What is “GT-Free” in one sentence

We estimate real-world GPS delivery noise (and optionally calibrated uncertainty) **without paired real ground-truth trajectories** by constructing pseudo ground truth $\mathring{X}_{GT}$ from a pseudo trip plan $\mathring{X}_{tp}$ and a synthetic-trained pseudo human behavior model, then learning the residual pseudo GPS noise to train a predictor.

---

1. Motivation

Learning a GPS denoiser typically requires paired supervision (X_{GT}, X_T), where X_T is a noisy GPS trajectory and X_{GT} is the corresponding ground truth (e.g., RTK/INS). In real deployments, such paired datasets are scarce, while most available data is either:

1. high-accuracy trajectories with limited scale,
2. commercial-grade GPS-only trajectories without ground truth,
3. synthetic paired data that may not match real human behavior or real GPS noise.

This proposal studies a GT-Free approach that learns a GPS delivery noise predictor **from real GPS trajectories** $\{X_T\}$ by building pseudo ground truth $\mathring{X}_{GT}$ from map-based trip-plan backbones and a synthetic-trained human behavior prior.

---

2. Datasets and Base Definitions

2.1 Datasets

- Synthetic dataset:

  $$S = \{(s_0, s_T)\}$$

  (Interpretation: synthetic trajectory pairs; in practice we treat s_0 as synthetic clean / ground truth, and s_T as synthetic noisy GPS.)
- Real-world GPS dataset:

  $$R = \{r_T\}$$

  (Real noisy GPS trajectories.)

we use a generic noisy trajectory symbol X_T when describing the method.

2.2 Conceptual decomposition (non-identifiability acknowledged)

Given an arbitrary real GPS human moving trajectory X_T, we define:

1. Trip plan X_{tp}: the background/backbone route a human planned or a navigation system suggests.
2. Human behavior noise ϵ_h: deviations from X_{tp} due to human motion choices.
3. GPS delivery noise ϵ_{GPS}: distortions added by GPS signal delivery and measurement processes.

We hypothesize:

$$X_T = X_{GT} + \epsilon_{GPS}$$

$$X_{GT} = X_{tp} + \epsilon_h$$

We do not claim we can uniquely recover $(X_{tp}, \epsilon_h, \epsilon_{GPS})$ from X_T; instead, we construct a pseudo supervision signal to learn a useful predictor of GPS delivery noise.

---

3. GT-Free Construction: $\mathring{X}_{tp}$, $\mathring{X}_{GT}$, $\mathring{\epsilon}_h$, $\mathring{\epsilon}_{GPS}$

3.1 Navigation / trip plan function

Define a deterministic navigation (trip-plan) function:

$$f_{tp}(X, \mathrm{map}) = \mathring{X}_{tp}$$

where X is a trajectory on the map (synthetic or real), and $\mathring{X}_{tp}$ is the pseudo trip plan.

Implementation examples:

- routing planner between endpoints (optionally with waypoints),
- map matching + smoothing,
- hybrid routing + matching.

3.2 Pseudo human behavior model

We assume existence of a synthetic human trajectory generator ω (used only to produce synthetic behavior patterns), and we learn a pseudo human behavior predictor:

$$\phi(\mathring{X}_{tp}) \sim p(\mathring{\epsilon}_h \mid \mathring{X}_{tp})$$

Training signal for ϕ (from synthetic data):
Given synthetic human moving trajectory s_0 (treated as synthetic ground truth), compute:

$$\mathring{X}_{tp}^{s} = f_{tp}(s_0, \mathrm{map})$$

Define pseudo human behavior residual:

$$\mathring{\epsilon}_{h}^{s} = s_0 - \mathring{X}_{tp}^{s}$$

Train ϕ to model:

$$\phi(\mathring{X}_{tp}) \approx p(\mathring{\epsilon}_h \mid \mathring{X}_{tp})$$

3.3 Pseudo ground truth on real data

For a real GPS trajectory X_T:

1. Pseudo trip plan:

$$\mathring{X}_{tp} = f_{tp}(X_T, \mathrm{map})$$

2. Pseudo human behavior (sampled or predicted):

$$\mathring{\epsilon}_{h}^{(k)} \sim \phi(\mathring{X}_{tp}), \quad k = 1, \dots, K$$

3. Pseudo ground truth:

$$\mathring{X}_{GT}^{(k)} = \mathring{X}_{tp} + \mathring{\epsilon}_{h}^{(k)}$$

4. Pseudo GPS noise:

$$\mathring{\epsilon}_{GPS}^{(k)} = X_T - \mathring{X}_{GT}^{(k)}$$

This constructs multiple pseudo-GT candidates per real trajectory, which yields a noise distribution (or interval) rather than a single point estimate.

---

4. Target Model: Pseudo GPS Noise Predictor θ

Our target is a model that predicts real GPS delivery noise from GPS-only input:

$$\theta(X_T) \sim p(\epsilon_{GPS} \mid X_T)$$

In GT-Free training, we instead supervise with pseudo noise:

$$\theta(X_T) \sim p(\mathring{\epsilon}_{GPS} \mid X_T)$$

where $\mathring{\epsilon}_{GPS}$ is computed from pseudo ground truth.

4.1 Training

Train θ on:

$$\left(X_T, \mathring{\epsilon}_{GPS}^{(k)}\right)$$

for all real trajectories $X_T \in R$ and samples $k = 1, \dots, K$.

4.2 Inference (map-free)

At inference time, θ takes only X_T:

- point prediction (mean):

$$\hat{\epsilon}_{GPS} = \theta_\mu(X_T)$$

- denoised trajectory:

$$\widehat{X}_{GT} = X_T - \hat{\epsilon}_{GPS}$$

Optionally, if θ predicts quantiles or a distribution, we return a confidence band for ϵ_{GPS} and thus for X_{GT}.

---

5. Outputs: Two Directions

This framework supports:

1. **GPS noise prediction**: output $\hat{\epsilon}_{GPS}$ (or its distribution).
2. **GPS noise range at confidence** k: produce an interval or region such that the model explains/tolerates $k\%$ of pseudo-GT samples:
   - sample-based empirical quantiles from $\{\mathring{\epsilon}_{GPS}^{(k)}\}_{k=1}^{K}$, or
   - directly predict quantiles via θ.

---

6. Evaluation Idea (GT-free, using $\mathring{\ }$ notation)

Goal. Evaluate whether θ learns **GPS delivery noise realism** (and optional calibration) under pseudo supervision, not trajectory accuracy.

A. Pseudo-noise alignment (primary)

For each X_T, construct K pseudo noise samples

$$\mathring{\epsilon}_{GPS}^{(k)} = X_T - \mathring{X}_{GT}^{(k)}, \quad k = 1, \dots, K.$$

Run inference to get predicted noise (or samples)

$$\hat{\epsilon}_{GPS} = \theta_\mu(X_T) \quad \text{or} \quad \hat{\epsilon}^{(m)}_{GPS} \sim \theta(X_T).$$

Compare the **distributional structure** between $\hat{\epsilon}_{GPS}$ and $\{\mathring{\epsilon}^{(k)}_{GPS}\}$ on fixed windows:

- tail / heavy-tail, skewness
- temporal correlation of $\Delta\epsilon$
- anisotropy (tangent vs normal)

B. Conditional consistency (primary)

Bucket windows by coarse context c (map- or metadata-derived proxy).
Check that both pseudo and predicted noises exhibit consistent heteroscedastic regimes:

$$p(\hat{\epsilon}_{GPS} \mid c_1) \neq p(\hat{\epsilon}_{GPS} \mid c_2), \quad p(\mathring{\epsilon}_{GPS} \mid c_1) \neq p(\mathring{\epsilon}_{GPS} \mid c_2).$$

C. Interval / quantile calibration to pseudo sampling (optional)

If θ outputs quantiles $\widehat{Q}_\alpha(X_T)$, test calibration **against the induced pseudo-noise distribution**:

$$\Pr\left(\mathring{\epsilon}_{GPS} \in \widehat{Q}_\alpha(X_T)\right) \approx \alpha.$$

D. Negative controls (sanity)

Include simple smoothers as baselines to ensure the above metrics do not merely reward oversmoothing.

---

7. Hypotheses

- $\mathring{X}_{tp}$ provides a useful low-frequency backbone for pseudo supervision.
- ϕ transfers human behavior patterns from synthetic generation into real backbones sufficiently to support pseudo-GT construction.
- training θ on real X_T with pseudo labels $\mathring{\epsilon}_{GPS}$ yields a better match to real GPS noise statistics than synthetic-only training.

---

8. Expected Deliverables

- A GT-Free pipeline to learn θ from real GPS-only trajectories without paired ground truth.

- A calibrated uncertainty estimator for ϵ_{GPS} and thus for trajectory denoising.
- A GT-Free evaluation protocol to compare denoisers and trajectory generators without real ground truth.