Improving GNSS Performances in Location-Based Services through Synthetic Carrier-Phase Measurements

Paolo Dabove[1], Milad Bagheri[1], Neil Gogoi[1]

[1] Dept. of Environment, Land, and Infrastructure Engineering (DIATI), Politecnico di Torino-
(paolo.dabove, milad.bagheri, neil.gogoi)@polito.it

Keywords: GNSS, carrier phase gap prediction, signal gap bridging, cycle slip repair, machine learning, satellite geometry.

Abstract

Carrier phase observations enable millimeter-level GNSS positioning, but their continuity is frequently disrupted by signal blockages and cycle slips. This limitation is particularly critical for low-cost and smartphone receivers, where weak antennas, urban multipath, and duty cycling cause frequent phase gaps that prevent reliable ambiguity resolution. Before addressing the full complexity of mass-market observations, the prediction methodology must be validated under controlled conditions. In this work we investigate whether machine learning, supported by precise satellite orbits and clocks, can predict carrier phase observations during signal gaps with millimeter-level accuracy. Twenty-four hours of Galileo data from the TORI permanent station (SPIN3 network, Torino, Italy) are processed at 30 s sampling using GFZ final SP3 and CLK products. After forming the ionosphere-free combination, an iterative carrier-phase based estimator removes the receiver clock, tropospheric delay, and ambiguity, reducing the residuals to a median standard deviation of 60 mm. Synthetic gaps from 60 s to 1800 s are introduced (1045 gaps total) and four prediction strategies are compared: polynomial fitting (degrees 3 and 5), Fourier-augmented polynomial, Gradient Boosting Regression with satellite geometry features, and Gaussian Process Regression. The Gradient Boosting model achieves the best overall performance, reaching 4.4 mm RMS for 60 s gaps, 9.4 mm for 5 min gaps, and 21 mm for 30 min gaps, well below the half-wavelength threshold required for cycle slip repair. These results demonstrate that geometry-aware gap prediction is feasible at the sub-wavelength level, providing a validated foundation for extending the approach to low-cost and smartphone GNSS receivers.

1. Introduction

Location-Based Services (LBS) have become a fundamental component of modern digital ecosystems, supporting applications such as navigation, mobility management, asset tracking, and context-aware services. In most outdoor scenarios, Global Navigation Satellite Systems (GNSS) represent the primary positioning technology for mass-market devices, including smartphones and Internet of Things (IoT) platforms. However, the positioning accuracy and reliability achievable with low-cost GNSS receivers remain limited, especially in challenging environments such as urban canyons, where multipath, signal attenuation, and non-line-of-sight conditions significantly degrade performance (Geng et al., 2021).

Despite this widespread adoption, low-cost GNSS observations are strongly affected by hardware limitations, including low-gain linearly polarized antennas, higher noise, frequent signal blockage, and severe multipath. In particular, carrier-phase measurements — which are the basis for high-precision positioning techniques such as Precise Point Positioning (PPP) — remain difficult to exploit reliably in mass-market devices due to frequent cycle slips and unstable ambiguity resolution (Pesyna et al., 2014, Kim et al., 2022, Bagheri et al., 2025). As a result, there is a growing interest in developing alternative approaches that can enhance positioning performance without requiring fully reliable carrier-phase tracking.

Within this context, predicting or reconstructing carrier-phase observations during signal gaps can be viewed as a promising strategy to bridge the gap between noisy code-based observables and full-quality carrier-phase processing. The basic idea is that, by removing the deterministic components of the GNSS observation equation using precise satellite orbits and clocks, the remaining carrier phase residual becomes small and smooth enough to be predicted by data-driven models. Such an approach is particularly attractive because it leverages temporal correlations and satellite geometry to maintain carrier-phase continuity even when the signal is temporarily lost.

However, before this strategy can be deployed on low-cost and smartphone receivers, the prediction methodology must first be validated under controlled conditions where the ground truth is known and the observation quality is not itself a confounding factor. For this reason, the present study addresses the carrier phase gap prediction problem on a geodetic permanent station equipped with a high-quality receiver and antenna, establishing a performance baseline under ideal conditions.

Specifically, we present a complete pipeline to: (1) compute observed-minus-computed carrier phase residuals using precise orbits, precise clocks, and the ionosphere-free combination; (2) remove the receiver clock and tropospheric delay through an iterative carrier-phase based estimator; (3) inject synthetic gaps of 60 s to 1800 s in continuous arcs; and (4) predict the missing observations using four strategies of increasing sophistication, including a Gradient Boosting model that exploits satellite elevation and azimuth as prediction features. The pipeline is applied to twenty-four hours of Galileo data from the TORI permanent station of the SPIN3 network in Torino, Italy, processed at 30 s sampling.

2. State of the Art

Research on low-cost and smartphone GNSS has expanded rapidly since raw observations became accessible on Android devices. A recent review shows that the field has evolved from early quality assessment of smartphone observables toward increasingly sophisticated positioning methods, including precise point positioning, relative positioning, and GNSS/INS integration (Geng et al., 2021). At the same time, persistent weaknesses in carrier-to-noise ratio, carrier-phase continuity, and overall observation quality directly affect the feasibility of precise carrier-phase-based solutions on mass-market devices. Studies on both smartphones and low-cost dedicated GNSS boards consistently report lower observation quality than geodetic-grade systems, especially in terms of multipath resistance, cycle-slip occurrence, and code noise (Geng et al., 2021, Hamza et al., 2024). For smartphones specifically, the antenna is a critical limiting factor, because poor multipath suppression and irregular gain patterns produce large time-correlated phase errors and destabilize ambiguity resolution (Pesyna et al., 2014). In the broader low-cost domain, dual-frequency receivers such as the u-blox ZED-F9P have demonstrated that affordable hardware can approach high-end performance under favorable conditions (Hamza et al., 2024), but observation quality, antenna choice, and environmental conditions remain decisive. Across all these platforms, the central bottleneck is the same: carrier phase observations are frequently interrupted by signal blockages and cycle slips, and these interruptions degrade or prevent precise positioning.

Conventional cycle slip detection and repair techniques rely on observation combinations such as the geometry-free combination, the Melbourne–Wübbena combination, and the ionosphere-free combination (Blewitt, 1990, de Lacy et al., 2012). These methods perform well for slips of a few cycles when the gap is short and the ionosphere is quiet. They become unreliable when the gap is longer than a few seconds, when ionospheric activity is high, or when only a single frequency is available — conditions that are common in low-cost and smartphone receivers. For static permanent stations the situation is more favorable because the receiver position is known and can be treated as a constant, but even in this case gaps exceeding a few tens of seconds exceed the prediction horizon of classical combination-based repair methods.

Recent advances in machine learning have opened new possibilities for GNSS signal processing. Several authors have applied artificial neural networks, recurrent networks, and tree-based methods to GNSS-related tasks such as multipath mitigation, ionospheric prediction, troposphere modeling, and PPP convergence acceleration (Siemuri et al., 2022, Tao et al., 2021, Zhang et al., 2023). Gaussian Process Regression has been used to model atmospheric delays with uncertainty quantification (Rasmussen and Williams, 2006). In parallel, data-driven approaches have begun to model smartphone-specific GNSS errors and learn positioning corrections directly from raw observations (Mohanty et al., 2023), and optimization-based processing has been shown to improve smartphone positioning performance significantly (Suzuki, 2023). However, the use of machine learning for direct prediction of carrier phase residuals during signal gaps has received limited attention, especially when combined with the physical structure provided by precise satellite orbits and clocks.

The motivation for such a prediction approach rests on two observations. First, by exploiting the deterministic part of the GNSS observation equation — geometric range, satellite clock, and ionospheric delay via the dual-frequency combination — most of the variability of the raw carrier phase signal can be removed analytically. The remaining residual is dominated by tropospheric delay, multipath, and noise, all of which evolve smoothly in time and are strongly correlated with satellite elevation and azimuth. Second, the cleaned residual is small enough (centimeter level) that even a moderately accurate predictor produces values useful for cycle slip repair, since one Galileo E1 wavelength is approximately 19 cm.

The present study addresses this gap. Rather than attempting to solve the full smartphone carrier phase reconstruction problem directly, we first validate the gap prediction methodology on a geodetic permanent station, where the known receiver coordinates and high-quality observations provide a controlled environment to isolate and evaluate the prediction accuracy. This approach establishes a performance baseline under ideal conditions, before the method is extended to low-cost and smartphone receivers in future work.

3. Data and Methodology

3.1 Data and Station

The data set consists of one full day (DOY 014, 14 January 2026) of multi-GNSS RINEX 3.04 observations recorded at the TORI station (Politecnico di Torino, SPIN3 network). The receiver is a Leica GR30, the antenna is a LEIAR25.R3 NONE, and the sampling interval is 1 s. The geodetic coordinates are 45° 03’ 48” N, 7° 39’ 41” E, ellipsoidal height 310.8 m. For satellite orbits and clocks the GFZ final products are used: 15 min SP3 orbits and 30 s CLK clocks. The native 1 s observation file is decimated to 30 s during parsing, yielding 2880 epochs and 25 Galileo satellites tracked over the day.

3.2 Observation Model and Residual Computation

The dual-frequency carrier phase observation at frequency i is

$$L_i = \rho + c\,\delta t_r - c\,\delta t^s + T - \frac{f_1^2}{f_i^2}\,I_1 + \lambda_i N_i + m_i + \varepsilon_i \quad (1)$$

where ρ is the geometric range, c is the speed of light, δt_r and δt^s are the receiver and satellite clock offsets, T is the tropospheric delay, I_1 is the slant ionospheric delay at the first frequency, $\lambda_i N_i$ is the integer ambiguity in meters, m is the multipath, and ε is the measurement noise.

The first-order ionospheric delay is removed by forming the ionosphere-free (IF) combination

$$L_{IF} = \frac{f_1^2\,L_1 - f_2^2\,L_2}{f_1^2 - f_2^2} \quad (2)$$

with $f_1 = 1575.42$ MHz (Galileo E1) and $f_2 = 1176.45$ MHz (Galileo E5a).

The satellite position at signal transmission time is obtained from the SP3 file by degree-10 Lagrange interpolation. The signal travel time τ is estimated iteratively: starting from $\tau = 0.07$ s, the geometric range is computed as

$$\rho = \sqrt{(X^s - X_r)^2 + (Y^s - Y_r)^2 + (Z^s - Z_r)^2} \quad (3)$$

and corrected for the Sagnac effect:

$$\rho_{\text{corrected}} = \rho + \frac{\omega_E}{c}\left(X^s\,Y_r - Y^s\,X_r\right) \quad (4)$$

Two iterations are sufficient for millimeter convergence. The satellite clock δt^s is interpolated linearly from the CLK file, and the ionosphere-free observed-minus-computed (O–C) residual is formed as

$$v_{IF} = L_{IF} - \rho + c\,\delta t^s = c\,\delta t_r + T + \lambda_{IF} N_{IF} + m_{IF} + \varepsilon_{IF} \quad (5)$$

3.3 Iterative Clock and Troposphere Estimation

Estimating the receiver clock from pseudorange residuals injects approximately 1 m of noise into the carrier phase residuals. To overcome this, an iterative scheme based exclusively on carrier phase is implemented.

A first guess of the receiver clock is obtained as the per-epoch median of the pseudorange residuals:

$$\delta t_r^{(0)} = \mathrm{median}_s \left[P_{IF}^s - \rho^s + c\,\delta t^s \right] \quad (6)$$

For each iteration k, two steps are alternated. In Step A, a per-arc tropospheric model is fitted:

$$\hat{T}^s(t) = a^s\, m(\varepsilon^s(t)) + b^s + c_1^s\,\Delta t + c_2^s\,\Delta t^2 + c_3^s\,\Delta t^3 \quad (7)$$

where $m(\varepsilon)$ is a Niell-like wet mapping function

$$m(\varepsilon) = \frac{1}{\sin\!\big(\varepsilon + 0.00143/(\tan\varepsilon + 0.0455)\big)} \quad (8)$$

ε is the satellite elevation, $\Delta t = (t - t_0)/3600$ is the elapsed time in hours, a^s absorbs the zenith total delay, b^s absorbs the constant ambiguity, and the cubic polynomial models the slow temporal variation of ZTD. The five coefficients are estimated by ordinary least squares for each arc independently.

In Step B, the receiver clock is re-estimated from carrier phase residuals after removing the per-arc model:

$$\delta t_r^{(k+1)} = \mathrm{median}_s \left[v_{IF}^s - \hat{T}^s \right] \quad (9)$$

The median provides robustness against multipath outliers. Four iterations are sufficient for convergence. The final cleaned residual for each arc is

$$\tilde{v}^s(t) = v_{IF}^s(t) - c\,\delta t_r(t) - \hat{T}^s(t) \quad (10)$$

containing only multipath, higher-order tropospheric variation, and receiver noise — typically at the centimeter level.

3.4 Synthetic Gap Creation and Prediction Models

For each cleaned arc, gaps of duration $\Delta \in \{60, 120, 300, 600, 900, 1800\}$ s are randomly inserted, avoiding the first and last 15 % of the arc and requiring an average elevation above 15° within the gap. Five gaps per duration per arc are generated, yielding 1045 synthetic gaps over the full data set. For each gap, the truth values are stored separately and the surrounding context (up to three times the gap duration on both sides) is provided to the prediction models.

Four prediction strategies are compared. The **polynomial interpolation** fits a degree-d polynomial ($d = 3$ or 5) to the context residuals:

$$\tilde{v}(t) \approx \sum_{k=0}^{d} a_k\, t^k \quad (11)$$

The **Fourier-augmented polynomial** extends this with $K = 4$ harmonics to capture multipath periodicities:

$$\tilde{v}(t) \approx \sum_{k=0}^{d} a_k\, t^k + \sum_{k=1}^{K} \big[b_k \sin(k\omega t) + c_k \cos(k\omega t) \big] \quad (12)$$

where $\omega = 2\pi/T$ and T is the context window span.

The **Gradient Boosting Regressor** (GBR) builds an additive ensemble of 200 regression trees of depth 4 with learning rate 0.05 and Huber loss (Friedman, 2001). Unlike the previous methods, the GBR operates on a feature vector that includes the satellite geometry:

$$\mathbf{x}(t) = \left[\, \tilde{t},\ \sin\varepsilon,\ \cos\varepsilon,\ \sin\alpha,\ \cos\alpha,\ \sin 2\varepsilon,\ \dot{\varepsilon},\ \dot{\alpha} \,\right] \quad (13)$$

where ε is the elevation, α is the azimuth, and $\dot{\varepsilon}$, $\dot{\alpha}$ are their time derivatives. All features remain available inside the gap because the satellite position can be computed from the SP3 file at any epoch.

The **Gaussian Process Regression** (GPR) models the residual as a sample from a Gaussian Process with a composite RBF + White noise kernel (Rasmussen and Williams, 2006):

$$k(t, t') = \sigma_f^2 \exp\!\left(-\frac{(t - t')^2}{2\ell^2} \right) + \sigma_n^2\, \delta_{tt'} \quad (14)$$

Hyperparameters are optimized by marginal log-likelihood. Due to the $O(n^3)$ cost, the context is subsampled to 500 points and only three gaps per duration are evaluated with GPR.

For each gap and method, the prediction error is $e(t) = \tilde{v}_{\text{truth}}(t) - \tilde{v}_{\text{predicted}}(t)$, and the RMS, MAE, and maximum absolute error are reported, aggregated over the 1045 gaps and grouped by duration.

4. Results and Discussion

The pipeline described in Section 3 was applied to 24 hours of Galileo data from the TORI permanent station. The results are presented in three stages: first, the quality of the satellite geometry and data coverage is assessed; second, the effectiveness of the iterative clock and troposphere estimation is evaluated through the reduction of the carrier phase residuals; third, the four prediction strategies are compared across the full range of synthetic gap durations.

A total of 25 Galileo satellites were tracked during the analysis day, most of them completing a full visibility arc above the 10° elevation cutoff. The mean arc elevations range between 15° and 55°, and the sky distribution is dense and well balanced (Figure 1), providing a strong geometric configuration throughout the day.

The quality of the carrier phase residuals before and after the iterative cleaning is the critical indicator of how well the deterministic part of the observation equation has been removed. Before the iterative procedure, the IF residuals are dominated by the receiver clock offset and the per-arc ambiguity, reaching several meters in amplitude. After four iterations the residuals collapse to the centimeter level (Figure 2). The mean cleaned residual standard deviation is 68 mm and the median is 60 mm,

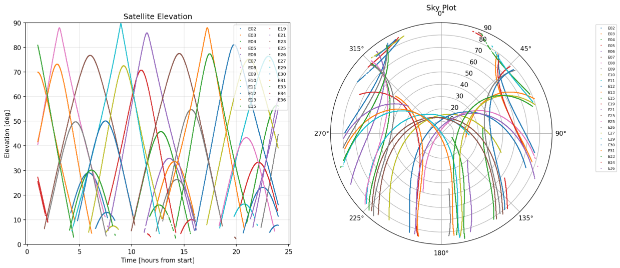

Figure 1. Satellite elevation time series (left) and sky plot (right) for 25 Galileo satellites tracked at TORI on 14 January 2026.

with the best arcs reaching 9–15 mm. The convergence is well-behaved, with the per-iteration receiver clock change decreasing monotonically from approximately 825 mm at iteration 1 to 46 mm at iteration 4.

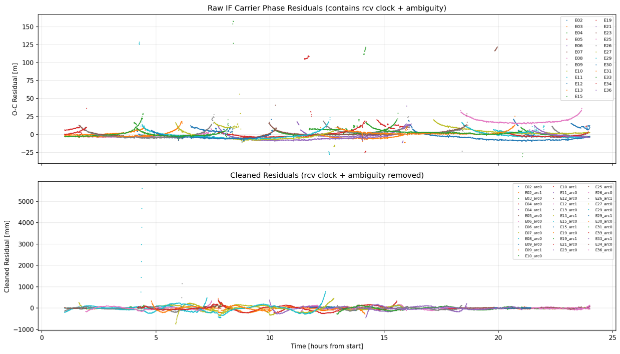

Figure 2. Top: raw IF carrier phase residuals. Bottom: cleaned residuals after iterative clock and troposphere estimation.

The estimated receiver clock (Figure 3) exhibits a total dynamic range of approximately 18 m, consistent with the typical behavior of an unsteered geodetic receiver. This smooth but large signal underlines the importance of the carrier-phase based iterative estimator, which a single-pass pseudorange clock would leave meter-level residuals that mask the centimeter-level structure relevant for gap prediction.

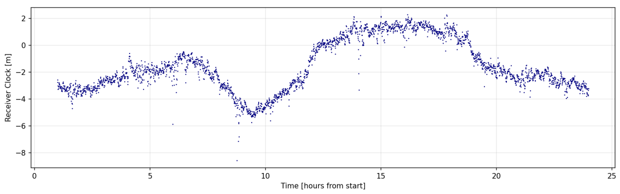

Figure 3. Estimated receiver clock from the iterative carrier-phase based estimator.

With the cleaned residuals established at the centimeter level, the four prediction strategies were applied to the 1045 synthetic gaps. Representative examples for the six tested durations (60 s to 1800 s) are shown in Figure 4. For short gaps all methods follow the truth closely. As the gap duration increases, differences become apparent: the Fourier+Polynomial method tends to overshoot for the longest gaps, while Gradient Boosting and Gaussian Process remain stable.

The quantitative comparison of the four methods is summarized in Table 1, which reports the mean RMS, its standard deviation, the MAE, and the mean maximum error for each method

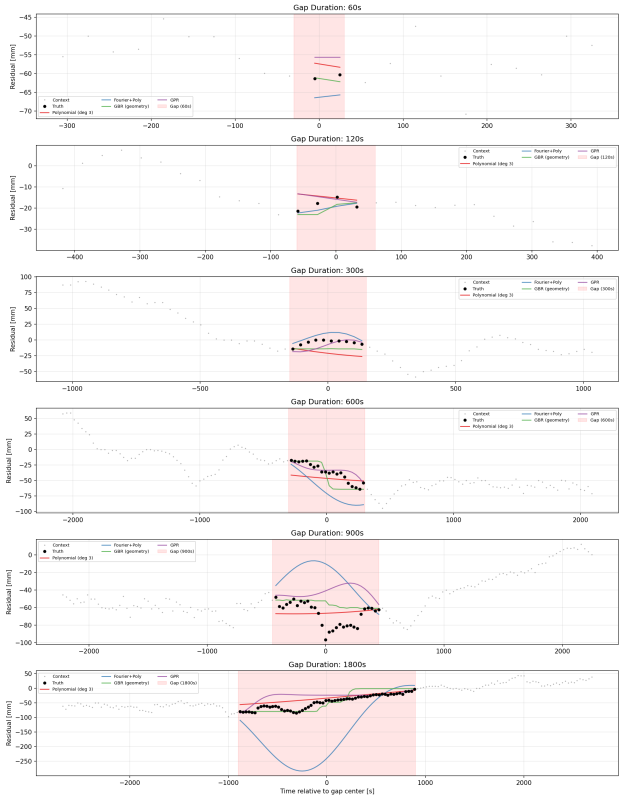

Figure 4. Examples of carrier phase gap prediction for six durations. The shaded area marks the gap; colored curves are the four prediction methods.

and gap duration. The Gradient Boosting Regressor with geometry features (GBR) is the most consistently accurate method from 60 s to 900 s, reaching 4.4 mm RMS at 60 s, 9.4 mm at 5 min, and 14.4 mm at 15 min. For the longest gap (1800 s), GBR is essentially tied with the degree-3 polynomial (21.4 mm vs 20.3 mm). The Gaussian Process Regression, evaluated on only 3 gaps per duration due to its computational cost, delivers the lowest RMS for the shortest gaps (3.5 mm at 60 s) but degrades for longer gaps and is computationally prohibitive at scale. The Fourier-augmented polynomial diverges at 1800 s (151 mm mean RMS), confirming that Fourier harmonics amplify when extrapolated beyond the fitted interval.

The growth of the RMS error with gap duration is shown in Figure 5. For all methods except Fourier+Polynomial, the error increases approximately linearly with the logarithm of the gap duration, confirming that the dominant residual components (multipath and unmodeled troposphere) evolve smoothly over the time scales considered. The error bars indicate that the GBR model produces the most compact spread across all durations. The Fourier+Polynomial, shown separately in the bottom panel, diverges at 1800 s (151 mm mean RMS) due to the well-known instability of harmonic extrapolation outside the fitted interval. The dashed green line marks the half-wavelength threshold ($\lambda/2 \approx 95$ mm), which all methods except Fourier+Polynomial remain well below.

The comparison across methods reveals a clear hierarchy. The polynomial methods (degree 3 and 5) provide a robust baseline whose accuracy is acceptable up to approximately 5 minutes, after which the lack of physical constraints causes the fitted curve to drift. The Fourier-augmented polynomial offers no

Gap [s]	Method	RMS [mm]	±std [mm]	MAE [mm]	Max [mm]
60	Poly. deg 3	5.8	4.9	5.5	7.0
60	Poly. deg 5	5.3	4.3	4.9	6.5
60	Fourier+Poly.	5.3	4.0	5.0	6.4
60	GBR (geom.)	4.4	3.3	4.0	5.4
60	GPR	3.5	1.4	3.4	4.1
120	Poly. deg 3	7.2	5.0	6.4	10.4
120	Poly. deg 5	6.7	5.5	6.0	9.7
120	Fourier+Poly.	7.8	6.2	7.0	11.2
120	GBR (geom.)	6.6	4.9	5.6	10.1
120	GPR	3.4	0.8	2.5	6.0
300	Poly. deg 3	10.9	10.0	9.7	17.8
300	Poly. deg 5	10.6	9.4	9.3	17.4
300	Fourier+Poly.	10.6	7.4	9.2	17.8
300	GBR (geom.)	9.4	7.0	7.7	17.1
300	GPR	7.5	4.0	6.1	13.0
600	Poly. deg 3	14.4	15.0	12.7	24.3
600	Poly. deg 5	12.5	11.7	10.9	21.6
600	Fourier+Poly.	14.0	11.2	12.3	24.0
600	GBR (geom.)	11.4	8.8	9.4	21.5
600	GPR	9.1	1.6	7.3	20.3
900	Poly. deg 3	16.6	16.2	14.5	29.7
900	Poly. deg 5	15.2	14.9	13.1	27.5
900	Fourier+Poly.	23.9	25.1	20.8	40.7
900	GBR (geom.)	14.4	14.0	11.5	29.4
900	GPR	19.0	8.7	15.2	38.4
1800	Poly. deg 3	20.3	19.5	17.4	39.7
1800	Poly. deg 5	22.9	25.0	19.7	42.9
1800	Fourier+Poly.	151.1	314.7	124.7	254.2
1800	GBR (geom.)	21.4	17.2	17.6	43.3
1800	GPR	22.8	5.1	17.2	57.5

Table 1. Prediction error statistics [mm] for five methods and six gap durations at 30 s sampling (175 gaps each; 170 for 1800 s).

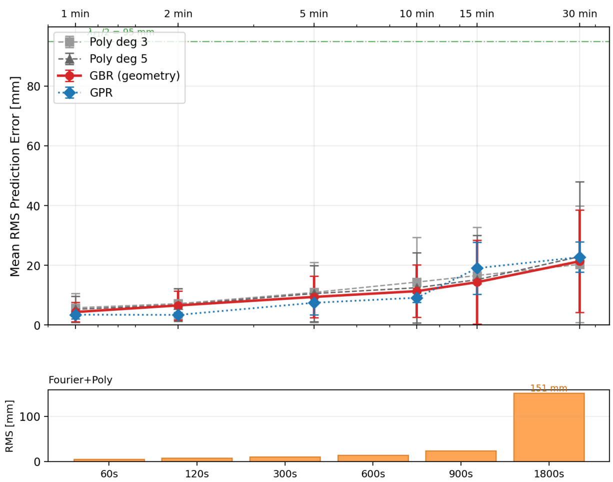

Figure 5. Top: mean RMS prediction error as a function of gap duration for polynomial, GBR, and GPR methods; error bars show the standard deviation across gaps. The dashed green line indicates the half-wavelength threshold. Bottom: Fourier+Polynomial shown separately due to its divergence at 1800 s.

systematic advantage over the simple polynomial and diverges for the longest gaps, a known weakness of harmonic extrapolation when the empirical period does not match the true multipath periodicity of the specific arc. The Gradient Boosting Regressor owes its superior performance to the use of elevation, azimuth, and their time derivatives as input features: because these quantities can be computed for any epoch from the SP3 orbit file — including inside the gap — the model effectively transfers the physical configuration learned from the surrounding context into the missing interval. For gaps exceeding 15 minutes, however, the local context-based GBR begins to lose its geometric advantage as the gap approaches the correlation length of the multipath signal, and its performance converges to that of the polynomial baseline.

From an operational perspective, the Galileo E1 carrier wavelength is approximately 19 cm. A prediction error smaller than half a wavelength is sufficient to resolve the integer number of cycles lost during the gap. The GBR model satisfies this condition for all tested durations up to and including 1800 s (mean RMS of 21 mm, corresponding to approximately 0.11 wavelengths), suggesting that AI-assisted gap prediction is a viable strategy for cycle slip repair in static permanent stations, even for outages of half an hour. For kinematic applications the same approach can be adopted as soon as the trajectory is approximately known from a complementary sensor such as an IMU or a low-rate code-only PPP solution.

Several limitations should be noted. The 30 s sampling rate is a deliberate compromise between temporal resolution and computational cost; it does not allow the analysis of gaps shorter than 60 s, which represent the most common cycle slip events. The cleaned residuals still contain a residual ZTD signal that could be further suppressed by adopting a random-walk tropospheric parameterization and the Vienna Mapping Function 3 (Landskron and Böhm, 2018), together with solid Earth tide, ocean loading, and phase wind-up corrections. A small fraction of arcs with cleaned residual standard deviations above 100 mm indicates that undetected cycle slips prevented the per-arc model from converging properly; a more sophisticated detector based on the wide-lane and geometry-free combinations would mitigate this. Finally, the current GBR model is trained per gap on its surrounding context; a globally trained model operating on the full 24 h data set, potentially based on Temporal Fusion Transformers or LSTM networks, is the natural evolution of this approach.

5. Conclusions

This study has demonstrated that the prediction of GNSS carrier phase observations during signal gaps is feasible at the millimeter level when precise orbits and clocks are combined with an iterative carrier-phase based estimator and a geometry-aware machine learning model. Twenty-four hours of Galileo data from the TORI permanent station were processed end-to-end, including ionosphere-free combination, iterative travel time correction, receiver clock and troposphere estimation, synthetic gap injection, and four-method gap prediction.

The iterative carrier-phase clock estimation reduces the cleaned residual standard deviation by a factor of approximately 5 compared to a single-pass pseudorange-based initialization, bringing the median residual from approximately 280 mm (iteration 1, pseudorange clock) down to 60 mm (iteration 4, carrier-phase clock) at 30 s sampling. Among the four prediction strategies, the Gradient Boosting Regressor with satellite geometry features proved the most accurate and robust method across the gap duration range from 60 s to 900 s, reaching 4.4 mm RMS at 60 s, 9.4 mm at 5 min, and 14.4 mm at 15 min. For 30 min gaps it is essentially tied with the degree-3 polynomial baseline (21.4 mm vs 20.3 mm). All methods achieve sub-wavelength prediction accuracy for gaps of up to 30 minutes, supporting the use of AI-assisted bridging for cycle slip repair and short-term carrier phase reconstruction in permanent GNSS stations.

Future work will extend the pipeline to the native 1 s sampling rate to address cycle-slip-relevant gaps shorter than 60 s, incorporate multi-constellation processing, adapt the method to kinematic platforms, and develop a globally trained machine learning model capable of predicting gaps without per-arc retraining.

6. Acknowledgment

The research activity has been carried out within the SPACE IT UP! project founded by the Italian Space Agency’s Call for ”Space Activities” (topic 15), related to MUR call n°341 of 03/15/2022, for ”Partnerships extended to universities, research centers, companies for the financing of basic research projects” (Prot. CI-2022-DSR-042 of 18 July 2022) financed by Decree of Award of the Italian Space Agency n. 53 of 01/26/2024 - CUP E53C24000300006. “Earth Observation Fund - Financing from the Presidency of the Council of Ministers pursuant to Art. 1, paragraph 254, of law 160/2019. The reference year 2024”.

References

Bagheri, M., Gogoi, N., Dabove, P., Pietra, V. d., 2025. Evaluating dual-frequency, multi-constellation gnss performance of modern smartphones for atmospheric monitoring across mountainous and urban areas. *2025 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, 495–504.

Blewitt, G., 1990. An automatic editing algorithm for GPS data. *Geophysical Research Letters*, 17(3), 199–202.

de Lacy, M. C., Reguzzoni, M., Sansò, F., 2012. Real-time cycle slip detection in triple-frequency GNSS. *GPS Solutions*, 16(3), 353–362.

Friedman, J. H., 2001. Greedy function approximation: a gradient boosting machine. *The Annals of Statistics*, 29(5), 1189–1232.

Geng, J., Li, G., Zhang, Q. et al., 2021. GNSS smartphone positioning: advances, challenges, opportunities, and future perspectives. *Satellite Navigation*, 2(24).

Hamza, V., Stopar, B., Sterle, O., 2024. Observations and positioning quality of low-cost GNSS receivers: a review. *GPS Solutions*, 28, 81.

Kim, H., Lee, J. et al., 2022. Inherent Limitations of Smartphone GNSS Positioning and Effective Methods to Increase the Accuracy Utilizing Dual-Frequency Measurements. *Sensors*, 22(24), 9879.

Landskron, D., Böhm, J., 2018. VMF3/GPT3: refined discrete and empirical troposphere mapping functions. *Journal of Geodesy*, 92(4), 349–360.

Mohanty, A., Gao, G. X. et al., 2023. Learning GNSS Positioning Corrections for Smartphones Using Graph Convolutional Neural Networks. *NAVIGATION*, 70(4).

Pesyna, K. M., Heath, R. W., Humphreys, T. E., 2014. Centimeter positioning with a smartphone-quality gnss antenna. *Proceedings of the 27th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS+ 2014)*, 1568–1577.

Rasmussen, C. E., Williams, C. K. I., 2006. *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, MA.

Siemuri, A., Selvan, K., Kuusniemi, H., Välisuo, P., Elmusrati, M. S., 2022. A systematic review of machine learning techniques for GNSS use cases. *IEEE Transactions on Aerospace and Electronic Systems*, 58(6), 5043–5077.

Suzuki, T., 2023. Precise Position Estimation Using Smartphone Raw GNSS Data Based on Two-Step Optimization. *Sensors*, 23(3), 1205.

Tao, Y., Liu, C., Chen, T., Zhao, X., Liu, C., Hu, H., Zhou, T., Xin, H., 2021. Real-time multipath mitigation in multi-GNSS short baseline positioning via CNN-LSTM method. *Mathematical Problems in Engineering*, 2021, Article 6573230.

Zhang, Q., Zhao, L., Zhao, L., Zhou, J., 2023. An improved robust adaptive Kalman filter for GNSS precise point positioning. *IEEE Sensors Journal*, 23(4), 3737–3747.