GEODESIC SPACE SYSTEMS · TECHNICAL NOTE

Verifying the Geodesic engine v1.8.6 against independent implementations

> **Scope.** Every result below is a statement about **Geodesic v1.8.6**. The references are **Orekit 13.1.9** and **GMAT R2026a**, named wherever they are used.

Abstract

An astrodynamics engine that agrees with itself proves nothing. There is a kind of anchor that a closed-form check and a Monte Carlo cannot supply: **an independent implementation of the same physics**, written by other people, which can disagree with you.

This note reports what happened when Geodesic v1.8.6 was compared, term by term, against **Orekit 13.1.9** across libration-point orbits, a matched Earth-centred force model, collision probability, relativistic precession and the non-conservative forces, with **GMAT R2026a** as a third opinion.

The agreement figures are given below. They are not the interesting part. The comparison found defects on both sides — and, as the published summary of it puts it: *a second implementation is worth having because it can disagree, and here it disagreed with us where we were wrong*.

1. Why a second implementation, rather than more of our own tests

Our engine carries structural self-checks — orbit closure, Jacobi-constant drift, an internal consistency comparison against a separately derived analytic reference. In one documented case **every one of them passed on an answer that was wrong**: asked for planar Lyapunov orbits about the Earth–Moon L1 and L2 points, the corrector returned a real, closed, periodic orbit from a *different family*. Closure passed. Jacobi drift passed. Internal consistency passed. The acceptance gate checked that the orbit closed and that it started near the libration point, and both are true of the orbit it wrongly returned.

What caught it was comparison against a published external catalogue — and, separately, Orekit, which reproduced the same defect from its own independent implementation.

The same shape recurs twice more below: in §3.3, where an independent implementation exposed a frame error our own tests passed, and in §3.4, where it falsified a derivation before we had compiled it.

2. Method: five rules, each of which has caught something

2.1 A reference that has never reproduced something known is not a reference. Before Orekit 13.1.9 judged anything in the force-model comparison, its configuration was validated twice offline. Against an exact Kepler solution over a full day it closed to **6.95e-9 km**. Its plumbing was checked by a limit test: the closed-form nodal rate is first-order in J2 and uses mean elements, so against an osculating propagation it differs by order J2 — measured 4.259e-3, which alone proves nothing.

Scaling J2 down shrank the difference in proportion — 4.259e-3, 4.248e-4, 4.244e-5, 4.211e-6 — each decade dividing by about 10.03.

2.2 Decompose until a disagreement is attributable. A single end-to-end drag figure mixes the force law, the relative velocity, the geodesy and the atmosphere model; a disagreement in it tells you nothing about which is wrong. The drag and SRP comparison was therefore split into six arms, with the density *injected* — constant in one arm, analytic in geodetic altitude in the next — removing the atmosphere model from the force-law arms entirely, and with the Sun supplied as a fixed vector to both sides. Each arm can then fail for one reason, and the failure signatures are distinctive: a unit error shows as a clean factor of 1000, ignoring co-rotation as about 13%, a flattening error as about 10 km of altitude.

2.3 Verify the constants before comparing the physics, or a result is just a constant difference. Orekit's solar radius and its IAU 2012 astronomical unit were confirmed equal to ours to the digit beforehand. One constant is *not* shared — Earth's rotation rate, ours against Orekit's sidereal value, 1.2e-8 relative — so the harness was made to return our value, making that arm a test of the force law rather than of whose constant is right.

2.4 A planted fault must actually land. In the force-model comparison, deliberately doubling the relativistic term trips the relativity arm and **not** the zonal arm. A check that cannot be made to fail is not evidence.

2.5 Predictions are registered and committed before the run, with the value that would refute them, so a result is graded against what was expected rather than against what turned out convenient.

3. Results — Geodesic v1.8.6 against Orekit 13.1.9

3.1 Libration-point orbits

In the cislunar routes the comparison **agrees with JPL's published halo families to 3e-10** — and Orekit independently reproduced the planar-Lyapunov defect we had already found against JPL, which is the more useful half of the result.

That branch was subsequently repaired. In v1.8.6 it returns a genuine member of those same JPL families or it refuses to answer — never an orbit of another family — within published amplitude bounds.

3.2 A matched Earth-centred force model

Orekit 13.1.9, run at a matched force model over 50 cases in five regimes — sun-synchronous, ISS-like, eccentric LEO, GTO and geostationary — agrees with the engine on every one to within the engine's own integrator floor.

The residual is not a model difference but our fixed-step truncation error, and we show that rather than assert it: it runs **1.26e-5 km** over a LEO revolution to **9.85e-1 km** over a GTO revolution, where Orekit's own floor is **6.95e-9 km** — five orders tighter, so our side limits the comparison throughout, and a Richardson estimate of the GTO error agrees with the measured disagreement to **1%**.

Differencing the rungs isolates single terms: the relativistic correction and the J3–J8 zonals each agree to about **8e-5 relative** — 8.476e-5 and 8.414e-5 respectively, and 6.2e-9 to 8.6e-8 on the five non-GTO cases. The isolated terms themselves range from 0.17 m to 11.7 m (relativistic) and 0.85 m to 2.79 km

(zonal), so that is eight or nine significant figures on a perturbation term wherever the step resolves the dynamics.

The control held. With the zonal field off and relativity disabled there is no force model left to disagree about, so those cases were checked against the exact analytic Kepler solution instead: none of the ten exceeded its own floor. Had that failed, nothing downstream would have meant anything.

3.3 Collision probability, and the defect this comparison found

v1.8.6 carries each object's position covariance out of **its own** radial/along-track/cross-track frame into inertial coordinates **before** the two are summed. That assembly agrees with Orekit to **1.3e-14 across 234 real conjunctions.**

It is assembled that way because this comparison proved the earlier arrangement wrong. A previous generation summed the two covariances without first rotating each out of its own frame. On those same 234 conjunctions the consequence was not a rounding difference: **43 cases were mis-tiered — 41 reported one risk tier safer than the truth and 2 one tier more dangerous** — and 21 returned a probability of exactly zero, one of which was a case whose correct probability placed it a tier higher. Our own tests passed throughout. The comparison is what found it.

What v1.8.6 does, measured on the deployed path — 234 cases through the deployed modules and the real enclave:

```
median ratio against the reference    1.000000
mis-tiered                            0 of 234      was 43
exact zeros                           0 of 234      was 21
worst deviation from offline replay   6.181e-14     median 2.435e-15
```

A second arm put 60 discrete-hard-body cases through the same path, with the control that the *old* assembly through the enclave reproduces the previously recorded answers **bit-identically** — median and maximum difference both exactly zero — so the comparison is between two known states and not between two unknowns.

> **The power of that deployed test, stated as it was registered: it validates the DEPLOYMENT, not the physics.** The deployed file and the replayed file are the same source, so an error shared by both would pass both. The physics evidence is the Orekit agreement above; the deployment evidence is the 234-case run. They are two different claims and neither substitutes for the other.

3.4 Relativistic precession

The published verification of the de Sitter term checks it against the known geodetic precession rate, because it has no convenient closed form in this configuration. This is the independent check alongside it.

The de Sitter term agrees with Orekit to **4.441e-16 relative on every component of five cases — two units in the last place, sign included.**

The more useful part is the order of events. Orekit's own de Sitter implementation was built into a probe and fed the same inputs *before* our term was compiled, and it returned **the same magnitude with the**

opposite sign — a plane precessing backwards. The external oracle falsified our derivation before a line of it had been built. The sign was traced to the Coriolis factor carried inside the IERS 2010 expression and corrected before anything was compiled.

3.5 Drag and solar radiation pressure

Six arms, each isolating one mechanism:

```
G1   geodesy                          altitude 3.711e-10 m, latitude/longitude 4.512e-13 deg
D1   drag force law, density fixed    4.441e-16 relative  (two units in the last place)
D2   density analytic in altitude     2.653e-14
S1   SRP cannonball, full sunlight    7.772e-16
S2   dual-cone shadow                 full sun exact, umbra exact, and all ten penumbra
                                      samples agreeing to six decimal places in the fraction
M1a  NRLMSISE-00, 414 of 477 points   1.427e-13
M1b  the remaining 63 points          bounded by one unit in the last place
```

The one disagreement, and why it is the model's own conditioning rather than an error. A registered gate of 1e-9 failed at 7 of 47 altitudes, worst 3.107e-04, while the other 40 agreed to 1e-14. The scatter was sporadic in altitude and concentrated at mid-latitudes, which rules out a wrong coefficient — that would show everywhere — and rules out the altitude-growing hypothesis that had been registered in advance.

A trichotomy was registered before measuring: ill-conditioning, a coefficient or branch difference, or neither. **Ill-conditioning is supported, and not marginally.** At each failing point, perturbing a single input by one unit in the last place moves our own density by essentially exactly the amount the two implementations disagree by:

```
altitude  latitude   density          one-ulp move    disagreement
160 km      +45      1.296212e-09      3.106e-04       3.107e-04
300 km      +45      3.440369e-11      8.163e-05       8.163e-05
450 km      +45      1.807045e-12      2.437e-05       2.437e-05
250 km      -60      8.544892e-11      2.491e-06       2.491e-06
170 km      +45      8.725044e-10      8.882e-16       0.000e+00
400 km      +45      5.836149e-12      7.772e-16       1.499e-14
```

The last two rows are the control: where the model is well conditioned, a one-ulp move does nothing and the two implementations agree to 1e-14 or exactly.

4. A third opinion: GMAT R2026a

GMAT R2026a was added as a second opinion and corroborates Orekit exactly at two-body. Beyond two-body that comparison is not yet matched, and the reason is identified rather than waved at: the two sides reduce the body-fixed field axis differently, which accounts for **99.93%** of the gap. The remaining **3.0e-5 km — about three centimetres — has no named cause**: GMAT's nutation staircase, its own zonal integration floor and the segment rule were each measured and excluded. It is ten times GMAT's floor and five orders below our own fixed-step truncation at GTO, so it is reported rather than chased.

5. What this does not establish

- **Not the customer's half.** For drag and SRP the caller supplies the ballistic and reflectivity coefficients and the space-weather indices. Those inputs cannot be verified by us or by anyone else, and no attempt was made to.
- **Not every regime.** The force-model comparison covers five orbit regimes at one step size per case. The GTO case is limited by our own step, not by agreement.
- **Not matched beyond two-body against GMAT**, as §4 states, with 3.0e-5 km unexplained.
- **Not an endorsement.** Orekit and GMAT are independently maintained tools that were measured against. Neither project has reviewed, validated or endorsed this engine or this document.
- **Not a claim about the atmosphere model's accuracy** — only that two implementations of it agree where it is well conditioned, and that where they do not, the cause is located.
- **Not flight certified**, and carrying no airworthiness, launch or operational authority.

6. Reproduction

The libration-point comparison uses JPL's published Earth–Moon families. The collision-probability work is graded against a publicly released conjunction dataset with its own answer key. Orekit 13.1.9 and GMAT R2026a are both openly available and the versions are stated above. Where a figure depends on our own configuration, the configuration is described rather than referenced.

The service is advisory. The trajectory and manoeuvre decisions stay with the operator's own flight dynamics team.