Supplementary Information: State-conditioned residual transfer reduces TLE/SGP4 satellite-orbit error

Xurui Zhang[1], Yiting Tong[2], Yingxi Ding[3], Tianyu Li[4*], Yingjie Zhang[1,4*], Zhengwen Zhang[5]

[1]Detroit Green Technology Institute, Hubei University of Technology, Wuhan, Hubei, 430068, China.
[2]School of Computer Science and Artificial Intelligence, Hubei University of Technology, Wuhan, Hubei, 430068, China.
[3]National Key Laboratory of Radar Signal Processing, Xidian University, Xi'an, China.
[4]Department of Applied Physics, The Hong Kong Polytechnic University, Hong Kong, China.
[5]School of Electrical and Electronic Engineering, Hubei University of Technology, Wuhan, Hubei, 430068, China.

*Corresponding author(s). E-mail(s): tianyu.li@connect.polyu.hk; yingjieyj.zhang@connect.polyu.hk;

The supplementary tables provide numerical and procedural detail for the main-text evidence. Tables S1–S3 report arc-level correction, component and transfer-history results. Tables S4 and S5 document the information boundary and frozen implementation, while Tables S6 and S7 provide the source-mapping, fusion and seed controls. Pointwise temporal and directional analyses remain in main-text Figure 6.

Supplementary Table S1. Arc-level 3D position RMSE before and after CROSS correction

Table S1 provides the complete arc-level values underlying Figure 1. Repeating each held-out arc across the three target-history budgets makes the within-arc response to additional pre-test support directly readable without compressing the results into a wide matrix.

Evaluation arc	Target history (h)	Uncorrected RMSE (m)	CROSS RMSE (m)	Reduction (%)
6–7 June 2026	24	793.65	235.71	70.3
	72	793.65	271.54	65.8
	120	793.65	217.01	72.7
8–9 June 2026	24	785.66	189.01	75.9
	72	785.66	145.84	81.4
	120	785.66	143.57	81.7
10–11 June 2026	24	688.49	465.76	32.4
	72	688.49	390.96	43.2
	120	688.49	261.99	61.9
12–13 June 2026	24	750.67	405.04	46.0
	72	750.67	233.78	68.9
	120	750.67	191.96	74.4
14–15 June 2026	24	889.04	323.48	63.6
	72	889.04	333.80	62.5
	120	889.04	303.25	65.9
16–17 June 2026	24	823.74	306.67	62.8
	72	823.74	226.13	72.5
	120	823.74	205.18	75.1
18–19 June 2026	24	769.65	317.82	58.7
	72	769.65	264.02	65.7
	120	769.65	235.88	69.4

Note. Each row represents one non-overlapping 48 h held-out arc at the stated target-history budget. Reduction is $100 \times (\mathrm{RMSE}_{raw} - \mathrm{RMSE}_{CROSS})/\mathrm{RMSE}_{raw}$.

Supplementary Table S2. Paired 3D RMSE changes after CROSS component removal

Table S2 expands the component responses plotted in Figure 3. Each contrast is first calculated within the same held-out arc and is then summarized across seven arcs, so changes remain paired to the frozen complete-CROSS prediction at the same target-history budget.

Target history (h)	Component removed	Mean RMSE increase (m)	95% paired interval (m)	Arcs favouring full CROSS
24	Temporal refinement	54.57	[-16.84, 125.98]	6/7
24	Causal GRU branch	41.65	[-26.87, 110.17]	6/7
24	Orbit-frame branch	8.45	[-2.29, 19.19]	5/7
72	Temporal refinement	17.95	[-10.76, 46.65]	6/7
72	Causal GRU branch	4.27	[-26.64, 35.18]	4/7
72	Orbit-frame branch	1.24	[-13.49, 15.96]	5/7
120	Temporal refinement	20.52	[-1.49, 42.54]	5/7
120	Causal GRU branch	5.20	[-16.76, 27.15]	4/7
120	Orbit-frame branch	22.22	[-10.69, 55.13]	6/7

Note. Positive values indicate that removing the named component increased 3D RMSE. Intervals are two-sided 95% t intervals for paired same-arc differences; $n = 7$ arcs per budget.

Supplementary Table S3. 3D position RMSE by source contribution and target-history budget

Table S3 supplies the numerical estimates behind the source-transfer and extended-history contrasts in Figure 2c,d. The 24, 72 and 120 h comparisons share the common 1–20 June archive. The 240 h row comes from the separately locked, deduplicated May–June archive and therefore marks a sensitivity boundary rather than another point on the common response curve.

After duplicate resolution, the extended archive contains 135 unique satellite–epoch pairs and 9,229 aligned 300 s rows. Its 5,707 rows overlapping the common June table are identical.

Target history (h)	Source-only prior mean [95% interval] (m)	Target-only prior mean [95% interval] (m)	Source + target prior mean [95% interval] (m)	Full CROSS mean [95% interval] (m)	Protocol role
0	604.98 [531.21, 678.75]	–	–	–	Source-only fallback; full CROSS structurally inapplicable
24	604.98 [531.21, 678.75]	578.03 [451.66, 704.39]	375.07 [301.24, 448.90]	320.50 [233.63, 407.37]	Common comparison budget
72	604.98 [531.21, 678.75]	306.80 [226.25, 387.34]	284.53 [218.88, 350.18]	266.58 [193.65, 339.52]	Common comparison budget
120	604.98 [531.21, 678.75]	249.48 [190.06, 308.90]	243.22 [185.80, 300.64]	222.69 [175.30, 270.08]	Common comparison budget
240	604.98 [531.21, 678.75]	233.44 [165.13, 301.75]	229.06 [167.55, 290.56]	221.88 [163.05, 280.70]	Separately locked extended-history boundary

Note. Values are seven-arc mean 3D position RMSE with two-sided 95% t intervals. The source-only estimate is repeated to show its fixed zero-history reference. The uncorrected SGP4/TLE mean is 785.84 m [728.43, 843.26].

Supplementary Table S4. Data provenance, temporal partition and model visibility

Table S4 records the information boundary used throughout the rolling-origin evaluation. The permitted target history B ends strictly before the held-out origin t_0. Source data remain available during full-history refitting, whereas future target POD becomes visible only after the corrected trajectory has been generated.
The common archive contains 5,707 aligned 300 s rows and 171,204 native 10 s rows. The extended archive contains 9,229 aligned rows and supplies 2,880 pre-test rows for each 240 h arc. Its TLE merge gives June records priority at four duplicate satellite–epoch keys; exactly repeated AUX_POEORB epochs retain the first filename-sorted record.

Partition	Satellite / product	Cadence	Relation to t_0	Model use	POD visible?
Source fitting archive	GRACE-C/D TLE/SGP4 + GNV1B	300 s	Historical source interval	Fit transfer prior, RTN refiner and causal GRU	Yes, before target test
Source validation	GRACE-C/D TLE/SGP4 + GNV1B	300 s	Historical interval disjoint from source fitting rows	GRU early stopping only	Yes, before target test
Common-budget target archive	Sentinel-1A TLE/SGP4 + AUX_POEORB	300 s and 10 s	1–20 June support; 24, 72 and 120 h conditions	Supply pre-test blocks and the native held-out scoring archive	Yes only before each test; withheld during inference
Extended-history target archive	Sentinel-1A TLE/SGP4 + AUX_POEORB	300 s	19 May–20 June support; 240 h condition only	Supply 240 h pre-test blocks; held-out arcs remain unchanged	Yes, pre-test only
Target fitting block	Sentinel-1A TLE/SGP4 + AUX_POEORB	300 s	First 60% of permitted B	Candidate-component fitting	Yes, pre-test only
RTN selection block	Sentinel-1A TLE/SGP4 + AUX_POEORB	300 s	Middle 20% of permitted B	Select refiner family and branch activation	Yes, pre-test only
Fusion calibration block	Sentinel-1A TLE/SGP4 + AUX_POEORB	300 s	Final 20% of permitted B	Calibrate $\alpha_R, \alpha_T, \alpha_N$	Yes, pre-test only
Full-history refit	Source archive + complete permitted target history	300 s	Completed before t_0	Refit prior, selected RTN refiner and causal GRU	Yes, pre-test only
Frozen held-out evaluation	Sentinel-1A TLE/SGP4; AUX_POEORB withheld until scoring	10 s	$[t_0, t_0 + 48\,\mathrm{h})$	Frozen position correction and post-hoc scoring	No during inference; yes after

Note. POD denotes precise-orbit determination; B is the permitted pre-test target POD history and t_0 is the held-out origin.

Supplementary Table S5. Model, optimization and selection settings

Table S5 collects the model families, causal inputs and frozen optimization settings used by CROSS and its registered comparators. Settings shown as fixed were chosen before outer evaluation; only the explicitly identified RTN configuration and fusion weights were selected from permitted pre-test target blocks.
Three neural seeds (2026076201, 2026076202 and 2026076203) were used for the direct neural comparators. The full CROSS implementation used deterministic seeds derived separately for each arc, budget and component fit; these remain encoded in the archived scripts and lock manifests.

Component or system	Primary role	Model and input context	Fitting / optimization	Selection or fixed setting
CROSS transfer prior	Source-derived state-dependent ECEF residual correction	ExtraTrees; current TLE/SGP4 state, norms, phase and age features	80 trees; unrestricted depth; minimum leaf 2; source weight 1; target weight 5	Family and hyperparameters fixed before outer evaluation
CROSS RTN refiner	Residual-of-residual refinement	Current and two 300 s-lag prior estimates, finite differences and state/orbit features; ECEF or RTN target representation	Ridge penalty 10 or ExtraTrees with 180 trees, unrestricted depth, minimum leaf 8; source weight 1; target weight 10	Representation, family and activation selected on middle 20% block
CROSS auxiliary GRU	Complementary full ECEF residual estimate	Three epochs at t, $t-300$ s and $t-600$ s; shared causal state features	One GRU layer; hidden 64; dropout 0.05; AdamW, learning rate 10^{-3}, weight decay 10^{-4}; batch 128; at most 25 epochs; patience 4; target weight 5	Source validation controls early stopping; refitted on complete permitted history
Direction-wise fusion	Combine refiner and GRU in RTN	Three convex weights $\alpha_R, \alpha_T, \alpha_N$	Prior centre $\alpha_0 = 0.5$; prior strength $N_0 = 24$; clipped to [0,1]	Calibrated on final 20% block and frozen at t_0
Linear residual calibration	Linear comparator	Standardized common state features	Ridge penalty 10; source weight 1; target weight 5	Fixed configuration
State-space residual filter	Target-history forecasting comparator	Three-axis residual state	Six pre-specified random-walk, constant-velocity and damped-velocity configurations	Configuration selected on pre-test target gate
Feedforward neural correction	Direct neural comparator	Flattened three-epoch causal window	Hidden width 64; dropout 0.05; optimizer and early stopping as GRU; three fixed seeds	Mean over seeds within arc
Recurrent neural correction	Direct recurrent comparator	Three-epoch causal state sequence	One GRU layer; hidden 64; dropout 0.05; optimizer and early stopping as CROSS GRU; three fixed seeds	Mean over seeds within arc; no CROSS prior, RTN refiner or fusion

Note. ECEF denotes the Earth-fixed frame; RTN denotes radial, transverse/along-track and normal coordinates. Hyperparameters not identified as selected were held fixed across outer arcs.

Supplementary Table S6. Post-hoc controls of source state–residual correspondence

Table S6 tests whether the source benefit depends on its original state–residual correspondence. Random pairing uses a zero-fixed-point derangement within satellite and UTC day, while the temporal disruption shifts each intact three-component residual vector by nine samples on the 300 s grid. Both controls preserve feature rows, sample counts, residual marginals, sample weights, tree architecture and matched deterministic fit seeds.

B (h)	Destroyed mapping	Favourable arcs	Mean difference (m)	Median difference (m)	Leave-one-arc-out mean range (m)	Consecutive-three-arc mean range (m)
0	Random within day	7/7	-190.51	-182.33	-194.68 to -185.08	-208.71 to -176.26
0	Shift by 45 min	7/7	-424.30	-414.66	-430.56 to -414.58	-457.24 to -397.94
24	Random within day	7/7	-205.25	-226.88	-227.99 to -190.95	-264.96 to -189.41
24	Shift by 45 min	7/7	-149.78	-134.53	-170.56 to -114.51	-201.54 to -104.87
72	Random within day	7/7	-45.53	-51.04	-53.08 to -38.99	-68.56 to -35.66
72	Shift by 45 min	7/7	-40.11	-37.94	-46.60 to -34.01	-54.28 to -32.02
120	Random within day	5/7	-6.78	-6.69	-8.11 to -4.19	-8.00 to -0.39
120	Shift by 45 min	6/7	-9.27	-7.32	-10.98 to -6.84	-11.73 to -1.94

A matched weighted-Ridge control produced positive source-plus-target minus target-only median differences at every positive budget in both ECEF and RTN, with only 1–3 of seven arcs favourable. The aligned tree prior outperformed Ridge on all seven arcs at every budget and in both representations; mean tree-minus-Ridge differences ranged from -287.97 to -447.13 m.

Note. Differences are correctly aligned minus disrupted mapping; negative values favour correct correspondence. Five predeclared mappings or matched fits were averaged within arc. All controls were evaluated after the formal method was frozen.

Supplementary Table S7. Post-hoc fusion and recurrent-model contrasts

Table S7 compares the frozen direction-wise shrinkage rule with equal, global and unshrunk alternatives, then reports the seed-averaged CROSS-minus-recurrent contrasts. Fusion alternatives reuse the same frozen branch predictions and pre-test calibration rows; neural seeds are averaged within arc and are not additional experimental units.

Contrast	B (h)	Favourable arcs	Mean difference (m)	Median difference (m)	Leave-one-arc-out mean range (m)	Consecutive-three-arc mean range (m)
Direction-wise shrunk minus fixed 0.5	24	4/7	-1.36	-1.30	-10.38 to 6.22	-21.76 to 25.92
Direction-wise shrunk minus fixed 0.5	72	1/7	9.83	3.49	5.85 to 12.55	0.77 to 22.03
Direction-wise shrunk minus fixed 0.5	120	3/7	-3.49	0.51	-6.63 to 1.61	-10.60 to 8.37
Direction-wise shrunk minus global shrunk	24	7/7	-3.59	-3.74	-4.18 to -2.26	-4.17 to -1.57
Direction-wise shrunk minus global shrunk	72	6/7	-0.51	-0.72	-0.92 to -0.24	-0.69 to 0.01
Direction-wise shrunk minus global shrunk	120	4/7	-0.07	-0.08	-0.24 to 0.09	-0.50 to 0.36
Direction-wise shrunk minus direction-wise unshrunk	24	6/7	-2.63	-1.18	-3.31 to -0.72	-5.55 to -0.29
Direction-wise shrunk minus direction-wise unshrunk	72	6/7	-0.49	-0.44	-0.67 to -0.27	-0.90 to -0.19
Direction-wise shrunk minus direction-wise unshrunk	120	5/7	-0.04	-0.06	-0.21 to 0.07	-0.31 to 0.10
CROSS minus recurrent correction	24	5/7	-13.19	-8.39	-17.64 to -2.98	-32.73 to -1.87
CROSS minus recurrent correction	72	5/7	-7.41	-6.85	-15.54 to -2.18	-30.54 to 9.48
CROSS minus recurrent correction	120	6/7	-30.21	-20.16	-35.31 to -16.61	-56.29 to -10.74

Across the three CROSS seeds, seven-arc mean RMSE ranged from 319.58 to 323.10 m at 24 h, from 263.39 to 266.58 m at 72 h and from 215.31 to 222.69 m at 120 h. Mean calibrated $(\alpha_R, \alpha_T, \alpha_N)$ values were (0.528, 0.979, 0.451), (0.229, 0.671, 0.507) and (0.137, 0.538, 0.481) at 24, 72 and 120 h.

Note. Differences are method named first minus comparator; negative values favour the first method. Arc-level comparisons use seed-averaged values. Leave-one-arc-out and consecutive-three ranges are descriptive sensitivity summaries over the seven evaluated arcs.