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.

Real-Time Continuous Neural Correction of Reduced- Dynamic LEO Orbit Predictions for LEO-PNT

Amir Allahvirdi-Zadeh

amir.allahvirdizadeh@curtin.edu.au

Curtin University

Ahmed El-Mowafy

Curtin University

Sheik Fattah

Curtin University

Research Article

Keywords: LEO-PNT, LEO orbit prediction, Real-time neural correction, Reduced-dynamic orbit prediction, Gated Recurrent Unit (GRU), Hybrid physics-machine learning

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10119188/v1

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

Additional Declarations: No competing interests reported.

Real-Time Continuous Neural Correction of Reduced-Dynamic LEO Orbit Predictions for LEO-PNT

Amir Allahvirdi-Zadeh[1], Ahmed El-Mowafy[1], Sheik Fattah[2]

[1] School of Earth and Planetary Sciences, Curtin University, GPO Box U1987, Perth, Western Australia 6845

[2] School of Electrical Engineering, Computing & Mathematical Sciences (EECMS), Curtin University, Australia

Correspondence: Amir Allahvirdi-Zadeh, School of Earth and Planetary Sciences, Curtin University, GPO Box U1987, Perth, WA 6845. Email: amir.allahvirdizadeh@curtin.edu.au

ORCID: 0000-0002-3722-4417

Abstract

Low Earth orbit positioning, navigation, and timing (LEO-PNT) services broadcast predicted orbits whose errors grow since the last orbit fit, mainly due to hard-to-model atmospheric drag, so a physics-only prediction soon becomes too inaccurate for broadcast. We present continuous neural correction of reduced-dynamic orbit prediction (CNC-RDOP), a deep neural network added on top of a geodetic-standard reduced-dynamic prediction, leaving the orbit determination unchanged. A gated recurrent unit network reads the last 30 minutes of prediction error, the predicted orbit minus a real-time kinematic position solution in the along-track, cross-track, and radial frame, plus the predicted position and velocity, and forecasts the error over the next 60 minutes. The forecast is refreshed at every measurement epoch, so the broadcast orbit is updated continuously rather than once per fit. The network is trained on GRACE-FO C orbits (2022-2023), validated on 2024, and tested on an independent 2025 set; the kinematic solution serves only as input, while a precise post-processed orbit provides the reference, so the reported accuracy reflects real-time operation. Under a hybrid broadcast rule, the 5-minute error stays at about 2-31 m for up to 6 hours after a fit, against a physics-only growth from about 2 to 463 m. Within the first 10 minutes the physics prediction is already about 2 m and broadcast directly; the neural correction applies beyond it. Applied to the companion satellite GRACE-FO D without retraining, the same network reproduces these results to within about 1%, so one network can serve co-orbital LEO-PNT satellites in real time.

Keywords: LEO-PNT; LEO orbit prediction; Real-time neural correction; Reduced-dynamic orbit prediction; Gated Recurrent Unit (GRU); Hybrid physics-machine learning

1. Introduction

Precise orbit prediction for low Earth orbit (LEO) satellites underpins a wide range of mission-critical applications, including Earth observation, space situational awareness, satellite-based augmentation systems, and precision geodesy. The emergence of LEO-based positioning, navigation, and timing (LEO-PNT) systems, whether opportunistic (e.g., Starlink, OneWeb, and Iridium) or dedicated (e.g., Xona Pulsar), has intensified the demand for high-accuracy real-time orbit determination and prediction capabilities (Ge et al. 2022). Recent studies demonstrate that LEO satellites can provide positioning accuracy at the decimeter to meter level when combined with precise orbit and clock products (El-Mowafy et al. 2023; Ge et al. 2018; González et al. 2022), while Doppler-based techniques have shown promise for instantaneous positioning under challenging observation conditions (Allahvirdi-Zadeh et al. 2025a, b). However, gaps in observation data and intermittent tracking, where LEO satellites pass over the same tracking location at short intervals, remain significant challenges for LEO-based navigation services, making robust orbit prediction essential for service continuity. The broadcast ephemerides delivered to users must be timely, continuously refreshed, and as accurate as possible. Unlike legacy GNSS satellites in near-stationary medium Earth orbit, which benefit from gravity-field dominance and predictable slowly varying perturbations, LEO satellites are immersed in the upper thermosphere and experience atmospheric drag, geomagnetic-storm-driven density variations, and solar radiation pressure on timescales that present a challenge for classical force-model-based orbit prediction.

For LEO satellites, the dominant perturbations during a prediction interval are atmospheric drag in the along-track direction, geopotential resonances and tesseral harmonics on time scales of one orbital revolution, third-body luni-solar perturbations on the radial component, and solar radiation pressure (Allahvirdizadeh 2022). Among these forces, atmospheric drag is the principal source of prediction-error growth; it scales with the local thermospheric density, which itself responds non-linearly to solar extreme ultraviolet (EUV) flux (parameterised by solar radio flux at 10.7 cm wavelength or $F10.7$) and geomagnetic activity (parameterised by planetary K-index or K_p). Empirical density models such as NRLMSISE (Picone et al. 2002) carry typical uncertainties of 10% to 30%, depending on solar conditions, and even at nominal

accuracy they cannot fully resolve the short-timescale density structure relevant to a 60-min prediction (Liu et al. 2025). Consequently, the prediction error of a reduced-dynamic orbit accumulates to a 3D root mean square error (3D RMSE) of several hundred meters at a 60-min prediction length, far below the broadcast ephemeris quality required by LEO-PNT users.

Within the geodetic community, reduced-dynamic precise orbit determination (RDPOD) is the de facto standard for high-accuracy LEO orbit estimation. As implemented in Bernese GNSS processing software (Dach et al. 2015), RDPOD fits a 24-h arc of GNSS observations using dynamic models. Within the arc, the residual unmodelled forces are absorbed by pseudo-stochastic velocity pulses estimated every 6 min in the along-track, cross-track, and radial (SWR) frame, together with three constant empirical accelerations and six once-per-revolution coefficients per arc (Allahvirdi-Zadeh et al. 2022; Wang et al. 2020). This is a well-established geodetic-standard choice by analysis centres, including the Centre for Orbit Determination in Europe (CODE), as part of their routine precise-orbit products. The same parameterisation, however, is optimised for fitting observations within the arc rather than for extrapolation. When the fitted orbit is propagated beyond the arc end into a 60-min or longer prediction interval, the pulses cannot extrapolate, and the empirical accelerations alone cannot capture the full variability in drag and solar radiation pressure (SRP) within the prediction window.

Notably, machine learning (ML) has emerged as an alternative framework for orbit prediction, with several studies exploiting archives of two-line elements and precise orbit products from CelesTrak®, International GNSS Service (IGS), and CODE (Kozhaya et al. 2021; Selvan et al. 2024; Thammawichai and Luangwilai 2024). Purely data-driven approaches, however, do not enforce physical consistency, extrapolate poorly outside the training distribution, and are particularly vulnerable to geomagnetic-storm conditions. Such purely data-driven methods reduce prediction errors from kilometres to several metres, but as single-shot, physically unconstrained forecasts, they cannot deliver the continuously refreshed metre-level accuracy that real-time LEO-PNT broadcast requires.

Hybrid physics-ML frameworks retain a physically consistent reduced-dynamic propagation as a baseline and delegate only residual error correction to the neural component, improving both accuracy and generalisation. Recent work along these lines includes the long short-term memory (LSTM)-based reduced-dynamic orbit prediction (RDOP) correction (Zhang et al.

2025), the warm-up-window architecture (Li et al. 2025), the particle swarm optimised RBF network (Guo et al. 2025), and the dynamics-regularised neural networks (Scorsoglio et al. 2023). These studies share two design principles: (1) the physics baseline sets the lower bound on achievable prediction error, and (2) capturing residual temporal structure requires explicit sequence modelling. They also share a common operational assumption that constrains their direct deployment in a real-time LEO-PNT context: the neural correction is issued as a single forecast when the physics-based fit concludes, and no further observations are ingested during the prediction interval. Moreover, where a warm-up window is employed, the prediction errors that populate that window must be computed against a precise reference orbit available after the fit boundary, a product that is not available in real time and must be derived offline from post-processed precise orbits. Consequently, these frameworks cannot respond to unmodelled events such as geomagnetic storms, attitude anomalies, or drag spikes that occur after the fit. In addition, the single forecast ages monotonically with elapsed time since the last fit, so a user operating several hours past the most recent fit is in effect relying on a several-hour-old prediction rather than a continuously refreshed ephemeris. The present work removes both constraints: it refreshes the forecast at every measurement epoch and populates the trailing window with a real-time kinematic position solution, accepting its decimetre-to-metre noise in exchange for real-time availability, while reserving the precise post-processed orbit for offline evaluation only.

This work presents continuous neural correction of reduced-dynamic orbit prediction (CNC-RDOP), a deep neural network enhancement of geodetic-standard reduced-dynamic orbit prediction designed for real-time operation of LEO-PNT constellations. Instead of a single forecast issued at the end of the physics-based fit, the framework uses a continuously sliding 30-min trailing window of prediction error computed in real time as the difference between the propagated orbit and a real-time kinematic GNSS position solution to forecast the next 60 min of error trajectory, re-issuing the forecast at each new observation epoch so that the corrected ephemeris is continuously refreshed. The underlying reduced-dynamic fit is expensive and infrequent, whereas this neural refresh is cheap and continuous; the framework's role is to extend the usable life of each fit from minutes to hours. Because the propagator's own accuracy is already at a few-meter level for the first 10 minutes past a fresh fit, the broadcast architecture applies the neural correction only beyond that interval; for the first 10 minutes, the

propagator's orbit is broadcast directly. This hybrid bypass is a deliberate design choice that exploits the propagator's strength at short time-since-fit and the network's strength at long time-since-fit and is one of the framework's contributions. A single network trained on a heavily tracked reference satellite is deployable across the entire LEO-PNT constellation with no retraining (zero-shot), and each satellite can autonomously refresh its broadcast correction in flight, scaling to tens or hundreds of co-orbital satellites without per-satellite retraining or ground-segment latency.

The contributions of this work are: (i) introducing a continuously refreshing neural correction layer over a geodetic-standard reduced-dynamic propagator, deployable zero-shot across co-orbital LEO-PNT satellites; (ii) presenting a stratified evaluation protocol that reports accuracy separately for each combination of time-since-fit and prediction length (the cells of a 2D grid), rather than as a single averaged figure; (iii) developing a sequence of architectural variants that pinpoints the accuracy limit just after a fresh fit and traces it to a training-data imbalance; and (iv) presenting evidence that, beneath this imbalance, a feature-limited noise floor set by the real-time kinematic input precludes decimetre-level ephemeris in real time, together with the system-level changes needed to overcome it.

The remainder of this paper is organised as follows. Section 2 describes the reduced-dynamic prediction baseline, the GRU sequence-to-sequence (Seq-to-Seq) architecture, and the training procedure. Section 3 describes the dataset, experimental configurations, and implementation details. Section 4 presents the prediction accuracy results, architectural studies, zero-shot transfer taking GRACE-FO D as an example, and limitations. Section 5 summarises the findings.

2. Methodology

2.1 Physics-Based Reduced-Dynamic Orbit Prediction

The second-order differential equation of motion for a LEO satellite in an Earth-centred inertial (ECI) frame is (Montenbruck and Gill 2000):

$$\ddot{\mathbf{r}} = -\frac{GM}{|\mathbf{r}|^3}\mathbf{r} + \mathbf{a}_{grav} + \mathbf{a}_{drag} + \mathbf{a}_{srp} + \mathbf{a}_{emp} + \boldsymbol{\epsilon} \tag{1}$$

where $\mathbf{r}$ is the satellite position vector, GM is the Earth's gravitational parameter, $\mathbf{a}_{grav}$ represents higher-order gravitational accelerations from the geopotential model and third-body perturbations. $\mathbf{a}_{drag}$ and $\mathbf{a}_{srp}$ are the accelerations due to atmospheric drag and SRP, respectively. $\mathbf{a}_{emp}$ denotes empirical accelerations estimated during orbit determination, and $\boldsymbol{\epsilon}$ subsumes all unmodeled perturbations. Equation (1) represents the ECI conceptual formulation. For numerical implementation, the equations of motion are expressed in the Earth-centred Earth-fixed (ECEF) rotating frame, where the right-hand side is augmented by Coriolis ($-2\boldsymbol{\omega}_E \times \dot{\mathbf{r}}$) and centrifugal ($-\boldsymbol{\omega}_E \times (\boldsymbol{\omega}_E \times \mathbf{r})$) pseudo-force terms, which read:

$$\ddot{\mathbf{r}} = \nabla U(\mathbf{r}) + \mathbf{f}_{ng} - 2\boldsymbol{\omega}_E \times \dot{\mathbf{r}} - \boldsymbol{\omega}_E \times (\boldsymbol{\omega}_E \times \mathbf{r}) \tag{2}$$

where $\boldsymbol{\omega}_E$ and $\dot{\mathbf{r}}$ are the Earth's angular velocity and satellite velocity vectors, respectively. $\nabla U(\mathbf{r})$ is the gradient of the full geopotential, including the Earth's geopotential together with the third body (luni-solar and planetary) terms, and $\mathbf{f}_{ng}$ collects the non-gravitational accelerations, including atmospheric drag, SRP, and empirical accelerations. Equation (2) is integrated using the classic fourth-order Runge-Kutta method (RK4) at a 30-second step.

The force model used to solve the equation of motion is implemented to geodetic-standard fidelity. The static geopotential is EGM2008 (Pavlis et al. 2012), truncated to degree and order 120, with solid-Earth tide corrections to the degree-2 Stokes coefficients per IERS Conventions 2010 using anelastic Love numbers, ocean tide corrections via the FES2014b spherical-harmonic constituents (Lyard et al. 2021), and pole-tide and ocean-pole-tide corrections per IERS 2010. The third-body accelerations are computed from JPL DE421 ephemerides. Atmospheric density is obtained from NRLMSIS 2.1 (Emmert et al. 2022), driven by daily $F10.7$ and 3-hour a_p indices linearly interpolated to the current epoch. SRP uses a cannonball model with a conical shadow function. The Schwarzschild post-Newtonian point-mass relativistic acceleration is included. The ECI-to-ECEF transformation uses the full IAU 2000A/IAU 2006 chain (precession, nutation, Earth rotation angle, polar motion), with Earth orientation parameters from the IERS C04 series. For the tested satellites, GRACE-FO, the nominal drag area-to-mass ratio is 0.00415 m^2/kg with drag coefficient $C_D = 2.25$, and the SRP area-to-mass ratio is 0.0207 m^2/kg with SRP coefficient $C_R = 1.5$ (Allahvirdi-Zadeh et al. 2024a, b). This full-fidelity force model is adopted as the current best available geodetic practice to ensure that the prediction error contains no fidelity-induced systematic bias. As will be shown in Section 4.4,

the resulting baseline accuracy is in fact insensitive to these non-drag choices and is limited by atmospheric drag, but retaining the geodetic-standard model guarantees that the residuals the network learns are drag-dominated rather than contaminated by avoidable modelling error.

The reduced-dynamic POD produces, in addition to the satellite state, three constant empirical accelerations and six once-per-revolution coefficients in the SWR frame, together with pseudo-stochastic velocity pulses estimated every 6 minutes. Within the prediction interval, the constants and once-per-revolution terms are continued unchanged. The pseudo-stochastic pulses, which absorb residual unmodelled drag and SRP within the arc, do not extrapolate naturally; we hold the final estimated pulse constant for one pulse interval (6 min) past the arc boundary and then linearly decay it to zero over the subsequent 30 minutes. This continuation extends the in-arc drag, and SRP fits within the prediction window while respecting the temporal correlation length of the pulse estimates without imposing a discontinuity at the arc boundary.

In reduced-dynamic POD, the empirical accelerations $\mathbf{a}_{emp}$ are parameterised as piecewise-constant or harmonic corrections in the SWR frame and are estimated together with the satellite state from GNSS carrier-phase measurements. Bernese GNSS processing software implements this approach with a weighted least-squares estimator with iterative outlier screening, producing a fitted state at the arc reference epoch ($\hat{\mathbf{r}}_0, \hat{\dot{\mathbf{r}}}_0$) together with empirical force parameters spanning the arc. RDOP consists of numerically integrating the equations of motion forward using the estimated force parameters (Jäggi et al. 2006):

$$\begin{aligned} \hat{\mathbf{r}}_{RDOP}(t) &= \hat{\mathbf{r}}_0 + \int_{t_0}^{t} \hat{\dot{\mathbf{r}}}(\tau) d\tau \\ \hat{\dot{\mathbf{r}}}(t) &= \hat{\dot{\mathbf{r}}}_0 + \int_{t_0}^{t} \hat{\ddot{\mathbf{r}}}(\tau) d\tau \end{aligned} \tag{3}$$

where τ is the integration variable spanning $[t_0, t]$ and the force model is held fixed at the values estimated during fitting. The RDOP position error at prediction time t is then

$$\mathbf{e}_t = \mathbf{r}_{true}(t) - \hat{\mathbf{r}}_{RDOP}(t) \tag{4}$$

where $\mathbf{r}_{true}$ is the precise post-processed reference, used to define the error for training and evaluation. In real-time operation, this reference is unavailable; the network instead observes $\tilde{\mathbf{e}}_t = \mathbf{r}_{kin}(t) - \hat{\mathbf{r}}_{RDOP}(t)$, where $\mathbf{r}_{kin}$ is the real-time kinematic solution. The two differ by the

kinematic error $\mathbf{r}_{kin} - \mathbf{r}_{true}$ (dm-to-m), negligible against the propagation error except immediately after a fit. The reference error $\mathbf{e}_t$ has a systematic structure dominated by secular growth in the along-track direction from drag mismodelling, with smaller periodic contributions in the cross-track and radial directions from geopotential resonances and luni-solar perturbations. Even with this geodetic-standard fidelity, the predicted-orbit 3D RMSE grows to several hundred metres at longer prediction lengths and long after the last fit (cf. Tables 2 and 3), dominated by atmospheric-drag mismodelling that is not fully absorbed by the empirical pulses. This residual error is the target of the neural correction described in the next sections.

2.2 SWR Coordinate Frame

Orbit prediction errors and neural network corrections are expressed in the local SWR orbital frame. Given satellite position and velocity in the ECEF frame, the SWR frame unit vectors are defined as (Montenbruck and Gill 2000):

$$\hat{r} = \frac{\mathbf{r}}{|\mathbf{r}|} \qquad \widehat{w} = \frac{\mathbf{r} \times \dot{\mathbf{r}}}{|\mathbf{r} \times \dot{\mathbf{r}}|} \qquad \hat{s} = \widehat{w} \times \hat{r} \tag{5}$$

By defining the transformation matrix $R_{SWR} = [\hat{s}^T \quad \widehat{w}^T \quad \hat{r}^T]$ as an orthonormal direction-cosine matrix, a position error vector $\mathbf{e}_{ECEF}$ is expressed in SWR components as

$$\mathbf{e}_{SWR} = R_{SWR}\mathbf{e}_{ECEF} = [e_S \quad e_W \quad e_R]^T \tag{6}$$

The SWR frame is the Gauss (RSW) orbital frame with axes ordered along-track (S), cross-track (W), and radial (R); equivalently, it is the radial/transverse/normal (RTN) frame with T = S (along-track) and N = W (cross-track). This ordering is used throughout. SWR decomposition is physically interpretable and directly relevant to mission operations: along-track errors are dominated by atmospheric drag mismodeling, cross-track errors reflect deficiencies in the out-of-plane force model, and radial errors are most sensitive to geopotential truncation. Operating the neural network in the SWR frame therefore aligns the correction targets with physically distinct error sources. The SWR basis is constructed from the ECEF state for consistency with the integration frame. This differs from the conventional radial/transverse/normal definition built from inertial state by sub-degree rotations that are immaterial to the residual decomposition at the magnitudes considered here.

2.3 Neural Network Architecture: GRU Seq-to-Seq

The neural component of CNC-RDOP is a Seq-to-Seq architecture with a GRU-based encoder and decoder (Cho et al. 2014; Sutskever et al. 2014). The encoder is a two-layer GRU with hidden size 128. At each encoder step t, it receives the nine-dimensional (9-D) input vector

$$\mathbf{x}_t = [\mathbf{r}_t^{ECEF}, \dot{\mathbf{r}}_t^{ECEF}, \tilde{\mathbf{e}}_t^{SWR}] \in \mathbb{R}^9 \tag{7}$$

where $\tilde{\mathbf{e}}_t^{SWR}$ is the real-time residual of Equation (4), formed against the kinematic solution as the only error component observable in real time. The GRU hidden state $\mathbf{h}_{t-1} \in \mathbb{R}^{128}$ carries a compressed representation of how the prediction error has evolved over all preceding epochs. The cell update proceeds as:

$$\mathbf{z}_t = \sigma(\mathbf{W}_z\mathbf{x}_t + \mathbf{U}_z\mathbf{h}_{t-1} + \mathbf{b}_z) \tag{8}$$

$$\mathbf{g}_t = \sigma(\mathbf{W}_r\mathbf{x}_t + \mathbf{U}_r\mathbf{h}_{t-1} + \mathbf{b}_r) \tag{9}$$

$$\tilde{\mathbf{h}}_t = \tanh(\mathbf{W}_h\mathbf{x}_t + \mathbf{U}_h(\mathbf{g}_t \odot \mathbf{h}_{t-1}) + \mathbf{b}_h) \tag{10}$$

$$\mathbf{h}_t = (1 - \mathbf{z}_t) \odot \mathbf{h}_{t-1} + \mathbf{z}_t \odot \tilde{\mathbf{h}}_t \tag{11}$$

where $\tanh$ is the hyperbolic tangent, $\sigma(\cdot)$ is the sigmoid function, $\odot$ denotes element-wise multiplication, and $\mathbf{W}$, $\mathbf{U}$, and $\mathbf{b}$ are learnable weight matrices and bias vectors. The update gate $\mathbf{z}_t$ determines what fraction of the accumulated error-pattern memory is retained from $\mathbf{h}_{t-1}$ versus replaced by new information derived from the current orbital state and prediction error $\mathbf{x}_t$. The reset gate $\mathbf{g}_t$ controls how much of the prior error trajectory is allowed to influence the candidate state $\tilde{\mathbf{h}}_t$. When the prediction error pattern changes abruptly, for example, following a geomagnetic storm that alters thermospheric drag, a small $\mathbf{r}_t$ allows the network to suppress stale memory and adapt to the new error dynamics. The encoder ingests its input sequence at the 30-s observation sampling interval, so the 30-min trailing window corresponds to $T_{enc} = 60$ steps. After these 60 steps, the final hidden state $\mathbf{h}_{60}$ encodes a summary of the error trajectory and is used to initialise the decoder. The decoder then autoregressively generates the sequence of SWR corrections $\hat{\mathbf{e}}_t^{\mathrm{SWR}} = [\hat{e}_S, \hat{e}_W, \hat{e}_R]_t^{\mathrm{T}}$ over the 60-min prediction lengths, with each step conditioned on the correction predicted at the preceding epoch.

The decoder is a single-layer GRU with hidden size 128. Its initial hidden state is formed from the encoder's final 128-D hidden vector through a three-layer feed-forward network of

dimensions 128 → 512 → 256 → 128, each layer an affine map $y = Wx + b$ (as in a design-matrix transformation, but with coefficients learned during training) followed by an element-wise ReLU, $ReLU(x) = max(0, x)$. The intermediate widths 512 and 256 are layer sizes set by convention; only the 128-D endpoints are fixed by the encoder and decoder hidden-state sizes. This learned non-linear mapping re-expresses the encoded error trajectory as the decoder initial state, preventing direct propagation of encoder-specific bias into the prediction phase. At each decoder step $t \in \{1, \dots, T_{\mathrm{dec}}\}$, where $T_{\mathrm{dec}} = 120$ (60 min at 30-s sampling), the cell receives the previously predicted SWR correction $\hat{\mathbf{e}}_{t-1}^{\mathrm{SWR}} \in \mathbb{R}^3$ and updates its hidden state through the same gating mechanism described above. The forecast correction at step t is obtained via a linear output projection:

$$\hat{\mathbf{e}}_t^{\mathrm{SWR}} = \mathbf{W}_{out}\mathbf{h}_t^{\mathrm{dec}} + \mathbf{b}_{out} \tag{12}$$

where $\mathbf{h}_t^{\mathrm{dec}} \in \mathbb{R}^{128}$ is the decoder hidden state at step t. The corrected orbit position at each prediction epoch is then obtained by adding the forecast SWR correction, rotated back to the ECEF frame, to the reduced-dynamic propagation.

The encoder input dimension and the decoder per-step input dimension are configurable. The operational model uses the 9-D/3-D configuration described above. Section 4 reports controlled variants that widen the encoder to 16-D (space-weather and Sun-geometry features) and the decoder per-step input to 17-D (propagated state and physics-context features), and add a learned per-sample gate on the decoder output. Figure 1 illustrates the complete architecture of CNC-RDOP.

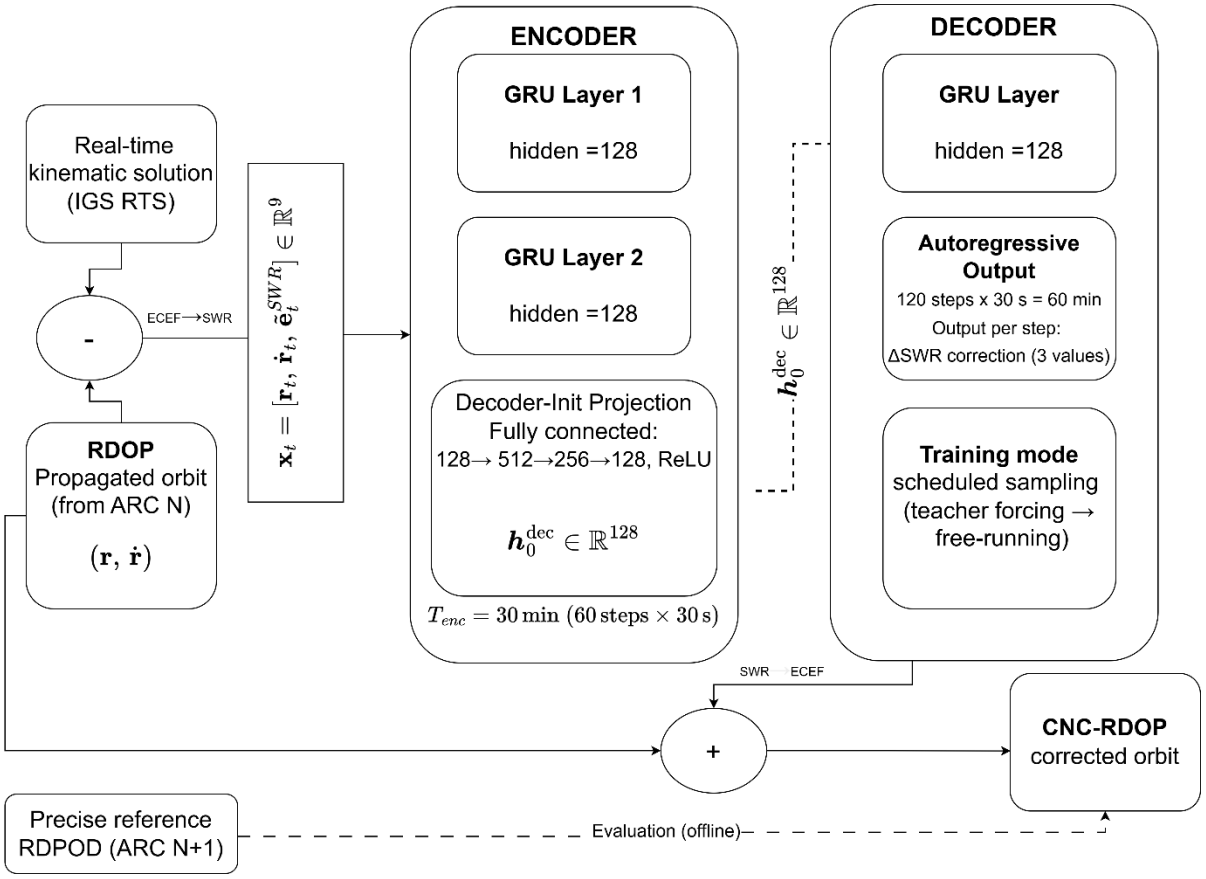


Fig. 1 GRU Seq-to-Seq architecture of CNC-RDOP. The encoder input residual $\tilde{\mathbf{e}}_t^{SWR}$ is formed in real time against the kinematic solution; the precise post-processed reference (dashed, offline) is used only for training targets and evaluation.

2.4 Training Strategy

Because the decoder generates the 60-min correction sequence autoregressively, a choice must be made during training about what to supply as the previous value. There are two options for that choice: the known ground truth correction derived from the reference orbit (teacher forcing), or the decoder's own prediction from the preceding step (free-running, exactly as at deployment). The teacher forcing stabilises early training because the decoder always receives correct inputs, but it does not match inference. Operationally, the ground truth is unavailable, so the decoder must consume its own possibly erroneous outputs. A model trained only with teacher forcing is therefore unprepared for how its small errors accumulate step by step at deployment; a mismatch known as exposure bias. We address this with scheduled sampling (Bengio et al. 2015), which begins with pure teacher forcing and progressively replaces the ground-truth input with the decoder's own prediction as training proceeds, so that by convergence the decoder has been trained under the same free-running condition it faces in operation. Concretely, in early epochs, pure teacher forcing is used: the decoder receives the

ground-truth SWR correction at the preceding epoch ($\lambda = 1$); the teacher-forcing ratio λ is then reduced linearly to zero over $E_{decay} = 100$ epochs as:

$$\lambda(n) = \max\left(0, 1 - \frac{n}{E_{decay}}\right) \tag{13}$$

where n is the current training epoch. At each decoder step, the actual decoder input is selected stochastically, where:

$$\mathbf{x}_t^{\text{dec}} = \begin{cases} \mathbf{e}_{t-1}^{\text{SWR,true}} & \text{with probability } \lambda \\ \hat{\mathbf{e}}_{t-1}^{\text{SWR}} & \text{with probability } 1 - \lambda \end{cases} \tag{14}$$

This gradual reduction mitigates exposure bias by progressively exposing the decoder to its own prediction errors during training. The training objective is the mean squared error (MSE) between predicted and true SWR corrections averaged over all prediction steps:

$$L = \frac{1}{T_{dec}} \sum_{t=1}^{T_{dec}} \left\| \hat{\mathbf{e}}_t^{\text{SWR}} - \mathbf{e}_t^{\text{SWR,true}} \right\|_2^2 \tag{15}$$

where $\| \|_2^2$ is the squared-L2-norm and $\mathbf{e}_t^{\text{SWR,true}}$ is the SWR error against the precise reference orbit, so the network is trained to recover the true error from the noisier real-time input. The network is optimised with the Adam optimiser with $\beta_1 = 0.9$, $\beta_2 = 0.999$ (Kingma and Ba 2014), an initial learning rate of 1×10^{-3}, and weight decay of 1×10^{-5}. The learning rate is halved whenever the validation loss does not improve for 10 consecutive epochs. Mini-batches of 256 samples are drawn at random each epoch, and gradients are clipped to $\|\nabla\|_2 \leq 1$ to stabilise late-stage training. The maximum budget is 200 epochs with early stopping after 40 epochs without validation improvement. Because $E_{decay} = 100$, the final 100 epochs are processed fully free-running ($\lambda = 0$), ensuring the saved checkpoint has been exposed to autoregressive inference before convergence. Validation loss is evaluated in free-running mode throughout, matching operational inference conditions to ensure that model selection rewards deployment-time performance. All stochastic components are seeded for reproducibility. These settings are applied uniformly across the 18,135 training samples (see Section 3).

2.5 Operational Broadcast Rule

The propagator described in Section 2.1, in the immediate post-fit interval, produces a 3D RMSE of ~2 m at the 5-min prediction length for a freshly fitted GRACE-FO arc; already meter-class without neural correction. Beyond ~10 min past the arc boundary, the propagation error exceeds the required accuracy, and the network's contribution becomes dominant (cf. Section 4). The broadcast architecture exploits this transition by applying a time-since-fit gate: for prediction epochs within 10 min of the most recent fit, the broadcast ephemeris is the propagator orbit directly; beyond 10 min, the Neural-corrected orbit is broadcast. This hybrid rule preserves the propagator's strength at fresh fit and the network's strength as the fit ages, and is treated as a deliberate design feature of the framework. This bypass does not mean the propagator supersedes the network in operation; the two act on very different timescales. The reduced-dynamic fit (Section 2.1) is computationally expensive, requires a full observation arc and ground-segment processing, and scales poorly across a large constellation, so it is performed only occasionally (e.g., on a daily cycle). The neural refresh, by contrast, re-issues the 60-min forecast from the trailing kinematic window at every 30-s epoch, onboard and at millisecond cost, independently for each satellite. The bypass governs only the first 10 minutes after each fit; for the many hours until the next fit, the broadcast is the neural-corrected orbit. The network's operational role is therefore to extend the usable lifetime of each expensive fit from the ~10 minutes over which the propagator alone remains meter-class to the multi-hour span documented in Section 4.1, over which the propagator-only error would otherwise grow by more than two orders of magnitude (Table 2). Maintaining user accuracy without it would require continuous re-fitting and re-uplinking, which is operationally infeasible. The motivation for this rule, including the training-data imbalance that prevents the network from directly improving the fresh-fit case, is analysed in Section 4.4.

3. Data and Experiments

3.1 Satellite and Data Sources

GRACE-FO C (GFCN) satellite, flying in a near-circular polar orbit at ~500 km altitude, with orbital inclination of ~89°, is selected for training purposes. GFCN carries a dual-frequency GNSS receiver providing continuous tracking of GPS satellites, enabling cm-level precise orbit determination using code and phase observations with ambiguity resolution. The GFCN nominal orbital period is ~95 minutes, and the orbital velocity is ~7.6 km/s. All orbit products,

which include reference orbits, are generated using Bernese GNSS processing software. The real-time kinematic position solution that forms the encoder input is computed epoch-wise (kinematic PPP) using IGS Real-Time Service (RTS) orbit and clock products, emulating the information available in flight, and carries dm-to-m-level noise. The precise post-processed reduced-dynamic orbit serves as the reference (truth) for both training labels and evaluation. The reduced-dynamic POD employs a comprehensive force model configuration detailed in Table 1. The propagator used to generate the prediction-phase trajectory in this paper additionally includes explicit atmospheric drag via NRLMSIS 2.1, Schwarzschild relativistic acceleration, IAU 2000A frame transformation, and pseudo-stochastic-pulse continuation past the arc boundary, as described in Section 2.1. Two configurations used in the analysis are intentionally distinct: the RDPOD step in Bernese absorbs unmodelled physics into pseudo-stochastic pulses and empirical accelerations to fit observations, and the propagator uses explicit physics-based force terms to extrapolate beyond the fit.

Table 1 Processing models and parameters for reduced-dynamic POD

Force Component	**Model/Configuration**
Earth's gravity field	EGM 2008 (120 × 120) (Pavlis et al. 2012)
Solid Earth tides	IERS conventions (Petit and Luzum 2010)
Ocean tides	FES2014b (Lyard et al. 2021), IERS conventions
Third-body perturbations	Sun, Moon, Jupiter, Venus, Mars (JPL DE421)
Atmospheric drag	No explicit atmospheric drag model; drag absorbed by the empirical accelerations.
Relativistic effects	IERS conventions
Empirical accelerations	3 constant empirical accelerations and 6 once-per-revolution coefficients per arc, plus 6-min pseudo-stochastic velocity pulses
Observation sampling interval	30 s
Observation combination	GPS dual-frequency ionospheric-free combination with ambiguity resolution
Elevation cutoff angle	5 degrees

The dataset spans four years from 2022 to 2025 (1,461 calendar days). Arcs for which Bernese failed to produce a converged output due to observation data gaps, processing failures, or other quality issues are excluded; the retained arcs are grouped into consecutive (N, N+1) pairs for sample construction (Section 3.2), with 644, 291, and 263 retained arcs entering the 2022–2023, 2024, and 2025 GFCN partitions. Each retained arc is sampled at 30-s intervals, producing 2,880 epochs per arc. The dataset is partitioned per year to prevent leakage across

solar activity conditions: arcs from 2022 and 2023 form the training set, arcs from 2024 constitute the validation set used for early stopping and learning-rate scheduling, and arcs from 2025 are held out as the test set. Normalisation statistics are computed exclusively from the training set and applied without modification to the validation and test sets.

Cross-spacecraft generalisation is evaluated on GRACE-FO D (GFDN), the companion satellite of GFCN flying in a near-identical orbit (~500 km altitude, ~89° inclination, ~220 km along-track separation). The 2025 GFDN arcs (226 consecutive (N, N+1) pairs spanning 260 retained daily arcs) are processed identically but independently in Bernese. They are used solely as a zero-shot test set, with the GFCN-trained model deployed on GFDN without any retraining or fine-tuning; the trained network is the only element shared between the two chains. This protocol mirrors the intended operational scenario: a single model is trained once on a heavily tracked reference satellite and deployed zero-shot, with no periodic or per-satellite retraining, across a co-orbital cohort sharing the same altitude, inclination, and drag environment. Each satellite then refreshes its own broadcast correction in flight at every 30-s observation epoch, so a fixed network serves the whole cohort continuously without ground-segment retraining cycles.

3.2 Sample Construction

Each training sample is constructed from a pair of consecutive 24-h fitting arcs (N, N+1) belonging to the same year partition (train, validation, or test). The Bernese-estimated initial state and force-model parameters for arc N are propagated 6 hours beyond the arc-N fit boundary using the full-physics RK4 integrator of Section 2.1. Two SWR position-error series are formed at matching epochs. The encoder-input residual is computed against the real-time kinematic solution (propagated minus kinematic; the only error component available in real time), while the training target and the evaluation error are computed against the arc-N+1 precise reduced-dynamic orbit. Each yields 720 samples per arc pair (6 h at 30 s sampling). A 60-min decoder prediction window (120 epochs) is then slid across this trajectory in 10-min steps (a stride of 20 epochs), producing 31 windows per arc pair. Each decoder window is paired with the preceding 30-min encoder window (60 epochs), which supplies the real-time, kinematic-derived post-fit SWR prediction error. For the first window, whose encoder span ends at the arc boundary, no post-fit residual exists yet, and this channel is set to zero; the broadcast rule (Section 2.5) bypasses the neural correction in that interval regardless. More generally, the first few encoder windows straddle the arc boundary (Figure 2): epochs

preceding the boundary lie within the fitted arc and carry no post-fit prediction error, so the encoder's residual channel transitions from zero at the fresh fit to fully post-fit prediction error once the window lies entirely beyond the boundary (time-since-fit ≥ 30 min). This is the same situation that makes the cold-start (k=0) case hard (Section 3.3): near the fresh fit, the encoder has little or no post-fit drift to read.

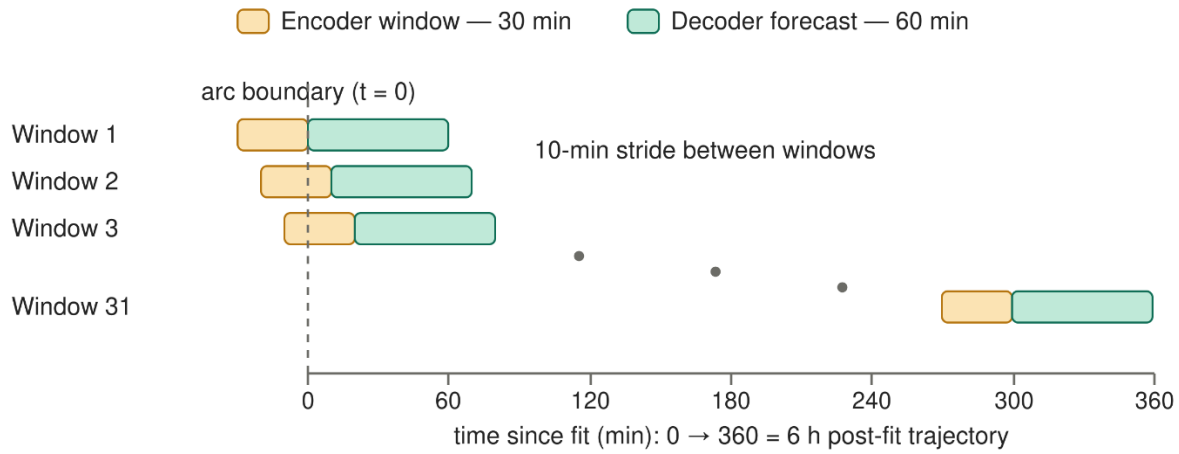


Fig. 2 Sliding-window sample construction over a consecutive arc pair.

The 30-min encoder window length was set in advance from three physical and operational considerations, rather than tuned on validation data. First, the Bernese pseudo-stochastic-pulse interval is 6 minutes; a 30-min window therefore covers five pulse intervals, capturing the systematic residual structure left among adjacent pulse epochs. Second, the empirical residual autocorrelation in the RDOP post-fit SWR error is observed to decay over 10-15 minutes; a 30-min window encompasses 2-3 correlation lengths, which is sufficient for the encoder to learn the local residual dynamics while remaining short enough that the window's leading edge does not become dominated by outdated information. Third, 30 minutes is about one-third of the 95-min GRACE-FO orbital period, ensuring that the encoder observes a representative orbit-phase segment without averaging over a full revolution. A formal sensitivity analysis across window lengths is deferred to future work; the Limitations section (4.8) notes that the encoder window length is one of the parameters that would need to be re-determined for satellites in substantially different orbits (different altitudes, inclinations, or drag environments). Applying this strategy to the screened dataset yields 18,135 training samples (from 585 GFCN 2022-2023 arc pairs after excluding boundary pairs), 7,843 validation samples (GFCN 2024), and 7,099 and 7,006 test samples for GFCN 2025 and GFDN 2025 (zero-shot), respectively. Each sample is built from a consecutive (N, N+1) arc pair, which

exists only when both arcs pass screening; because the retained arcs are fragmented into runs of consecutive days, the number of pairs equals the number of retained arcs minus the number of runs. The 585, 253, 229, and 226 pairs (training, validation, GFCN-test, GFDN-test) span 644, 291, 263, and 260 retained arcs in 59, 38, 34, and 34 runs, respectively; at 31 windows per pair, these give the sample counts above.

3.3 Cold-Start Single-Shot Experiment

A controlled three-experiment study evaluates whether the cold-start operational case is learnable by the GRU Seq-to-Seq architecture with the available training set size. Cold-start is a single-shot 60-min prediction issued at the end of each arc fit, with no post-fit observation history available to the encoder (window offset k=0). It represents the limiting case at a fresh fit, when no post-fit observation history is yet available to the encoder, so the SWR-error channel is set to zero. In real time, the kinematic solution would supply a residual here, but one carrying only measurement noise and no drift; Section 3.4 tests whether providing it helps. These three configurations isolate the contribution of static and per-epoch deterministic physics features at the encoder input. A complementary set of experiments presented in Section 3.4 tests whether adding a time-varying residual and physics conditioning can improve performance in the cold-start case. The two sets are interpreted jointly in Section 4. The three configurations differ only in the encoder feature vector at each time step:

- **Geometry encoder (9-D):** $\mathbf{x}_t$ as defined in Equation (7).
- **Physics-augmented encoder (21-D):**

$$\mathbf{x}_t = \left[\mathbf{r}_t^{ECEF}, \dot{\mathbf{r}}_t^{ECEF}, \tilde{\mathbf{e}}_t^{SWR}, \mathbf{a}_{emp}, \mathbf{w}^{\mathrm{SW}}\right] \in \mathbb{R}^{21} \tag{16}$$

  where $\mathbf{a}_{emp} \in \mathbb{R}^9$ comprises the three constant and six once-per-revolution empirical acceleration coefficients estimated for arc N, and $\mathbf{w}^{\mathrm{SW}} = [F_{10.7}^{\mathrm{daily}},\ F_{10.7}^{81\mathrm{d}},\ K_p]^{\mathrm{T}} \in \mathbb{R}^3$ are arc-level space-weather indices, all broadcast as constants across the 60 encoder time steps. Here, $F_{10.7}^{\mathrm{daily}}$ is the daily 10.7-cm solar radio flux and $F_{10.7}^{81\mathrm{d}}$ is its centred 81-day running average (spanning three solar rotations); both are the standard inputs the NRLMSIS 2.1 density model requires to separate short-term solar variability from the slowly varying background level.
- **Drag-solar-augmented encoder (12-D):**

$$\mathbf{x}_t = \left[\mathbf{r}_t^{ECEF}, \dot{\mathbf{r}}_t^{ECEF}, \tilde{\mathbf{e}}_t^{SWR}, \log \rho_t\,, K_{p,t}, \cos \theta_t^*\right] \in \mathbb{R}^{12} \tag{17}$$

where ρ_t is the NRLMSIS 2.1 atmospheric mass density evaluated at the propagated state, $K_{p,t}$ is the same planetary geomagnetic index as in Equation (16), but here interpolated from the 3-h table to each epoch t rather than held constant over the arc. θ_t^* is the solar zenith angle (the angle at the satellite between the zenith direction and the vector to the Sun) at the satellite position, all varying at each encoder time step. The cosine of this angle is used because it is the direct geometric driver of solar EUV energy deposition into the thermosphere, and hence of the diurnal modulation of upper-atmosphere density at LEO altitudes.

To quantify how much the cold-start results vary from one training run to the next, we repeat the training five times with rotating validation folds (a five-fold repeated-training protocol used here for variance estimation, not for test-set cross-validation). The 2022-2023 training arc pairs are shuffled and divided into five equal groups; each run trains on four groups and uses the fifth for early stopping, in place of the fixed 2024 validation set used elsewhere. The 2025 GFCN and GFDN test sets (Section 3.1) are never used in any run and serve as the same independent evaluation set throughout, so run-to-run variability is reported as the mean and standard deviation of the 2025 test-set errors across the five trained networks.

3.4 Architectural Variants Chain

Beyond the encoder-feature experiments of the previous section, three architectural variants are built on the geodetic-standard propagator baseline (Section 2.1) to test the accuracy limit imposed by the network architecture. All three are trained and evaluated on the same GFCN 2022-2025 partition, with an identical optimiser, batch size, gradient clipping, and random seed. The separate effect of propagator fidelity (low-fidelity vs. the Section 2.1 configuration) is reported in Section 4.4 and is not one of the three variants.

(i) Kinematic cold-start input: At the strict fresh-fit window (k=0), both the operational baseline and the Section 3.3 single-shot diagnostic leave the encoder's post-fit residual at zero, because no post-fit observation exists yet at the arc boundary. This is a property of how that window is constructed, not of real-time operation, in which the kinematic solution is available. Built on the operational baseline of Section 2, this variant fills the first 60 epochs past the arc boundary with the real-time-available kinematic SWR residual (kinematic solution minus

reduced-dynamic fit). It therefore tests directly whether the cold-start degradation diagnosed in Section 3.3 is caused by the absence of time-varying residual information at the encoder.

***(ii) Decoder-aware physics conditioning*:** The encoder is widened from 9-D to 16-D by appending space-weather and Sun-geometry features, and the decoder per-step input is widened from 3-D to 17-D by appending the propagated state, elapsed time, $F10.7$, K_p, Sun-Earth-satellite geometry, eclipse flag, and beta angle at each rollout epoch (the angle between the Sun direction and the satellite's orbital plane; cf. the eclipse model in Allahverdi-Zadeh et al. 2016). The decoder is no longer a pure latent extrapolator; at each future step, it is re-grounded on physics. This variant uses a denormalised MSE loss, in which the squared error at each step is weighted to increase with the prediction step index, so that longer-lead errors contribute more, together with a jerk-smoothness penalty on the third time-derivative of the predicted position that discourages non-physical high-frequency oscillation.

(iii) Learned residual gate: A small per-step gate $\alpha \in [0, 1]$, computed by a two-layer fully connected network over the encoder's last hidden state and the decoder's per-step physics features, multiplies the network's correction output. The gate bias is initialised to $\sigma^{-1}(0.9)$ so the variant behaves like decoder-aware physics conditioning at initialisation and learns to suppress the correction only where it is harmful. This tests whether the cold-start over-correction can be self-calibrated through a learned gate.

The three variants are trained sequentially, each on top of the previous one's improvements; variant *(ii)* is also evaluated independently on the propagator baseline (without *(i)*) for clean attribution. Results are explained in Section 4.

3.5 Pure ML Experiment

To isolate the contribution of the SWR prediction-error component, that is the encoder's direct view of the recent residual between the observed and propagated position, a pure-ML experiment is constructed in which the last three dimensions of the encoder input vector in Equation (7) are set to zero, leaving only the ECEF position and velocity. The GRU Seq-to-Seq architecture, training procedure, optimiser, normalisation, dataset partitioning, and evaluation protocol are identical to those of the full CNC-RDOP configuration. This experiment tests whether the network can predict the next 60 minutes of post-fit error trajectory from the propagated state alone (i.e., ECEF position and velocity at the encoder steps), without

observing the recent residual that the physics backbone (the reduced-dynamic propagator of Section 2.1, which generates the predicted orbit and therefore the SWR prediction error supplied to the encoder) produces against the truth. Despite the conventional "pure-ML" label retained here for continuity with the prior literature, the propagator-derived position and velocity remain in the encoder input; only the explicit residual component is removed.

3.6 Evaluation Protocol

All 3D RMSE values compare the corrected (or propagated) orbit against the precise post-processed reference; the real-time kinematic solution enters only through the encoder input, not in the evaluation. Prediction accuracy is reported under two complementary protocols. The *lumped* protocol averages 3D RMSE across all 31 sliding windows per arc pair at each of the prediction lengths reproducing the convention used in prior single-shot prediction studies. The *stratified* protocol bins each (window, decoder-step) pair by the time elapsed between the start of arc N+1 and the prediction epoch (the *time-since-fit*), into six bins {0-10, 10-30, 30-60, 60-120, 120-240, 240-360} min, and reports 3D RMSE within each (time-since-fit × prediction-length) cell.

The stratified protocol is needed because the baseline RDOP error at the 5-min prediction length varies by more than 2 orders of magnitude depending on how outdated the fit is, from ~2 m immediately after a fresh fit to ~463 m at 4-6 h past the fit. A single lumped 5-min figure is dominated by these large, long-stale errors and does not reflect the accuracy a LEO-PNT user actually experiences just after an upload. We therefore define two reference cells: the post-upload cell (time-since-fit $\in$ [0, 30) min, prediction lengths = 5 min), the situation a user occupies immediately after each broadcast, and the fresh-fit cell (time-since-fit $\in$ [0, 10) min, prediction lengths = 5 min), which isolates the literal fresh-fit case (the encoder window ends exactly at the arc boundary, k=0). The fresh-fit cell is the single most operationally relevant number in a continuous-refresh broadcast architecture. All numbers in Section 4 are reported under both protocols unless noted otherwise.

3.7 Computational Cost

The trained network comprises 434,179 parameters across a two-layer encoder GRU, a three-layer fully connected bottleneck network, a single-layer decoder GRU, and a linear output layer. Computation time per sample is 0.54 ms on an NVIDIA RTX 2000 Ada Generation GPU (CUDA

12.1) at batch size 32, and 4.3 ms on a single core of an Intel Core Ultra 9 285 CPU. Relative to the 30-s (30,000 ms) GNSS observation interval of broadcast processing chain, the single-core CPU step leaves a margin of about 7,000× (more than three orders of magnitude), so the network can be embedded directly into the broadcast generation process and run once per 30-second epoch without any real-time bottleneck, even on a single-threaded embedded processor onboard the satellite. On the GPU, the 0.54 ms per-sample cost (batch size 32) extends this margin to ~56,000× for simultaneous ground-segment processing of multiple spacecraft. These figures cover neural inference only. The geodetic-standard propagator that generates the predicted state (Section 2.1) is computationally heavier than the network, and its real-time cost is not assessed here. The Section 4.4 result that baseline accuracy is largely insensitive to non-drag force-model fidelity suggests a reduced-fidelity onboard propagator (retaining the drag model but simplifying geopotential, tides, frame, and relativity) could supply the predicted state at much lower cost, but a full onboard timing budget (propagator plus network) is left to a future implementation study.

4. Results and Discussion

4.1 Main Results: Stratified and Lumped Performance

Prediction accuracy is reported under the two protocols defined in Section 3.6. The stratified protocol (Table 2) is the primary result. It presents the 3D RMSE on a 2D grid: the rows bin each epoch by the time elapsed since the most recent orbit fit (from 0-10 min down to 4-6 h), and the columns bin by the prediction length (5 to 60 min). Each cell reports three values: the propagator-only baseline RDOP error, the CNC-RDOP-corrected error, and the percentage improvement between them, so that the table shows both the absolute accuracy and the gain contributed by the network in every operating condition. This layout deliberately separates the two situations a continuous-refresh user encounters: performance just after a fresh fit (the top rows) and performance several hours after the most recent fit (the bottom rows). The separation matters because the baseline error depends strongly on how stale the orbit fit is: at the 5-min prediction length, the baseline RDOP 3D RMSE grows from ~2 m immediately after a fresh fit to ~460 m at 4-6 h since the fit; a spread of more than two orders of magnitude across the time-since-fit bins, not across the prediction-length columns. A single lumped average would mask this variation, which is why the stratified grid is adopted as the primary result rather than a lumped figure. Against this strongly varying baseline, the CNC-RDOP-corrected error

stays within roughly 4-31 m at the 5-min prediction period across the entire 0-6 h range (Figure 3). The lumped all-windows protocol (Table 3) is retained only for direct comparison with prior single-shot cold-start studies. All numbers below refer to the CNC-RDOP framework of Section 2 on the 2025 GFCN test set; the architectural-experiment and cross-spacecraft results follow in Sections 4.4 and 4.5, respectively.

Table 2 CNC-RDOP stratified 3D RMSE (in m) on the 2025 GFCN test set. Each cell shows three values: baseline RDOP 3D RMSE / CNC-RDOP-corrected 3D RMSE / improvement %, respectively. 'low-N' marks cells with fewer than 30 samples, which are not reported; blank '—' marks cells with no samples. (positive => improvement, negative => degradation)

		Prediction Length	
		5 min	**10 min**
Time since fit	0-10 min	2.06 / 3.98 / -93%	low-N
	10-30 min	30.17 / 8.57 / +72%	low-N
	30-60 min	110.72 / 10.60 / +90%	94.69 / 16.81 / +82%
	1-2 h	171.10 / 14.24 / +92%	167.28 / 23.56 / +86%
	2-4 h	322.72 / 19.31 / +94%	313.55 / 32.11 / +90%
	4-6 h	463.27 / 31.01 / +93%	463.11 / 43.33 / +91%
		Prediction Length	
		30 min	**60 min**
Time since fit	0-10 min	—	—
	10-30 min	—	—
	30-60 min	94.69 / 62.00 / +35%	low-N
	1-2 h	167.28 / 71.79 / +57%	167.28 /147.70 / +12%
	2-4 h	313.55 / 106.76 / +66%	313.55 / 164.81 / +47%
	4-6 h	478.37 / 139.59 / +71%	497.79 / 227.36 / +54%

Within a time-since-fit row the baseline RDOP RMSE is nearly constant across the prediction-length columns, as expected when stratifying by time since the most recent fit. The small residual column-to-column variation (e.g., 110.72 m vs 94.69 m in the 30-60 min row, ~17%) is a stride-induced artifact (windows slide in 10-min steps), so different columns sample slightly shifted prediction-epoch subsets within the same bin, and the steeply growing baseline magnifies the shift. This affects only the cross-column baseline comparison; within each cell the baseline and corrected errors are computed on the same epochs, so every improvement percentage is unaffected.

Figure 3 visualises the 5-min prediction length column of Table 2: baseline error grows by more than two orders of magnitude across the six time-since-fit bins, whereas the CNC-RDOP-

corrected error grows by less than one order of magnitude and stays below 32 m even at the 4-6 h end. Two cells in Table 2 deserve special emphasis, as they capture the two situations in which a continuous-refresh LEO-PNT user most often operates. The *post-upload cell* (time-since-fit ∈ [0, 30) min, 5-min prediction length) is the situation a downstream user occupies immediately after each broadcast upload. This cell aggregates the 0-10 min and 10-30 min sub-bins by sample-weighted RMS combination, giving a baseline RDOP 3D RMSE of 24.66 m and CNC-RDOP-corrected 3D RMSE of 7.36 m, a 70% improvement. The *fresh-fit cell* (time-since-fit ∈ [0, 10) min, 5-min prediction length) isolates the literal k=0 fresh-fit case. The encoder window ends exactly at the arc boundary, where the orbit has just been fitted to the GNSS observations, so no prediction drift has yet accumulated; the SWR error channel therefore holds only kinematic measurement noise, and the propagator alone has likewise had no time to drift. In this cell, the propagator achieves 2.06 m 3D RMSE, already meter-class without any neural correction, and the network does not improve upon it. The operational broadcast rule defined in Section 2.5 therefore bypasses the neural correction for the first 10 minutes past each fresh fit and broadcasts the propagator orbit directly; the diagnostic causes are examined in Section 4.4.

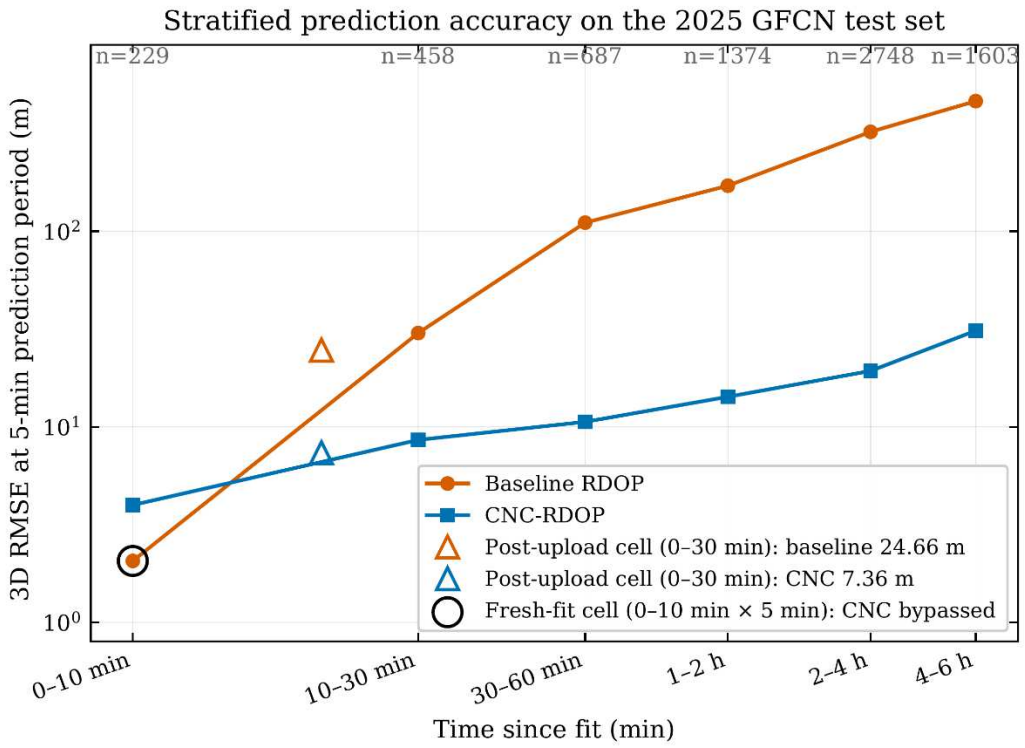


Fig. 3 Stratified prediction accuracy at the 5-min prediction length on the 2025 GFCN test set. Sample counts per bin (n) are shown at the top of the panel.

Across the remaining cells of Table 2, the network reduces baseline RDOP by roughly 70-95% at the 5- and 10-min prediction lengths and by 12-71% at the 30- and 60-min prediction lengths. At the 5-min prediction length, the CNC-RDOP-corrected error stays at or below ~20m out to 2-4 h time-since-fit and below 32 m even at 4-6 h. This pattern reflects the framework's design: the encoder reads the systematic structure of the accumulated post-fit residual from a 30-min trailing window, and the decoder extrapolates it forward. The richer that residual structure (i.e., the more the orbit has drifted since the fit), the more predictive value the network extracts, which is why its contribution grows with time-since-fit.

The lumped all-windows view in Table 3 aggregates the cells of Table 2 into single per-prediction length numbers, weighted by the sample count. Because the long time-since-fit cells contribute the bulk of the samples and carry the largest absolute baseline errors, the lumped figures are dominated by that tail. Under this protocol the network reduces 3D RMSE from 309.34 m to 20.42 m at the 5-min prediction length (93.4% reduction), from 343.74 m to 109.85 m at the 30-min prediction length (68.0%), and from 388.96 m to 191.68 m at the 60-min prediction length (50.7%), with improvement decreasing monotonically with prediction length as the prediction extends beyond the range over which the trailing 30-min encoder window remains predictive of the future error trajectory. These lumped figures are sample-weighted averages of the Table 2 cells and therefore fall between the smallest and largest of them; they are reported for compatibility with the cold-start prediction literature and should not be read as the accuracy a continuous-refresh LEO-PNT user experiences in any single operational point.

Table 3 Lumped all-windows 3D RMSE on the 2025 GFCN test set.

Prediction Length	RDOP 3D (m)	CNC-RDOP 3D (m)	Improvement
5 min	309.34	20.42	93.4%
10 min	315.78	32.09	89.8%
30 min	343.74	109.85	68.0%
60 min	388.96	191.68	50.7%

4.2 Directional Error Analysis

The per-component RMSE values across the along-track, cross-track, and radial components of the SWR frame (Table 4, lumped protocol) reveal a prediction-length-dependent reorganisation of the residual error budget.

Table 4 CNC-RDOP per-component RMSE (m) and inter-component ratios, 2025 GFCN test set, lumped protocol.

Prediction Length	S (m)	W (m)	R (m)	S/W	S/R	W/R
5 min	13.69	10.42	11.00	1.31	1.24	0.95
10 min	17.09	18.80	19.61	0.91	0.87	0.96
30 min	97.45	36.59	35.11	2.66	2.78	1.04
60 min	184.91	37.93	33.33	4.87	5.55	1.14

Two prediction-length-dependent patterns are visible. At short prediction length (5-10 min), the three SWR components are of comparable magnitude, and the along-track component does not yet dominate; at 10 min, the residual is in fact mildly cross-track- and radial-leaning. This near-isotropy (comparable error in all three directions) is consistent with the post-fit residual being dominated by the network's autoregressive error rather than by any single physical perturbation. At short prediction length, the encoder's 30-min trailing window has already absorbed most of the drag and once-per-revolution structure, and what remains is approximately uncorrelated across the three directions.

At longer prediction length (30-60 min), the along-track component grows substantially relative to W and R: S/W rises from 1.31 at 5 min to 4.87 at 60 min, and S/R from 1.24 to 5.55; the signature of atmospheric drag emerges as the principal uncorrected perturbation. The once-per-revolution empirical accelerations and the 6-min pseudo-stochastic pulse continuation absorb most of the in-plane periodic force-model misfit during and just after the arc, but cannot extrapolate the slowly evolving drag-density mismodelling beyond the range where the 30-min trailing window stays predictive. The network recovers part of this growth from the encoder window, but its skill fades as the prediction length extends. A structural factor compounds this: in the operational 3D configuration, the decoder consumes only its own previous SWR correction at each rollout step (Eq. 12), so it extrapolates the error pattern without observing the satellite's position along its orbit over the 60-min prediction. This blindness to future orbit phase is the likely mechanism behind the long-lead skill fade; variant (ii) (Section 4.4) addresses it by re-grounding the decoder on the propagated state at each step, which is why that variant stabilises the long-prediction-length rollout.

The cross-track and radial components remain close across all prediction lengths (W/R between 0.95 and 1.14 under the neural-corrected lumped protocol of Table 4). A force-model sensitivity test (Section 4.4) shows that this balance is not produced by the propagator-fidelity

upgrades: the same W/R structure is present in a low-fidelity propagator with GMST-only rotation and no tides, and the full IAU 2000A transformation, FES2014b ocean tides, pole tide, and Schwarzschild term change W and R by ≤0.14 m (sub-percent) at every prediction length. The W≈R balance is therefore an intrinsic property of the drag-dominated error geometry: along-track error accumulates in the orbit plane while the cross-track and radial residuals reflect the smaller, once-per-revolution structure that the empirical-acceleration parameterisation absorbs roughly symmetrically, rather than a consequence of force-model fidelity. The cross-spacecraft GFDN results (Section 4.5) reproduce this pattern to within a few tens of centimetres at every prediction length, confirming it is a property of the orbital environment rather than of a particular spacecraft.

4.3 Cold-Start Single-Shot Prediction: First-Pass Experiment

Section 3.3 defined three encoder configurations: geometry-only (9-D), physics-augmented (21-D), and drag–solar-augmented (12-D), and trained each under the single-shot cold-start protocol (k=0, no post-fit observation history at the encoder). Table 5 summarises the resulting CNC-RDOP 3D RMSE on the 2025 GFCN test set at 5- and 60-min prediction lengths, and the propagator-only RDOP baseline included as a reference.

Table 5 Single-shot cold-start (k=0) CNC-RDOP 3D RMSE on the 2025 GFCN test set, by encoder configuration. Values are mean ± standard deviation across the five networks of the repeated-training protocol (Section 3.3)

Encoder configuration	Encoder dim.	5-min RMSE (m)	60-min RMSE (m)
Geometry encoder	9	3.22 ± 1.61	159.6 ± 2.2
Physics-augmented encoder	21	5.49 ± 5.07	156.5 ± 1.9
Drag-solar-augmented encoder	12	3.86 ± 1.81	158.3 ± 2.2
Propagator-only RDOP baseline	-	2.06	153.5

None of the three configurations improves the propagator-only baseline at the 5-min prediction length: the geometry encoder degrades 3D RMSE by ~56%, the physics-augmented encoder by ~166%, and the drag-solar-augmented encoder by ~87%. At the 60-min prediction length, all three are close to the baseline (within ~2-4%, i.e. 3-6 m of the 153.5 m baseline), neither clearly improves nor strongly degrades it. The physics-augmented configuration inflates the cross-fold standard deviation at 5 min from 1.61 m (geometry encoder) to 5.07 m without any corresponding reduction in error, indicating that the added input dimensions contributed optimisation noise rather than useful information for learning.

A feature-attribution test was applied to the two augmented configurations to measure how much the added inputs actually affect the output. For each fold-trained model, prediction was run twice on identical encoder inputs: once with the added features held fixed at their training-fold mean, and once with their true values. The L2 norm of the difference between the two predictions, $\|\Delta\|_2$, fell below 10^{-3} in all five folds at both prediction lengths for the physics-augmented configuration, and was indistinguishable from zero for the drag-solar-augmented configuration. The network therefore extracts no usable information from either set of added features under the cold-start protocol. The natural first interpretation is that the encoder simply has no informative input. Section 4.4 tests this directly and shows the explanation is incomplete; even when the kinematic residual is supplied (variant i), the cold-start cell does not improve.

Two conclusions follow. First, single-shot cold-start prediction is not a learnable operational scenario for a generic GRU Seq-to-Seq at the available training-set size, with or without static or per-epoch physics features at the encoder. Second, the propagator-only baseline at k=0 is already at meter-class (2.06 m at the 5-min prediction length), and the operational broadcast rule of Section 2.5 therefore bypasses the neural correction for the first 10 min past each fresh fit. This is not a workaround introduced after the fact; it is the operational consequence of the result in Table 5 and is integral to the framework as deployed.

4.4 Architectural Studies: Kinematic Encoder, Decoder Conditioning, and Residual Gate

Three further architectural variants, defined in Section 3.4, were trained and evaluated on the same GFCN partition using the same optimiser, batch size, gradient clipping, and random seed. Each variant inherits the previous variant's improvements except where noted.

Baseline-physics comparison: A force-model sensitivity test shows that the propagator-only 3D RMSE is essentially invariant to non-drag force-model fidelity (the NRLMSIS 2.1 drag model is held fixed throughout this test). Truncating EGM2008 from 120×120 to 30×30, removing solid-Earth/ocean/pole tides, replacing the full IAU 2000A/2006 transformation with a GMST-only rotation, and dropping the Schwarzschild term together change the baseline 3D RMSE by less than 0.2% at every prediction length from 5 to 60 min (≤0.14 m per component). The atmospheric-drag mismodelling, as the dominant error source (identical in both configurations), therefore sets the propagator floor, not geopotential, frame, or tidal fidelity.

This confirms that the limit the architectural variants are measured against is aerodynamic, consistent with the feature-limited fresh-fit floor. Table 6 summarises the key fresh-fit and post-upload cells, and the lumped 5- and 30-min prediction lengths.

Table 6 Additional architectural variants on the 2025 GFCN test set. Each cell shows the CNC-RDOP 3D RMSE (m) and its change relative to the propagator-only RDOP baseline (positive indicates improvement, negative indicates degradation). The 'baseline' row is the operational configuration of Section 2; each subsequent variant builds on the one above it, beginning with the kinematic cold-start input of variant (i).

Variant	fresh-fit cell	post-upload cell	All 5-min	All 30-min
Baseline	3.98 m (-93%)	7.36 m (+70%)	20.42 m (+93.4%)	109.85 m (+68.0%)
+ Kinematic encoder	8.62 m (-319%)	31.77 m (-28%)	19.89 m (+93.6%)	80.1 m (+76.7%)
+ Decoder physics features	8.28 m (-302%)	8.79 m (+64%)	16.98 m (+94.5%)	99.07 m (+71.2%)
+ Learned residual gate	8.16 m (−296%)	9.27 m (+62%)	16.98 m (+94.5%)	99.07 m (+71.2%)

Kinematic encoder: The operational baseline already supplies the kinematic SWR residual $\tilde{e}_t^{SWR}$ (Eq. 7) at every encoder window except the k=0 fresh-fit window, which is zeroed because no post-fit residual exists at the arc boundary. This variant is identical to that baseline except that it fills that single window with the genuine kinematic residual (cm-to-dm scale) available in real time. It is not a null change but a targeted modification of the fresh-fit behaviour. The variant improves the lumped 30-min accuracy (109.85 m → 80.1 m), but the post-upload and fresh-fit cells regress severely (-28% and -319%): the encoder must reconcile the cm-to-dm-scale kinematic residual just after a fit with the metre-scale propagation residual at stale-fit windows, and learns the mean correction for the dominant stale case, over-applying it to the small-residual minority.

Decoder physics features: Re-grounding the decoder at each rollout step on the propagated state and physics-context features (space-weather indices and Sun-Earth-satellite geometry) recovers most of the post-upload cell loss introduced by the kinematic encoder (-28% → +64%), though the lumped 30-min accuracy falls back from +76.7% to +71.2% (still ~3 percentage points above the baseline). The fresh-fit cell remains 3-4 times worse than the propagator baseline. The decoder-side physics anchoring stabilises the long-prediction-length rollout but does not resolve the cold-start miscalibration.

Learned residual gate: The gate does recognise the approach of a fresh fit: its value averages 0.991 just after a fresh fit, versus 0.998 at 4-6 h, a 0.7% shift in the right direction (a lower value means more of the correction is held back). But this is far too little: multiplying the correction by ~0.99 barely changes it, when at the fresh fit it needs to be turned off almost entirely. The fresh-fit cell therefore stays at 8.16 m (-296% vs the baseline), essentially unchanged from the decoder-physics variant.

The three variants together establish that no architectural or feature-level change attempted here improves the fresh-fit cell over the propagator-only baseline. The kinematic-encoder result rules out a starved input as the cause: supplying the genuine, time-varying kinematic residual (cm-to-dm scale) in the cold-start window, as the same residual available in real time, does not repair the fresh-fit cell but degrades it further (-319%). The network therefore does not fail for lack of input; it fails even when the real-time-available residual is present. The decoder-physics result rules out an unstable long-prediction-length rollout. The residual-gate result rules out a simple output-scaling fix: even when the network is given an explicit per-step control to switch its own correction off, it does not do so strongly enough. We attribute the common failure to a training-data imbalance. The fresh-fit case accounts for only ~3% of the training samples, so under standard MSE training, the loss is dominated by the 97% of large-error stale-fit samples, and the network learns a correction sized for that majority. Applied to the ~2 m fresh-fit baseline, that correction overshoots, and the gate cannot turn it off against the weight of the majority. The proposed operational broadcast rule (propagator-only for the first 10 min after each fresh fit) is the design response to this imbalance and is unaffected by failures in the architectural variants.

The 3% training share of fresh-fit samples raises the question of whether standard class-imbalance remedies, e.g., oversampling the minority class, focal loss, stratified mini-batch construction, or per-bin loss weighting, would break the floor. They would not, for two reasons. First, the fresh-fit degradation is not an underfitting failure: in this cell the network overcorrects relative to a near-zero baseline residual (see Table 2, 0-10 min row), so upweighting the cold-start case with focal loss or loss reweighting would amplify the overcorrection rather than reduce it. Second, at the fresh fit the encoder has no post-fit drift to read: the SWR-error channel is zeroed in the operational and diagnostic configurations, and even when it is filled with the real-time kinematic residual (variant i) that residual carries only dm-to-m

measurement noise, not drift. The remaining inputs (propagated state, $F10.7$, K_p, $\cos\theta_t^*$) carry no information that would let the network predict the next 5 minutes of residual below that noise level. The obstacle is therefore a feature limitation set by the kinematic-input noise floor, not an optimisation problem. A prediction-length-weighted loss is in fact tested in variant (ii), with no improvement at the fresh-fit cell. The hybrid bypass is the operationally optimal response to a situation where the propagator already achieves meter-class accuracy and no learnable residual is available.

4.5 Cross-Spacecraft Generalisation: GRACE-FO D

Cross-spacecraft generalisation is evaluated by deploying the GFCN-trained network on the 2025 GFDN arcs in a zero-shot setting, with no retraining or per-satellite adjustment (as described in Section 3.1). The test asks whether the network has learned a satellite-specific residual pattern or a generalisable post-fit error structure that transfers across co-orbital twins. Table 7 reports the lumped all-windows comparison between GFCN (the training satellite) and GFDN (zero-shot) on the 2025 test partition.

Table 7 Cross-spacecraft comparison of CNC-RDOP performance on the 2025 test partition, lumped all-windows protocol. GFDN values are zero-shot.

Prediction Length	Satellite	RDOP 3D (m)	CNC-RDOP 3D (m)	Reduction	S (m)	W (m)	R (m)
5 min	GFCN	309.34	20.42	93.40%	13.69	10.42	11.00
	GFDN	309.27	20.66	93.32%	14.06	10.34	11.05
10 min	GFCN	315.78	32.09	89.84%	17.09	18.80	19.61
	GFDN	315.68	32.21	89.80%	17.36	18.66	19.70
30 min	GFCN	343.74	109.85	68.04%	97.45	36.59	35.11
	GFDN	343.37	110.26	67.89%	97.94	36.34	35.27
60 min	GFCN	388.96	191.68	50.72%	184.91	37.93	33.33
	GFDN	388.55	192.03	50.58%	185.34	37.56	33.40

The agreement between GFCN and GFDN is striking under both protocols. At every lumped prediction length, the GFDN zero-shot result tracks GFCN closely. The RDOP baseline and the improvement percentage agree to within ~0.2%, and the neural-corrected 3D RMSE to within ~1% (per-component values to within a few tens of centimetres at the longest prediction length). The transfer holds equally under the stratified protocol. Just after an upload of broadcast ephemeris (0-30 min since fit, 5-min prediction length) GFDN gives 7.34 m against GFCN's 7.36 m; in the first 10 minutes past a fresh fit (5-min prediction length) GFDN is 3.97 m

against GFCN's 3.98 m; and in the multi-hour stale-fit bins the two satellites agree to within 1% at every prediction length.

The per-component (S, W, R) decomposition (Figure 4) illustrates the cross-spacecraft agreement and confirms the prediction-length-dependent component reorganisation documented in Section 4.2. At short prediction length, the S, W, and R errors are of comparable magnitude, while along-track dominance grows progressively at longer prediction length. GFDN reproduces this pattern almost identically. This rules out the alternative interpretation that the GFCN training has fitted a receiver- or observation-history-specific pattern; the network has instead learned a post-fit error structure characteristic of the GRACE-FO orbital environment, transferable to a co-orbital twin of the same spacecraft class.

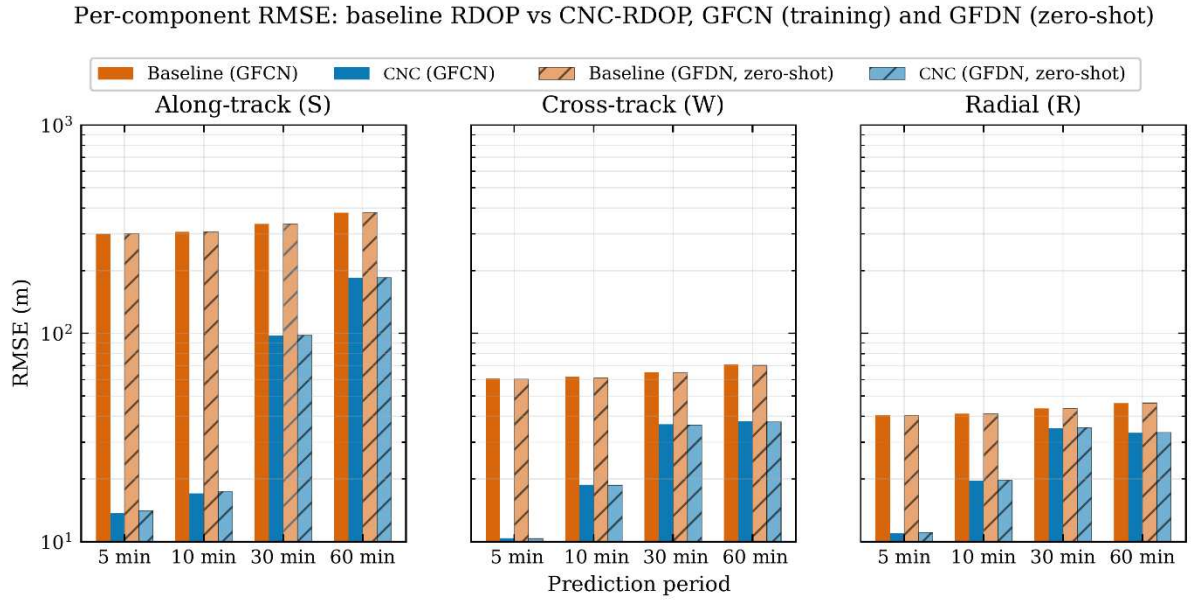


Fig. 4 Per-component RMSE at the four prediction lengths reported in the lumped protocol, baseline reduced-dynamic propagator versus CNC-RDOP, for GFCN and GFDN

For a LEO-PNT operator whose constellation includes co-orbital satellites sharing the same altitude, inclination, and drag environment, this result has direct operational consequences: a single network, trained on a heavily tracked reference satellite, can be deployed across the cohort without the need for per-satellite retraining or parameter adjustment. Transferability to satellites at substantially different altitudes, inclinations, or drag environments is not tested here and is left for future work. GRACE-FO C and D are near-identical co-orbital twins separated by approximately 220 km in-track; the present result is therefore best read as a proof

of concept for transfer within an orbital cohort, not as a guarantee of transfer to satellites in a substantially different orbit (different altitude, inclination, or drag environment).

4.6 Pure-ML Experiment: Removing the Physics Backbone

Section 3.5 defined the pure-ML experiment, in which the SWR prediction-error component is set to zero, leaving only the ECEF position and velocity at the encoder, with the architecture, hyperparameters, normalisation, partitioning, and evaluation protocol otherwise identical to the full CNC-RDOP configuration. The comparison against Table 3 is therefore directly interpretable. Table 8 summarises the results.

Table 8 Pure-ML experiment 3D RMSE on the 2025 GFCN test set, lumped all-windows protocol.

Prediction Length	RDOP (m)	Pure-ML (m)	CNC-RDOP (m)
5 min	309.34	274.5	20.42
10 min	315.78	278.9	32.09
30 min	343.74	306.0	109.85
60 min	388.96	353.3	191.68

The pure-ML experiment converges to a non-trivial solution and achieves a modest reduction in 3D RMSE relative to the propagator-only baseline (11.3% at 5 min, 11.7% at 10 min, 11.0% at 30 min, and 9.2% at 60 min prediction lengths). The reduction indicates that the state (position, velocity, and the orbital phase they encode) carries weak but non-zero predictive information about the deterministic, orbit-phase-dependent component of the post-fit prediction error. The pure-ML experiment, however, cannot approach the accuracy of the full CNC-RDOP framework: at a 5-min prediction, the pure-ML 3D RMSE is ~13 times larger, and at a 30-min prediction, ~3 times larger than CNC-RDOP. The differential narrows at longer prediction length, not because pure-ML improves, but because CNC-RDOP's autoregressive correction loses skill as the prediction extends beyond the regime where the trailing 30-min encoder window remains predictive,

The result establishes that the SWR prediction-error component is the essential ingredient that enables the CNC-RDOP correction, not the GRU architecture itself. A recurrent network receiving only orbital kinematics is unable to track the current RDOP error state and therefore cannot generate corrections at the magnitude required for any LEO-PNT broadcast use case. The physics backbone is therefore not a convenience layer that the network could in principle

replace given enough data; it is a structural prerequisite, supplying the recent residual on which all subsequent neural correction is conditioned.

4.7 Operational Distinction from Single-Forecast Neural Correction Architecture

CNC-RDOP shares its GRU Seq-to-Seq core with recent neural reduced-dynamic correction methods (Li et al. 2025; Zhang et al. 2025), but differs from them conceptually rather than architecturally. Those methods issue a single forecast when the daily fit concludes and, where a warm-up window is used, populate it with a post-processed precise orbit that is available only offline. CNC-RDOP instead re-issues the forecast at every observation epoch from a refreshed 30-min trailing window built on a real-time kinematic position solution. These two choices, continuous refresh and real-time availability of input, are what make the framework deployable for live LEO-PNT broadcast. Combined with the proposed broadcast rule, they yield a continuously refreshing architecture with three operational consequences:

1. The freshness of the broadcast ephemeris is decoupled from the daily fit cycle. Because the fit and the refresh operate on different timescales (Section 3.6), the forecast is always "the next 60 minutes from this moment," not "the next 60 minutes from a fit several hours ago." The maximum age of the correction a downstream user encounters is bounded by the observation epoch interval (30 s here) rather than by the upload interval.

2. The validity window can be extended without an abrupt quality cliff at its end, because the operator can re-upload at any moment during the day rather than only at the daily fit boundary. The ground segment is therefore free to optimise upload schedules around station-contact opportunities rather than around the prediction-quality cliff at the end of the validity window (user-side broadcast validity, not the fit cycle).

3. The framework provides an in-built anomaly-detection channel: sudden growth in the observed prediction error is a real-time indicator that the underlying force model has been disturbed, and the operator can flag the satellite for review or an unscheduled re-fit, independent of any external space-weather monitoring service.

The most capable of these methods reports centimetre-level prediction accuracy (Li et al. 2025). We deliberately do not benchmark against it numerically, because the two methods

target different operating conditions and a side-by-side error figure would compare quantities that are not like-for-like: Li et al. operate on offline precise orbits with a single post-fit forecast, and their evaluation spans a different solar-activity condition than the active 2022-2025 interval studied here. The meaningful comparison is conceptual. Offline single-forecast methods define the accuracy attainable once precise products are available after the fit; CNC-RDOP defines the accuracy attainable in real time, when only a kinematic solution is available. Metre-class prediction is insufficient for carrier-phase precise point positioning, but is sufficient for coarse acquisition, integrity monitoring of broadcast ephemerides, and Doppler-based positioning at the sub-metre level (Allahvirdi-Zadeh et al. 2025a, b), where the orbit error contributes only a fraction of the positioning error budget (cf. Montenbruck et al. 2018). The two approaches are therefore complementary rather than competing.

4.8 Limitations and Future Directions

Several limitations bound the present results, each pointing to a specific direction for future research.

Single orbital configuration: The empirical evaluation is conducted on GRACE-FO formation at approximately 500 km altitude and 89° inclination. The cross-spacecraft result of Section 4.5 establishes structural rather than satellite-specific learning within this approach, but generalisation to lower-altitude, higher-eccentricity, or sun-synchronous orbits has not been investigated. The methodology is transferable in principle, but would require retraining on orbit-specific data, analysing and re-determining the encoder window length and decoder prediction length against the new orbit's residual autocorrelation timescale.

Storm-time performance: During geomagnetic storms, thermospheric density rises by factors of 2 to 10 above climatological values, producing along-track residuals that may lie outside the training distribution. The continuous-correction architecture observes storm-induced residual growth in real time through the trailing encoder window, but the forward forecast over the storm interval is unvalidated. Extending the method to storm conditions, via density-targeted augmentation of the training data, conditioning the network on real-time geomagnetic indices, or detecting when the encoder inputs fall outside the training range, is identified as future work.

Fresh-fit cell ceiling: As analysed in Section 4.4, no intervention attempted here improves the fresh-fit cell over the propagator's 2.06 m; the limit is a feature-limited noise floor set by the

real-time kinematic input, so the broadcast rule of Section 2.5 bypasses the neural correction for the first 10 min after each fit. Reaching decimetre accuracy would require additional in-flight information, most directly an onboard accelerometer that measures non-gravitational acceleration; sub-minute refresh and improved propagator-side density modelling are independent paths to the same target.

***Solar-cycle coverage**:* The dataset spans 2022-2025, covering the rising phase and peak of Solar Cycle 25, but no solar-minimum data. Performance under solar-minimum conditions, when the post-fit residual is weak in magnitude, may differ from the figures reported here and warrants separate evaluation in future work.

***Uncertainty quantification**:* The framework produces a single forecast (single best estimate, no error bounds) and does not provide a calibrated uncertainty estimate. Operational orbit-prediction systems require a formal covariance or prediction-interval characterisation to support principled fall-back logic on the user side, which is the rule by which a user decides what to do when the broadcast correction is flagged as unreliable, such as down-weighting the satellite, switching to an alternative orbit source, or excluding it from the positioning solution. Several routes to calibrated uncertainty alongside the point prediction, such as Bayesian GRU variants, Monte Carlo dropout, and deep ensembles, are identified as a priority for follow-on work. This also has a direct broadcast interface implication. An operational LEO-PNT message would carry a reference epoch and issue-of-data flag, analogous to the GPS time-of-ephemeris and IODE, so that the user can determine the age of the received correction. More importantly, the user needs a broadcast accuracy indicator (similar to the GPS User Range Accuracy) to weight the orbit term in their error budget and to decide when to down-weight or exclude the satellite. Because the present framework outputs only a point forecast, it cannot yet predict such uncertainties; the calibrated-uncertainty methods above are precisely what would enable a trustworthy broadcast accuracy indicator. We note, however, that the continuous-refresh design already bounds the age of the correction the user holds by the observation-epoch interval (30 s here) rather than by the upload interval, so the staleness the user must account for is small; what the accuracy indicator must convey is the residual prediction uncertainty, which still depends on the time-since-fit

***Reference-orbit independence**:* The training labels are derived from arc-N+1 reduced-dynamic orbits, fitted independently to new GNSS observations and approximating the true trajectory. However, the predicted and reference orbits share the same force-model structure and processing assumptions, so any systematic modelling bias common to both cancels in the reference-minus-predicted difference and does not appear in the error used for training. The network therefore learns to correct the stochastic, arc-to-arc component of the prediction error rather than any systematic bias consistent across arcs. A geometrically independent reference orbit determination approach would help isolate this systematic component and is identified as a necessary step before operational deployment.

***Orbit-plus-clock prediction**:* User-side navigation accuracy in a LEO-PNT service is governed by the signal-in-space range error (SISRE; Montenbruck et al. 2018), which combines the radial orbit prediction error with the clock prediction error. For a LEO transmitter at GRACE-FO altitude, these two contributions are of comparable magnitude (El-Mowafy et al. 2023), so an orbit-only correction is not a complete operational deliverable. Therefore, the integrated orbit-plus-clock prediction framework will be reported in a companion paper.

***Alternative sequence backbones**:* The choice of GRU is principled but has not been benchmarked against modern alternatives under the propagator configuration. Transformer self-attention (Vaswani et al. 2017), temporal convolutional networks, and state-space models each bring different modelling strengths that may suit either the short-prediction approach, where local autocorrelation dominates, or the long-prediction approach, where global structure matters. A protocol-matched architectural comparison is identified as a methodological follow-up.

Onboard implementation: By design, CNC-RDOP operates on the real-time kinematic position solution, differencing it against the propagated orbit to form the encoder's error input. This is what makes the framework deployable in flight without any post-processed product, and is the central operational distinction from offline cold-start methods (Section 4.7). In the present study, the encoder residual was formed from a real-time kinematic solution based on IGS RTS products, while accuracy was evaluated against the precise reduced-dynamic reference. Generating the kinematic solution onboard in flight (rather than from ground-

streamed RTS products) and the accompanying onboard propagation error budget remain the subject of a dedicated implementation study.

5. Conclusions

This paper presents CNC-RDOP, a real-time, continuous-refresh residual-extrapolation system in which a GRU Seq-to-Seq network consumes a 30-min trailing window of real-time, kinematic-derived SWR prediction error, together with the propagated ECEF state, and forecasts the next 60 minutes of error trajectory. On the 2025 GRACE-FO C test set with the geodetic-standard propagator, the framework holds the 5-min 3D RMSE to ~2-31 m across the entire 0-6 h time-since-fit range, against a baseline of ~2-463 m; single- to tens-of-metre accuracy where the propagator alone degrades by more than two orders of magnitude. Transfer to GRACE-FO D, which corresponds to using the trained network of GRACE-FO C and using it for GRACE-FO D without retraining, reproduces these results closely. The improvement percentage matches GFCN and the corrected 3D RMSE to within ~1% at every prediction length, confirming that a single trained network can serve every satellite in a co-orbital cohort without per-satellite retraining. An operational hybrid bypass, which broadcasts the propagator orbit directly for the first 10 minutes after each fit and the neural-corrected orbit thereafter, exploits the propagator's natural meter-class accuracy just after a fit and the network's strength as the fit ages.

The fresh-fit ceiling is shown to be a feature-limited noise floor rather than an architectural one. At the fresh fit, the 5-min error is set by the arc-boundary state consistency and the dm-to-m noise of the real-time kinematic input, not by 5-min drag, which is at the mm level, so the inputs available in real time cannot constrain the residual below that level. A state-feature-only network therefore has an estimated floor of order 0.5 m for the fresh-fit case; decimetre accuracy requires additional in-flight information beyond what the present framework assumes. Lifting this ceiling is system-dependent rather than an algorithmic problem and is left to future work.

Competing Interests

The authors declare that they have no competing financial or non-financial interests relevant to the content of this article.

Author Contributions

AA proposed the main idea, designed and performed the experiments, and wrote the manuscript; AE and SF helped with the writing and reviewing the paper.

Data Availability

The GRACE-FO GNSS observation data are publicly available from NASA's Physical Oceanography Distributed Active Archive Centre (PO.DAAC, https://podaac.jpl.nasa.gov). IGS orbit and clock products, including the Real-Time Service streams, are available from the International GNSS Service (https://igs.org). Precise orbits were generated with the Bernese GNSS Software, available under license from the Astronomical Institute of the University of Bern.

Acknowledgements

This research was supported by the Curtin University Faculty of Science and Engineering Research Grant ("Machine Learning-Enhanced Orbit Prediction for Low Earth Orbit Satellites") and by the Australian Research Council Discovery Project DP240101710.

References

Allahverdi-Zadeh A, Asgari J, Amiri-Simkooei A R (2016) Investigation of GPS draconitic year effect on GPS time series of eliminated eclipsing GPS satellite data. J Geod Sci, 6(1), 93–102. https://doi.org/10.1515/jogs-2016-0007

Allahvirdizadeh A, El-Mowafy A, Mcclusky S, Allgeyer S, Hammond A (2024a) Ginan Supporting Future LEO-PNT. IGNSS2024, Sydney, NSW.
https://doi.org/10.13140/RG.2.2.22771.30248

Allahvirdi-Zadeh A, El-Mowafy A, Wang K (2024b) Leveraging Future LEO Constellations for the Precise Orbit Determination of Lower Small Satellites. Proceedings of the 2024 International Technical Meeting of The Institute of Navigation, Long Beach, California, January 2024, pp. 756-769. https://doi.org/10.33012/2024.19485

Allahvirdi-Zadeh A, El-Mowafy A, Wang K (2025a) Doppler Positioning Using Multi-Constellation LEO Satellite Broadband Signals as Signals of Opportunity. NAVIGATION J Inst Navig, 72(2), navi.691. https://doi.org/10.33012/navi.691

Allahvirdi-Zadeh A, El-Mowafy A, Wang K (2025b) Network-based precise orbit determination of broadband LEO satellites using Doppler-shift measurements. GPS Solut, 30(1), 16. https://doi.org/10.1007/s10291-025-01963-0

Allahvirdi-Zadeh A, Wang K, El-Mowafy A (2022) Precise Orbit Determination of LEO Satellites Based on Undifferenced GNSS Observations. J Surv Eng, 148(1), 03121001. https://doi.org/10.1061/(ASCE)SU.1943-5428.0000382

Allahvirdizadeh A (2022) Precise Orbit Determination of CubeSats [PhD, Curtin University]. Curtin Theses. https://doi.org/10.13140/RG.2.2.15681.43369

Bengio S, Vinyals O, Jaitly N, Shazeer N (2015) Scheduled sampling for sequence prediction with recurrent neural networks. Adv Neural Inf Process Syst, 28. https://doi.org/10.48550/arXiv.1506.03099

Cho K, Van Merriënboer B, Gulçehre Ç, Bahdanau D, Bougares F, Schwenk H, Bengio Y (2014) Learning phrase representations using RNN encoder–decoder for statistical machine translation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) (pp. 1724-1734). https://doi.org/10.48550/arXiv.1406.1078

Dach R, Lutz S, Walser P, Fridez P (2015). Bernese GNSS Software Version 5.2. Bern Open Publishing. https://doi.org/10.7892/boris.72297

El-Mowafy A, Wang K, Li Y, Allahvirdi-Zadeh A (2023) The Impact of Orbital and Clock Errors on Positioning from LEO Constellations and Proposed Orbital Solutions. Int Arch Photogramm Remote Sens Spat Inf Sci. https://doi.org/10.5194/isprs-archives-XLVIII-1-W2-2023-1111-2023

Emmert J T, Jones Jr M, Siskind D E, Drob D P, Picone J M, Stevens M H, Bailey S M, Bender S, Bernath P F, Funke B, Hervig M E, Pérot K (2022) NRLMSIS 2.1: An Empirical Model of Nitric Oxide Incorporated Into MSIS. J Geophys Res Space Phys, 127(10), e2022JA030896. https://doi.org/10.1029/2022JA030896

Ge H, Li B, Ge M, Zang N, Nie L, Shen Y, Schuh H (2018) Initial Assessment of Precise Point Positioning with LEO Enhanced Global Navigation Satellite Systems (LeGNSS). Remote Sens, 10(7), 984. https://doi.org/10.3390/rs10070984

Ge H, Li B, Jia S, Nie L, Wu T, Yang Z, Shang J, Zheng Y, Ge M (2022) LEO Enhanced Global Navigation Satellite System (LeGNSS): progress, opportunities, and challenges. Geo-Spat Inf Sci, 25(1), 1-13. https://doi.org/10.1080/10095020.2021.1978277

González A, Rodriguez I, Navarro P, Sobrero F, Carbonell E, Calle D, Fernández J (2022) LEO satellites for PNT, the next step for precise positioning applications. Proceedings of the 35th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2022). https://doi.org/10.33012/2022.18436

Guo M, Lu T, Chen S, Song X, Liu D, Wu X (2025) Orbital prediction accuracy improvement method based on particle swarm optimization RBF neural network. Adv Space Res, 75. https://doi.org/10.1016/j.asr.2025.03.031

Jäggi A, Hugentobler U, Beutler G (2006) Pseudo-Stochastic Orbit Modeling Techniques for Low-Earth Orbiters. J Geod, 80(1), 47-60. https://doi.org/10.1007/s00190-006-0029-9

Kingma D P, Ba J (2014) Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980. https://doi.org/10.48550/arXiv.1412.6980

Kozhaya S E, Haidar-Ahmad J A, Abdallah A A, Kassas Z M, Saab S S (2021) Comparison of neural network architectures for simultaneous tracking and navigation with LEO satellites. Proceedings of the 34th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2021). https://doi.org/10.33012/2021.18110

Li B, Wu T, Ge H (2025) Deep neural network based precise orbit prediction for low earth orbit (LEO) satellites. J Geod, 99(9), 75. https://doi.org/10.1007/s00190-025-01999-7

Liu J, Wang K, Chen B, El-Mowafy A, Allahvirdi-Zadeh A, Yang X (2025) Accuracy Degradation Rate of LEO Satellite Predicted Orbits Using Different POD Methods. Proceedings of the 2025 International Technical Meeting of The Institute of Navigation. https://doi.org/10.33012/2025.20020

Lyard F H, Allain D J, Cancet M, Carrère L, Picot N (2021) FES2014 global ocean tide atlas: design and performance. Ocean Sci, 17(3), 615-649. https://doi.org/10.5194/os-17-615-2021

Montenbruck O, Gill E (2000) Satellite Orbits: Models, Methods and Applications. Springer. https://doi.org/10.1007/978-3-642-58351-3

Montenbruck O, Steigenberger P, Hauschild A (2018) Multi-GNSS signal-in-space range error assessment–Methodology and results. Adv Space Res, 61(12), pp.3020-3038. https://doi.org/10.1016/j.asr.2018.03.041

Pavlis N K, Holmes S A, Kenyon S C, Factor J K (2012) The development and evaluation of the Earth Gravitational Model 2008 (EGM2008). J Geophys Res Solid Earth, 117(B4). https://doi.org/10.1029/2011JB008916

Petit G, Luzum B (2010) IERS conventions. Verlag des Bundesamts für Kartographie und Geodäsie. https://www.iers.org/IERS/EN/Publications/TechnicalNotes/tn36.html

Picone J M, Hedin A E, Drob D P, Aikin A C (2002) NRLMSISE-00 empirical model of the atmosphere: Statistical comparisons and scientific issues. J Geophys Res Space Phys, 107(A12), SIA 15–11–SIA 15–16. https://doi.org/10.1029/2002JA009430

Scorsoglio A, Ghilardi L, Furfaro R (2023) A Physic-Informed Neural Network Approach to Orbit Determination. J Astronaut Sci, 70(4), 25. https://doi.org/10.1007/s40295-023-00392-w

Selvan K, Siemuri A, Prol F S, Välisuo P, Kuusniemi H (2024) Machine Learning for LEO and MEO Satellite Orbit Prediction. Proceedings of the 37th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2024). https://doi.org/10.33012/2024.19765

Sutskever I, Vinyals O, Le Q V (2014) Sequence to sequence learning with neural networks. Adv Neural Inf Process Syst, 27. https://doi.org/10.48550/arXiv.1409.3215

Thammawichai M, Luangwilai T (2024) Data-driven satellite orbit prediction using two-line elements. Astron Comput, 46, 100782. https://doi.org/10.1016/j.ascom.2023.100782

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser Ł, Polosukhin I (2017) Attention is all you need. Adv Neural Inf Process Syst, 30. https://doi.org/10.48550/arXiv.1706.03762

Wang K, Allahvirdi-Zadeh A, El-Mowafy A, Gross J N (2020) A Sensitivity Study of POD Using Dual-Frequency GPS for CubeSats Data Limitation and Resources. Remote Sens, 12(13):2107. https://doi.org/10.3390/rs12132107

Zhang W, Zhang K, Li X, Huang J (2025) Improving LEO short-term orbit prediction using LSTM neural network. Adv Space Res, 76(1), 481-496. https://doi.org/10.1016/j.asr.2025.04.067