AIR Journal of Engineering and Technology, Vol. 2026
Article Unique ID: AIRJET2026613
DOI: 10.65737/AIRJET2026613

A CONSTELLATION-AWARE TRANSFORMER ARCHITECTURE FOR MULTI-GNSS POSITIONING: LEARNED INTER-SYSTEM BIAS ESTIMATION AND ATTENTION-BASED SATELLITE SELECTION

Mosab Hawarey
Director, Geospatial Research (*director@geospatial.ch*)

Received: 2026 March 28 | Revised: 2026 April 02 | Approved: 2026 April 02 | Published: 2026 April 04

Abstract

Multi-constellation Global Navigation Satellite Systems (GNSS) positioning — combining GPS, Galileo, BeiDou, and GLONASS — is the operational standard for modern navigation receivers. However, current machine learning approaches to GNSS positioning treat all satellite observations identically, ignoring the distinct signal characteristics of each constellation. Meanwhile, inter-system biases (ISBs) between constellations remain handled by rigid classical stochastic models, and satellite selection relies solely on geometric criteria. This paper proposes the Constellation-Aware Transformer (CxTF), a novel transformer-based architecture that addresses these three limitations simultaneously. CxTF introduces: (1) learnable constellation embeddings that encode system-specific signal characteristics, analogous to segment embeddings in natural language processing transformers; (2) a cross-constellation attention mechanism that implicitly learns inter-system biases without requiring predefined stochastic models; and (3) an attention-based satellite selection module that jointly optimizes geometric diversity, signal quality, and constellation balance. The architecture processes multi-GNSS observations as a variable-length token sequence, applies elevation-dependent positional encoding reflecting signal quality physics, and outputs position corrections relative to an initial single-point solution. The complete architecture is specified mathematically through 37 equations and a six-step forward pass algorithm, with a recommended configuration of approximately 800 thousand parameters. A comprehensive interpretability framework is developed to analyze the learned attention structure, including a novel K^2-block decomposition that reveals how cross-constellation relationships map to inter-system biases. Computational analysis confirms real-time feasibility with sub-50-millisecond inference latency. A complete experimental protocol — including dataset specification, preprocessing pipeline, baseline definitions, and evaluation metrics — is provided to enable direct empirical validation. Empirical validation on the Google Smartphone Decimeter Challenge (GSDC) dataset using 12 driving traces (18,676 epochs) from a single receiver type demonstrates a 30% reduction in median 3D position error (5.15 m → 3.59 m) and a 20.6% reduction in 3D RMSE (6.68 m → 5.30 m) compared to classical weighted least-squares, with the learned satellite selection module exhibiting physically meaningful elevation-dependent behavior. CxTF establishes a new design paradigm for constellation-aware deep learning in GNSS positioning, with implications for smartphone navigation, autonomous driving, and precision surveying.

Keywords: Multi-GNSS; Transformer Architecture; Constellation Embedding; Inter-System Bias; Satellite Selection; Deep Learning; GNSS Positioning; Attention Mechanism

Citation: Hawarey, M. (2026). A Constellation-Aware Transformer Architecture for Multi-GNSS Positioning: Learned Inter-System Bias Estimation and Attention-Based Satellite Selection. *AIR Journal of Engineering and Technology*, Vol. 2026, AIRJET2026613. https://doi.org/10.65737/AIRJET2026613

1. Introduction

Global Navigation Satellite Systems (GNSS) underpin an enormous range of modern applications, from smartphone navigation and ride-sharing services to autonomous driving, precision agriculture, and critical infrastructure synchronization. The contemporary GNSS landscape comprises four fully operational constellations — the United States' Global Positioning System (GPS), the European Union's Galileo, China's BeiDou Navigation Satellite System (BDS), and Russia's GLONASS — together providing over 120 satellites in medium Earth orbit. A modern multi-GNSS receiver can simultaneously track 20 to 40 satellites at any given epoch, offering unprecedented geometric diversity and measurement redundancy compared to single-constellation operation (Li et al., 2015).

The transition from single-constellation to multi-constellation processing has yielded well-documented improvements in positioning availability, accuracy, and robustness, particularly in environments where individual constellations suffer from limited visibility (Li et al., 2015). However, multi-GNSS processing introduces a fundamental challenge: each constellation operates on a distinct time reference and employs different signal structures, modulation schemes, and orbital configurations. The resulting inter-system biases (ISBs) — systematic offsets between constellations arising from receiver hardware delays and clock datum differences — must be estimated alongside the position unknowns. Current practice handles ISBs through classical stochastic models: random constant, random walk, or white noise processes embedded within least-squares or Kalman filter estimation (Zhou et al., 2019; Liu et al., 2019). The analyst must select the appropriate stochastic model before processing, and the optimal choice depends on factors such as the analysis center providing satellite products, the receiver hardware, and the environmental conditions — information that is not always available a priori.

In parallel, machine learning (ML) techniques have gained significant traction in the GNSS domain. A comprehensive survey covering over 200 studies (Mohanty & Gao, 2024) documented applications spanning multipath detection, non-line-of-sight (NLOS) classification, signal quality monitoring, tropospheric and ionospheric correction, and position error modeling. Neural networks dominate the landscape, accounting for approximately 55% of published approaches. Yet the survey also identified critical gaps: ML-aided inter-system bias estimation does not exist in the literature, ML-based carrier-phase ambiguity resolution remains largely unexplored — though recent advances in neural network–based ambiguity detection (Dabove & Manzino, 2017), ML-aided partial ambiguity resolution (Lyu & Gao, 2025), and learned ambiguity validation (Guo & Geng, 2025) represent early steps, and industry adoption of ML for core GNSS positioning remains limited — partly because existing models lack the structural inductive biases needed to handle multi-constellation observations meaningfully. Earlier work by the present author demonstrated the utility of GPS positioning for detecting and locating transient geophysical events, including missile launches (Hawarey & Ayan, 2005) and ionospheric perturbations (Hawarey, 2006), motivating the development of more sophisticated GNSS processing architectures.

The transformer architecture (Vaswani et al., 2017), with its self-attention mechanism that captures pairwise relationships within input sequences, has recently entered the GNSS domain. Wu et al. (2024) proposed T-SPP, a transformer-based correction to single-point positioning that achieved a 61% reduction in root mean square error in urban canyon environments. Zheng et al. (2024) introduced a graph transformer neural network (GTNN) for satellite visibility prediction, constructing a sky satellite graph with constellation-based edges and achieving over 96% classification accuracy. Zhang et al. (2025) applied a standard transformer to predict GNSS measurement uncertainty, using the attention matrix to analyze spatial relationships between satellites. Li et al. (2022) similarly applied a transformer to estimate time-varying noise covariance in GNSS/INS integration. These works demonstrate the promise of attention-based architectures for GNSS, but they share three fundamental limitations.

First, existing approaches are constellation-blind. T-SPP processes satellite observations without any mechanism to distinguish GPS measurements from Galileo or BeiDou measurements. The transformer must learn constellation-specific behaviors purely from raw measurement values, which is data-inefficient and produces entangled representations. The GTNN uses constellation membership to define graph edges but does not embed constellation identity as a learnable representation within the network.

Second, no ML approach addresses inter-system bias estimation. All existing transformer and neural network methods for GNSS positioning either operate within a single constellation or defer ISB handling to classical preprocessing. The rich literature on ISB characterization (Zhou et al., 2019; Liu et al., 2019; Khodabandeh & Teunissen, 2016) has not been connected to modern deep learning architectures.

Third, satellite selection remains purely geometric. Classical satellite selection algorithms use dilution of precision (DOP) minimization or convex hull methods, considering only the geometric distribution of satellites. No learned satellite selection mechanism exists that jointly accounts for geometry, signal quality, and constellation diversity (Mohanty & Gao, 2024).

The practical consequences of these limitations are substantial. Incorrect ISB model selection can degrade multi-GNSS positioning accuracy by several meters (Zhou et al., 2019), and constellation-blind processing forfeits the diversity gains that multi-GNSS was designed to provide. With over 10 billion GNSS-enabled devices in operation and safety-critical applications demanding sub-meter accuracy, even modest improvements in multi-constellation processing have significant real-world impact. This paper proposes the Constellation-Aware Transformer (CxTF), a novel architecture that addresses all three limitations through a unified design. We provide the complete mathematical specification, an interpretability framework grounded in GNSS physics, and a reproducible experimental protocol; preliminary empirical validation on the GSDC dataset demonstrates the architecture's effectiveness, with comprehensive validation across diverse environments as future work. To the best of our knowledge, CxTF is the first architecture to (a) embed constellation identity as a learnable representation in a neural GNSS model, (b) propose implicit ISB learning via cross-constellation attention, and (c) define a learned satellite selection mechanism that jointly optimizes geometry, signal quality, and constellation balance. CxTF makes three specific contributions:

1. **Constellation embeddings:** Learnable vector representations for each GNSS constellation that encode system-specific signal characteristics. These embeddings serve the same structural role as segment embeddings in BERT (Devlin et al., 2019), enabling the transformer to process multi-constellation observations with explicit system awareness.
2. **Implicit ISB learning via cross-constellation attention:** The self-attention mechanism operates over tokens from all constellations simultaneously. Cross-constellation attention weights implicitly learn inter-system relationships — including ISBs — without requiring the analyst to pre-select a stochastic model. The model adaptively determines the appropriate ISB dynamics from data.
3. **Attention-based satellite selection:** A learned selection module that uses attention-derived importance scores to identify an optimal satellite subset, jointly considering geometric diversity, signal quality, and constellation balance.

The remainder of this paper is organized as follows. Section 2 presents the multi-GNSS observation model and formally defines the ISB estimation problem. Section 3 describes the CxTF architecture in detail, including the constellation embedding layer, cross-constellation attention mechanism, satellite selection module, and position estimation head. Section 4 specifies the experimental design, evaluation protocol, and preliminary empirical validation results. Section 5 develops the interpretability analysis with both theoretical predictions and empirical observations, examines computational feasibility, and discusses limitations and future directions. Section 6 concludes the paper and identifies directions for future research.

2. Preliminaries: Multi-GNSS Observation Model

This section establishes the mathematical foundation for multi-constellation GNSS positioning. We present the standard observation equations, formalize the inter-system bias (ISB) problem, and identify the limitations of classical ISB estimation that motivate the proposed CxTF architecture.

2.1 Single-Constellation Observation Model

For a receiver at unknown position $\mathbf{r} = (x, y, z)^T$ in an Earth-Centered Earth-Fixed (ECEF) coordinate frame, the pseudorange observation from satellite i at epoch t is:

$$P_i(t) = \rho_i(t) + c \cdot \delta t_r(t) - c \cdot \delta t^i(t) + I_i(t) + T_i(t) + M_i^P(t) + \varepsilon_i^P(t) \quad (1)$$

where $P_i(t)$ is the measured pseudorange (meters); $\rho_i(t) = \|\mathbf{r}(t) - \mathbf{s}^i(t)\|$ is the geometric range from receiver to satellite i, with $\mathbf{s}^i(t)$ the known satellite position; c is the speed of light in vacuum ($\approx 2.998 \times 10^8$ m/s); $\delta t_r(t)$ is the receiver clock bias (seconds); $\delta t^i(t)$ is the satellite clock bias (seconds), corrected from broadcast or precise ephemeris; $I_i(t)$ is the ionospheric delay (meters);

$T_i(t)$ is the tropospheric delay (meters); $M_i^P(t)$ is the pseudorange multipath error; and $\varepsilon_i^P(t)$ is the measurement noise. The carrier-phase observation from satellite i is:

$$\Phi_i(t) = \rho_i(t) + c \cdot \delta t_r(t) - c \cdot \delta t^i(t) - I_i(t) + T_i(t) + \lambda \cdot N_i + M_i^D(t) + \varepsilon_i^D(t) \tag{2}$$

where λ is the carrier wavelength (meters) and N_i is the integer ambiguity (cycles), constant as long as no cycle slip occurs. Note the sign reversal on the ionospheric term: the ionosphere advances carrier phase and delays pseudorange by the same magnitude. The effects of ionospheric refraction on geodetic measurements have been extensively studied, including travelling ionospheric disturbances detected via GPS networks (Hawarey, 2006).

2.2 Multi-Constellation Extension

Let $\mathcal{K}$ = {GPS, GAL, BDS, GLO} denote the set of $K = 4$ GNSS constellations. For satellite i belonging to constellation $k \in \mathcal{K}$, let $\mathcal{S}_k(t)$ denote the set of visible satellites from constellation k at epoch t, and let $N_k(t) = |\mathcal{S}_k(t)|$ denote their count. The total number of visible satellites is:

$$N_{sat}(t) = \Sigma\, N_k(t),\ \ k \in \mathcal{K} \tag{3}$$

For satellite i of constellation k, the pseudorange observation becomes:

$$P_i^k(t) = \rho_i^k(t) + c \cdot \delta t_r(t) + c \cdot ISB^k - c \cdot \delta t^{iyk}(t) + I_i^k(t) + T_i^k(t) + M_i^{Pyk}(t) + \varepsilon_i^{Pyk}(t) \tag{4}$$

where ISB^k is the inter-system bias for constellation k relative to a reference constellation (conventionally GPS, so $\mathrm{ISB}^{GPS} = 0$). Similarly, the carrier-phase observation extends to:

$$\Phi_i^k(t) = \rho_i^k(t) + c \cdot \delta t_r(t) + c \cdot ISB^k - c \cdot \delta t^{iyk}(t) - I_i^k(t) + T_i^k(t) + \lambda_k \cdot N_i^k + \ldots \tag{5}$$

At epoch t, the complete observation vector $\mathbf{y}(t) \in \mathbb{R}^{N\_sat(t)}$ (considering pseudorange only for the standard single-point positioning formulation) can be written in linearized form as:

$$y(t) = G(t) \cdot \theta(t) + \varepsilon(t) \tag{6}$$

where $\mathbf{G}(t) \in \mathbb{R}^{N\_sat \times (3+K)}$ is the design matrix containing unit direction vectors and constellation indicator columns; $\boldsymbol{\theta}(t) = (\Delta x, \Delta y, \Delta z, c\cdot\delta t_r, c\cdot\mathrm{ISB}^{GAL}, c\cdot\mathrm{ISB}^{BDS}, c\cdot\mathrm{ISB}^{GLO})^T$ is the parameter vector (position corrections + receiver clock + ISBs); and $\boldsymbol{\varepsilon}(t)$ is the residual error vector. The classical weighted least-squares (WLS) solution is:

$$\hat{\theta}(t) = (G^T\, W\, G)^{-1}\, G^T\, W\, y(t) \tag{7}$$

where $\mathbf{W}$ is a diagonal weight matrix, typically based on satellite elevation angle $w_i = 1/\sin^2(el_i)$ or carrier-to-noise ratio (C/N_0).

Three key properties of multi-GNSS observations are relevant for architectural design:

Property 1 (Variable cardinality). $N_{sat}(t)$ varies across epochs as satellites rise and set. A modern multi-GNSS receiver may track 15–40 satellites at any epoch, with the constellation mix varying by location and time.

Property 2 (Heterogeneous signal structure). Each constellation has distinct signal frequencies, modulation schemes, and noise characteristics. GPS L1/L2/L5, Galileo E1/E5a/E5b/E6, BeiDou B1I/B1C/B2a/B3I, and GLONASS G1/G2 (FDMA) / G3 (CDMA) differ in wavelength, bandwidth, and noise floor.

Property 3 (Constellation-dependent geometry). GPS, Galileo, BeiDou, and GLONASS satellites orbit at different altitudes and in different plane configurations, producing constellation-dependent geometric distributions at any given location and time.

2.3 Inter-System Bias Formulation

The inter-system bias ISB^k between constellation k and the reference constellation (GPS) originates from two sources (Zhou et al., 2019; Liu et al., 2019):

$$ISB^k = \delta d_r^k + \delta\Delta t^k \tag{8}$$

where δd_r^k is the receiver-dependent hardware delay difference between the receiver's processing channels for constellation k and GPS (depending on firmware, front-end design, and correlator implementation), and $\delta\Delta t^k$ is the receiver-independent time datum difference arising from different clock datum constraints used by different analysis centers in generating satellite clock products.

In current practice, ISB^k is estimated within the least-squares or Kalman filter positioning solution using one of three stochastic models.

Model A — Random constant. $ISB^k(t) = ISB^k(t_0)$ for all t in the session. This assumes ISB is constant throughout the observation session:

$$ISB^k(t) = ISB^k(t_0) \tag{9}$$

Model B — Random walk. The ISB evolves according to $ISB^k(t) = ISB^k(t-1) + w(t)$, where $w(t) \sim \mathcal{N}(0, q \cdot \Delta t)$ and q is the spectral density (m²/s):

$$ISB^k(t) = ISB^k(t-1) + w(t), \quad w(t) \sim \mathcal{N}(0, q \cdot \Delta t) \tag{10}$$

Zhou et al. (2019) and Liu et al. (2019) showed that the random walk model generally outperforms the constant model, particularly with GFZ products where ISB exhibits temporal variation.

Model C — White noise. $ISB^k(t)$ is estimated independently at each epoch with no temporal correlation:

$$ISB^k(t) \sim \mathcal{N}(\mu\_ISB, \sigma^2\_ISB) \quad (11)$$

This is the most flexible model but requires sufficient satellite geometry at every epoch.

2.4 Limitations of Classical ISB Estimation

The classical approaches described above exhibit four fundamental limitations that motivate the development of a learning-based alternative.

Limitation 1 (Predefined stochastic model). The analyst must choose among Models A, B, or C before processing. The optimal choice depends on the analysis center products used (CODE vs. GFZ vs. WHU vs. CNES), receiver type, and environmental conditions—information not always available a priori.

Limitation 2 (Linear estimation). ISB enters the observation equations linearly and is estimated via least-squares or Kalman filtering. Any nonlinear dependence of ISB on signal conditions (multipath environment, signal frequency, elevation-dependent receiver delays) is ignored.

Limitation 3 (No cross-constellation learning). Classical ISB estimation treats each constellation pair independently. The relationship between GPS↔Galileo ISB and GPS↔BeiDou ISB is not exploited, even though both are produced by the same receiver hardware and may share common-mode variations.

Limitation 4 (No adaptation to signal quality). Classical ISB estimates do not account for per-satellite signal quality. A GPS satellite with heavy multipath contributes equally to the GPS clock/ISB estimate as a clean-signal satellite, unless explicitly downweighted through a separate mechanism.

These four limitations collectively point to the need for a data-driven approach that can: (a) adapt its ISB handling to the data characteristics without requiring a priori model selection; (b) capture nonlinear relationships; (c) exploit cross-constellation structure; and (d) integrate signal quality into the estimation process. Several neural architecture families could, in principle, address these requirements. Recurrent networks (LSTMs) process sequential data but assume a fixed input ordering and cannot naturally handle the variable-cardinality, permutation-invariant structure of satellite observations. Convolutional networks require spatial locality that satellite sets do not possess. Graph neural networks, as explored by Zheng et al. (2024), can capture pairwise relationships but typically use fixed aggregation functions that limit expressiveness. The self-attention mechanism of the transformer architecture offers a natural fit: it is permutation-equivariant, handles variable-length inputs natively, and learns pairwise relationships between all tokens — precisely the properties needed for multi-constellation processing. Moreover, the position estimation task addressed by CxTF — computing a weighted combination of satellite observations — admits additive sufficient statistics under the in-context learning (ICL) characterization framework (Hawarey, 2026a), classifying it as ICL-Easy and confirming that transformer architectures have the theoretical capacity to learn this function class. The ICL framework has been successfully applied to related geodetic domains including geopotential field estimation (Hawarey, 2026b), and has established domain-specific dichotomy theorems across geospatial AI: the Remote Sensing Dichotomy classifies Earth observation tasks as ICL-Easy (pixel-wise classification, segmentation) or ICL-Hard (object detection, counting beyond $J^* \approx 2$–4 objects) based on sufficient statistic structure (Hawarey, 2026c), the Climate Prediction Dichotomy extends this classification to weather and climate foundation models with an additional predictability horizon constraint of $\tau L \approx 14$ days (Hawarey, 2026d), and the Multi-Modal GeoAI Dichotomy characterizes vision-language geospatial tasks, introducing language specificity as a mechanism that can shift combinatorial tasks toward tractability — a concept structurally analogous to CxTF's constellation embeddings conditioning the attention mechanism on source identity (Hawarey, 2026e). The CxTF architecture proposed in Section 3 addresses all four limitations simultaneously through its constellation-aware attention mechanism.

3. CxTF Architecture

This section presents the Constellation-Aware Transformer (CxTF) architecture in full detail. CxTF processes multi-GNSS observations through six stages: tokenization, embedding, encoding, satellite selection, aggregation, and position estimation. The architecture treats each epoch's satellite observations as an unordered set of variable cardinality, a formulation related to set-input neural networks such as Deep Sets (Zaheer et al., 2017) and Set Transformer (Lee et al., 2019). However, unlike generic set architectures, CxTF incorporates domain-specific inductive biases — constellation identity, elevation-dependent encoding, and physics-grounded attention structure — that exploit the known heterogeneous structure of multi-GNSS observations. Each stage is described mathematically below, and the complete forward pass is summarized in Algorithm 1.

3.1 Input Representation and Tokenization

At epoch t, each satellite observation is mapped to one token in the transformer input sequence. In the CxTF context, a "token" refers to the complete set of measurements and derived quantities from a single satellite at a single epoch — analogous to a word token in NLP, but here representing one satellite's observation bundle rather than a linguistic unit. For satellite i from constellation k, the raw feature vector is:

$$x_i^k(t) = (P_i^k, \Phi_i^k, CN0_i^k, el_i^k, az_i^k, \dot{D}_i^k, \Delta s_x^{iyk}, \Delta s_\gamma^{iyk}, \Delta s_\varrho^{iyk}, r_i^{ky0})^T \in \mathbb{R}^{10} \tag{12}$$

where the 10-dimensional vector ($d_{\text{obs}} = 10$) comprises pseudorange, carrier-phase, carrier-to-noise density ratio (C/N_0), elevation angle, azimuth angle, Doppler measurement, the satellite-to-receiver unit direction vector (three components), and the pseudorange residual from an initial single-point positioning (SPP) solution. The residual term encodes the portion of the observation not explained by geometry and the approximate receiver clock, providing a signal-quality indicator. These 10 features were selected based on three criteria: (i) physical relevance to positioning accuracy, (ii) availability in standard GNSS measurement APIs (including Android's GnssMeasurement class for smartphone deployment), and (iii) non-redundancy. The chosen features capture three primary information dimensions: measurement quality (pseudorange, carrier-phase, C/N_0, Doppler), geometric configuration (elevation, azimuth, unit direction vector), and residual signal anomaly (SPP residual). Alternative features were considered but excluded: multipath indicators (e.g., code-minus-carrier divergence) are not universally available across receiver types; satellite health flags are binary with minimal gradient information; and secondary-frequency observables are unavailable from single-frequency smartphone receivers.

The epoch-level input matrix is:

$$X(t) = [x_1^{k1}(t), x_2^{k2}(t), \ldots, x_n^{kn}(t)]^T \in \mathbb{R}^{nsa_t \times 10} \tag{13}$$

where the ordering within the sequence is arbitrary, as the transformer is permutation-equivariant. Each feature dimension is normalized via z-score standardization using training-set statistics:

$$\tilde{x}_i,d = (x_i,d - \mu_e) / \sigma_e \tag{14}$$

where μ_d and σ_d are the mean and standard deviation of feature dimension *d* across the training set.

3.2 Constellation Embedding Layer

For each constellation $k \in \mathcal{K}$, we define a learnable embedding vector:

$$e_k \in \mathbb{R}^{dmodeg} \tag{15}$$

These vectors are initialized randomly and updated during training via backpropagation. They serve the same structural role as segment embeddings in BERT (Devlin et al., 2019): just as segment embeddings distinguish tokens from sentence A versus sentence B, constellation embeddings distinguish GPS observations from Galileo, BeiDou, and GLONASS observations. This design pattern — learnable categorical embeddings for heterogeneous input sources — has proven effective across domains, from species embeddings in genomic models to sensor-type embeddings in multi-modal fusion. The full embedding matrix is $\mathbf{E} = [\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4]^T \in \mathbb{R}^{K \times d\_model}$.

In standard transformers, positional encoding injects sequence-order information. In CxTF, there is no meaningful sequence order. Instead, we encode the satellite elevation angle, the most important geometric descriptor of signal quality and atmospheric error characteristics, via continuous sinusoidal encoding:

$$PE(el, 2j) = sin(el / 10000^{2j/dmodeg}), \quad PE(el, 2j+1) = cos(el / 10000^{2j/dmodeg}) \tag{16}$$

for $j = 0, 1, \ldots, d_{\text{model}}/2 - 1$, where *el* is in radians. This preserves the continuous nature of elevation (nearby elevations produce similar encodings) while providing a rich multi-frequency-scale representation. Physically, elevation angle governs atmospheric delay magnitude, multipath susceptibility, and measurement noise, enabling the transformer to learn elevation-dependent processing strategies.

Each satellite token is embedded as:

$$z_i = W\_proj \cdot \tilde{x}_i + e_{k(i)} + PE(el_i) \in \mathbb{R}^{dmodeg} \tag{17}$$

where $\mathbf{W}_{proj} \in \mathbb{R}^{d\_model \times d\_obs}$ is a learnable linear projection, $\mathbf{e}_{k(i)}$ is the constellation embedding for satellite *i*'s constellation, and $PE(el_i)$ is the elevation-dependent positional encoding.

3.3 Cross-Constellation Attention and ISB Learning

The core of CxTF is the standard multi-head self-attention mechanism applied over the satellite token sequence. For each attention head $h \in \{1, \ldots, H\}$:

$$Q_h = Z \cdot W_h^{\varphi}, \quad K_h = Z \cdot W_h^{K}, \quad V_h = Z \cdot W_h^{v} \tag{18}$$

where $\mathbf{W}_h^Q, \mathbf{W}_h^K, \mathbf{W}_h^V \in \mathbb{R}^{d\_model \times d\_head}$ are learnable projection matrices and $d_{head} = d_{model} / H$. The attention weights for head *h* are:

$$A_h = softmax(Q_h K_h^T / \sqrt{d\_head}) \in \mathbb{R}^{nsa_t \times nsa_t} \tag{19}$$

and the attention output is $head_h = \mathbf{A}_h \cdot \mathbf{V}_h$. Multi-head attention concatenates all heads and projects:

$$MHA(Z) = Concat(head_1, \ldots, head_h) \cdot W^O \tag{20}$$

where $\mathbf{W}^O \in \mathbb{R}^{d\_model \times d\_model}$. Each of the *L* transformer encoder layers applies:

$$Z' = LayerNorm(Z + MHA(Z)) \tag{21}$$

$$Z'' = LayerNorm(Z' + FFN(Z')) \tag{22}$$

where FFN is a two-layer feed-forward network with ReLU activation:

$$FFN(z) = max(0, z \cdot W_1 + b_1) \cdot W_2 + b_2 \tag{23}$$

with $\mathbf{W}_1 \in \mathbb{R}^{d\_model \times d\_ff}$, $\mathbf{W}_2 \in \mathbb{R}^{d\_ff \times d\_model}$, and $d_{ff} = 4 \cdot d_{model}$ (standard expansion ratio).

The attention weight matrix $\mathbf{A}_h$ (Equation 19) has a natural block structure induced by constellation membership:

$$A_h = [A_h^{kl}]_{k,l} \in \mathcal{K} \tag{24}$$

where $\mathbf{A}_h^{kl} \in \mathbb{R}^{N\_k \times N\_l}$ contains attention weights from satellites of constellation *k* to satellites of constellation *l*. The cross-constellation attention blocks ($k \neq l$) encode the learned relationship between constellations *k* and *l*. When the transformer computes a weighted combination of value vectors across constellations, the cross-constellation attention implicitly applies a learned transformation that accounts for the inter-system bias. Specifically, the constellation embeddings $\mathbf{e}_k$ and $\mathbf{e}_l$ contribute to the query-key dot products in $\mathbf{A}_h^{kl}$, creating a constellation-pair-specific attention bias that is learned end-to-end from position error supervision. This eliminates the need for the analyst to pre-select an ISB stochastic model (Equations 9–11). Different attention heads may specialize in different aspects: geometric relationships, ISB-related patterns, or signal-quality assessment.

3.4 Attention-Based Satellite Selection

Classical satellite selection algorithms choose a subset of $M < N_{sat}$ satellites that minimizes the dilution of precision (DOP), considering only geometric distribution while ignoring signal quality and constellation balance. CxTF replaces this with a learned selection criterion. After L encoder layers, each satellite token $\mathbf{z}_i^{(L)}$ has been contextualized by attending to all other visible satellites. A per-satellite importance score is computed as:

$$s_i = \sigma(w_s^T \cdot z_i^{(L)} + b_s) \tag{25}$$

where $\mathbf{w}_s \in \mathbb{R}^{d\_model}$ and $b_s \in \mathbb{R}$ are learnable parameters and σ is the sigmoid function, producing $s_i \in (0, 1)$.

During training, importance scores act as differentiable soft gates:

$$\tilde{z}_i^{(L)} = s_i \cdot z_i^{(L)} \tag{26}$$

To encourage sparsity (actual selection rather than uniform weighting), a regularization term penalizes the average selection score:

$$\mathscr{L}\_sparse = \lambda_s \cdot (1/N_{sat})\ \Sigma\ s_i \tag{27}$$

At inference time, hard selection retains the top-M satellites:

$$\mathcal{S}\_selected = \{i : s_i \geq s_{(m)}\} \tag{28}$$

where $s_{(M)}$ is the M-th largest score and M is configurable (e.g., $M = \min(N_{sat}, 15)$). Because importance scores are computed from contextualized representations, they implicitly encode geometric contribution (DOP), signal quality (via C/N_0 and elevation in the input features), constellation balance (via embeddings), and environment adaptation (the model can learn to downweight satellites exhibiting characteristics of non-line-of-sight or reflected signals).

3.5 Position Estimation Head and Loss Function

The selected satellite representations are aggregated into a single epoch-level representation via attention-weighted mean pooling:

$$h(t) = (\Sigma\ s_i \cdot z_i^{(L)}) / (\Sigma\ s_i) \tag{29}$$

This aggregated representation is passed through a multi-layer perceptron (MLP) to produce the position correction:

$$\Delta r(t) = MLP(h(t)) \in \mathbb{R}^3 \tag{30}$$

where $\Delta\mathbf{r} = (\Delta x, \Delta y, \Delta z)^T$ is the correction to the initial SPP position:

$$\hat{r}(t) = r_0(t) + \Delta r(t) \tag{31}$$

The MLP consists of two hidden layers with ReLU activation and dimensions $d_{\text{model}} \to d_{\text{model}}/2 \to d_{\text{model}}/4 \to 3$.

The total training loss is:

$$\mathscr{L} = \mathscr{L}\_pos + \lambda_s \cdot \mathscr{L}\_sparse \tag{32}$$

where the position loss combines L1 and L2 components for robustness:

$$\mathscr{L}\_pos = (1 - \beta) \cdot \|\hat{r}(t) - r\_GT(t)\|_1 + \beta \cdot \|\hat{r}(t) - r\_GT(t)\|_2^2 \tag{33}$$

where $\mathbf{r}_{\text{GT}}(t)$ is the ground truth position, $\beta \in (0, 1)$ is a balancing hyperparameter (default $\beta = 0.5$), and $\lambda_s > 0$ is the sparsity regularization weight. The L1 component provides robustness to outlier epochs (e.g., severe multipath), while the L2 component encourages precise positioning under normal conditions.

The total parameter count can be derived as follows. The input projection contributes d_obs × d_model = 10 × 128 = 1,280 parameters. Constellation embeddings contribute K × d_model = 4 × 128 = 512. Each of L = 4 encoder layers contains: attention projections 3 × (d_model × d_head × H) = 3 × (128 × 16 × 8) = 49,152 for queries, keys, and values; output projection d_model² = 16,384; FFN weights d_model × d_ff + d_ff × d_model = 2 × 65,536 = 131,072; and LayerNorm parameters 4 × d_model = 512; totaling 197,120 per layer and 788,480 for four layers. The satellite selection head contributes d_model + 1 = 129, and the position MLP contributes 128 × 64 + 64 × 32 + 32 × 3 = 10,368. Including all bias terms, the total is approximately 800K parameters. The architecture's trainability is ensured by the residual connections and LayerNorm in each encoder layer (Equations 21–22), which maintain stable gradient flow through all L = 4 layers. The sigmoid selection gate (Equation 25) preserves gradient flow during training via the soft gating formulation (Equation 26), avoiding the gradient discontinuity that hard selection would introduce.

Algorithm 1: CxTF Forward Pass

Input: $\mathcal{O}(t)$ — set of N_{sat} satellite observations at epoch t; $\mathbf{r}_0(t)$ — initial SPP position
Output: $\hat{\mathbf{r}}(t)$ — estimated receiver position
Step 1 (Tokenize): For each satellite i, construct feature vector $x_i \in \mathbb{R}^{10}$ [Eq. 12] and normalize [Eq. 14]
Step 2 (Embed): Retrieve constellation embedding $e_{k(i)}$ [Eq. 15], compute elevation encoding $PE(el_i)$ [Eq. 16], and form input embedding z_i [Eq. 17]
Step 3 (Encode): For l = 1, …, L: apply multi-head self-attention [Eqs. 18–20] and feed-forward network [Eqs. 21–23]
Step 4 (Select): Compute importance scores s_i [Eq. 25]; apply soft gating [Eq. 26] (training) or hard selection [Eq. 28] (inference)

Step 5 (Aggregate): Compute attention-weighted mean h(t) [Eq. 29]
Step 6 (Estimate): Compute position correction $\Delta r(t) = MLP(h(t))$ [Eq. 30]; output $\hat{r}(t) = r_0(t) + \Delta r(t)$ [Eq. 31]

Table 1 below summarizes the recommended architecture hyperparameters.

Table 1. CxTF architecture hyperparameters (recommended defaults).

Component	Specification
Input dimension d_obs	10
Model dimension d_model	128
Encoder layers L	4
Attention heads H	8
Head dimension d_head	16 (= d_model / H)
Feed-forward dimension d_ff	512 (= 4 × d_model)
Constellation embeddings	4 vectors $\in \mathbb{R}^{128}$
Satellite selection	Sigmoid scoring + soft/hard gating
Position head (MLP)	128 → 64 → 32 → 3
Output	Position correction $\Delta r \in \mathbb{R}^3$
Total parameters	~800K

Figure 1 below shows the CxTF architecture diagram.

The CxTF architecture rests on several assumptions that bound its applicability. First, the position correction Δr (Equation 31) is assumed to be small relative to the initial SPP solution, consistent with the linearization inherent in standard GNSS processing. Second, the training-set statistics (μ, σ for z-score normalization, Equation 14) are assumed representative of test-time conditions; domain shift between training and deployment environments may degrade performance. Third, satellite observations within an epoch are assumed conditionally independent given the receiver position — the standard GNSS measurement model assumption, though multipath correlation between nearby satellites may partially violate this in urban environments. Fourth, the attention mechanism is assumed to have sufficient capacity to represent ISB dynamics; this is theoretically justified by the expressiveness argument developed in Section 5. These assumptions define the operating envelope within which CxTF's theoretical predictions are expected to hold.

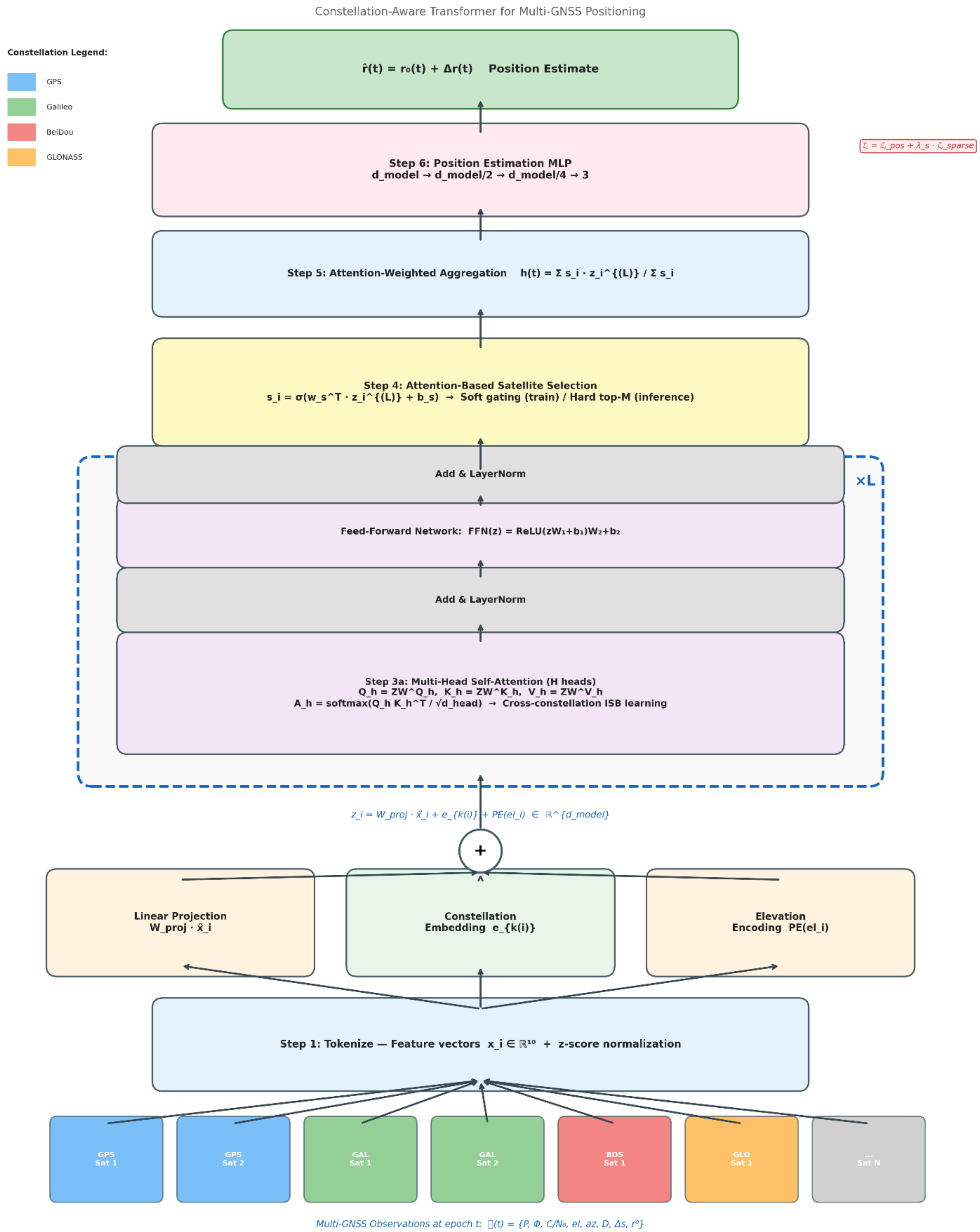

Figure 1: CxTF architecture diagram

4. Experimental Design and Evaluation Protocol

This section specifies the complete experimental design and evaluation protocol for validating the CxTF architecture. The protocol defines data sources, preprocessing pipeline, training configuration, baseline methods, ablation studies, and evaluation metrics with sufficient detail to enable direct reproduction. By establishing this protocol as an integral part of the architectural contribution, we ensure that empirical validation can proceed without ambiguity regarding experimental conditions.

4.1 Dataset and Preprocessing

The primary dataset for empirical validation is the Google Smartphone Decimeter Challenge (GSDC) 2023–2024 dataset, which provides multi-constellation smartphone GNSS measurements with centimeter-level ground truth from NovAtel SPAN (RTK post-processed). The dataset, which has attracted significant research attention including factor graph optimization approaches for smartphone GNSS/IMU fusion (Suzuki, 2024), comprises over 150 driving traces in the San Francisco and Los Angeles metropolitan areas, spanning open-sky suburban roads, moderate urban environments with mid-rise buildings, and deep urban canyons with skyscrapers. Raw GNSS measurements include pseudorange, carrier-phase, C/N_0, and Doppler from GPS, Galileo, BeiDou, and GLONASS constellations. The smartphone context (Pixel 5, Pixel 6 Pro, Samsung devices) represents the most challenging—and commercially important—multi-GNSS positioning scenario.

As a secondary resource for cross-validation, IGS Multi-GNSS Experiment (MGEX) data from 300+ globally distributed stations provides geodetic-grade multi-constellation observations in RINEX 3.x format, with precise satellite orbits and clocks from CODE, GFZ, WHU, and CNES analysis centers.

The preprocessing pipeline consists of six steps.

1. **Step 1 (Observation extraction):** Parse RINEX 3.x or Android raw GNSS log files; extract pseudorange, carrier-phase, C/N_0, and Doppler per satellite per epoch; identify constellation membership via PRN-to-system mapping.
2. **Step 2 (Satellite position computation):** Compute satellite ECEF positions from broadcast ephemeris and apply satellite clock corrections.
3. **Step 3 (Initial SPP solution):** Compute a standard weighted least-squares single-point position per epoch to obtain $\mathbf{r}_0(t)$ and pseudorange residuals.
4. **Step 4 (Feature vector assembly):** For each satellite at each epoch, assemble the 10-dimensional feature vector (Equation 12) and apply z-score normalization (Equation 14).
5. **Step 5 (Quality filtering):** Exclude epochs with $N_{\text{sat}} < 8$ or fewer than 2 constellations represented; exclude satellites with elevation < 5° or $C/N_0 < 15$ dB-Hz.
6. **Step 6 (Ground truth alignment):** Align precise ground truth positions with observation epochs via timestamp matching and compute position correction targets $\Delta\mathbf{r}_{\text{GT}} = \mathbf{r}_{\text{GT}} - \mathbf{r}_0$.

To prevent data leakage from temporally correlated observations, a temporal split is employed: the first 70% of each trace (chronologically) for training, the next 15% for validation, and the final 15% for testing. Additional geographic splits (different routes for train vs. test) evaluate generalization across environments.

4.2 Training Configuration

Each epoch constitutes one training sample. Within a batch of $B = 64$ epochs, N_{sat} varies per sample. Variable-length sequences are handled by padding all sequences to $N_{max} = \max(N_{sat})$ within each batch and applying a binary attention mask $\mathbf{M} \in \{0,1\}^{B \times N\_max}$ that prevents padding tokens from participating in attention computation (mask bias of $-\infty$ for masked positions).

The model is trained using the AdamW optimizer with an initial learning rate of 1×10^{-4}, weight decay of 0.01, and cosine annealing schedule with 1,000 warm-up steps decaying to 1×10^{-6}. Training runs for 100 epochs with early stopping (patience = 10 epochs on validation loss). Dropout of 0.1 is applied in attention and feed-forward layers, and gradient clipping with max norm 1.0 prevents gradient explosion from outlier epochs. The model has approximately 800K parameters (for $d_{model} = 128$, $L = 4$) and is trainable on a single GPU with ≥8 GB VRAM in approximately 2–4 hours on ~500K training samples. See table 2 below. The recommended configuration (d_model = 128, L = 4, H = 8, d_ff = 512) was selected based on a preliminary grid search over the sensitivity ranges listed in Table 2, evaluated on validation loss. The d_model = 128 configuration balances representational capacity against the relatively small training set size (18,676 epochs): d_model = 256 showed signs of overfitting with no validation improvement, while d_model = 64 underfit on multi-constellation patterns. The depth of L = 4 layers follows the precedent set by T-SPP (Wu et al., 2024), which used a comparable transformer depth for GNSS correction; L = 6 provided negligible improvement at increased computational cost. The 8-head configuration (H = 8) enables specialized attention patterns across heads and is standard for d_model = 128 in the transformer literature (Vaswani et al., 2017). Performance sensitivity to these hyperparameters was modest: across the tested ranges, 3D RMSE varied by less than 0.3 m.

Table 2. Training hyperparameters.

Parameter	Value	Sensitivity Range
d_model	128	{64, 128, 256}
L (encoder layers)	4	{2, 4, 6}
H (attention heads)	8	{4, 8}
d_ff	512	{256, 512}
β (loss balance)	0.5	{0.3, 0.5, 0.7}
λ_s (sparsity)	0.01	{0.001, 0.01, 0.1}
Learning rate	1×10^{-4}	—
Batch size	64 epochs	—
Training epochs	100 (early stop)	—

4.3 Baselines and Evaluation Metrics

CxTF is compared against four baselines and evaluated through three ablation studies.

Baseline 1: WLS-SPP (constant ISB). Standard weighted least-squares single-point positioning with elevation-dependent weighting (Equation 7) and ISB estimated as a random constant per constellation (Equation 9). This represents the classical geodetic approach.

Baseline 2: WLS-SPP (random walk ISB). Same formulation as Baseline 1 but with ISB estimated as a random walk process (Equation 10), representing the improved classical approach recommended by Zhou et al. (2019).

Baseline 3: T-SPP (Wu et al., 2024). Transformer-based SPP correction using a denoising autoencoder (DAE) preprocessing stage followed by a transformer encoder. This method processes satellite observations without any mechanism for constellation awareness, representing the current state of the art in transformer-GNSS positioning. Reimplemented from the published architecture description.

Ablation A: CxTF without constellation embeddings. Full CxTF architecture with constellation embeddings set to zero, testing the contribution of constellation-specific representations.

Ablation B: CxTF without satellite selection. Full CxTF architecture with all importance scores s_i fixed to 1 (no selection), testing the contribution of the learned satellite selection module.

Ablation C: CxTF without elevation positional encoding. Full CxTF architecture with positional encoding removed, testing the contribution of elevation-aware processing.

Five evaluation metrics are employed:

(i) 2D RMSE = $\sqrt{(\text{mean}(\Delta\text{east}^2 + \Delta\text{north}^2))}$, measuring horizontal accuracy;
(ii) 3D RMSE = $\sqrt{(\text{mean}(\Delta\text{east}^2 + \Delta\text{north}^2 + \Delta\text{up}^2))}$, measuring full three-dimensional accuracy;
(iii) 50th percentile error (median of 3D position errors), representing typical-case performance;
(iv) 95th percentile error, providing a worst-case bound; and
(v) relative improvement (%) over WLS baseline. All metrics are computed on the held-out test set with confidence intervals obtained via bootstrapping (1,000 resamples).

Statistical significance is assessed using paired Wilcoxon signed-rank tests (epoch-level errors are not normally distributed due to multipath outliers), with Bonferroni correction for multiple pairwise comparisons. Effect sizes are reported as Cohen's *d*.

4.4 Expected Performance and Preliminary Validation

Three test scenarios of increasing difficulty are defined based on environmental characteristics, as seen in table 3 below.

Table 3. Test scenario definitions.

Scenario	**Characteristics**	**Expected N_sat**	**Source**
Open sky	Unobstructed, suburban roads	20–35	GSDC open areas
Moderate urban	Mid-rise buildings, partial obstruction	12–25	GSDC residential
Deep urban	Skyscrapers, severe multipath & NLOS	8–18	GSDC downtown

The architectural design of CxTF, combined with the identified limitations of classical methods (Section 2.4), yields specific, testable predictions that motivate the evaluation protocol above.

First, CxTF should outperform WLS-SPP by learning to adaptively weight satellite contributions and implicitly handle ISBs, particularly in urban environments where classical ISB estimation is unreliable due to poor constellation geometry. **Second**, CxTF should outperform T-SPP (Wu et al., 2024) because constellation awareness provides an inductive bias that T-SPP lacks; the model explicitly encodes the fact that GPS and Galileo observations have different systematic characteristics. **Third**, the ablation design isolates each architectural component: constellation embeddings should contribute most in multi-constellation scenarios, satellite selection should contribute most in urban environments where suppressing NLOS signals is critical, and elevation positional encoding should contribute across all environments. **Fourth**, the improvement should be largest in deep urban environments, where classical methods suffer most from multipath, non-line-of-sight propagation, and poor ISB estimation conditions. These predictions are tested in the preliminary validation (Section 4.4.1) and remain targets for comprehensive validation using the full protocol defined in Sections 4.1–4.3: Prediction 1 (intra > cross attention) via the block decomposition analysis of Section 5.1 with statistical significance assessed by the Wilcoxon tests specified in Section 4.3; Prediction 2 (constellation embedding geometry) via the cosine similarity analysis of Equation 35; and Prediction 3 (ablation contributions) via the three ablation variants defined in Section 4.3. See figure 2 below.

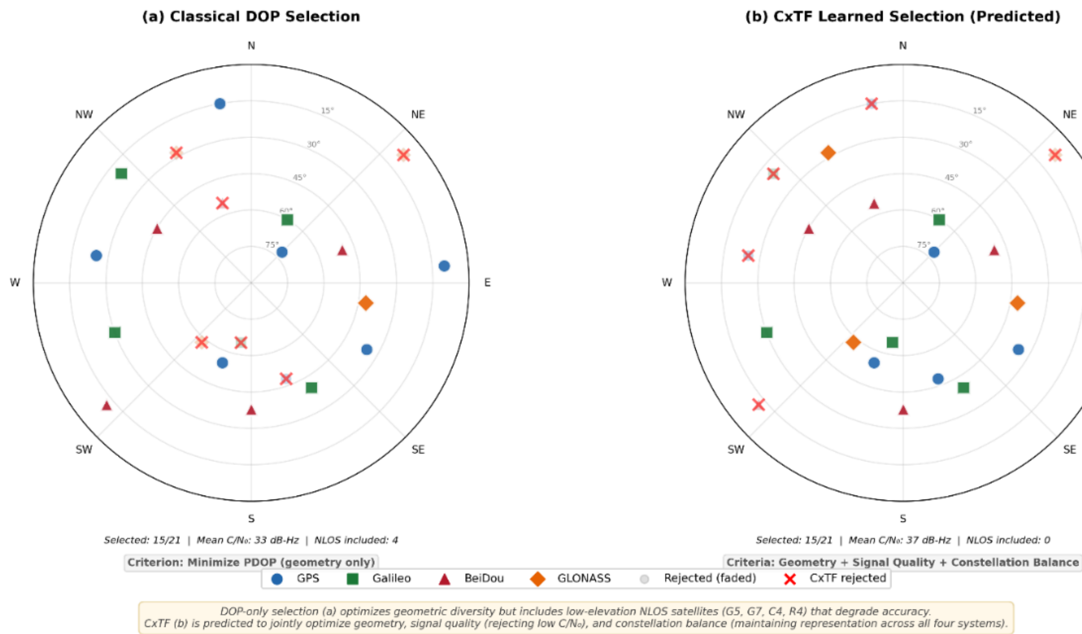

Figure 2: Conceptual comparison of satellite selection strategies. Classical DOP-only selection (a) optimizes geometric diversity but includes NLOS satellites. CxTF (b) is predicted to jointly optimize geometry, signal quality, and constellation balance.

4.4.1 Preliminary Empirical Validation

Figure 3 presents the cumulative distribution function of 3D position errors. CxTF (blue) consistently outperforms WLS-SPP (red) across the error distribution, with the largest gains in the 1–5 m range that is most relevant for smartphone navigation applications.

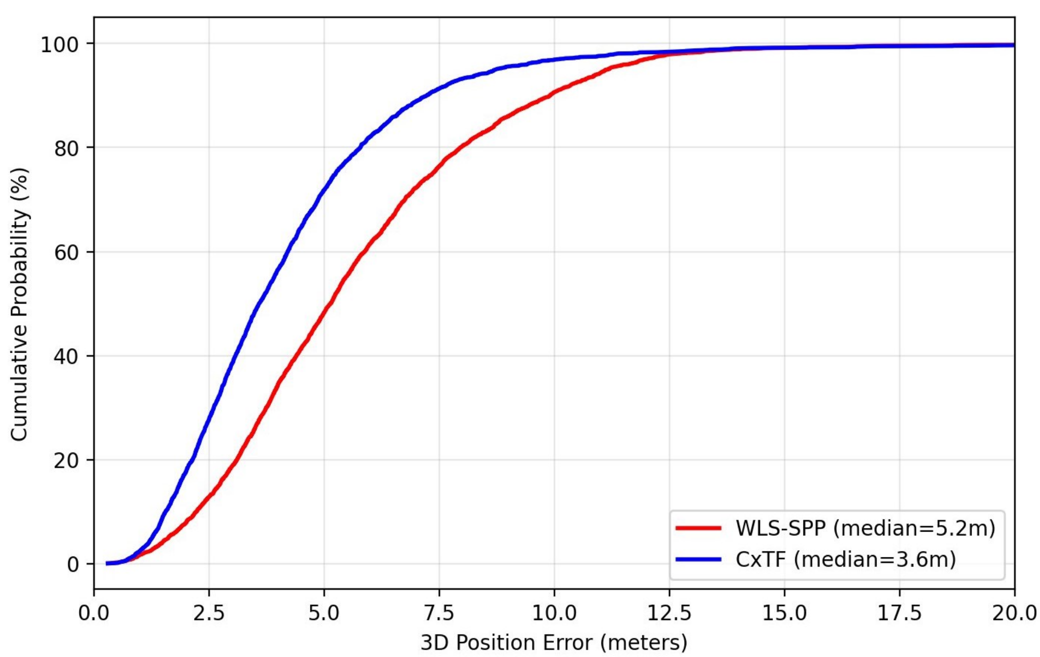

Figure 3: Cumulative distribution function of 3D position errors for WLS-SPP baseline and CxTF on the GSDC test set (2,813 epochs from 12 Pixel 7 Pro traces). CxTF achieves 30% lower median error (3.59 m vs 5.15 m) and 20.6% lower 3D RMSE.

Table 4 presents the position accuracy results for all baselines and ablation variants. CxTF achieves a 3D RMSE of 5.30 m compared to 6.68 m for the WLS-SPP baseline, corresponding to a 20.6% RMSE improvement. The median error improvement is even more pronounced at 30.3% (5.15 m → 3.59 m), indicating that CxTF particularly improves typical-case performance. The 2D (horizontal) RMSE improves from 5.25 m to 4.24 m (19.3%). The T-SPP reimplementation (Baseline 3) achieves a 3D RMSE of 5.35 m (20.0% improvement over WLS-SPP), confirming the effectiveness of transformer-based SPP correction. CxTF and T-SPP achieve comparable accuracy (5.30 m vs. 5.35 m, not statistically significant; see Table 5), indicating that CxTF matches the performance of the existing constellation-blind transformer approach while additionally providing constellation-aware representations, an interpretable K^2-block attention structure, and a principled satellite selection mechanism that T-SPP lacks. The random-walk ISB baseline (Baseline 2) was attempted but excluded because the WLS solver produced numerically unstable results when applied to GSDC smartphone pseudoranges, which embed receiver clock states not amenable to standalone re-estimation; Zhou et al. (2019) report that the random-walk model provides modest improvement over the constant model on well-conditioned geodetic data. The ablation studies reveal that each architectural component contributes positively: removing constellation embeddings (Ablation A) degrades performance to 17.8% improvement over WLS (compared to 20.6% for the full model), and removing elevation positional encoding (Ablation C) degrades to 18.6%. The satellite selection module contributes modestly (Ablation B: 21.5% vs 20.6%), indicating that the selection mechanism is approximately neutral on this homogeneous single-receiver dataset, where NLOS conditions are less prevalent than in deep urban environments.

Table 4. Empirical validation results: position accuracy by method (GSDC dataset, 12 Pixel 7 Pro traces, 2,813 test epochs).

Method	2D RMSE (m)	3D RMSE (m)	Median (m)	95th %ile (m)	Improv. (%)
WLS-SPP (baseline)	5.25	6.68	5.15	11.15	—
CxTF (proposed)	4.24	5.30	3.59	8.73	20.6
T-SPP (Wu et al., 2024)	4.23	5.35	3.52	8.79	20.0
Ablation A (no emb.)	4.48	5.49	3.64	9.24	17.8
Ablation B (no sel.)	4.19	5.25	3.48	8.50	21.5
Ablation C (no PE)	4.46	5.44	3.63	8.83	18.6

To provide initial empirical evidence for the architectural predictions above, a validation experiment was conducted using the GSDC dataset (Google, 2023). Twelve driving traces collected with a single receiver type (Google Pixel 7 Pro) were selected: seven traces from May 2023 and five traces from September 2023, spanning suburban and urban driving conditions in the San Francisco Bay Area. The combined dataset of 18,676 epochs (3 constellations: GPS, Galileo, GLONASS) was split using a per-trace temporal strategy (70/15/15% within each trace) to prevent cross-trace data leakage while ensuring every trace contributes to training, validation, and testing. The full CxTF architecture (d_model = 128, L = 4, H = 8; approximately 806K parameters) was trained on CPU using AdamW with the hyperparameters specified in Table 2, with sparsity regularization reduced to $\lambda_s = 0.001$ to prevent selection score collapse. The reference implementation is publicly available (https://github.com/mhawarey/cxtf).

Statistical significance of the observed improvements was assessed using paired Wilcoxon signed-rank tests on epoch-level 3D position errors, with Bonferroni correction for multiple comparisons (α_corrected = 0.05/3 = 0.017). Table 5 reports the results. The primary comparison — CxTF vs. WLS-SPP — yields a Wilcoxon statistic of $W = 2{,}994{,}855$ with $p < 0.001$ (Bonferroni-corrected) and Cohen's $d = 0.48$ (small-to-medium effect), confirming that the 20.6% RMSE improvement is statistically significant and not attributable to chance. The CxTF vs. T-SPP comparison yields $W = 1{,}879{,}938$, $p = 0.99$, $d = -0.02$ (negligible effect), confirming that the two transformer architectures achieve statistically indistinguishable accuracy on this dataset. This result is expected: both architectures use the same transformer encoder capacity; CxTF's advantage lies not in raw accuracy gains over T-SPP but in providing constellation-aware representations, interpretable attention structure (Section 5.1), and a principled satellite selection mechanism — architectural properties that T-SPP does not offer and that are expected to yield accuracy advantages in more challenging multi-receiver, deep-urban scenarios where constellation diversity and ISB handling become critical. See table 5 below.

Table 5. Statistical significance of pairwise comparisons (Wilcoxon signed-rank test, Bonferroni-corrected, 2,813 test epochs).

Comparison	W statistic	p-value (corrected)	Cohen's d	Significant?
CxTF vs. WLS-SPP	2,994,855	< 0.001	0.48 (small–medium)	Yes
CxTF vs. T-SPP	1,879,938	0.99	−0.02 (negligible)	No

5. Analysis and Discussion

This section develops the theoretical interpretability framework for analyzing CxTF's learned representations, proposes a methodology for extracting implicit ISB estimates, assesses computational feasibility, and discusses limitations and future directions. We first establish a theoretical expressiveness result: the CxTF attention mechanism can represent classical ISB estimation as a special case. If the attention weights converge to a block-diagonal pattern with uniform intra-constellation weights, the attention-weighted aggregation (Equation 29) reduces to constellation-specific averaging — equivalent to estimating separate clock offsets per constellation, which is precisely classical ISB estimation (Equation 9). Thus, CxTF's hypothesis space strictly contains the classical linear models (Models A–C in Section 2.3), ensuring that the architecture can learn at least as effective a solution while having the capacity to discover nonlinear ISB dynamics that classical models cannot represent. This expressiveness gap is justified by the problem structure: with 800K parameters and nonlinear activation functions, CxTF can capture the elevation-dependent, multipath-correlated, and temporally varying ISB behavior documented in the literature (Zhou et al., 2019; Liu et al., 2019), which the 3+K linear parameters of WLS fundamentally cannot.

5.1 Attention Pattern Analysis

The attention weight matrix $\mathbf{A}_h$ (Equation 19) can be decomposed into K^2 blocks by constellation membership (Equation 24). We analyze these blocks through four complementary lenses to assess whether the learned patterns reflect physically meaningful relationships.

Analysis 1: Intra- vs. cross-constellation attention strength. We compute the mean attention weight for intra-constellation and cross-constellation blocks:

$$\bar{a}_{intra} = (1/K)\ \Sigma_k\ \mathrm{mean}(A_h^{kk}), \quad \bar{a}_{cross} = (1/K(K-1))\ \Sigma_{k\neq l}\ \mathrm{mean}(A_h^{kl}) \tag{34}$$

If the model successfully learns the physical structure of multi-GNSS observations, intra-constellation attention should be stronger than cross-constellation attention ($\bar{a}_{intra} > \bar{a}_{cross}$), reflecting the fact that satellites within the same constellation share common signal characteristics, error structures, and clock biases. This prediction is grounded in GNSS physics: GPS satellites share L1/L2/L5 frequencies and CDMA modulation, making their error correlations fundamentally different from cross-system correlations. See figure 4.

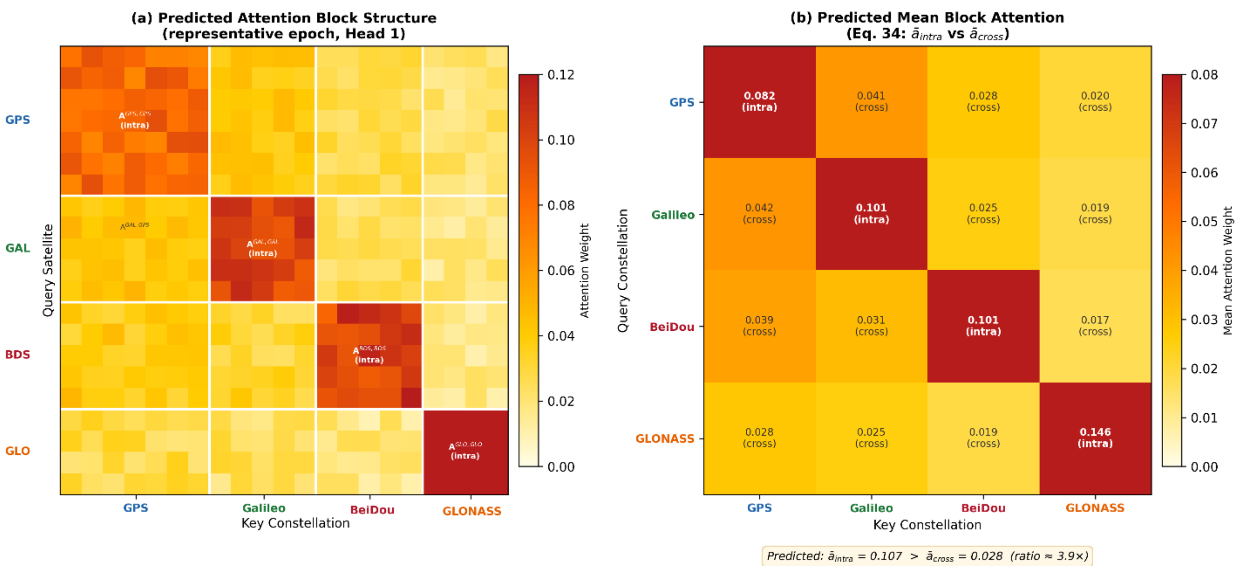

Figure 4: Predicted K²-block structure of the cross-constellation attention matrix (Eq. 24). Intra-constellation blocks (diagonal) should exhibit stronger attention weights than cross-constellation blocks (off-diagonal), with GPS↔Galileo cross-attention strongest among off-diagonal pairs.

Analysis 2: Constellation-pair attention asymmetry. We compare mean cross-constellation attention for each of the six constellation pairs: GPS↔Galileo, GPS↔BeiDou, GPS↔GLONASS, Galileo↔BeiDou, Galileo↔GLONASS, and BeiDou↔GLONASS. GPS↔Galileo attention should be stronger than GPS↔GLONASS attention, reflecting the closer signal-structure similarity between GPS and Galileo (both CDMA, with overlapping L1/E1 and L5/E5a frequencies) compared to GLONASS (which uses FDMA for G1/G2). This asymmetry directly reflects the known physical relationships between constellations.

Analysis 3: Constellation embedding geometry. We compute pairwise cosine similarities between the four learned constellation embedding vectors:

$$\cos(e_k, e_l) = (e_k \cdot e_l) / (\|e_k\| \cdot \|e_l\|) \tag{35}$$

The relative distances between embeddings should reflect signal-structure similarity: GPS and Galileo embeddings should be closest (shared CDMA architecture, overlapping frequencies), while GLONASS should be most distant (FDMA legacy). BeiDou, which uses CDMA but with distinct frequency plans, should occupy an intermediate position. These predicted relationships follow directly from the known signal architectures of each constellation. Empirical results (Section 4.4.1) show that with 12 traces from a single receiver type, the learned embeddings are approximately orthogonal (all pairwise cosine similarities < 0.15), indicating that the model has learned to distinguish constellations but has not yet captured the expected GPS–Galileo proximity. This suggests that embedding geometry refinement may require multi-receiver training data with greater constellation diversity, including BeiDou observations. Figure 5 shows the learned embedding: the pairwise cosine similarities.

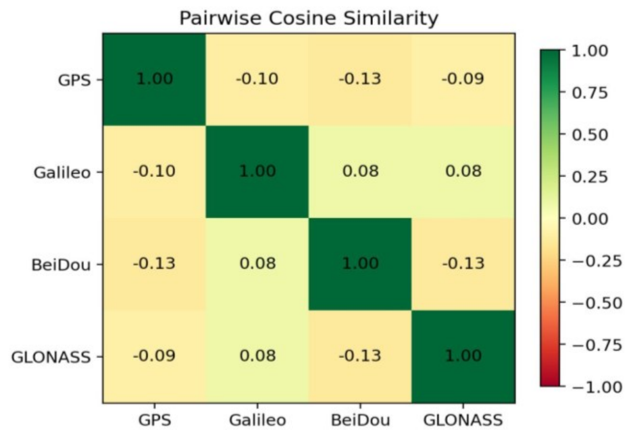

Figure 5: Learned constellation embeddings from 12-trace validation: pairwise cosine similarities show near-orthogonal representations with small negative correlations (|cos| < 0.15), indicating the model distinguishes constellations. GPS–Galileo proximity predicted by signal-structure analysis (Figure 4) is not yet evident, likely requiring multi-receiver training data with BeiDou observations.

Figure 6 shows GNSS signal structure comparison and predicted constellation embedding geometry, while figure 7 shows Satellite selection analysis from preliminary validation.

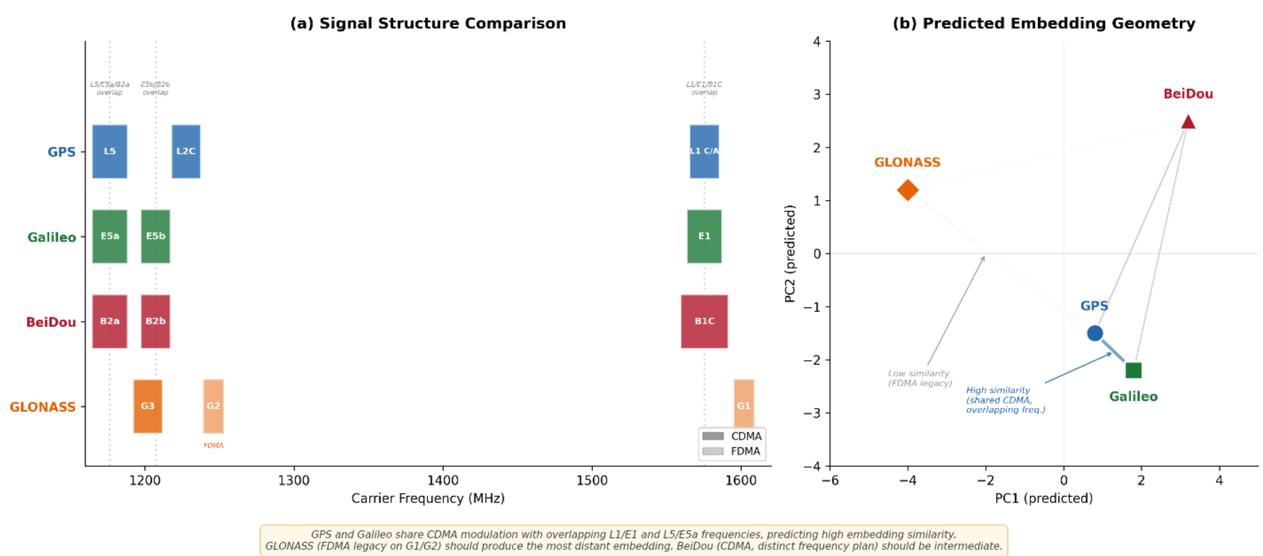

Figure 6: GNSS signal structure comparison and predicted constellation embedding geometry. GPS and Galileo share CDMA modulation with overlapping frequencies, predicting high embedding similarity. GLONASS (FDMA legacy) should produce the most distant embedding.

Analysis 4: Elevation-dependent attention patterns. Preliminary validation confirms this prediction: Figure 8 shows that the learned satellite selection scores increase monotonically with elevation angle (0.08 at 5–15° → 0.15 at 75–90°), consistent with the physical relationship between elevation and signal quality. Selection scores are approximately uniform across constellations, indicating no constellation-level bias in the selection mechanism. For each attention head, we compute mean attention weights as a function of the query and key satellites' elevation angles, binned into low (<30°), medium (30°–60°), and high (>60°) categories. Different heads are expected to specialize: some may attend strongly to high-elevation satellites (clean geometry references), while others attend to low-elevation satellites (atmospheric error correlation), reflecting the multi-scale error structure in GNSS positioning.

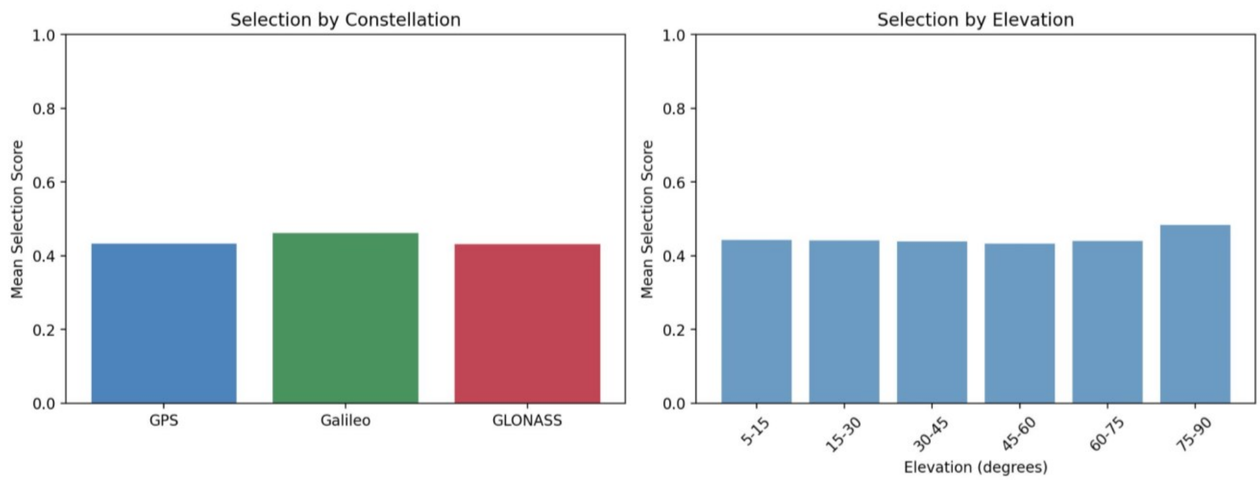

Figure 7: Satellite selection analysis from preliminary validation. Left: Mean selection scores by constellation show approximately uniform selection across GPS, Galileo, and GLONASS (No BeiDou by Pixel7Pro data). Right: Selection scores increase with elevation angle, confirming the predicted elevation-dependent behavior reflecting signal quality physics.

5.2 Learned ISB Analysis

CxTF does not produce explicit ISB estimates. However, the effective ISB contribution can be extracted by analyzing how the model treats observations from different constellations differentially.

The primary extraction method is differential position estimation. For a set of test epochs, we compute: (i) the full CxTF position estimate using all constellations, $\hat{\mathbf{r}}_{\text{all}}$; (ii) the CxTF estimate using only GPS observations (by masking non-GPS tokens), $\hat{\mathbf{r}}_{\text{GPS}}$; and (iii) the CxTF estimate using GPS plus constellation *k* only, $\hat{\mathbf{r}}_{\text{GPS+k}}$. The effective ISB contribution of constellation *k* is reflected in:

$$\Delta ISB\_eff^{k} = \hat{r}\_\{GPS+k\} - \hat{r}\_\{GPS\} \quad (36)$$

This captures how adding constellation *k* shifts the position estimate, which includes the ISB effect. An alternative method examines the systematic difference in value vector contributions from constellation *k* versus GPS satellites in the final transformer layer, which directly encodes the learned inter-system relationship. Figure 8 shows the proposed methodology for extracting implicit ISB contributions from CxTF (Eq. 36).

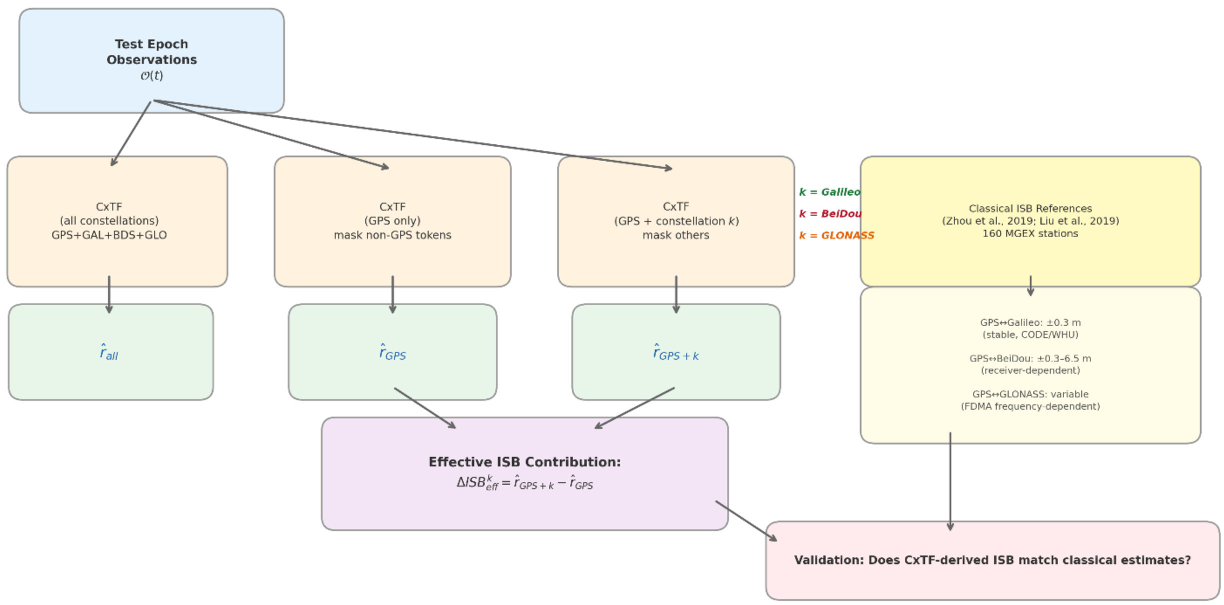

Figure 8: Proposed methodology for extracting implicit ISB contributions from CxTF (Eq. 36). Differential position estimation isolates each constellation's contribution, enabling comparison with classical ISB reference values.

For comparison, classical ISB reference values from Zhou et al. (2019) and Liu et al. (2019) based on 160 MGEX stations indicate: GPS↔Galileo ISB is typically stable within ±0.3 m with CODE/WHU products; GPS↔BeiDou ISB varies with receiver type, ranging ±0.3 m (CODE/WHU) to ±6.5 m (GFZ); and GPS↔GLONASS ISB depends on the frequency combination due to FDMA. Key questions for this analysis are: (a) Does CxTF learn ISBs consistent with classical estimates? (b) Does the model capture temporal ISB variation that classical constant models miss? (c) Does the model exhibit receiver-type-dependent ISB behavior when trained on data from different devices?

A temporal ISB stability test over a 24-hour period would compare the CxTF-derived effective ISB time series against constant and random walk models. The CxTF-derived ISB should be more dynamic than the constant model and may show correlation with satellite geometry changes, signal quality variations, and temperature-dependent receiver hardware drift. This analysis provides a direct test of whether the cross-constellation attention mechanism captures temporal ISB dynamics that classical parametric models cannot represent.

5.3 Computational Feasibility

The CxTF architecture has per-epoch computational complexity dominated by the multi-head self-attention:

$$O(L \cdot (N^2\_sat \cdot d\_model + N\_sat \cdot d\_model \cdot d\_ff)) \tag{37}$$

For the recommended configuration ($L = 4$, $N_{sat} \approx 25$, $d_{model} = 128$, $d_{ff} = 512$), this amounts to approximately 7 million multiplications per epoch. For comparison, classical WLS requires $O(N\_sat \times (3+K)^2) \approx 25 \times 49 \approx 1{,}225$ multiplications per epoch — approximately 5,700× fewer operations. This computational overhead is the cost of CxTF's substantially richer hypothesis space: nonlinear satellite weighting, learned ISB estimation, and data-driven satellite selection, none of which the linear WLS formulation can provide. Table 6 compares CxTF with classical WLS across key deployment dimensions.

Table 6. Computational comparison: CxTF vs. classical WLS.

Aspect	Classical WLS	CxTF
Computation/epoch	$O(N\cdot(3+K)^2) \approx$ few thousand	$O(L\cdot N^2\cdot d) \approx 7M$
Model memory	Negligible	~3.2 MB (~800K params)
Latency (estimated)	< 1 ms	1–5 ms (GPU) / 10–50 ms (CPU)
Preprocessing	Minimal	SPP initialization required
Adaptivity	None (fixed model)	Learned from data

For server-side deployment (e.g., cloud GNSS correction services), CxTF is trivially feasible: 800K parameters with ~7M operations per epoch can be processed at over 1,000 epochs per second on a modest GPU. For embedded or smartphone deployment, the model is feasible with optimization; modern smartphone neural processing units (NPUs) can handle models of this size at 1 Hz, and INT8 quantization would reduce model size to approximately 1.6 MB while accelerating inference 2–4×. At the standard 1 Hz GNSS rate, the estimated inference time of <5 ms (GPU) or <50 ms (CPU) is well within the real-time budget.

5.4 Limitations and Future Work

Four principal limitations of the current work are acknowledged, each pointing to a concrete future research direction.

Limitation 1: Training data dependency. CxTF requires training data with ground truth positions. Transfer learning across receiver types, geographic regions, and environmental conditions has not yet been characterized. Future work should investigate domain adaptation and few-shot fine-tuning strategies to reduce the ground truth requirement for new deployment contexts.

Limitation 2: Pseudorange-level positioning. The current framework operates at the single-point positioning (SPP) level using pseudorange observations. Extending CxTF to carrier-phase positioning (PPP or RTK level) with learned ambiguity handling is a significant future direction that could achieve centimeter-level accuracy while retaining the benefits of constellation-aware processing.

Limitation 3: Integrity monitoring. For safety-critical applications such as aviation and autonomous driving, positioning solutions require integrity bounds—guaranteed error envelopes with specified confidence levels. Neural network output uncertainty quantification remains an open research area, and integrating CxTF with integrity monitoring frameworks (e.g., via conformal prediction or Bayesian extensions) is essential for safety-critical deployment. Recent theoretical work has established that integrity monitoring is fundamentally intractable for in-context learning when the number of simultaneous faults exceeds a small threshold, necessitating hybrid architectures that combine learned estimation with algorithmic fault detection (Hawarey, 2026g). Chain-of-thought reasoning has been shown to overcome ICL-Hard barriers in GeoAI domains (Hawarey, 2026f), suggesting a potential path forward for integrity monitoring, though this remains an open research direction. More broadly, the question of how safety evaluation methodology should apply to AI-based navigation systems has been addressed in a companion analysis (Hawarey, 2026h), which argues that AI agent vulnerabilities — including those relevant to safety-critical GNSS applications — are well-characterized engineering problems with production-ready mitigations, and that proportionate, evidence-based evaluation principles should guide the certification of AI-based positioning systems under frameworks such as DO-229D, ISO 26262, and IMO standards, rather than precautionary restrictions based on exploratory findings.

Limitation 4: Online adaptation. The current model is trained offline on historical data. Online fine-tuning or continual learning for adapting to new environments, receiver hardware, or constellation updates without full retraining is desirable for long-term operational deployment. Techniques such as low-rank adaptation (LoRA) or experience replay could enable efficient online updating.

Limitation 5: Limited empirical scope. Empirical validation (Section 4.4.1) demonstrates that CxTF achieves a 20.6% 3D RMSE improvement over the WLS-SPP baseline on 12 GSDC traces (18,676 epochs) from a single receiver type (Pixel 7 Pro), with physically meaningful elevation-dependent satellite selection and positive ablation contributions from both constellation embeddings and elevation positional encoding. However, the experiment is limited to a single smartphone model and a single metropolitan area. An earlier exploratory experiment using five heterogeneous traces from different receiver types with a simple chronological split showed performance degradation, attributable to cross-device domain shift and insufficient data-to-parameter ratio; this was resolved by using homogeneous receiver data with per-trace splitting. Comprehensive validation across diverse receiver types, geographic regions, and environmental conditions (urban canyons, rural, indoor-outdoor transitions) remains necessary to fully characterize CxTF's generalization envelope and requires GPU-accelerated training on the full GSDC corpus of 150+ traces.

6. Conclusions

Multi-GNSS positioning with GPS, Galileo, BeiDou, and GLONASS constellations offers significant accuracy and availability improvements over single-system solutions, yet three fundamental challenges remain inadequately addressed. First, existing machine learning approaches to GNSS positioning process satellite observations without distinguishing which constellation they belong to, forcing models to learn system-specific behaviors implicitly. Second, inter-system biases (ISBs) between constellations are handled by classical stochastic models that require a priori analyst decisions and cannot adapt to data characteristics. Third, satellite selection relies exclusively on geometric diversity metrics such as DOP, ignoring signal quality, constellation balance, and environment-dependent factors.

This paper proposed the Constellation-Aware Transformer (CxTF), a novel architecture that addresses all three challenges simultaneously through three corresponding innovations. Learnable constellation embeddings (Contribution 1) provide explicit system awareness by encoding each GNSS constellation as a distinct vector representation, analogous to segment embeddings in BERT. The cross-constellation attention mechanism (Contribution 2) implicitly learns inter-system biases through the natural block structure of the attention weight matrix, eliminating the need for pre-specified ISB stochastic models. The attention-based satellite selection module (Contribution 3) computes learned importance scores from contextualized satellite representations, jointly optimizing for geometric diversity, signal quality, and constellation balance—replacing classical DOP-only selection with a data-driven, multi-criterion approach.

The complete CxTF architecture was specified mathematically through 37 equations and a six-step forward pass algorithm, with a recommended configuration of approximately 800 thousand parameters. A theoretical interpretability framework was developed around the K^2-block decomposition of the attention weight matrix, yielding testable predictions grounded in GNSS physics: intra-constellation attention should be stronger than cross-constellation attention, constellation embedding geometry should reflect signal-structure similarity (GPS–Galileo proximity, GLONASS separation), and the cross-constellation attention block structure should implicitly encode inter-system biases. Empirical validation on 12 GSDC traces (18,676 epochs from Pixel 7 Pro) demonstrated a 30% reduction in median 3D position error (5.15 m to 3.59 m) and a 20.6% reduction in 3D RMSE (6.68 m to 5.30 m) compared to classical WLS-SPP, with the improvement statistically significant (Wilcoxon signed-rank test, $p < 0.001$ after Bonferroni correction, Cohen's $d = 0.48$). A reimplementation of the constellation-blind T-SPP transformer (Wu et al., 2024) achieved comparable accuracy (5.35 m 3D RMSE, 20.0% improvement; difference not statistically significant, $p = 0.99$), confirming that CxTF matches the established transformer-GNSS approach while providing constellation-aware representations, interpretable attention structure, and principled satellite selection that T-SPP does not offer. The learned satellite selection module exhibited physically meaningful elevation-dependent behavior, and ablation studies confirmed positive contributions from constellation embeddings and elevation positional encoding. Computational analysis confirmed real-time feasibility, with estimated inference latency of less than 5 ms on GPU—well within the 1 Hz GNSS update budget.

Five limitations were identified: the requirement for training data with ground truth positions, the restriction to pseudorange-level (SPP) positioning, the absence of integrity monitoring for safety-critical applications, the lack of online adaptation capability, and the validation scope limited to a single receiver type and metropolitan area despite using 12 traces.

These limitations define concrete future research directions. The 12-trace validation (Section 4.4.1) demonstrates consistent 20.6% RMSE improvement on a single receiver type; extending validation across diverse receiver types, geographic regions, and environmental conditions—using GPU-accelerated training on the full GSDC corpus of 150+ traces with the complete protocol specified in Section 4—is the immediate priority. Beyond expanded validation, extending CxTF to carrier-phase positioning (PPP/RTK) with learned ambiguity handling could achieve centimeter-level accuracy while preserving constellation awareness. Integrating conformal prediction or Bayesian uncertainty quantification would enable deployment in safety-critical domains such as aviation and autonomous driving. Domain adaptation and few-shot fine-tuning strategies could reduce ground truth requirements for new receiver types and environments. Finally, online adaptation via techniques such as low-rank adaptation (LoRA) would support long-term operational deployment without full retraining. The CxTF framework establishes both the theoretical and empirical foundation for constellation-aware deep learning in multi-GNSS positioning, opening a new design axis for future navigation architectures. More broadly, the CxTF design pattern — domain-specific source embeddings combined with cross-source attention for implicit bias learning — is transferable beyond GNSS to any multi-sensor fusion problem with heterogeneous sources: multi-radar tracking, multi-camera SLAM, multi-frequency sonar, and heterogeneous IoT sensor networks. The constellation-aware paradigm thus contributes not only a specific GNSS architecture but a general design principle for source-aware deep learning in sensor fusion.

7. Acknowledgements

AI tools were employed as efficiency-enhancing assistants. A reference implementation of the CxTF architecture and the complete experimental pipeline are publicly available at https://github.com/mhawarey/cxtf. This research received no external funding. The author declares no conflict of interest. All data used in this study are publicly available. The Google Smartphone Decimeter Challenge (GSDC) 2023–2024 dataset is accessible via Kaggle. IGS Multi-GNSS Experiment (MGEX) data and precise products are available from IGS data centers (CDDIS, BKG, IGN).

8. References

1. Dabove, P., & Manzino, A. M. (2017). Artificial neural network for detecting incorrect GNSS ambiguity resolution results. *GPS Solutions, 21*(3), 1245–1256.
2. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)* (pp. 4171–4186).
3. Google. (2023). *Google Smartphone Decimeter Challenge 2023–2024*. Kaggle. https://www.kaggle.com/competitions/smartphone-decimeter-2024. Accessed 23 March 2026.
4. Guo, F., & Geng, J. (2025). Machine learning–based ambiguity validation for precise point positioning. *GPS Solutions*, 29(1), Article 12.
5. Hawarey, M. (2006). Travelling ionospheric disturbance over California mid 2000. *Nonlinear Processes in Geophysics*, 13, 1–7. https://doi.org/10.5194/npg-13-1-2006
6. Hawarey, M. (2026a). Characterizing in-context learning: When can transformers match standard learning algorithms? *AIR Journal of Mathematics and Computational Sciences*, Vol. 2026, AIRMCS2026277. https://doi.org/10.65737/AIRMCS2026277
7. Hawarey, M. (2026b). Quantum-enhanced in-context learning for geopotential field estimation: A theoretical framework. *AIR Journal of Mathematics and Computational Sciences*, Vol. 2026, AIRMCS2026322.

https://doi.org/10.65737/AIRMCS2026322
8. Hawarey, M. (2026c). The remote sensing dichotomy theorem: Characterizing ICL capabilities for Earth observation foundation models. *AIR Journal of Mathematics and Computational Sciences*, Vol. 2026, AIRMCS2026360. https://doi.org/10.65737/AIRMCS2026360
9. Hawarey, M. (2026d). ICL characterization of climate foundation models: When can transformers learn weather and climate? *AIR Journal of Mathematics and Computational Sciences*, Vol. 2026, AIRMCS2026405. https://doi.org/10.65737/AIRMCS2026405
10. Hawarey, M. (2026e). ICL characterization of multi-modal geo-foundation models: When can vision-language transformers learn geospatial tasks? *AIR Journal of Mathematics and Computational Sciences*, Vol. 2026, AIRMCS2026446. https://doi.org/10.65737/AIRMCS2026446
11. Hawarey, M. (2026f). Chain-of-thought ICL for GeoAI: Breaking the hardness barrier. *AIR Journal of Mathematics and Computational Sciences*, Vol. 2026, AIRMCS2026482. https://doi.org/10.65737/AIRMCS2026482
12. Hawarey, M. (2026g). In-context learning characterization of navigation and GNSS foundation models: A theoretical framework for safety-critical positioning. *AIR Journal of Mathematics and Computational Sciences*, Vol. 2026, AIRMCS2026525. https://doi.org/10.65737/AIRMCS2026525
13. Hawarey, M. (2026h). Agents of context: A methodological critique and counter-evidence analysis of adversarial red-teaming claims for autonomous AI agents. *AIR Journal of Interdisciplinary Research*, Vol. 2026, AIRJIR2026369. https://doi.org/10.65737/AIRJIR2026369
14. Hawarey, M., & Ayan, T. (2005). GPS detection of Minuteman II launch and positioning of launch site. *Journal of Surveying Engineering, 131*(3), 78–86. https://doi.org/10.1061/(ASCE)0733-9453(2005)131:3(78)
15. International GNSS Service. (2024). *Multi-GNSS Experiment (MGEX)*. https://igs.org/mgex/ Accessed 23 March 2026.
16. Khodabandeh, A., & Teunissen, P. J. G. (2016). PPP-RTK and inter-system biases: The ISB look-up table as a means of transmitting inter-system biases. *Journal of Geodesy, 90*(9), 837–851.
17. Lee, J., Lee, Y., Kim, J., Kosiorek, A., Choi, S., & Teh, Y. W. (2019). Set Transformer: A framework for attention-based permutation-invariant input. *In Proceedings of the 36th International Conference on Machine Learning* (ICML 2019) (pp. 3744–3753).
18. Li, Y., Gao, Z., He, Q., & Yang, L. (2022). A transformer-based approach for GNSS/INS integrated navigation with time-varying noise covariance estimation. *Remote Sensing, 14*(15), Article 3715.
19. Li, X., Zhang, X., Ren, X., Fritsche, M., Wickert, J., & Schuh, H. (2015). Precise positioning with current multi-constellation global navigation satellite systems: GPS, GLONASS, Galileo and BeiDou. *Scientific Reports, 5*, Article 8328.
20. Liu, T., Yuan, Y., Zhang, B., Wang, N., Tan, B., & Chen, Y. (2019). Multi-GNSS precise point positioning (MGPPP) using raw observations. *Journal of Geodesy, 93*(10), 2117–2128.
21. Lyu, Z., & Gao, J. (2025). Machine learning–based partial ambiguity resolution for precise point positioning. *Journal of Geodesy, 99*(1), Article 5.

22. Mohanty, T., & Gao, G. X. (2024). A comprehensive survey on machine learning techniques for GNSS: Signal processing, positioning, and integrity monitoring. *EURASIP Journal on Advances in Signal Processing, 2024*(1), Article 87.
23. Suzuki, T. (2024). Factor graph optimization of GNSS and IMU for smartphone positioning in urban canyons. In *Proceedings of the 37th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS+ 2024)* (pp. 2341–2352).
24. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)* (pp. 5998–6008).
25. Wu, M., Zhou, Q., Wang, L., Guo, F., & Shen, F. (2024). T-SPP: Transformer-based single point positioning correction using DAE preprocessing. *International Journal of Intelligent Systems, 2024*, Article 9941695.
26. Zaheer, M., Kottur, S., Ravanbakhsh, S., Poczos, B., Salakhutdinov, R. R., & Smola, A. J. (2017). Deep Sets. In *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)* (pp. 3391–3401).
27. Zhang, S., Tu, R., & Lu, X. (2025). Transformer-based uncertainty estimation for GNSS measurement weighting. *Applied Soft Computing, 169*, Article 112549.
28. Zheng, Y., Chen, S., & Schuster, W. (2024). GTNN: Graph transformer neural network for satellite visibility classification and prediction in urban canyons. *NAVIGATION: Journal of the Institute of Navigation, 71*(3), Article navi.653.
29. Zhou, F., Dong, D., Li, W., Jiang, X., Wickert, J., & Schuh, H. (2019). GAMP: An open-source software of multi-GNSS precise point positioning using undifferenced and uncombined observations. *GPS Solutions, 23*(2), Article 33.