2011JB008916`. URL: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011JB008916`.

[48] Michael R Pearlman, John J Degnan, and John M Bosworth. “The international laser ranging service”. In: *Advances in space research* 30.2 (2002), pp. 135–143.

[49] G. Petit and B. Luzum. *IERS Conventions (2010)*. Frankfurt am Main: Verlag des Bundesamts für Kartographie und Geodäsie: IERS Technical Note No. 36, 2010.

[50] Gérard Petit, Brian Luzum, et al. *IERS conventions (2010)*. 2010.

[51] JM Picone et al. “NRLMSISE-00 empirical model of the atmosphere: Statistical comparisons and scientific issues”. In: *Journal of Geophysical Research: Space Physics* 107.A12 (2002), SIA–15.

[52] P Rebischung. “Reference Frame Committee Technical Report 2023”. In: *IGS Central Bureau* (), p. 237.

[53] J. C. Ries, S. V. Bettadpur, and R. J. Eanes. “GRACE Intermediate Field 48 (GIF48)”. In: *Technical Report* (2011).

[54] J.C. Ries et al. *The Development And Evaluation of The Global Gravity Model GGM05 - CSR-16-02.* May 2018.

[55] C. J. Rodriguez-Solano et al. “Impact of Albedo Radiation on GPS Satellites”. In: *Journal of Geodesy* 86 (2012), pp. 309–317. DOI: `10.1007/s00190-011-0517-4`.

[56] R. Savcenko and W. Bosch. “EOT11a - Empirical Ocean Tide Model from Multi-Mission Satellite Altimetry”. In: *Deutsches Geodätisches Forschungsinstitut (DGFI)* 89 (2012). DOI: `10.5676/DWD-POS2011`.

[57] Martin Schlüter, Jose A Egea, and Julio R Banga. “Extended ant colony optimization for non-convex mixed integer nonlinear programming”. In: *Computers & Operations Research* 36.7 (2009), pp. 2217–2229.

[58] Ralf Schmid et al. “Generation of a consistent absolute phase-center correction model for GPS receiver and satellite antennas”. In: *Journal of Geodesy* 81 (2007), pp. 781–798.

[59] PK Seidelmann. “1980 IAU theory of nutation: The final report of the IAU working group on nutation”. In: *Celestial mechanics* 27.1 (1982), pp. 79–106.

[60] Kun Shang et al. “GRACE time-variable gravity field recovery using an improved energy balance approach”. In: *Geophysical Supplements to the Monthly Notices of the Royal Astronomical Society* 203.3 (2015), pp. 1773–1786.

[61] Linus Shihora et al. “Non-Tidal Background Modeling for Satellite Gravimetry Based on Operational ECWMF and ERA5 Reanalysis Data: AOD1B RL07”. In: *Journal of Geophysical Research: Solid Earth* 127.8 (2022), e2022JB024360. DOI: `https://doi.org/10.1029/2022JB024360`. URL: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2022JB024360`.

[62] Rainer Storn and Kenneth Price. “Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces”. In: *Journal of global optimization* 11 (1997), pp. 341–359.

[63] Dariusz Strugarek et al. “Satellite laser ranging to GNSS-based Swarm orbits with handling of systematic errors”. In: *GPS solutions* 26.4 (2022), p. 104.

[64] Dražen Švehla and M Rothacher. “Kinematic and reduced-dynamic precise orbit determination of low earth orbiters”. In: *Advances in Geosciences* 1 (2003), pp. 47–56.

[65] João Teixeira da Encarnação et al. “Description of the multi-approach gravity field models from Swarm GPS data”. In: *Earth System Science Data* 12.2 (2020), pp. 1385–1417.

[66] João Teixeira da Encarnação et al. “Gravity field models derived from Swarm GPS data”. In: *Earth, Planets and Space* 68 (2016), pp. 1–15.

[67] Jose van den IJssel et al. “Precise science orbits for the Swarm satellite constellation”. In: *Advances in Space Research* 56.6 (2015), pp. 1042–1055. ISSN: 0273-1177. DOI: `https://doi.org/10.1016/j.asr.2015.06.002`.

[68] Martin Wermuth, Oliver Montenbruck, and Tom Van Helleputte. *GPS high precision orbit determination software tools (GHOST).* 2010.

[69] Wikipedia contributors. *Fourier series — Wikipedia, The Free Encyclopedia.* `https://en.wikipedia.org/w/index.php?title=Fourier_series&oldid=1295214720`. [Online; accessed 14-June-2025]. 2025.

[70] Wikipedia contributors. *Generalized least squares — Wikipedia, The Free Encyclopedia.* `https://en.wikipedia.org/w/index.php?title=Generalized_least_squares&oldid=1292245629`. [Online; accessed 7-June-2025]. 2025.

[71] Wikipedia contributors. *Inverse-variance weighting — Wikipedia, The Free Encyclopedia.* `https://en.wikipedia.org/w/index.php?title=Inverse-variance_weighting&oldid=1289542670`. [Online; accessed 9-June-2025]. 2025.

[72] Wikipedia contributors. *Solar cycle 25 — Wikipedia, The Free Encyclopedia.* [Online; accessed 24-May-2025]. 2025. URL: `https://en.wikipedia.org/w/index.php?title=Solar_cycle_25&oldid=1291422092`.

[73] Wikipedia contributors. *Unbiased estimation of standard deviation — Wikipedia, The Free Encyclopedia.* [Online; accessed 19-May-2025]. 2025. URL: `https://en.wikipedia.org/w/index.php?title=Unbiased_estimation_of_standard_deviation&oldid=1285799346`.

[74] Norbert Zehentner and Torsten Mayer-Gürr. "Precise orbit determination based on raw GPS measurements". In: *Journal of Geodesy* 90 (2016), pp. 275–286.

[75] Bingbing Zhang et al. "Precise orbit solution for swarm using space-borne GPS data and optimized pseudo-stochastic pulses". In: *Sensors* 17.3 (2017), p. 635.

[76] Rui Zhang et al. "Assessment of Swarm Kinematic Orbit Determination Using Two Different Double-Difference Methods". In: *Remote Sensing* 15.10 (2023). ISSN: 2072-4292. DOI: `10.3390/rs15102669`.

[77] P Zingerle et al. "The combined global gravity field model XGM2019e". In: *Journal of geodesy* 94.7 (2020), p. 66.

Appendix A

Project Plan and Implementation Reflection

Here the original project plan defined at the end of the literature review is presented. We discuss how the thesis deviated from this plan, the reasons why this occurred and reflect upon the the thesis went.

A.1 Original plan overview

At the outset of this research, a clear plan was defined, accompanied by a Gantt chart outlining the projected timeline and structure of the thesis. This plan involved five distinct work packages. The original plan is summarised as follows:

1. Retrieve kinematic orbits.
2. Combine kinematic orbits into a single combined orbit solution.
3. Apply a gravity field estimation method to recover a temporal gravity field model.
4. Compare this model to those derived from individual kinematic solutions, to established Swarm-based models and to the state of the art combined Swarm gravity field models.
5. Validate the results and evaluate the effectiveness of the combination strategy.

A.1.1 Work packages (original plan)

The original work packages are provided here.

WP1: Retrieve and pre-process kinematic orbits

The project would begin by collecting Swarm kinematic orbit solutions from three analysis centres: TU Delft (TUD), the Astronomical Institute of the University of Bern (AIUB), and the Institute of Geodesy Graz (IFG). These would include position time series and formal covariance information for each of the three Swarm satellites. Retrieval was to be carried out from the Aristarchos server at TU Delft, followed by preprocessing to align epochs and handle missing data.

WP2: Combine kinematic orbits

The next step would involve combining the independent kinematic orbit solutions using one or more combination strategies. This would be performed at the time-series level — rather than at the level of normal equations — to manage complexity within the time frame of a Master's thesis. Multiple methods would be tested where time permitted.

WP3: Apply the gravity field estimation method

Following orbit combination, a gravity field model would be estimated using the decorrelated acceleration approach, as detailed in subsection F.5.4. This would require using Tudat to compute non-gravitational accelerations, in combination with appropriate dynamical force models. The resulting product would be a set of spherical harmonic coefficients representing the time-variable gravity field.

WP4: Perform validation and answer research questions

The gravity field derived from the combined orbit would then be validated against superior external models. This would test whether orbit combination led to meaningful improvements in the estimated gravity field model, especially compared to combined Swarm models computed by the Multi-approach gravity field models from Swarm GPS data initiative, addressing the primary research objective.

WP5: Tudat development

It was expected that this approach would rely on Tudat's force modelling and orbit propagation framework. During implementation, any missing functionality would be identified and developed to enable full gravity field estimation within Tudat. This work package supported the research question on propagating orbit uncertainty into gravity field recovery.

A.1.2 Planned research outputs

The intended deliverables of the project were:

- Combined kinematic orbits for Swarm A/B/C, along with reusable software tools for performing orbit combination.
- A time-variable gravity field model estimated from the combined orbit product. And, if time permitted:
- A gravity field estimation pipeline – based on the decorrelated acceleration approach – implemented in Tudat.

A.2 Comparison to actual thesis work

While the general structure and objectives of the project remain consistent, several deviations from the original plan occurred during the course of the thesis. These arose both from practical challenges and from new insights that shifted the direction of the research. The most significant changes are summarised below:

- **Focus on orbit combination and validation:** The scope of the thesis was narrowed to concentrate on orbit combination and the validation of its accuracy and uncertainty. The gravity field estimation component was removed due to time constraints and the significant effort required to develop the necessary software infrastructure in Tudat, as well as the large amount of analysis which could be conducted on the combined orbits themselves.
- **Emphasis on uncertainty propagation:** The role of uncertainty became central to the analysis. A larger portion of the thesis was dedicated to investigating how uncertainty information could be propagated and validated using SLR residuals. This became mroe interesting,
- **Changes to the validation framework:** A comprehensive pipeline was developed for validating orbit combinations using SLR-based residuals as an independent validation source. This replaced validating the combination by comparing the estimated gravity field models to existing models.
- **Inclusion of optimisation and VCE methods:** While originally listed as optional extensions, optimisation-based and variance component estimation (VCE) methods more important to the thesis, due to their potential to address inconsistent uncertainty information across input orbits, and to account for systematic bias in the input orbits.

Essentially, work packages one, two and four were retained, with adaptation to the fourth work package to shift validation from at the gravity field level to at the level of the orbits themselves. Gravity field estimation and the implementation thereof in Tudat (Work packages three and five) were both dropped; while changes have not been made to Tudat, fully functional and independent code is implemented and ready to be provided to others. These changes are a shift from the full gravity field estimation pipeline to a specific analysis of Swarm orbit combination.

A.3 Reflection

It is clear that my original plan was overambitious. During this project, I encountered numerous challenges which resulted in the implementation of the combination taking more time than expected. Dealing with managing inconsistencies in the input data, such as mismatched epochs, missing values and outliers, was more challenging than I anticipated before starting. When the goal is the combination of multiple years of orbits, retrieving and managing large datasets from the Aristarchos server as well as the ESA and IFG dissemination servers is not trivial, which is something I did not consider in advance. Automating these processes was necessary for practical reasons which increased the complexity. Once the data was received, quite a bit of pre-processing was necessary, and issues such as time-stamp misalignment needed to be fixed.

A major shortcoming of my original plan was the absence of a validation method for the combined orbits without gravity field estimation. Originally, I had not planned to use SLR for validation. This left me with limited options – essentially only comparisons against reduced-dynamic orbits (RDO's) – to assess accuracy and verify whether the combination strategies were functioning as intended.

Around the time of the midterm review, it became evident that SLR validation was essential. Incorporating it, however, required a significant amount of additional work that had not been accounted for in the original plan. Although it remained technically possible to validate the final gravity field solutions – thanks to the generous support of Aleš Bezděk from the Astronomical Institute of the Czech Academy of Sciences, who offered to assist with the estimation of gravity models – the transition to using SLR was necessary and time-consuming. This included both the technical implementation of software and the methodological effort to understand how to use the SLR observations effectively.

Importantly, the availability of true external validation enabled the detection of methodological issues that would otherwise have remained hidden; issues that would be much more difficult to identify if validation were only performed at the gravity field level.

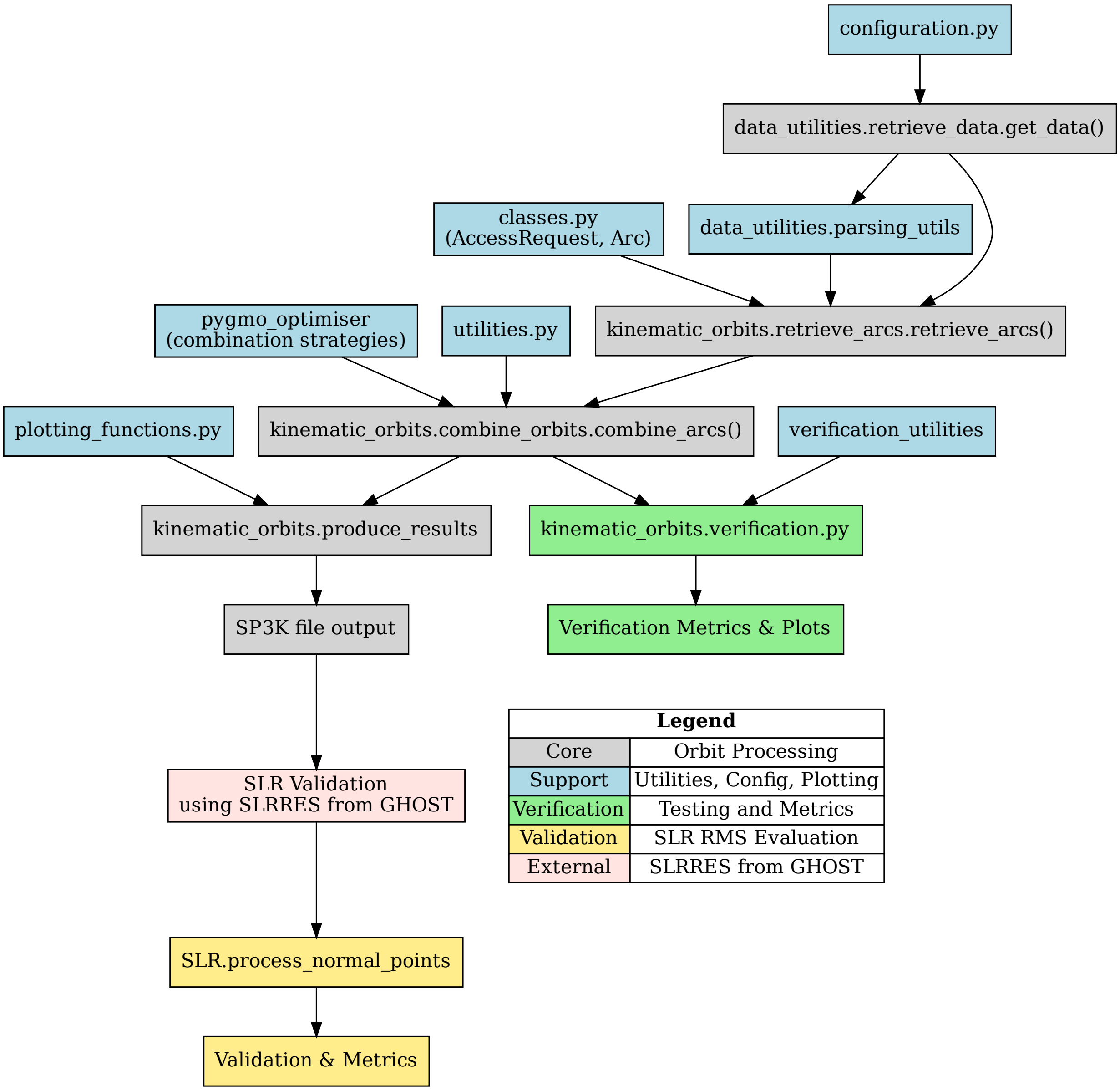
Fortunately, this delay lead to a number of lines of inquiry which were very interesting and which I hadn't thought of during the literature review, including the more objective assessment of the accuracy of the combined orbits, and especially the assessment of the covariance information using the SLR residuals for comparison. This culminated in the addition of the testing of whether or not the input orbit covariances could be rescaled using SLR validation data, which ended up quite an interesting concept with promising results.

Appendix B

Code Structure

A flow diagram of the implemented software is presented in Figure B.1. This includes the combination methods introduced as well as all infrastructure required to process data, save results and perform the verification and validation described in chapter 5. Modules have been implemented to facilitate orbit combination. Data is retrieved when necessary from external FTP servers and the Tu Delft Aristarchos server using the retrieve_data module and pre-processed and parsed using the parsing_utils module. Classes are defined to support data storage. The retrieve_arcs module connects these modules into a single pipeline which acts to retrieve necessary data.

Retrieved orbits are combined using combination strategies implemented in the combine_orbits module, with support from utilities and the pygmo_optimiser module containing the implementations of the optimisation functions. The produce_results module contains the pipeline which connects retrieval to pre-processing, combination and output (saving) of combined orbits. separate verification functionality is provided in the verification module. SLR validation is conducted externally on the Aristarchos server using the GHOST software. The result of this validation – a csv containing normal points – is then loaded in using process_normal_points, which processes these observations and computes and plots the many validation metrics.



Legend	
Core	Orbit Processing
Support	Utilities, Config, Plotting
Verification	Testing and Metrics
Validation	SLR RMS Evaluation
External	SLRRES from GHOST

Figure B.1: Data flow architecture of the `gravtools` package. This diagram illustrates the major processing modules and their interactions, including orbit retrieval, combination, verification, and validation via SLR. The colour-coded legend highlights the role of each component.

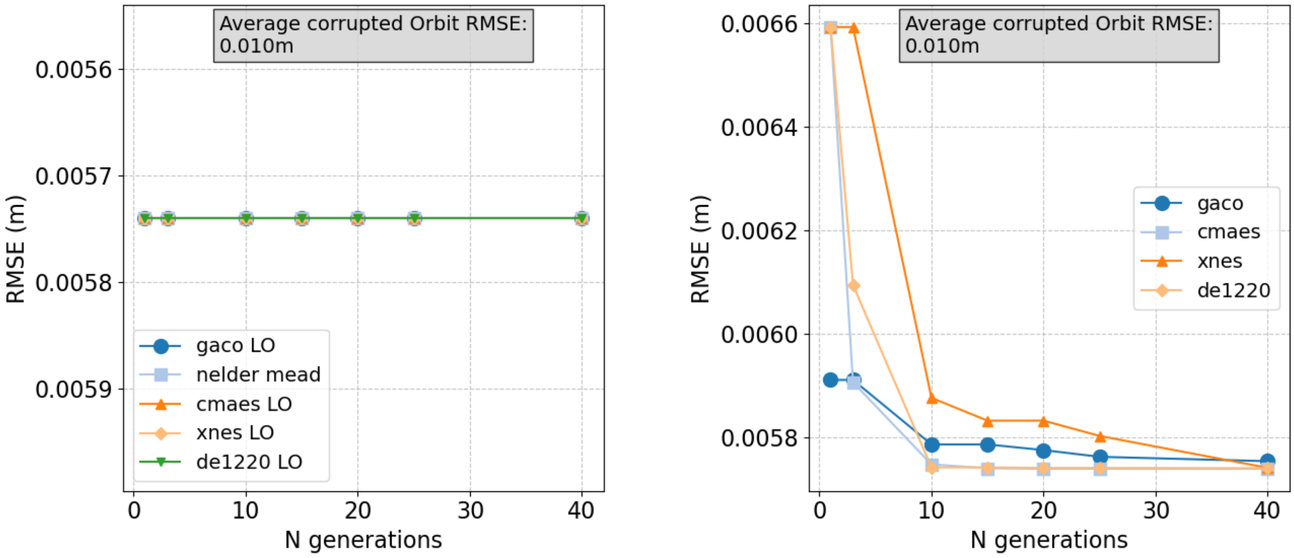
Appendix C

Additional verification and sensitivity analysis plots

In this chapter we introduce additional verification tests conducted to ensure that the results of chapter 5 are not dependent on the particular random seed.

C.1 Methodology verification and sensitivity study

C.1.1 Optimisation algorithm population and generation sensitivity analysis - Random seed 20

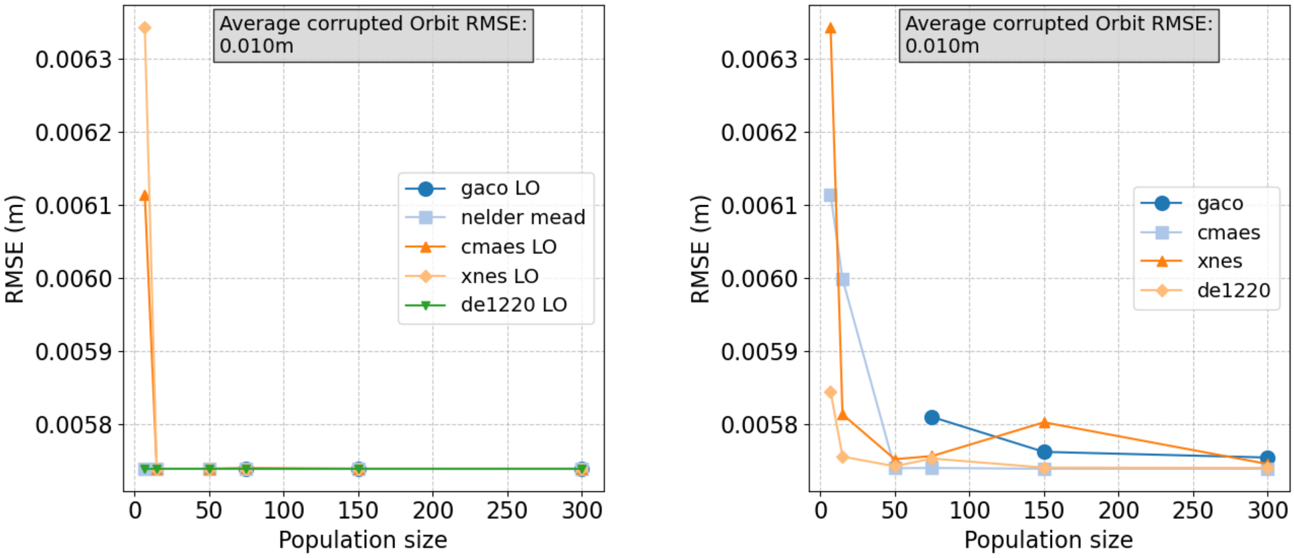


(a) Locally optimised methods.

(b) Globally optimised methods

Figure C.1: Impact of number of generations on RMS of the combined orbit with a random seed of 20.

These results confirm that the performance of optimiser-based combination methods is consistent across different random seeds. Increasing the number of generations improves the performance of globally optimised methods, though the benefit diminishes after a certain threshold. Locally optimised methods such as Nelder-Mead and residual-weighted tend to converge earlier, and their final performance is largely stable once a baseline number of generations is reached. This test demonstrates that the optimal number of generations selected in the main experiments is not dependent on random seed.



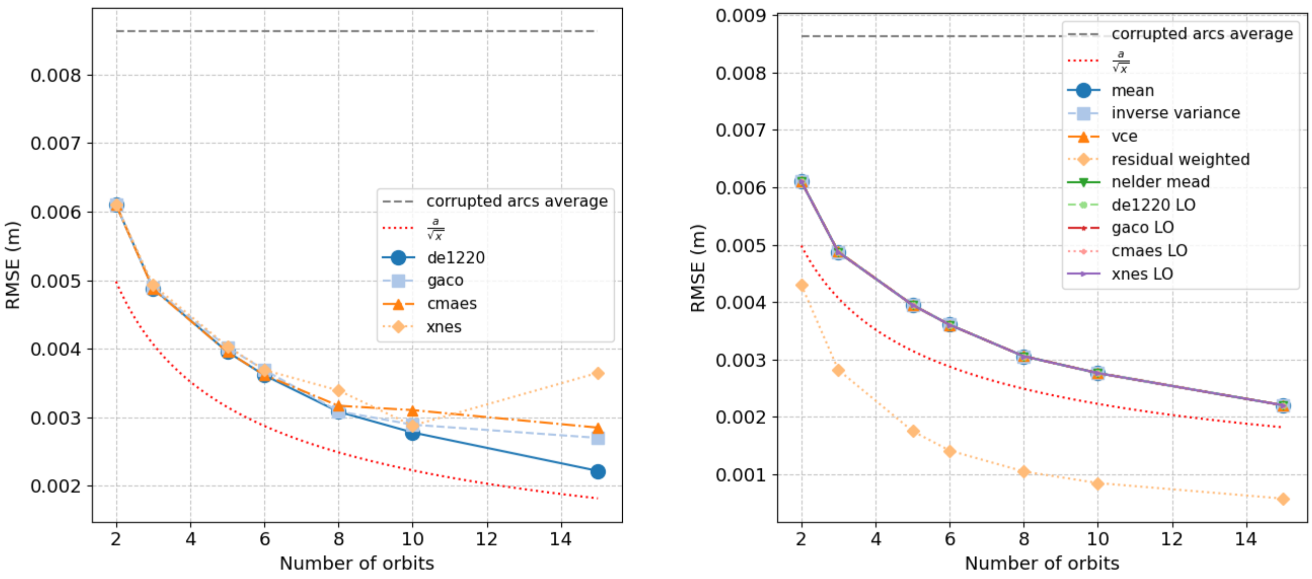
(a) Local and combined optimisers

(b) Global optimisers

Figure C.2: Change in optimiser performance with increasing population size with a random seed of 20.

Population size has a similar effect as generation count: performance of global methods improves with increasing population, but only up to a point. Local optimisers are less sensitive to this parameter. These results confirm that the optimiser parameter settings used in the main results strike an appropriate balance between accuracy and computational efficiency. The robustness of the outcome across seeds confirms that the chosen configuration is not over-fit to a particular noise realisation.

C.1.2 Non-mirrored corrupted orbits - Random seed 20



(a) Global methods.

(b) Non-global methods.

Figure C.3: Impact of increasing the number of input corrupt orbits for random seed 20.

These results use a different random seed to apply non-mirrored corruptions to input orbits. As before, global methods retain better robustness as the number of corrupted arcs increases. The residual-weighted method continues to degrade quickly as the proportion of biased inputs grows, confirming its sensitivity to the reference orbit and input outliers. The consistency with earlier mirrored-noise results further strengthens confidence in the observed ranking of method robustness.

C.1.3 50-run variance component analysis

Here we present the results for a different noise-realisation for the 50-run VCE analysis performed in section 5.4.

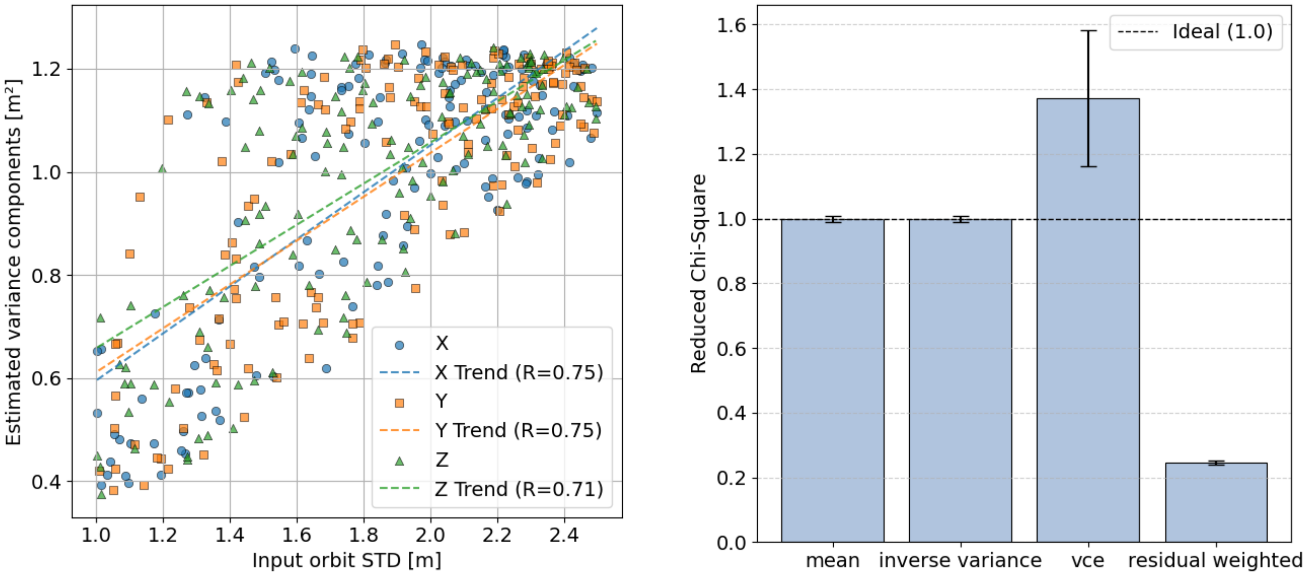


Figure C.4: The 50-run estimated variance components for noise realisations generated from random seeds 1000-1050, and the resulting reduced chi-square of the orbit solutions across different methods.

This test repeats the variance component estimation using a new batch of noise realisations (seeds 1000–1050). The similarity in the structure of the estimated weights and the resulting reduced chi-square statistics across methods confirms that the VCE method behaves consistently across different instances of the random seed. In particular, the bias of the VCE method towards down-weighting certain inputs remains visible, and the reduced χ^2 values follow the same ranking. This supports the conclusion that the previously reported variance behaviour is representative and not unique to a specific noise seed.

C.1.4 Threshold filtering results for an additional combination method

In order to demonstrate that the Pareto front of section 5.5 remains dominated by the combined solution regardless of combination method used, we present the results of a repeat test using the VCE combination method in Figure C.5.

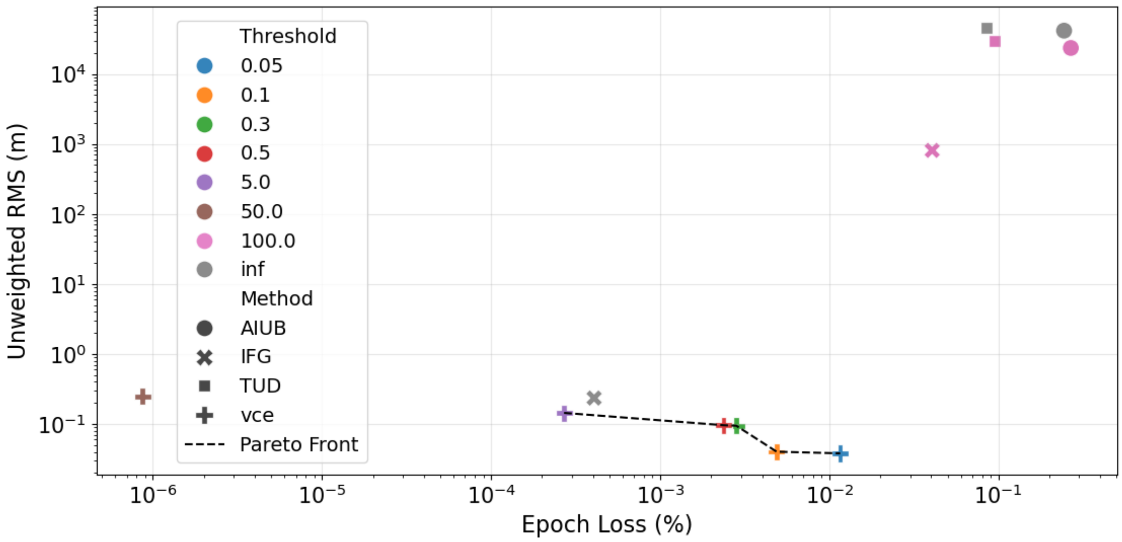


Figure C.5: Pareto front including all input orbits tested with various filtering thresholds. Linear-log scale. The threshold ceilings are denoted by the colour of the markers, while their shape refers to the method or analysis centre which produced the orbit. An infinite threshold refers to no pre-screening.

Despite switching to the VCE combination method, the combined orbit remains consistently on the Pareto front. This demonstrates that the advantage of screening based on performance thresholds is not dependent on the inverse-variance method used in the main text. Rather, it reflects a fundamental benefit of filtering out poorly performing input orbits. These results generalise the findings of section 5.5 and confirm that the screening strategy improves the combined solution without loss of epochs.

Appendix D

Results for Swarm B and C, 2023

Combined orbits for the full year of 2023 were produced for each Swarm satellite in order to enable the estimation of a gravity field model from the complete set for the purpose of determining how combining orbits affects their accuracy. The results for Swarm B and C are presented here. We observe a significant improvement in performance for Swarm B in Figure D.1, where the combined orbits achieve cm-level accuracy comparable to the reduced dynamic orbit. We also observe that residual-weighting produces the best combined orbit, likely due to the fact that the RDO is incredibly accurate for this orbit. For Swarm C, there remains an improvement, however it can be seen in Figure D.2 that IFG's input orbit is incredibly competitive and the improvement not as large as a result.

The superior results of Swarm B are explained by the superior orbit solutions which result from the higher orbit altitude (511 km vs. 462 km) leading to force model errors such as solar radiation pressure and atmospheric drag being lower [75]. Additionally, Swarm B has roughly twice the SLR observation data compared to Swarm A and C due to the fact that the pair in tandem compete for SLR station observations [44].

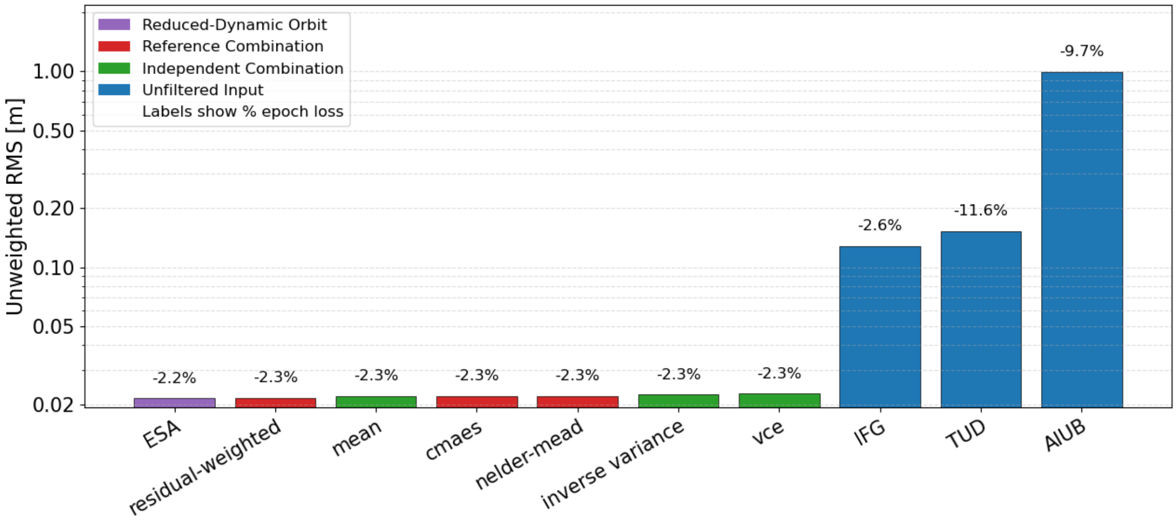


Figure D.1: RMS of the SLR residuals for each orbit for the year of 2022 for Swarm B. % epoch loss is annotated atop the bars.

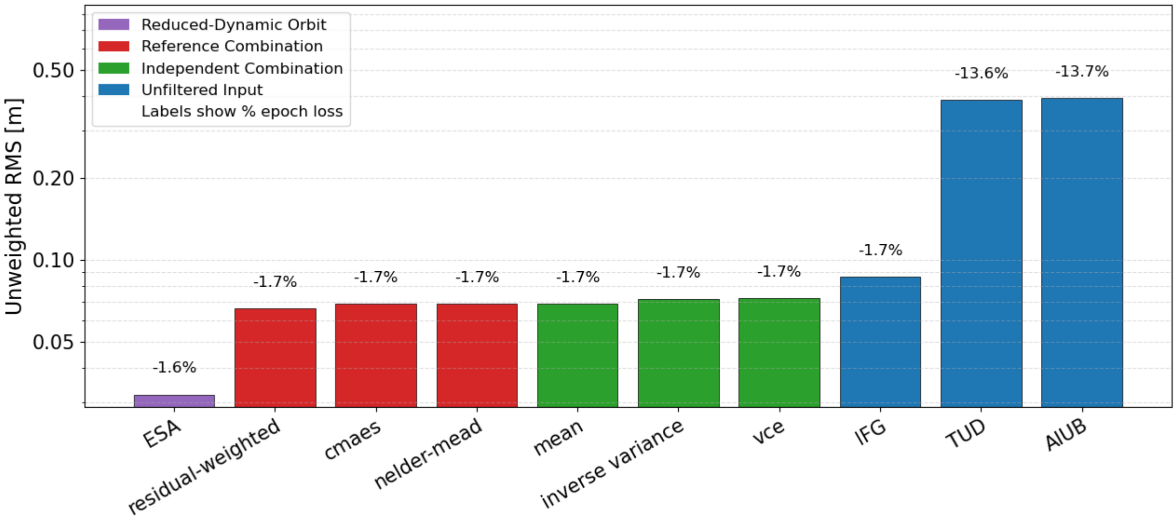


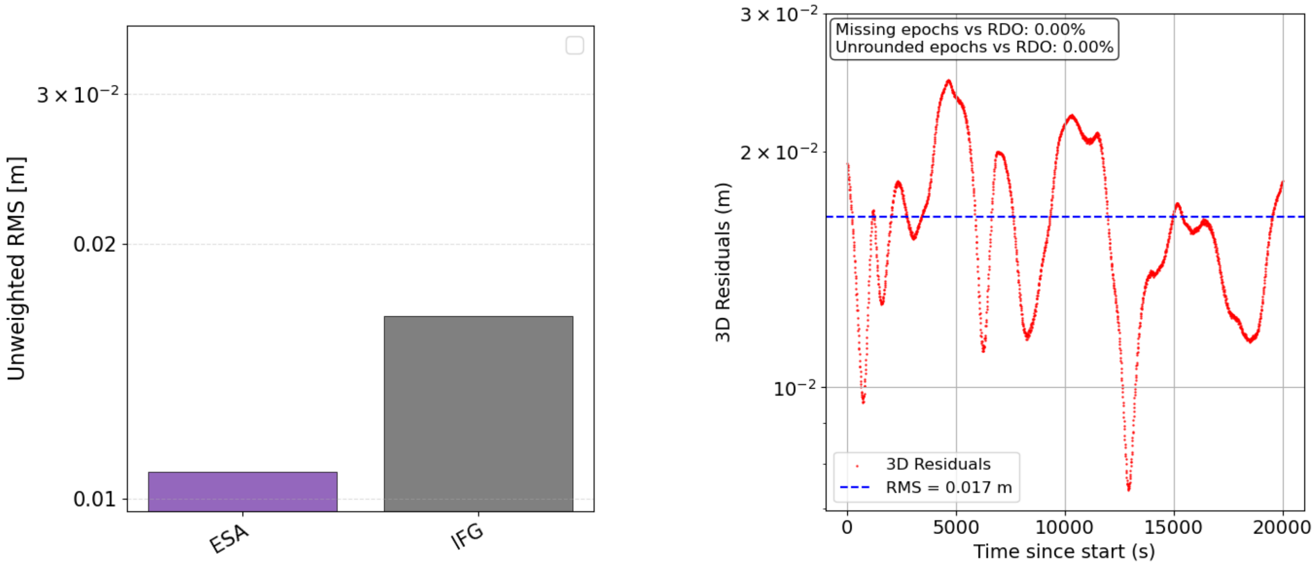
Figure D.2: RMS of the SLR residuals for each orbit for the year of 2023 for Swarm C. % epoch loss is annotated atop the bars.

Appendix E

Comparison of State-of-the-art Reduced-dynamic Orbits

As identified in section 2.3 there are two sources of state of the art Swarm RDO's; those produced by the TUD and disseminated by ESA (referred to as the ESA RDO), and those produced and disseminated by the IFG. In this thesis, we make use of the former.

In order to justify the selection of the ESA RDO we compare the two to demonstrate their relative performance by retrieving the SLR residuals of each using the GHOST software and computing the RMS. For this analysis, we use the period of January 2023, comparing the orbits over this timeframe. The results are presented in Figure E.1.



(a)

(b)

Figure E.1: RMS of the SLR residuals and 3D position difference for the ESA and IFG Swarm reduced-dynamic orbits for the first month of January, 2023.

The ESA RDO beats the IFG RDO by 0.7 cm; a 40% improvement. An investigation into this difference was conducted, as this difference appeared larger than expected. The IFG RDO is not provided in sp3 format. It is provided in celestial reference frame as opposed to the ITRF frame, requiring transformation using the Astropy astrodynamics library [1]. Additionally, the time stamps are given in Modified Julian Date (MJD) instead of Gregorian time. Therefore, we need to convert both of these to the standardised format described in chapter 4. Both the time conversion and reference frame transformation was validated through comparison with PhD student Frederik Jacobs' implementation of the Astropy transformations and time conversion, and no differences were found between our implementations, ruling this out as a source of error. After ruling out implementation error, we accept the results as they are and proceeded with the thesis using the ESA RDO. This is not of great importance to us, as the thesis required the use of only one RDO, the ESA orbit is state of the art and any differences between it and the RDO from IFG are minimal.

Appendix F

Gravity Field Estimation

Gravity field models of a celestial body describe the gravitational potential in the space surrounding that body. These models are essential tools in geodesy and planetary science as they provide insight into the distribution of mass within and on the surface of the body. In this appendix, we introduce key concepts related to gravity field modelling, focusing on the Earth.

F.1 Gravimetry

Gravimetry is broadly defined as the measurement of a gravitational field. In geophysics, it involves quantifying the variations in gravitational acceleration due to the distribution of mass.

F.1.1 Earth's Gravity Field

The gravitational forces around Earth are primarily the result of its own mass. While third-body effects from celestial bodies such as the Moon and Sun exist, they are typically modelled and removed when constructing gravity field models of Earth. These models aim to isolate the component of the gravitational field arising from the Earth's internal mass distribution.

Due to the heterogeneous distribution of mass within the Earth, the gravitational field is not uniform. The geoid is a conceptual surface representing mean sea level, assuming a global ocean in equilibrium under gravity and rotation alone. Variations in geoid height correspond to variations in gravitational acceleration. Earth's gravity field is commonly divided into static and time-varying components.

Static Component

The static gravity field represents the time-invariant portion of Earth's gravity field, typically derived from long-term averages of gravity observations.

Time-Varying Component

Time-varying gravity components reflect changes in Earth's mass distribution due to dynamic processes in the hydrosphere, cryosphere, and atmosphere. These are inferred by differencing time-tagged gravity solutions with a static reference field. Such variations are of significant interest in Earth sciences [26].

F.2 Spherical Harmonics

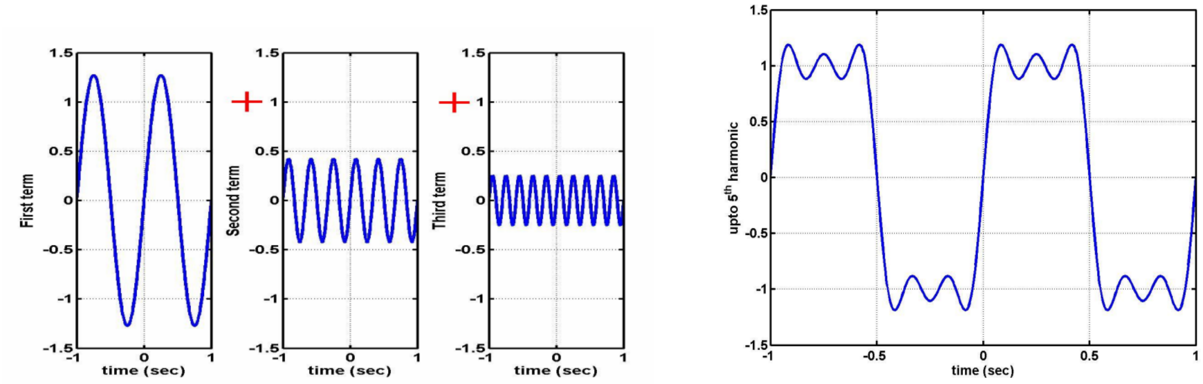
Gravity field models are commonly represented using spherical harmonic expansions. This is a natural basis for representing functions on the surface of a sphere, making it well suited to global potential fields.

F.2.1 Fourier Series

To understand spherical harmonics, it is helpful to begin with Fourier series. Any real, periodic function can be approximated as a sum of weighted sines and cosines [69]:

$$f(\theta) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos(n\theta) + b_n \sin(n\theta) \tag{F.1}$$

Here, a_n and b_n are the coefficients, and n determines the frequency of the component. As n increases, higher frequency details are captured, improving the approximation.



(a) First 3 Fourier series basis functions of a square wave

(b) Approximation of a square wave using the sum of basis functions

Figure F.1: Reconstruction of a square wave using a Fourier series. Image credit: BMS College of Engineering

F.2.2 Extension to Spherical Harmonics

Spherical harmonics extend this concept to the surface of a sphere. They are eigenfunctions of the Laplace operator in spherical coordinates and are used to represent scalar fields such as gravitational potential on spherical domains [45].

The gravitational potential U outside the Earth's mass distribution satisfies Laplace's equation:

$$\nabla^2 U = 0 \tag{F.2}$$

In spherical coordinates (r, θ, φ), the Laplacian becomes:

$$\nabla^2 U = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial U}{\partial r}\right) + \frac{1}{r^2 \sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial U}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 U}{\partial\varphi^2} \tag{F.3}$$

Separation of variables yields solutions involving associated Legendre functions and trigonometric functions. The resulting spherical harmonic expansion of the gravitational potential is [38]:

$$V(r,\theta,\varphi) = \frac{GM}{r}\sum_{l=0}^{\infty}\left(\frac{R}{r}\right)^l \sum_{m=0}^{l} \bar{P}_{l,m}(\cos\theta)\left[C_{l,m}\cos(m\varphi) + S_{l,m}\sin(m\varphi)\right] \tag{F.4}$$

Here:

- R is a reference Earth radius,
- l and m are degree and order,
- $\bar{P}_{l,m}$ are the fully normalised associated Legendre functions,
- $C_{l,m}$ and $S_{l,m}$ are the spherical harmonic coefficients.

F.2.3 Coefficient Interpretation

The coefficients $C_{l,m}$ and $S_{l,m}$ determine the spatial structure of the gravity field. Different classes of harmonics describe different spatial patterns:

- **Zonal harmonics**: $m = 0$, symmetric about the rotational axis.
- **Sectorial harmonics**: $l = m$, longitudinally varying.
- **Tesseral harmonics**: $l \neq m \neq 0$, general case with both latitudinal and longitudinal variation.

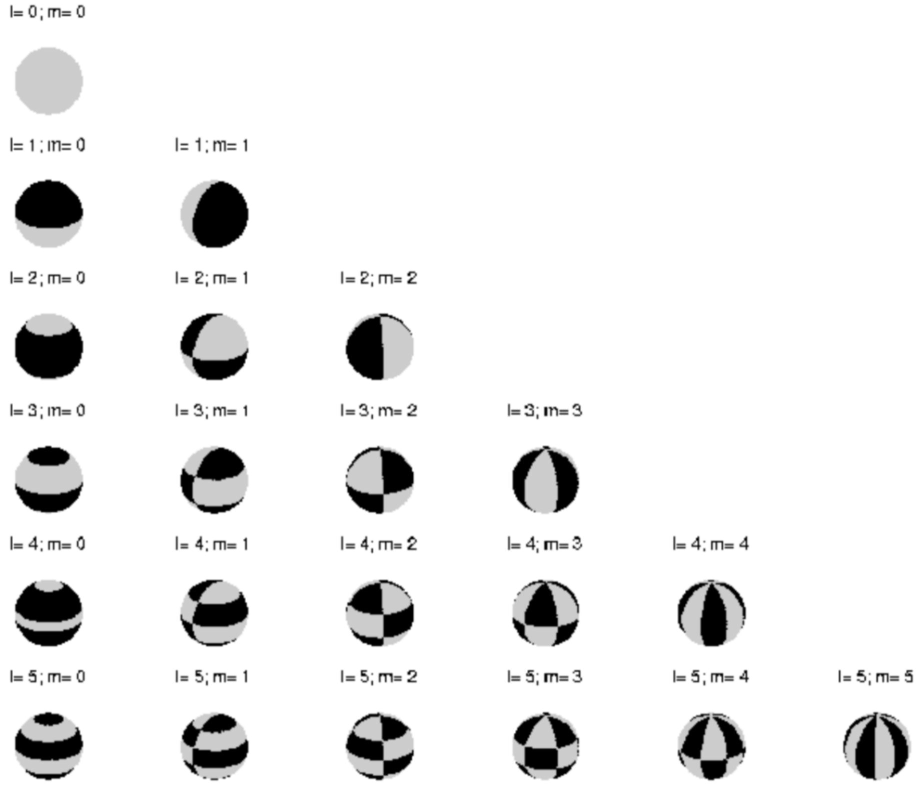


Figure F.2: Examples of individual spherical harmonics. Image credit: Massachusetts Institute of Technology (MIT) OpenCourseWare.

In gravity field recovery, the coefficients are estimated by applying least squares inversion to satellite observations. Increasing the degree l enables finer spatial features to be represented, at the cost of increased sensitivity to noise.

F.3 Data sources

The data used for global gravity field modelling comes from a combination of terrestrial and space-based sources. Combining multiple observation types, when done properly, improves the overall model fidelity by compensating for the weaknesses of individual data sets [65]. For instance, GRACE and GRACE-FO have difficulty accurately resolving the lowest-degree coefficients, such as $C_{2,0}$, which are therefore often substituted using SLR estimates [24].

F.3.1 Terrestrial data

Ground-based, airborne, and marine gravity measurements are available through institutions such as the National Centers for Environmental Information (NCEI) [20]. These include isolated point measurements, regional surveys (e.g., over Antarctica), and gridded data products. While such data offer high spatial resolution in well-surveyed areas, coverage is uneven, particularly in remote regions.

F.3.2 Space-based data

Space-based gravity data encompass several techniques. GNSS observations enable orbit reconstruction of low-Earth orbiters using either kinematic, dynamic, or reduced-dynamic methods. This thesis employs kinematic orbits — position time-series estimated independently of force models. SLR determines satellite positions by timing laser pulses sent from ground stations. Gravity gradiometry, such as that performed by the GOCE mission, directly measures gravity gradients.

Spaceborne sensors offer global coverage, particularly from dedicated gravity missions in LEO. Their ability to observe regions inaccessible to terrestrial surveys makes them indispensable for consistent global modelling.

F.4 Combined gravity field models

Combining independently estimated gravity field models is a powerful approach to improving quality and realism [65]. Techniques such as variance-based weighting and VCE help reconcile models derived from different data sources and assumptions.

F.4.1 Existing models

Several leading gravity field models are summarised below.

GOCE-DIR-R6 (Direct Approach Gravity Model)

This model, derived from GOCE gravity gradient and GPS tracking data, resolves spherical harmonics up to degree 300. It combines satellite and terrestrial data for improved precision in oceanography, geodynamics, and climate research [36].

GGM05C

Produced by CSR, GGM05C fuses data from GRACE, GOCE, and various terrestrial and altimetric datasets. It resolves harmonics up to degree 360 and is widely used for hydrology and sea level studies [54].

XGM2019e

Developed by TUM and partners, this model integrates satellite and terrestrial gravity with topographic information, resolving up to degree 719 [77].

EGM2008

Developed by the National Geospatial-Intelligence Agency (NGA), this high-resolution model extends to degree and order 2190. It combines satellite, surface, and altimetry [47].

ITSG-Grace2018

A time-variable model from TU Graz based on monthly and daily GRACE data. Solutions extend to degree 180 (monthly) and 40 (daily), and are valuable for mass change studies such as glacier melt and groundwater depletion [37].

F.5 Gravity field estimation techniques

Several advanced methods have been developed for estimating the Earth's gravity field from satellite tracking data. These techniques, applied by different research institutes, each reflect unique modelling philosophies and trade-offs. Their diversity in assumptions and processing schemes makes them complementary and, in many cases, suitable for combination.

F.5.1 Celestial mechanics approach

The Celestial Mechanics Approach (Celestial Mechanics Approach (CMA)), employed by the Astronomical Institute of the University of Bern [3], originates from classical orbit determination. It is based on Newtonian mechanics and treats gravity field estimation as a generalisation of orbit determination.

Satellite orbits are divided into arcs—typically 5 to 30 minutes long—and modelled using Keplerian elements and additional perturbative parameters. These include the unknown spherical harmonic coefficients, third-body effects, and empirical parameters for non-gravitational forces.

Input data consist of precise kinematic positions. A least-squares adjustment minimises residuals between observed and modelled satellite positions, iteratively refining the model parameters. The strength of CMA lies in its rigorous physical foundation and ability to incorporate a wide array of force models and perturbations.

F.5.2 Short-arc approach

The Short-Arc Approach (Short-Arc Approach (SAA)), implemented at the Institute of Geodesy, Graz [42], segments the orbit into short arcs and treats each as an independent boundary value problem. This localises errors and increases robustness.

Each arc, typically tens of minutes long, uses GNSS-derived positions and velocities—possibly augmented by inter-satellite ranging (as in GRACE). Initial conditions, empirical accelerations, and spherical harmonic coefficients are estimated via least-squares.

Because arcs are independently processed, SAA captures short-wavelength features and high-frequency gravity variations. It also allows for regional quality assessment and cross-validation of estimated parameters. The trade-off is the challenge of stitching consistent models across arcs.

F.5.3 Improved energy balance approach

The Improved Energy Balance Approach (Improved Energy Balance Approach (IEBA)), developed at Ohio State University [60], is based on energy conservation. It relates gravitational potential and satellite kinetic energy, forming an energy balance equation.

After modelling and subtracting non-conservative forces, satellite velocity (estimated from position data) and potential variations are used to solve for the geopotential:

$$V - \frac{1}{2}\|\dot{r}\|^2 + \int_{t_0}^{t} \frac{\partial V}{\partial t} dt - \int_{t_0}^{t} f \cdot \dot{r} dt - E^0 \tag{F.5}$$

Here, V is the gravitational potential, $\dot{r}$ the velocity, f non-conservative forces, and E^0 an integration constant. Since velocities are derived via numerical differentiation, this method is sensitive to noise and differentiation quality.

The potential V is separated into a known component V^R (from a reference model) and an unknown V^E. With sufficient force modelling and accelerometer data, V^E can be isolated and estimated.

F.5.4 Decorrelated acceleration approach

The Decorrelated Acceleration Approach (Decorrelated Acceleration Approach (DAA)), proposed by the Institute of Geodesy, Graz [6], derives gravity from satellite acceleration. It is particularly suitable for kinematic orbits, making it relevant to combined orbit processing.

Using Newton's second law, GNSS-derived positions are double-differentiated to obtain acceleration. After removing known non-gravitational and third-body accelerations, the residual is attributed to gravity. The fundamental equation is:

$$\frac{d^2 r}{dt^2} = a_{grav} + a_{LS} + a_{TID} + a_{REL} + a_{NG} \tag{F.6}$$

Where a_{grav} is the gravitational acceleration, and other terms represent lunisolar, tidal, relativistic, and non-gravitational effects.

The second derivative is approximated via a polynomial smoothing filter:

$$\frac{d^2 r}{dt^2} \approx \frac{d^2 Q(r_{gps})}{dt^2} = F \star r_{gps} \tag{F.7}$$

The gravity-induced acceleration is expressed as the gradient of the geopotential:

$$a_{grav}(r) = \nabla V(r) = \sum_{n,m} [C_{nm} \nabla V_{nm}^{(c)} + S_{nm} \nabla V_{nm}^{(s)}] \tag{F.8}$$

After subtracting the modelled accelerations:

$$\frac{d^2 Q(r_{gps})}{dt^2} - a_{other} = \sum_{n,m} [C_{nm} \nabla V_{nm}^{(c)} + S_{nm} \nabla V_{nm}^{(s)}] + \epsilon \tag{F.9}$$

This linear regression system is solved using least-squares. However, noise amplification from differentiation and autocorrelated GNSS noise must be addressed.

Dealing with noise amplification Differentiation of noisy GNSS positions introduces autocorrelated errors. These manifest as structured residuals, biasing parameter estimates. The regression system is transformed using a Generalised Least Squares (Generalised Least Squares (GLS)) matrix W derived from the differentiation filter F:

$$Var(\epsilon) = \sigma^2 F F^T \quad \Rightarrow \quad W = F^{-1} \tag{F.10}$$

This transformation reduces autocorrelation and restores white-noise assumptions.

Dealing with GNSS error correlation Additional GNSS error correlations arise due to satellite constellation geometry. A second GLS transformation W^* is computed by fitting an autoregressive model to the GNSS error process:

$$W^* = T^{-1}, \text{ with } Var(\epsilon) = TT^T \tag{F.11}$$

Applying both transformations yields residuals suitable for Ordinary Least Squares (OLS), preserving estimator optimality. This ensures valid uncertainty estimates for the recovered spherical harmonic coefficients.

Appendix G

Models for kinematic orbits and gravity field models from Swarm GPS data

These model definitions are retrieved from the 'Multi-approach gravity field models from Swarm GPS data' project standards and background models document [19].

G.1 Delft University of Technology

G.1.1 Kinematic orbit

Table G.1: Delft University of Technology Kinematic Orbit Solution [19]

Attribute	**Details**
Software	GPS High precision Orbit determination Software Tool (GHOST) [68]
Differencing Scheme	Undifferenced
Linear combination	Ionosphere-free
GPS observations	Code and carrier phase
Estimator	Bayesian weighted Least-Squares (LS)
Arc length	30 hours
Data weighting	a-priori weights equal to 1m and 1mm for code and phase observations (resp.)
Transmitter PCV	IGS08.atx model [58]
Receiver PCV	Empirical, derived from 70 days of data
Data screening	Minimum SNR of 10, minimum of 6 GPS satellites, code and phase outlier editing threshold of 2 m and 3.5 cm, respectively, 1 meter or larger difference between estimated KO positions and with Reduced-Dynamic Precise Science Orbit (PSO)
Earth precession model	International Astronomical Union (IAU) 1976 [39]
Earth nutation model	IAU 1980 [59]
Earth orientation model	Centre for Orbit Determination in Europe (CODE) final Earth Rotation Parameters (ERP)

G.1.2 Gravity field model

G.2 Astronomical Institute of the University of Bern

G.2.1 Kinematic orbit

Table G.2: Astronomical Institute of the University of Bern Kinematic Orbit Solution [19]

Attribute	**Details**
Software	Bernese v5.3 [14]
Differencing Scheme	Undifferenced
Linear combination	Ionosphere-free
GPS observations	Carrier phase
Estimator	Batch LS
Arc length	24 hours
Data weighting	N/A
Transmitter PCV	Official IGS08 ANTEX up to day 17/028, official IGS14 ANTEX from day 17/029 on
Receiver PCV	Stacking of residuals from reduced-dynamic POD of approx. 120 days, 9 iterations, 1° binning
Data screening	2 cm/s or larger time-differences of the geometry-free linear combination of L1B GPS carrier phase observations
Earth precession model	International Earth Rotation Service (IERS) 2010 Conventions [50]
Earth nutation model	IERS 2010 Conventions [50]
Earth orientation model	CODE final ERP

G.2.2 Gravity field model

Parameter	**Description**
Software	Bernese v5.3 [15]
Approach	Celestial Mechanics Approach (CMA) [3]
Reference GFM	AIUB GRACE-only static model, version 3 (AIUB-GRACE03S) [31]
Empirical Parameters	Daily piecewise-constant, 15 minutes piecewise-constant (constrained)
Drag Model	None
EARP and EIRP Models	None
Non-tidal Model	Atmosphere and Ocean De-aliasing Level 1B (AOD1B) product
Ocean Tidal Model	2011 Empirical Ocean Tide model (EOT11a) [56]
Permanent Tide System	tide-free

Table G.3: Gravity field modelling parameters for the Astronomical Institute of the University of Bern.

G.3 Institute of Geodesy Graz

G.3.1 Kinematic orbit

Table G.4: Institute of Geodesy Graz Kinematic Orbit Solution [19]

Attribute	**Details**
Software	Gravity Recovery Object Oriented Programming System (GROOPS)
Differencing Scheme	None
Linear combination	None (the ionospheric influence is co-estimated)
GPS observations	Code and carrier phase
Estimator	LS
Arc length	24 hours
Data weighting	Elevation and azimuth-dependent, epoch-wise Variance Component Estimation (VCE)
Transmitter PCV	Empirical, estimated from 5.5 years of data, including data from several LEO missions (GRACE, Jason 2 & 3, MetOp-A & -B, Sentinel 3A, Swarm, TanDEM-X, TerraSAR-X) [74]
Receiver PCV	Empirical, spherical harmonics (maximum D/O 60), derived from 38 months of data
Data screening	Implicit in VCE
Earth precession model	IAU 2006/2000A precession-nutation model [50]
Earth nutation model	IAU 2006/2000A precession-nutation model [50]
Earth orientation model	IERS Earth Orientation Parameter (EOP) 08 C04 [50]

G.3.2 Gravity field model

Parameter	**Description**
Software	GROOPS
Approach	Short-Arcs Approach (SAA) [74]
Reference GFM	GOCO release 05 satellite-only gravity field model (GOCO05S) [41]
Empirical Parameters	Piecewise linear for each arc (ranging from 15 to 45 minutes)
Drag Model	Jacchia-Bowman 2008 (JB2008) [11]
EARP and EIRP Models	[55]
Non-tidal Model	Atmosphere and Ocean De-aliasing Level 1B RL06 (AOD1B-RL06) product [17]
Ocean Tidal Model	2014 Finite Element Solution (FES2014) global tide model [13]
Permanent Tide System	zero tide

Table G.5: Gravity field modeling parameters for the Institute of Geodesy Graz.

G.4 Ohio State University

G.4.1 Gravity field model

Parameter	Description
Software	Developed in-house
Approach	Improved Energy Balance Approach (IEBA) [60]
Reference GFM	GRACE Intermediate Field 48 (GIF48) [53] up to Degree and Order (D/O) 200
Empirical Parameters	2nd order polynomial every 3 hours, 1-Cycle Per Revolution (CPR) sinusoidal every 24 hours
Regularization	None
Drag Model	(US) Naval Research Laboratory Mass Spectrometer and Incoherent Scatter Radar Atmospheric model (NRLMSISE) [51]
EARP and EIRP Models	[35]
Non-tidal Model	Atmosphere and Ocean De-aliasing Level 1B (AOD1B) product [61]
Ocean Tidal Model	2011 Empirical Ocean Tide model (EOT11a) [56]
Permanent Tide System	tide-free

Table G.6: Gravity field modelling parameters for Ohio State University.

G.5 Astronomical Institute Ondřejov

G.5.1 Gravity field model

Parameter	Description
Software	Developed in-house
Approach	Decorrelated Acceleration Approach (DAA) [5, 4]
Reference GFM	ITG GRACE-only static model, 2010 (ITG-GRACE2010s) [43]
Empirical Parameters	Daily constant-piecewise
Drag Model	(US) Naval Research Laboratory Mass Spectrometer and Incoherent Scatter Radar Atmospheric model (NRLMSISE) [51]
EARP and EIRP Models	[35]
Non-tidal Model	Atmosphere and Ocean De-aliasing Level 1B (AOD1B) product [61]
Ocean Tidal Model	2004 Finite Element Solution (FES2004) global tide model [40]
Permanent Tide System	tide-free

Table G.7: Gravity field modeling parameters for the Astronomical Institute Ondřejov.

G.6 Common

In this section the common model parameters shared between each approach are summarised.

G.6.1 Kinematic orbits

Table G.8: Common carrier phase and GPS settings for kinematic orbit solutions shared between each institute

Attribute	**Details**
Carrier phase ambiguities	Float
Receiver clock corrections	Co-estimated
Sampling rate	10 or 1 seconds (depending on L1B GPS data)
Elevation cut-off angle	0°
GPS orbits and clocks	Final orbits and 5-second clocks of Centre for Orbit Determination in Europe (CODE) (Dach et al., 2017)
Swarm attitude	L1B attitude data

G.6.2 Gravity field models

Parameter	**Description**
Atmospheric Tidal Model	Biancale and Bode (2006) [7]
Solid Earth Tidal Model	IERS2010 [49]
Pole Tidal Model	IERS2010 [49]
Ocean Pole Tidal Model	IERS2010 [49]
Third body perturbations	Sun, Moon, Mercury, Venus, Mars, Jupiter, and Saturn, following the JPL Planetary and Lunar Ephemerides [23]
C2,0 coefficient	Estimated alongside other coefficients

Table G.9: Common gravity field modelling parameters.