Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# TLE-driven numerical phase rate calibration of uncontrolled geostationary satellite forecasts

**E. Sergeev**

e.sergeev@surrey.ac.uk

University of Surrey

**A. Viquerat**

University of Surrey

**C. P. Bridges**

University of Surrey

**Article**

**Keywords:**

**Posted Date:** November 17th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-11447255/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** No competing interests reported.

# TLE-driven numerical phase rate calibration of uncontrolled geostationary satellite forecasts

E. Sergeev[1*], A. Viquerat[1] and C. P. Bridges[2]

[1*]School of Engineering, University of Surrey, Stag Hill, Guildford, GU2 7XH, Surrey, United Kingdom.
[2]Surrey Space Centre, University of Surrey, Stag Hill, Guildford, GU2 7XH, Surrey, United Kingdom.

*Corresponding author(s). E-mail(s): e.sergeev@surrey.ac.uk;
Contributing authors: a.viquerat@surrey.ac.uk; c.p.bridges@surrey.ac.uk;

**Abstract**

Small differences in orbital phase rate can accumulate into substantial position errors when TLE-derived states initialise numerical GEO forecasts. We estimate the local phase rate difference between numerical predictions and catalogue states from up to ten selected TLE epochs within the preceding 90 days. The fitted difference sets a semi-major axis correction bounded at 2 km, preserving the anchor mean longitude. On 160 matched held-out objects, calibration with an object-selected effective area-to-mass ratio reduces four-month median position error relative to later catalogue states from 76.79 km for SDP4 to 17.65 km, a 77.0% reduction. A fixed global area-to-mass ratio gives a comparable median of 17.79 km. Matched numerical controls show that calibration primarily reduces transverse position error. SDP4 performs better at seven days, while calibrated four-month 95th- and 99th-percentile error estimates are larger. These findings support phase rate calibration for reducing typical long-horizon position errors in the retrospectively classified population.

**Keywords:** Space Debris, Geostationary Earth Orbit, Phase Rate Calibration, Numerical Propagation

# 1 Introduction

The Geostationary Earth Orbit (GEO) region is a finite resource whose debris environment lacks the atmospheric deorbiting available in Low Earth Orbit (LEO) [1]. Uncontrolled spacecraft and upper stages traverse longitude slots occupied by active satellites, while public catalogues are lacking for small sized ($< 1$ m) debris [2]. Independent optical and infrared surveys indicate a substantial unresolved population [3], and conjunction studies motivate better characterisation of the evolving GEO environment [4]. Improving forecasts from publicly available orbital data would support wider access to studies of the long-term motion of catalogued debris.

Public Two-Line Element (TLE) histories obtained from Space-Track provide the orbital dataset for this study [5]. A TLE contains mean elements fitted for SGP4/SDP4, with a different interpretation from the osculating elements used by a numerical force model [6]. Evaluating a TLE with its associated propagator provides a Cartesian position and velocity, whose subsequent evolution depends on the propagation model. The semi-analytical SGP4/SDP4 theory and the numerical force model can therefore predict different orbital phase rates from the same starting state. A persistent difference in phase rate accumulates into transverse separation over the forecast horizon. Estimating this drift from preceding catalogue histories offers an accessible approach to improving long-horizon agreement with later catalogue states in the GEO region, using public TLEs as the orbital input.

Previous studies have fitted numerical trajectories to successive TLE-derived states using batch least squares and evaluated the resulting predictions against independent precision ephemerides [7]. Extended fitting windows and systematic along-track debiasing have also been investigated for pseudo-orbit determination in Medium Earth Orbit (MEO) and GEO, reducing post-fit position discrepancies for selected satellites [8]. Here, we examine a correction to the local numerical phase rate that preserves the latest catalogue mean longitude.

The study contributes a constrained numerical initialisation method that estimates historical phase rate drift from a small set of catalogue epochs. We refer to histories classified as drifting, librating or disposed as uncontrolled by Control Region Inclination (COREIN), which identifies control activity from changes in TLE-derived orbital elements and characterises the dynamical behaviour over each available history [9]. Comparisons of initial-state interpretation, force models and calibration show how each affects position error. Separate development, validation and test objects allow us to assess how well the correction transfers to other objects. The method improves typical long-horizon forecasts of catalogued GEO objects from public TLE histories, with potential applications for satellite operators and regulators.

# 2 Methods

Calibration and evaluation are applied to the population identified by COREIN. Each forecast starts from the Cartesian state obtained by evaluating the latest TLE at the anchor epoch. Backward numerical propagation over the preceding 90 days provides the historical phase differences used to estimate a local rate correction, which is applied through a bounded semi-major axis adjustment that preserves the anchor

mean longitude. The effective area-to-mass ratio used to calculate Solar Radiation Pressure (SRP) is either fixed globally or selected separately for each object using historical forecasts completed before the anchor epoch. The corrected state is then propagated forwards and evaluated against later TLE-derived states, with separate development, validation and test objects supporting method selection and assessment.

## 2.1 Catalogue population and classification

TLE histories were obtained from Space-Track [5]. Classification was applied once to each complete available history before constructing forecast folds. Eligibility based on terminal elements and history-wide dynamical labels can therefore use observations after the forecast anchor. We retained objects whose terminal mean elements met the conditions: $e < 0.2$, $i < 70^\circ$ and $0.9 < n < 1.1$ revolutions per day [9, 10]. Invalid epochs and Cartesian states were rejected, and for exactly duplicate epochs, the last source record was retained, while closely spaced epochs remained separate. Element outliers were retained to preserve possible genuine manoeuvres.

COREIN describes each history by dynamical behaviour, radial region and inclination [9]. Discrete changes in inclination, node, semi-major axis, eccentricity and perigee argument identify control activity. For an element $x$, the adaptive prominence is

$$p_x = \left[\sigma(\Delta x), 0.1\,\mathrm{IQR}(x), 0.01(x_{\max} - x_{\min})\right]_{\max} \tag{1}$$

where $\Delta x = x_{k+1} - x_k$, $\sigma$ is standard deviation and IQR denotes the interquartile range. Stationkeeping requires detected control, residence within the protected radial band ($42{,}164 \pm 200$ km), every unwrapped longitude within $1^\circ$ of its first value, and an absolute mean specific-energy gradient below $10^{-7}\,\mathrm{km^2\,s^{-2}}$ per observation. A prior stationkeeping interval followed by sustained departure from its slot is classified as disposal when the terminal radius remains in that band. Otherwise, uncontrolled histories are tested for libration about a geopotential stable longitude using

$$q = \sin^2\theta + (\dot{\theta}/k)^2, \tag{2}$$

with wrapped offset $\theta$ and libration-rate scale $k$; the implemented criterion is median $q < 1$ [9, 11]. Remaining uncontrolled objects are drifting. Radial labels describe the entire history as protected, above, below, graveyard ($r > 42,399$ km throughout) or crossing, while inclination is high if any recorded $i > 15^\circ$, as per the GEO Protected Region (GEOPR) definitions [12]. Only drifting, librating and disposed histories enter the propagation study.

We compared the broad classes with the European Space Agency (ESA)'s 2018 Classification of Geosynchronous Objects catalogue as a cross-epoch consistency check [13]. A separate DISCOS snapshot retrieved in October 2026 was matched to the test objects by COSPAR identifier to review the available activity flags [14].

## 2.2 Numerical propagation

A TLE contains mean elements fitted within SGP4/SDP4 theory, so numerical initialisation requires evaluation with its associated propagator [6]. Each TLE was evaluated with SDP4 at its own epoch and transformed to an EME2000 Cartesian position and velocity. The latest such state supplies the numerical forecast anchor. This preserves the catalogue's instantaneous phase while making the change of propagation theory explicit; direct numerical propagation of this state is also the uncalibrated control.

The Orekit numerical model [15] uses a degree-and-order $12 \times 12$ terrestrial gravity field, Sun and Moon point-mass attraction, and isotropic SRP, with atmospheric drag omitted. A Dormand–Prince 8(5,3) integrator uses a minimum step of 1 s and a maximum step of 300 s. Orekit constructs absolute and relative tolerance vectors consistent with the orbit representation from a position tolerance of 1 m. The Earth frame is ITRF with IERS 2010 conventions and the simple Earth orientation option, while the body shape uses WGS84 equatorial radius and flattening, with the Holmes–Featherstone model [16] used as a normalised gravity provider. Sun and Moon ephemerides are loaded through Orekit, and the SRP model includes Earth occultation through its ellipsoidal-body implementation. We fix the radiation coefficient at $C_R = 1.5$ [17] and infer an effective area-to-mass ratio $\alpha = A/m$; only $C_R\alpha$ is identifiable in this radiation model. A reference mass of 2000 kg is used solely to report an equivalent area.

For calibration, the anchor's osculating Cartesian state is temporarily expressed as

$$\boldsymbol{s} = [a, e_x, e_y, h_x, h_y, \lambda_M]^{\mathsf{T}}, \tag{3}$$

where the nonsingular equinoctial coordinates permit a bounded change to semi-major axis $a$ while holding the other five coordinates, including mean longitude $\lambda_M$, fixed. The corrected state is converted back to Cartesian position and velocity before propagation. Holding the other five coordinates fixed preserves mean longitude, orbital shape and orientation. Changing $a$ adjusts the radial scale and velocity magnitude, so the preserved phase refers specifically to mean longitude. Equinoctial coordinates provide a regular representation for this single-parameter correction in nearly circular, nearly equatorial GEO orbits.

## 2.3 Phase rate estimation and correction

Historical phase fitting uses only TLE epochs from the 90 days preceding the forecast anchor $t_0$. At least eight distinct epochs are required. When more than ten are available, ten observations are selected at approximately equal intervals through the chronologically ordered records, including the earliest and latest observations. The elapsed time between retained observations depends on catalogue cadence, here described as the inter-TLE gap. For each candidate effective area-to-mass ratio, the uncorrected numerical anchor state is propagated backwards to these epochs. The TLE-derived state at epoch $t_i$ provides $(\boldsymbol{r}_{o,i}, \boldsymbol{v}_{o,i})$, and the numerical trajectory provides $\boldsymbol{r}_{p,i}$. With the catalogue state's normal unit vector

$$\widehat{\boldsymbol{h}}_i = \frac{\boldsymbol{r}_{o,i} \times \boldsymbol{v}_{o,i}}{\|\boldsymbol{r}_{o,i} \times \boldsymbol{v}_{o,i}\|}, \tag{4}$$

the signed phase difference between the numerical and catalogue states in the catalogue orbital plane is

$$\epsilon_i = \operatorname{atan2}\Big(\widehat{\boldsymbol{h}}_i \cdot [\widehat{\boldsymbol{r}}_{o,i} \times \widehat{\boldsymbol{r}}_{p,i}], \widehat{\boldsymbol{r}}_{o,i} \cdot \widehat{\boldsymbol{r}}_{p,i}\Big). \tag{5}$$

All position and velocity vectors used in Equation (5) are expressed in EME2000 and evaluated at the same catalogue epoch. The resulting phase difference represents the azimuth of the numerical position vector projected onto the catalogue orbital plane, with positive values indicating that the numerical vector leads the catalogue vector in the direction of orbital motion. Projection and normalisation scale both arguments of atan2 by the same positive factor, preserving the calculated angle. This angle is undefined when the predicted position is perpendicular to the catalogue orbital plane, where its projection vanishes, and becomes sensitive to small positional changes near this configuration. The phase differences are then unwrapped chronologically by adding or subtracting multiples of $2\pi$ to remove jumps exceeding $\pi$, although changes in the underlying phase difference greater than $\pi$ between successive observations remain ambiguous.

With $\tau_i = (t_i - t_0)$ in days, a linear Huber fit [18] estimates the local phase rate

$$(\widehat{b}_0, \widehat{b}_1) = \arg\min_{b_0, b_1} \sum_i \rho_\delta(\epsilon_i - b_0 - b_1 \tau_i), \tag{6}$$

where the estimation uses robust least squares with equal weight assigned to each retained catalogue epoch and a Huber loss $\rho_\delta$ that is quadratic for residual magnitudes up to $\delta = 0.005^\circ$ and linear beyond this threshold. Optimisation uses a trust-region reflective method with convergence tolerances of $10^{-8}$. The fitted slope defines the mean-motion adjustment $\Delta n = -\widehat{b}_1$, which is converted to a semi-major axis correction through a first-order expansion of the Keplerian relation $n = \sqrt{\mu/a^3}$ [19],

$$\frac{\Delta n}{n} = -\frac{3}{2}\frac{\Delta a}{a}, \qquad \Delta a = \frac{2a}{3n}\widehat{b}_1, \tag{7}$$

where $n$ and $\widehat{b}_1$ use consistent angular and time units (radians per day). The correction is clipped to $|\Delta a| \leq 2{,}000$ m and applied at $t_0$ with $\Delta\lambda_M = 0$. Thus the latest TLE supplies the starting phase, while its local historical phase rate disagreement sets the numerical rate. The Keplerian conversion is a local first-order approximation to the phase rate response under the full force model. The correction acts as an effective adjustment to the numerical initial state and can absorb combined effects of force model mismatch, catalogue biases and state estimation.

## 2.4 Selection of the effective area-to-mass ratio

We compare two SRP settings, both using a constant effective area-to-mass ratio $\alpha$ during each forecast. The global setting uses $\alpha_G = 0.0125$ m$^2$ kg$^{-1}$ for all objects. The object-selected setting determines $\alpha$ separately for each object and forecast origin using nested forecasts wholly before $t_0$. The historical selection procedure is described below. Up to six historical profile anchors are spaced 15 days apart. At each, a 90-day phase fit is followed by a nominal 30-day forecast. Every available later TLE within $\pm 7$ days

of the inner target, and no later than $t_0$, contributes to the profile's median position disagreement $\widetilde{e}_j(\alpha)$, where the tilde denotes the median over the available comparison epochs. The initial search grid comprised 17 candidate values of $\alpha$ spanning the interval $[0.001, 0.200]$ $\mathrm{m^2\,kg^{-1}}$.

Historical profile anchors are chosen at nominal offsets of 30, 45, 60, 75, 90 and 105 days before $t_0$, using the latest available catalogue epoch on or before each nominal date. Each anchor requires a full preceding 90-day history, and repeated anchors are omitted. Within each profile, candidate disagreements are normalised by their median across candidates. Candidates successfully evaluated on the greatest number of profiles are then ranked by the median normalised error plus its median absolute deviation across those profiles. Candidates with equal selection scores are ranked successively by lower median disagreement, lower 90th-percentile disagreement, proximity to the global value $\alpha_G = 0.0125$ $\mathrm{m^2\,kg^{-1}}$ and, finally, smaller $\alpha$. If the best coarse candidate is an interior strict local minimum, the search is refined by adding two candidate values at equal logarithmic intervals between it and each immediate neighbour. The additional candidates are evaluated and ranked together with the successful coarse candidates using the same selection rule. At least three successful profiles are required, and the selected $\alpha$ is held fixed over the outer forecast. Objects with insufficient historical profiles remain unavailable for this variant.

## 2.5 Evaluation and verification

Objects were assigned to development, validation and reserved test sets in nominal proportions of 70%, 15% and 15%, with all forecast origins of each object assigned to the same set. Development comparisons used median four-month position error to select the phase-fit settings, and a sensitivity analysis informed the retention of the global and object-selected SRP settings. Settings were frozen before evaluating the reserved test. Development controls used three forecast origins separated by 30 days, while the later object-specific validation and final test used one origin per object.

For forecast evaluation, later TLEs were matched near 7, 30, 60 and 100 days with a maximum target mismatch of 3 days. The four-month reference was the last TLE at or before four calendar months after the anchor and within the preceding 7 days. All eligible endpoints were retained regardless of forecast disagreement. Predictions and catalogue states were evaluated at the same epoch, with position disagreement

$$e = \|\boldsymbol{r}_p - \boldsymbol{r}_o\|_2 = \sqrt{e_R^2 + e_T^2 + e_N^2}, \tag{8}$$

where radial, transverse and normal components were resolved in the catalogue state's local orbital frame.

Baseline SDP4 and uncalibrated numerical forecasts from the same Cartesian anchor provided the controls. Input comparisons varied the anchor between the SDP4-derived Cartesian state, its osculating equinoctial round trip, and printed TLE mean elements treated as osculating, with and without calibration using the global area-to-mass ratio. Force comparisons added the Earth geopotential, lunar gravity, solar gravity and SRP, refitting calibration for each model. Force allocations introduced the Earth geopotential first and used Shapley averaging over the addition orders of the remaining three forces.

Position and absolute component disagreements were allocated separately, retaining signed contributions.

Validation also compared the two constant area-to-mass ratio settings with a time-varying effective area-to-mass ratio. The time-varying experiment fitted the logarithm of historical profile winners using robust regression with exponentially decreasing weights for older profiles. Its forward slope was damped and the predicted ratio clipped to the tested range, with propagation updated in one-day segments. Cadence sensitivity was assessed separately by thinning the historical observations while holding the object, anchor, target and propagator fixed. The unthinned and thinned conditions in this experiment used a minimum of six fitting observations, compared with eight in the main method.

Four-month median position disagreement was the main comparison selected during development, with other horizons, tail quantiles and dynamical classes assessed secondarily. Reported 95% intervals were calculated from the 2.5th and 97.5th percentiles of differences obtained from 2,000 bootstrap resamples clustered by object, retaining all forecast origins of each sampled object and preserving the paired comparisons between methods. Exploratory associations between selected effective area-to-mass ratio and four-month position disagreement were assessed separately by dataset split using Spearman rank correlation and object-bootstrap intervals.

# 3 Results

We first assess forecast performance on the reserved test objects, then compare SRP settings and examine the contributions of initialisation, perturbations and phase correction. Analyses of extreme disagreements, dynamical classes and historical sampling describe the variation in performance, followed by forecasts across the full eligible cohort.

## 3.1 Held-out forecast performance

On the 160 common four-month endpoints, SDP4 had a median position disagreement of 76.79 km. Figure 1 shows that phase calibration reduced this to 17.79 km with the global area-to-mass ratio and 17.65 km with an object-selected value. The reduction with object-selected calibration was 59.14 km (95% interval, 44.10–76.45 km), corresponding to a 77.0% reduction in the population median, with 83.8% of individual forecasts improving, as depicted in Table 1.

To distinguish the contribution of phase calibration from the change in propagator, we compared calibrated and uncalibrated numerical forecasts using the same force models. Without calibration, the four-month medians were 608.83 km with the global area-to-mass ratio and 609.63 km with object-selected values. Applying the correction to the same force models reduced these medians by 591.04 km (95% interval, 538.91–620.43 km) and 591.99 km (95% interval, 542.94–625.08 km), respectively. Individual forecasts improved in 91.9% and 90.6% of the corresponding paired comparisons. The uncalibrated numerical forecasts had substantially greater median disagreement than SDP4.

**Table 1**: Matched test performance on endpoints shared by all five methods. G and E denote global and object-selected SRP settings for calibrated forecasts. The SDP4 baseline is shared between the two SRP settings. Reduction is the SDP4 median minus the calibrated median. Win is the percentage of forecasts improving on SDP4.

| Horizon ($N$) | SRP setting | Median (km) SDP4 | Median (km) Calibrated | Reduction [95% interval] (km) | Win (%) |
|---|---|---|---|---|---|
| 7 days (156) | G | 2.40 | 8.04 | -5.65 [-6.53, -4.67] | 18.6 |
| | E | | 7.91 | -5.51 [-6.55, -4.39] | 20.5 |
| 30 days (157) | G | 9.99 | 5.54 | 4.45 [2.85, 7.28] | 64.3 |
| | E | | 6.06 | 3.93 [2.72, 6.94] | 68.2 |
| 60 days (160) | G | 25.32 | 9.89 | 15.43 [9.11, 25.44] | 73.8 |
| | E | | 9.61 | 15.72 [9.96, 25.71] | 78.1 |
| 100 days (159) | G | 54.16 | 20.52 | 33.63 [20.16, 55.70] | 71.7 |
| | E | | 22.06 | 32.09 [19.87, 54.31] | 75.5 |
| 4 months (160) | G | 76.79 | 17.79 | 59.00 [43.54, 76.16] | 81.9 |
| | E | | 17.65 | 59.14 [44.10, 76.45] | 83.8 |

The corresponding component disagreements show that this improvement was predominantly transverse. For the object-selected numerical control, median absolute radial, transverse and normal disagreements changed from 6.49, 609.62 and 0.35 km before calibration to 4.36, 16.54 and 0.36 km after calibration. The corresponding SDP4 medians were 7.72, 68.48 and 2.85 km. The numerical model supplied smaller typical normal disagreement, while phase calibration substantially reduced transverse disagreement.

Figure 1a shows how the four-month improvement varied with forecast horizon. The median with object-selected calibration was smaller than SDP4 at 30, 60 and 100 days by 39.3%, 62.1% and 59.3%, respectively. At seven days, it was 7.91 km versus 2.40 km for SDP4, an increase of 5.51 km (95% interval, 4.39–6.55 km); 20.5% of individual forecasts improved. At four months, calibrated 95th- and 99th-percentile disagreement estimates were larger than those of SDP4, with the wide intervals examined in Section 3.4.

## 3.2 Global and object-selected radiation settings

On the matched four-month test endpoints, object-selected calibration reduced the median disagreement by 0.14 km relative to global calibration (95% interval, −2.59–3.54 km). The interval spans reductions favouring either variant. Object-selected calibration had the higher four-month 90th-percentile disagreement (135.12 versus 117.35 km). At 30 days, it also had the higher median (6.06 versus 5.54 km), although its 90th percentile was lower (19.24 versus 24.85 km). Table 1 presents both frozen area-to-mass variants to show their endpoint position error across horizons.

The global and object-selected settings differ in the effective area-to-mass ratio used to calculate SRP, with the radiation coefficient held fixed. Calibration with

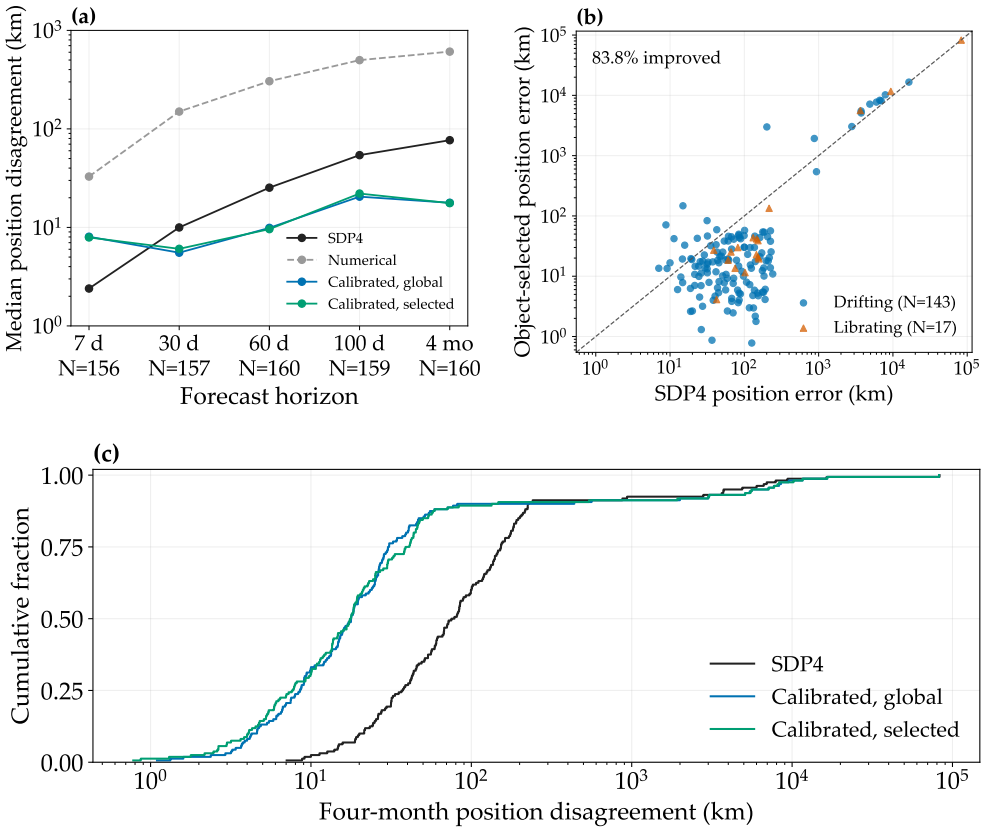



**Fig. 1**: Held-out forecast performance. **(a)** Median catalogue disagreement across horizons; Numerical denotes the uncalibrated control with the global area-to-mass ratio. **(b)** Paired four-month disagreements for 160 objects, coloured by dynamical class; points below the diagonal favour calibration. **(c)** Cumulative distributions for the same objects, comparing SDP4 with global and object-selected calibration.

the global value requires only the outer phase-fit history, whereas object-selected calibration also requires successful nested historical profiles. Their closely comparable four-month medians support the global value as a practical reference, with the value of object-specific selection dependent on the forecast horizon and summary statistic.

Calibration with the global area-to-mass ratio also provided wider coverage because it required no object-specific historical selection. Across all 163 available four-month endpoints shared by global calibration and SDP4, the respective medians were 17.88 and 78.48 km, with 81.6% of individual forecasts improving.

Before settings were frozen for the reserved test, the separate validation set was used to compare the results of fixed global, fixed object-selected and time-varying effective area-to-mass ratios, described in Section 2.5, on 171 common four-month endpoints.

Table 2 shows the time-varying experiment had the largest median and 90th-percentile disagreements, supporting retention of the two fixed variants for the final test.

**Table 2**: Four-month position disagreement for phase-calibrated forecasts using fixed global, fixed object-selected and time-varying effective area-to-mass ratios. The separate validation set informed variant selection before settings were frozen for evaluation on the reserved test.

| Cohort | $N$ | SRP setting | Median (km) | P90 (km) |
|---|---|---|---|---|
| Validation | 171 | Fixed global | 20.72 | 175.30 |
| | | Fixed object-selected | 17.48 | 149.46 |
| | | Time-varying object-selected | 29.09 | 192.18 |
| Test | 160 | Fixed global | 17.79 | 117.35 |
| | | Fixed object-selected | 17.65 | 135.12 |

## 3.3 Model contributions

Identifying the contributions to forecast improvement guides the choice of initialisation and force model. Figure 2 presents matched development controls that separate input interpretation, physical perturbations and phase correction on 2,503 four-month forecast cases from 852 objects.

The SDP4-derived Cartesian anchor and its osculating equinoctial round trip produce practically identical forecasts, with maximum endpoint separations of 0.35 mm before calibration and 3.02 mm after calibration. The equinoctial branch in Figure 2a verifies preservation of the same physical state through coordinate conversion. Treating printed TLE mean elements as osculating instead raises calibrated median position disagreement from 20.00 to 37.07 km. Median absolute transverse and normal disagreements rise from 19.22 to 35.62 km and from 0.48 to 6.44 km, while the radial median decreases from 4.21 to 4.00 km. These controls support obtaining the Cartesian anchor through SDP4, then using equinoctial coordinates to apply the bounded semi-major axis correction.

Figure 2b shows the four-month position median decreasing from 259.06 km for central gravity to 127.06 km after adding the $12 \times 12$ Earth geopotential, and to 20.00 km for the complete model, with phase calibration refitted for each force model. The central-to-full reduction is 239.07 km (95% interval, 216.95–269.61 km). The effects of individual perturbations depend on the reference model and interactions among forces.

Figure 2c estimates each force's contribution to the reduction in median disagreement, with phase calibration independently refitted for every configuration. Earth geopotential is introduced first, and the contributions of lunar gravity, solar gravity and SRP are averaged over their possible addition orders using the Shapley method. The component results show different roles for the perturbations, with Earth geopotential contributing 156.34 km transversely, lunar gravity 83.07 km normally and SRP

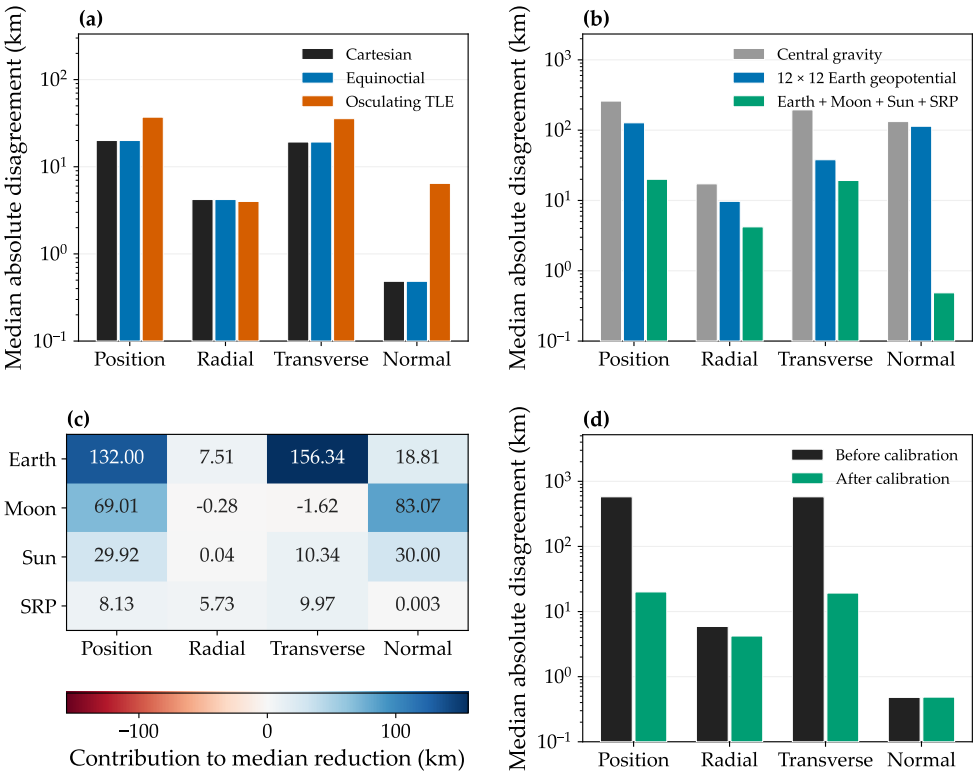



**Fig. 2**: Input, perturbation and phase contributions in matched four-month development controls. **(a)** Calibrated disagreement for the SDP4-derived Cartesian anchor, its equinoctial identity control, and printed TLE mean elements treated as osculating. **(b)** Central-gravity, full-geopotential and complete-force models, each independently calibrated. **(c)** Estimated force contributions to the reduction in median disagreement from calibrated central gravity to the complete model, with Earth geopotential added first and Shapley averaging over Moon, Sun and SRP. **(d)** The complete model before and after phase correction.

5.73 km radially. These contributions describe changes in population medians under the specified comparison procedure, with each position or component metric evaluated separately and negative values indicating increased error. The complete tested force model therefore addresses complementary components of catalogue disagreement.

Finally, Figure 2d isolates phase correction by holding the complete force model fixed. The four-month position median falls from 577.66 to 20.00 km, a reduction of 96.5%. Median absolute radial and transverse disagreements decrease by 28.6% and 96.7% from their respective uncalibrated medians of 5.90 and 577.62 km, while the normal median remains close to 0.48 km. Phase calibration primarily reduces transverse disagreement, complementing the radial and normal structure supplied by the force model.

### 3.4 Extreme errors, dynamical classes and sampling

Extreme-case, dynamical-group and sampling analyses assess how the typical forecast improvement varies across the error distribution and available histories. The four-month empirical distributions in Figure 1c show improved central agreement alongside persistent extreme discrepancies. Table 3a shows that, on the matched test objects, object-selected calibration lowered the 90th-percentile disagreement and raised the 95th and 99th percentiles. The wide 95% intervals span improvements and degradations at all three quantiles, highlighting substantial uncertainty in tail performance.

**Table 3**: Four-month robustness summaries. **(a)** Object-selected calibration and SDP4. **(b)** Exploratory test-group median reductions, with high inclination overlapping the dynamical groups.

**(a)** Test disagreement quantiles (km)

| Percentile | SDP4 | Calibrated | Reduction | 95% interval |
|---|---|---|---|---|
| 90th | 225.32 | 135.12 | 90.20 | [-2,118.82, 255.37] |
| 95th | 3,795.63 | 5,675.31 | -1,879.68 | [-2,261.17, 143.12] |
| 99th | 12,252.30 | 13,653.61 | -1,401.31 | [-2,308.60, 613.27] |

**(b)** Test-group median reductions (km)

| Group | $N$ | Reduction | 95% interval |
|---|---|---|---|
| Drifting | 143 | 53.84 | [39.98, 70.77] |
| Librating | 17 | 118.34 | [49.89, 130.42] |
| High inclination | 6 | -849.17 | [-1,718.85, 31.92] |

The exploratory group summaries in Table 3b show positive median reductions for drifting and librating objects. Each interval describes its group's reduction relative to SDP4. The high-inclination estimate favours SDP4, with a wide interval from only six objects. The small group sizes limit conclusions about performance across dynamical regimes, particularly at high inclination.

Across the full classified catalogue, 1,450 objects were matched by COSPAR identifier with the 2018 COGO catalogue [13]. Of 954 objects classified as uncontrolled by COGO, 943 were also classified as uncontrolled by COREIN, corresponding to 98.8%. Among the 1,119 matched objects classified as uncontrolled by COREIN, the corresponding agreement was 84.3%. The COGO catalogue classified 175 of these objects as controlled. The comparison describes cross-epoch agreement between catalogues with different classification criteria.

All successful test fits used ten observations. The 2,000-m correction bound was reached in 12 of 163 fits using the global area-to-mass ratio and 11 of 160 object-selected fits, with all cases retained. An exploratory review identified 14 object-selected four-month disagreements above 1,000 km, comprising 11 drifters and 3 librators, including 3 high-inclination objects. The threshold lies within an observed gap between 541.40 and 1,926.25 km; thresholds of 750 and 1,500 km identify the same objects. All 11 correction-bound hits occurred in this group, with all 160 objects retained in

the performance estimates. The concentration of correction-bound hits among large disagreements identifies correction saturation as a candidate forecast-quality indicator for evaluation on new data. Its use in screening forecasts would require prospective validation.

The separate comparison with the DISCOS snapshot retrieved in October 2026 matched 159 of the 160 test objects. Fourteen matched objects were flagged as active, including 12 of the 14 forecasts with disagreements above 1,000 km; activity was unspecified for the remaining 145 matches. The snapshot postdates the forecast anchors, so assessing control activity during the forecast interval requires dated operational evidence. All objects remained in the reported evaluation.

Comparison of the historically selected effective area-to-mass ratio with four-month position disagreement gave weak positive Spearman correlations of 0.138, 0.170 and 0.113 in the development, validation and test sets, respectively. Three of the five test objects at the upper search limit and five of the 21 at the lower limit had disagreements above 1,000 km. These results identify extreme fitted parameter values as candidates for further diagnosis, with their physical interpretation requiring independent information.

Controlled sampling sensitivity was assessed separately on matched development endpoints using the global area-to-mass ratio. Imposing a minimum 15-day observation spacing increased median disagreement while lowering the 90th percentile. Observed maximum inter-observation gaps had a weak association with four-month disagreement ($\rho_s = 0.012$). Together, these results support assessing sampling through both typical and upper-tail performance, with imposed thinning and naturally occurring gaps treated as distinct diagnostics.

## 3.5 Full-cohort forecasts

We next describe the frozen calibration across the pooled eligible dataset, separately from the held-out evidence above. The four-month cohort contains 818 development, 171 validation and 160 test objects. These pooled summaries characterise performance across the available classified population, complementing the reserved-test assessment in Section 3.1.

Figure 3 shows that, at four months, median position disagreement with later catalogue states decreases from 64.33 km for SDP4 to 17.22 km with global calibration and 16.37 km with object-selected calibration. The reduction in population medians with object-selected calibration is 74.5%, and 81.4% of individual forecasts improve. Its 90th-percentile disagreement decreases from 221.85 to 99.21 km, compared with 85.55 km for global calibration. These summary gains coexist with a broad spread of individual outcomes.

Table 4 summarises performance across forecast horizons, with the median changes illustrated in Figure 3a. Relative to SDP4, object-selected calibration increases the seven-day median disagreement by 252.2% to 7.14 km, while reducing the 30-, 60- and 100-day medians by 42.7%, 62.8% and 59.3%, respectively. The evaluated horizons therefore favour SDP4 at seven days and calibration from 30 days onwards. At four months, the calibrated 95th- and 99th-percentile disagreements increase by 21.8% to 3,926.06 km and 5.3% to 15,847.66 km, respectively. The full population exhibits the

same combination of improved typical agreement and persistent extreme discrepancies observed in the held-out test described in Section 3.1.

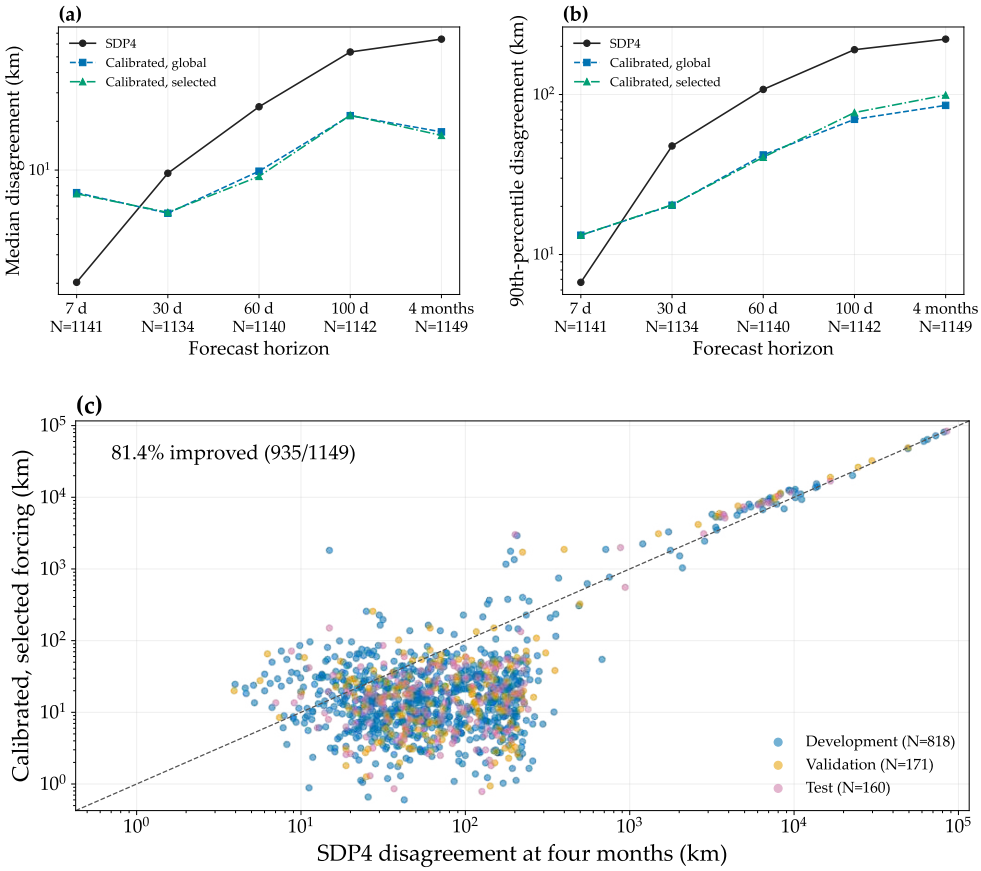



**Fig. 3**: Phase-calibrated forecasts across the pooled eligible dataset. **(a)** Median and **(b)** 90th-percentile position disagreement at each horizon, with matched object counts. **(c)** Paired four-month disagreements for 1,149 objects, coloured by development, validation and test membership. Points below the equality line favour object-selected calibration, with the percentage improved.

Median absolute radial, transverse and normal disagreements on the 1,149 objects are 7.67, 60.28 and 2.90 km for SDP4, compared with 3.84, 15.20 and 0.35 km for object-selected calibration. Global calibration gives 4.12, 15.99 and 0.35 km. These component summaries describe the full-cohort outcome, and the matched uncalibrated numerical controls above isolate which changes arise from phase calibration itself.

The calibrated four-month median is smaller than the 100-day median. Each horizon summarises its own eligible objects and reference epochs, so this ordering describes

**Table 4**: Full eligible cohort forecasts with frozen settings. The four-month cohort comprises 818 development, 171 validation and 160 test objects.

| Horizon | $N$ | SDP4 (km) | | Global (km) | | Object-Selected (km) | |
|---|---|---|---|---|---|---|---|
| | | Median | P90 | Median | P90 | Median | P90 |
| 7 days | 1141 | 2.03 | 6.70 | 7.25 | 13.21 | 7.14 | 13.21 |
| 30 days | 1134 | 9.56 | 47.70 | 5.41 | 20.33 | 5.48 | 20.46 |
| 60 days | 1140 | 24.59 | 107.60 | 9.84 | 41.90 | 9.14 | 40.52 |
| 100 days | 1142 | 53.58 | 190.56 | 21.68 | 69.98 | 21.82 | 77.12 |
| 4 months | 1149 | 64.33 | 221.85 | 17.22 | 85.55 | 16.37 | 99.21 |

population summaries at the evaluated horizons. Assessing changes along individual forecast trajectories requires a fixed cohort evaluated at consistent elapsed times.

# 4 Discussion

A bounded historical phase rate correction substantially improves typical long-horizon agreement with later catalogue states. Across the full eligible cohort, the four-month median decreases by 74.5% to 16.37 km, with 81.4% of individual forecasts improving. The reserved test gives a corresponding reduction of 77.0% to 17.65 km and an improvement rate of 83.8%, showing that the benefit extends to objects excluded from method selection. These results support using preceding public catalogue histories to calibrate the local numerical rate while preserving the latest catalogue-derived mean longitude.

The matched controls explain the source of this improvement. Figure 2d shows a 96.5% reduction in the development median when phase correction is applied to the same numerical force model, with a 96.7% reduction in median absolute transverse disagreement and little change in the normal component. This pattern is consistent with correcting a persistent rate difference that accumulates during propagation. The perturbation comparisons show complementary roles for Earth geopotential, lunar gravity and SRP. The coordinate conversion control produces practically identical forecasts from equivalent representations of the same state. The estimated force contributions depend on introducing Earth geopotential first and refitting calibration for each force model. The method complements numerical history fitting and TLE debiasing [7, 8] through a constrained adjustment of semi-major axis.

The radiation parameter comparisons show limited overall benefit from selecting a separate area-to-mass ratio for each object under the tested design. The global and object-selected four-month test medians differ by only 0.14 km, with a 95% interval spanning differences in either direction. Table 2 also shows that the time-varying area-to-mass ratio experiment produced a validation median of 29.09 km, compared with 20.72 km for the fixed global value and 17.48 km for fixed object-selected values. The larger median and 90th-percentile errors supported retaining the two fixed variants. One possible explanation is that extrapolating historical area-to-mass ratio estimates carries error forward, arising from catalogue discrepancies and interactions with phase

calibration. The comparison establishes the performance of the tested time-varying formulation; determining whether physical changes in effective illuminated area can be recovered requires independent evidence and further model comparisons. A fixed global area-to-mass ratio is therefore a practical reference for typical long-horizon catalogue agreement in this population.

It was initially thought that more frequent TLE updates during the 90 days preceding the forecast anchor would substantially reduce endpoint position errors across the studied population. The sampling experiments gave mixed support for this expectation; imposing a minimum spacing between retained TLE epochs increased the development median position error while lowering the 90th percentile. Separately, the maximum inter-TLE gap in the original histories showed almost no rank correlation with four-month position error ($\rho_s = 0.012$). These findings give little support to maximum inter-TLE gap alone as an indicator of endpoint disagreement over the histories examined. The timing of observations within the fitting window, the temporal structure of catalogue discrepancies and the use of a limited set of retained epochs may all influence the fitted slope. The present comparisons do not separate these possible explanations. Further sampling experiments could vary the number and placement of retained epochs on fixed objects and forecast origins to determine which features of historical coverage affect calibration.

The improvement also depends on forecast horizon and the part of the disagreement distribution being assessed. On the reserved test, the seven-day calibrated median is 7.91 km compared with 2.40 km for SDP4, while calibration gives smaller medians at the evaluated 30-day and longer horizons. A correction estimated over a historical window may provide a better long-term rate while retaining short-period discrepancies, although the present experiment does not establish their separate contributions. The larger calibrated upper-quantile estimates and wide intervals leave substantial uncertainty in extreme-case performance. All 11 object-selected correction-bound hits occur among the 14 four-month disagreements above 1,000 km. This association supports investigating saturation as a warning indicator, with prospective evaluation required before using it to screen forecasts. Increasing the correction bound would also require separate assessment, since saturation can reflect several forms of model or catalogue inconsistency.

Attitude-dependent variations in illuminated area can produce time-varying radiation forces, particularly for objects with high area-to-mass ratios [20, 21]. Such variations provide a possible limitation of the SRP model with a constant effective area-to-mass ratio. Establishing their contribution to individual extreme forecasts would require independent area-to-mass and attitude information.

Several limitations define the interpretation of these findings. Later TLE-derived states share modelling assumptions and potentially correlated errors with the histories used for fitting, so independent tracking or precision ephemerides are needed to quantify absolute orbital accuracy. Population eligibility and dynamical labels were assigned from complete available histories, whereas phase fitting and area-to-mass ratio selection obeyed the forecast epoch boundary. The results therefore describe forecasting within a retrospectively classified population. A further test should classify objects using only information available when each forecast begins. The small librating and high-inclination samples also restrict conclusions across dynamical regimes, and the

selected radiation parameter represents an effective model coefficient whose physical interpretation requires external validation. Within the tested 90 day histories, maximum inter-TLE gap had little association with four-month position error, while observation thinning affected median and upper-tail errors differently. Further tests with different history lengths and sampling patterns would establish how broadly these findings apply.

The present controls establish the contribution of phase correction within the tested numerical model. Direct comparisons with a historical phase correction to SDP4 and a conventional multi-parameter history fit would clarify the relative benefits of this constrained approach.

# 5 Conclusion

Bounded phase rate calibration provides a simple adjustment to numerical initialisation from public TLE histories while preserving the anchor mean longitude. Across 1,149 matched four-month objects, object-selected calibration reduces median position error relative to corresponding catalogue states by 74.5% to 16.37 km, with 81.4% of individual forecasts improving. The reserved test confirms a 77.0% reduction to 17.65 km on 160 objects, with 83.8% improving. Matched controls identify reduced transverse drift as the principal contribution. A fixed global area-to-mass ratio gives closely comparable typical performance, while the tested time-varying formulation and sampling changes provide no consistent improvement across the reported summaries.

The demonstrated benefit applies to typical agreement at the evaluated 30-day and longer horizons, with SDP4 retaining the smaller seven-day median and calibrated forecasts exhibiting persistent extreme discrepancies. Catalogue-based references, retrospective population selection and limited representation of some dynamical groups constrain generalisation. The 90-day historical sampling experiments showed a weak relationship between maximum inter-TLE gap and endpoint position error, with observation thinning producing different effects on median and upper-tail errors. For satellite operators, government agencies and regulators, the method could improve typical long-horizon forecasts of catalogued GEO objects using public TLE histories. Independent orbital validation and tests using objects classified from information available before forecasting are needed to establish its reliability for operational use.

# Acknowledgements

The authors acknowledge the University of Surrey Doctoral College for funding this PhD project under the DCSA 7 Breaking Barriers studentship.

# Author contributions

**E. Sergeev** conceived and designed the study, developed the methodology and software, curated the data, performed the simulations and formal analysis, and wrote and revised the manuscript. **A. Viquerat** and **C. P. Bridges** supervised the research, and reviewed and edited manuscript drafts.

## Data and code availability

The original TLE histories were obtained from Space-Track [5]. The derived data and code supporting this study are available from the corresponding author on reasonable request.

## Competing interests

The authors declare no competing interests.

## Use of Generative Artificial Intelligence

OpenAI Codex assisted with software editing and visualisation of existing simulation results, numerical consistency checks, editorial drafting and LaTeX restructuring. The authors retain responsibility for verifying the code, results, interpretations and manuscript before submission.

## References


[1] Kostelecký, J., Klokočník, J.: Investigation of the instability of graveyard orbits due to the earth's gravitational perturbations. Studia Geophysica et Geodaetica **36**(4), 297–301 (1992) https://doi.org/10.1007/BF01625482

[2] Schildknecht, T., Musci, R., Ploner, M., Preisig, S., Leon Cruz, J., Krag, H.: Optical Observation of Space Debris in the Geostationary Ring. In: Proceedings of the 3rd European Conference on Space Debris. ESA SP-473. ESA Publications Division, Darmstadt, Germany (2001). https://conference.sdo.esoc.esa.int/proceedings/sdc3/paper/126/SDC3-paper126.pdf

[3] Murray-Krezan, J., Meng, K., Seitzer, P.: Estimation of the GEO belt debris population by two independent remote sensing techniques. Optical Engineering **58**(04), 1 (2019) https://doi.org/10.1117/1.oe.58.4.041608

[4] Oltrogge, D.L., Alfano, S., Law, C., Cacioni, A., Kelso, T.S.: A comprehensive assessment of collision likelihood in Geosynchronous Earth Orbit. Acta Astronautica **147**, 316–345 (2018) https://doi.org/10.1016/j.actaastro.2018.03.017

[5] U.S. Space Command: Space-Track.org: Two-Line Element (TLE) Earth Satellite Orbit Data. Space-Track, https://www.space-track.org/ (2024). https://www.space-track.org/ Accessed 2026-07-01

[6] Vallado, D.A., Crawford, P., Hujsak, R., Kelso, T.S.: Revisiting Spacetrack Report #3. In: AIAA/AAS Astrodynamics Specialist Conference. American Institute of Aeronautics and Astronautics, Keystone, CO (2006). https://doi.org/10.2514/6.2006-6753 . https://doi.org/10.2514/6.2006-6753

[7] Levit, C., Marshall, W.: Improved orbit predictions using two-line elements. Advances in Space Research **47**(7), 1107–1115 (2011) https://doi.org/10.1016/j.asr.2010.10.017

[8] Hallgarten La Casta, M.I., Amato, D.: Debiasing of two-line element sets for batch least squares pseudo-orbit determination in MEO and GEO. Advances in Space Research **75**(10), 7259–7289 (2025) https://doi.org/10.1016/j.asr.2025.03.046

[9] Sergeev, E., Viquerat, A., Bridges, C.P.: Eccentricity-Driven Area and Mass Estimation of Geostationary Debris: A New Evaluation and Classification. In: 5th IAA Conference on Space Situational Awareness (ICSSA). IAA, Tres Cantos, Madrid (2026). https://openresearch.surrey.ac.uk/esploro/outputs/conferenceProceeding/Eccentricity-Driven-Area-and-Mass-Estimation-of/991114691602346?institution=44SUR_INST

[10] Anderson, P.V.: Characterizing Longitude-Dependent Orbital Debris Congestion in the Geosynchronous Orbit Regime. PhD, University of Colorado, Boulder, Colorado (2015). https://scholar.colorado.edu/concern/graduate_thesis_or_dissertations/1n79h450x

[11] Allan, R.R.: Perturbations of a geostationary satellite by the longitude-dependent terms in the earth's gravitational field. Planetary and Space Science **11**(11), 1325–1334 (1963) https://doi.org/10.1016/0032-0633(63)90239-3

[12] Inter-Agency Space Debris Coordination Committee (IADC): Space Debris Mitigation Guidelines. Technical Report IADC-02-01, Inter-Agency Space Debris Coordination Committee (IADC) (February 2025). https://www.unoosa.org/res/oosadoc/data/documents/2025/aac_105c_12025crp/aac_105c_12025crp_9_0_html/AC105_C1_2025_CRP09E.pdf

[13] Lemmens, S.: Classification of Geosynchronous Objects. Technical Note GEN-DB-LOG-00211-OPS-GR, ESA Space Debris Office, Darmstadt (May 2018)

[14] European Space Agency: DISCOSweb - Database and Information System Characterising Objects in Space, https://discosweb.esoc.esa.int/ (2026). https://discosweb.esoc.esa.int/

[15] CS-SI, Airbus Defense and Space, CS Group, Ansys Government Initiatives (AGI): CS-SI/Orekit: 13.1. Zenodo (2025). https://doi.org/10.5281/zenodo.16694949

[16] Holmes, S.A., Featherstone, W.E.: A unified approach to the Clenshaw summation and the recursive computation of very high degree and order normalised associated Legendre functions. Journal of Geodesy **76**(5), 279–299 (2002) https://doi.org/10.1007/s00190-002-0216-2

[17] Anderson, P.V., Schaub, H.: Local debris congestion in the geosynchronous environment with population augmentation. Acta Astronautica **94**(2), 619–628 (2014)

https://doi.org/10.1016/j.actaastro.2013.08.023

[18] Huber, P.J.: Robust Estimation of a Location Parameter. The Annals of Mathematical Statistics **35**(1), 73–101 (1964) https://doi.org/10.1214/aoms/1177703732

[19] Bate, R., Mueller, D., White, J.: Fundamentals of Astrodynamics. Dover Books on Aeronautical Engineering. Dover Publications Inc., New York (1971)

[20] Früh, C., Schildknecht, T.: Variation of the area-to-mass ratio of high area-to-mass ratio space debris objects: HAMR space debris objects. Monthly Notices of the Royal Astronomical Society **419**(4), 3521–3528 (2012) https://doi.org/10.1111/j.1365-2966.2011.19990.x

[21] Früh, C., Jah, M., Kelecy, T.: Coupled Orbit-Attitude Dynamics of High Area-to-Mass Ratio (HAMR) Objects: Influence of Solar Radiation Pressure, Earth's Shadow and the Visibility in Light Curves. Celestial Mechanics and Dynamical Astronomy **117**(4), 385–404 (2013) https://doi.org/10.1007/s10569-013-9516-5