Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Pilot-residual-driven adaptive robust fusion of inertial and Doppler velocity measurements for underwater navigation

Sobhan Toulabi
Amirkabir University of Technology
Seyyede Marzieh Mousavi
mmousavi9429@aut.ac.ir
Amirkabir University of Technology

Article

Keywords: underwater navigation, Doppler velocity log, inertial navigation, factor-graph optimization, robust sensor fusion, impulsive measurement errors

Posted Date: November 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11540344/v1

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

Additional Declarations: No competing interests reported.

Pilot-residual-driven adaptive robust fusion of inertial and Doppler velocity measurements for underwater navigation

Sobhan Toulabi[1] and Seyyede Marzieh Mousavi[2,*]

[1]Department of Mechanical Engineering, Amirkabir University of Technology, Tehran, Iran

[2]Department of Aerospace Engineering, Amirkabir University of Technology, Tehran, Iran

[*]Corresponding author: mmousavi9429@aut.ac.ir

Abstract

Doppler velocity log (DVL) observations constrain inertial drift in underwater navigation but may exhibit changing dispersion, impulsive errors and temporary loss. We examine a two-stage batch factor-graph estimator that uses posterior DVL residuals from a Gaussian pilot to construct separate scale and standardized-tail statistics. The former modulates DVL factor scale; the latter selects a generalized robust weight. The graph uses calibrated IMU–DVL geometry and preintegrated inertial measurements, and omits missing DVL factors. On two held-out trajectories under 96 injected-fault realizations, mean trajectory error was 1.128 m versus 1.196 m for Gaussian and 1.186 m for pilot-weighted Huber. The paired difference from a pilot-weighted fixed Geman–McClure-type baseline was not significant. Without retuning, transfer to the independent TANK dataset yielded a mean error of 2.837 m across 96 further perturbations, with a 6.09% paired gain over the Geman-type baseline under impulsive errors. Ablations identify tail adaptation, rather than broad-noise scaling, as the principal supported component. The evidence concerns controlled perturbations of four physical sequences, not independent missions, and does not establish real-time operation or calibrated uncertainty.

Keywords: underwater navigation; Doppler velocity log; inertial navigation; factor-graph optimization; robust sensor fusion; impulsive measurement errors.

1 Introduction

Underwater navigation relies on a combination of inertial and complementary measurements because continuous satellite positioning is unavailable after submergence [1]. A Doppler velocity log (DVL) constrains velocity drift in an inertial measurement unit (IMU)-based solution, but its observations can change in dispersion, develop impulsive errors or become unavailable. Bottom-lock instability, weak beam returns, acoustic interference and vehicle

motion can therefore invalidate a fixed Gaussian observation model. The measurement-fusion problem is to reduce the influence of corrupted velocity observations while retaining useful DVL information and a consistent prediction of the velocity measured at the sensor.

Adaptive and robust INS/DVL filters address this problem through dual adaptive factors [2], covariance matching and noise estimation [3], Lie-group Student's-t filtering [4], hybrid robust kernels [5], and adaptive-$\mathcal{H}_\infty$ variational-Bayesian estimation [6]. These approaches establish that time-varying noise and heavy-tailed corruption need more flexible treatment than a stationary quadratic penalty. Physics-guided adaptive filtering provides a further example of robustness under degraded underwater observations [7]. The present study examines the corresponding problem within an optimization-based IMU–DVL measurement chain.

Other methods detect faulty DVL observations or reconstruct missing information. Examples include recurrent-network outlier detection [8], velocity estimation during complete beam outage [9], BeamsNet measurement enhancement [10], and experimental comparisons of learning strategies in DVL-denied navigation [11]. Particularly relevant to the measurement context, Jin et al. [12] use a multi-output least-squares support vector regression predictor to reconstruct unavailable beams and combine them with the remaining measured beams. Fault-tolerant adaptive filtering also combines temporal learning with tightly coupled SINS/DVL integration [13]. These methods address diagnosis or replacement; the estimator studied here uses the available recorded velocity and adjusts its influence continuously, without a learned replacement model or discrete fault class.

Factor-graph optimization (FGO) represents inertial dynamics, states and sensor observations in a sparse nonlinear least-squares problem. Its temporal structure supports IMU preintegration and heterogeneous measurements [14–16]. Underwater FGO already includes robust INS/USBL/DVL fusion [17], windowed Mahalanobis-distance anomaly detection [18], confidence-dependent covariance adjustment [19], and preintegration-based INS/DVL alignment [20]. Robust graph construction alone is consequently an established capability.

The closest adaptive methods make a narrower novelty claim necessary. Cheng et al. [21] adapt correlation-entropy factors and the optimization window from posterior residual statistics. Liu et al. [22] introduce an adaptively reweighted virtual-velocity factor for deep-sea docking, while Cheng et al. [23] combine learned fault diagnosis with adaptive correntropy FGO. Adaptive robust INS/DVL/PS graphs with delayed marginalization [24] further demonstrate the maturity of this direction. Outside underwater navigation, AEROS adapts graph robustness through a latent parameter [25]; scale-variant robust-kernel analysis [26] and gradient-adaptive marine FGO [27] likewise motivate residual-dependent tuning. This work therefore does not claim the first adaptive or robust underwater factor graph.

The specific question is whether *residual dispersion* and *scale-standardized tail shape* can control different properties of a DVL factor. A large raw residual can arise from broad noise inflation or a sparse spike. Uniform covariance inflation weakens every observation, whereas a strong fixed kernel can suppress useful observations when corruption is absent. The two effects should be characterized separately. Their interpretation also depends on the physical model: incorrect frame transforms, gravity conventions, timing or sensor lever arms can mimic measurement degradation. DVL mounting calibration [28] and delay handling [29] are

therefore prerequisites for meaningful residual analysis. Song et al. [30] explicitly model the spatial inconsistency of SINS/DVL/pressure-sensor reference centers, reinforcing the need to compare measured and predicted velocities at the same physical point.

We address this question using a Gaussian physical pilot graph followed by an adaptive robust solve. Component-wise median centering and MAD scaling characterize broad residual dispersion; standardized tail fraction and tail-to-center ratio characterize impulsive behavior. The former controls effective covariance scaling, and the latter controls the scale and exponent of the robust weight. Missing DVL measurements are omitted. The residual statistics regulate factor influence; they are not treated as an instrument uncertainty budget.

Our specific methodological contribution is a *pilot-conditioned scale–shape weighting rule* for DVL velocity factors: a robust dispersion summary and scale-standardized tail descriptors drive different components of a fixed second-stage least-squares weight. This is a proposed combination and operationalization of established statistical ingredients, not a new robust loss family, factor-graph formalism or sensor-geometry law. Two further contributions are a matched ablation on held-out recorded trajectories and an unchanged-parameter transfer to the independent TANK benchmark [31]. The evidence supports tail-sensitive rejection of impulsive contamination within the tested graph more strongly than it supports the broad-noise scaling channel. A dropout diagnostic does not alter the optimizer; persistent bias and severe mixed degradation remain limitations. Fault realizations repeatedly perturb a small set of recorded physical trajectories and are not independent missions.

2 Measurement model and factor-graph formulation

2.1 Frames, state and inertial factors

World, body, IMU and DVL frames are denoted by W, B, M, D. The transform $\mathbf{T}_{AB}$ maps coordinates from B to A, with rotation $\mathbf{R}_{AB}$ and translation $\mathbf{t}_{AB}$; hence $\mathbf{R}_{BW} = \mathbf{R}_{WB}^{\mathsf{T}}$. Position and velocity refer to the body origin. At keyframe k, the state is

$$\mathbf{x}_k = [\mathbf{p}_k^W, \mathbf{v}_k^W, \mathbf{q}_{WB,k}, \mathbf{b}_{a,k}, \mathbf{b}_{g,k}, \mathbf{b}_{D,k}], \qquad \mathcal{X} = \{\mathbf{x}_0, \ldots, \mathbf{x}_K\}. \tag{1}$$

The IMU biases are $\mathbf{b}_a, \mathbf{b}_g$, and $\mathbf{b}_D$ is a tightly regularized DVL-offset nuisance state. Attitude is represented on SO(3), with three rotation parameters in the optimizer, giving 18 parameters per keyframe. The nuisance state is not adaptively released or interpreted as a validated persistent-bias estimator.

Body-frame angular rate and specific force satisfy

$$\tilde{\boldsymbol{\omega}} = \boldsymbol{\omega}^B + \mathbf{b}_g + \mathbf{n}_g, \tag{2}$$

$$\tilde{\mathbf{f}} = \mathbf{R}_{BW}(\mathbf{a}^W - \mathbf{g}^W) + \mathbf{b}_a + \mathbf{n}_a. \tag{3}$$

Raw IMU samples between graph states are preintegrated on SO(3) [14, 20]. The implementation retains the raw sub-intervals and recomputes increments at the candidate accelerometer and gyroscope biases during each residual evaluation. Angular rate and specific force are

midpoint-averaged, with the force rotated by the half-step attitude. Adjacent IMU biases have random-walk factors. Conservative diagonal position, velocity, rotation and bias scales are used rather than a fully propagated preintegration covariance. Supplementary Section S1 gives the recursions, residuals and frame interfaces; Supplementary Table S1 lists all frozen factor scales.

2.2 Velocity at the DVL point

Let $\mathbf{r}_{BD}^{B}$ be the DVL position relative to the body origin. The recorded sensor-frame velocity is rotated into the body frame without subtracting a lever-arm term. The factor predicts velocity at that same sensor point [30]:

$$\tilde{\mathbf{v}}_{D,k}^{B} = \mathbf{R}_{BD}\tilde{\mathbf{v}}_{D,k}^{D}, \tag{4}$$

$$\hat{\mathbf{v}}_{D,k}^{B} = \mathbf{R}_{BW,k}\mathbf{v}_{k}^{W} + (\tilde{\boldsymbol{\omega}}_{k} - \mathbf{b}_{g,k}) \times \mathbf{r}_{BD}^{B} + \mathbf{b}_{D,k}, \tag{5}$$

$$\mathbf{r}_{D,k} = \hat{\mathbf{v}}_{D,k}^{B} - \tilde{\mathbf{v}}_{D,k}^{B}. \tag{6}$$

The rotational contribution is thus introduced once, in the prediction. Subtracting it from the measurement is used only to form an approximate body-origin velocity for initialization, never as the factor observation. The same transforms and geometry are used by every compared estimator.

The auxiliary offset has fixed whitened residuals

$$\mathbf{r}_{b_D,k}^{0} = \frac{\mathbf{b}_{D,k}}{0.010\ \mathrm{m/s}}, \qquad \mathbf{r}_{b_D,ij}^{rw} = \frac{\mathbf{b}_{D,j} - \mathbf{b}_{D,i}}{0.004\ \mathrm{m/s}}. \tag{7}$$

These constraints do not change with fault type. Removing the local residual median in the adaptive statistics below reduces sensitivity to sustained location shifts, but does not estimate or correct an unknown persistent bias.

2.3 Common graph objective

Writing J_{phys} for the common initial-state, preintegrated IMU, IMU-bias and nuisance-state prior terms, the graph solves

$$\mathcal{X}^{\star} = \arg\min_{\mathcal{X}} \left[J_{\mathrm{phys}}(\mathcal{X}) + \sum_{k \in \mathcal{D}} \mathcal{L}_{D}(\mathbf{r}_{D,k}) \right]. \tag{8}$$

The Gaussian DVL term is $\|\mathbf{r}_{D,k}/\sigma_{D,0}\|_2^2$; robust variants change only its statistical weighting. The set $\mathcal{D}$ contains available valid DVL observations. All methods omit missing factors and retain inertial propagation during outages. No ground-truth pose, pseudo-depth or sonar observation is supplied to a graph factor. An onboard quaternion is used only for initialization. Reference measurements enter only the evaluation procedure described in Section 4.3.

3 Pilot-residual-driven adaptive DVL treatment

3.1 Pilot residual, robust centering and scale

A Gaussian solve using $\sigma_{D,0} = 0.07$ m/s gives the pilot trajectory $\hat{\mathcal{X}}^G$. Its posterior DVL residual is

$$\boldsymbol{\nu}_k = \hat{\mathbf{v}}_{D,k}^{B,G} - \tilde{\mathbf{v}}_{D,k}^{B}, \tag{9}$$

where the prediction uses Eq. (5) evaluated at the pilot state. These are posterior residuals after using the DVL data, rather than independent pre-fit innovations. No ground truth or fault label is used to compute them.

The trailing window $\mathcal{W}_k$ contains valid residuals from the most recent 8 s (20 graph keyframes at 2.5 Hz). With at least five valid residuals, component-wise median centering and robust scaling give

$$\mu_{d,k} = \mathrm{median}_{j\in\mathcal{W}_k}\, \nu_{j,d}, \qquad \bar{\nu}_{j,d} = \nu_{j,d} - \mu_{d,k}, \tag{10}$$

$$s_{d,k} = \max\left[1.4826\,\mathrm{median}_{j\in\mathcal{W}_k}\,|\bar{\nu}_{j,d} - \mathrm{median}_{\ell\in\mathcal{W}_k}\,\bar{\nu}_{\ell,d}|\,, 10^{-4}\ \mathrm{m/s}\right], \tag{11}$$

$$s_k = \sqrt{(s_{x,k}^2 + s_{y,k}^2 + s_{z,k}^2)/3}. \tag{12}$$

Median centering suppresses location shifts in the scale and shape summaries; component-wise standardization prevents a broad change in dispersion from automatically being interpreted as a tail-shape change. The frozen clean reference is $s_0 = 7.0746 \times 10^{-3}$ m/s. Evidence values remain zero when fewer than five valid residuals are available. Supplementary Section S2 provides the full-precision references and configuration.

3.2 Separate broad-noise and heavy-tail evidence

Let $C(x) = \mathrm{clip}_{[0,1]}(x)$. Broad-scale inflation controls the effective DVL standard deviation through

$$e_{N,k} = C\left(\frac{\ln[1 + \max(s_k/s_0 - 1, 0)]}{\ln(4.5)}\right), \tag{13}$$

$$m_{\sigma,k} = \mathrm{clip}_{[1,4]}(1 + 3e_{N,k}), \qquad \sigma_{D,k}^{\mathrm{eff}} = \sigma_{D,0} m_{\sigma,k}. \tag{14}$$

Clean-like scales give evidence near zero; increasing dispersion reduces the factor's inverse covariance. This quantity controls influence in the graph, not a calibrated estimate of the DVL instrument's measurement uncertainty.

For tail shape, define standardized residual norms and two window statistics:

$$\zeta_j = \left\|[\bar{\nu}_{j,x}/s_{x,k}, \bar{\nu}_{j,y}/s_{y,k}, \bar{\nu}_{j,z}/s_{z,k}]^{\mathsf{T}}\right\|_2, \tag{15}$$

$$f_{T,k} = \frac{1}{|\mathcal{W}_k|}\sum_{j\in\mathcal{W}_k} \mathbb{I}(\zeta_j > 3.6), \qquad r_{T,k} = \frac{Q_{0.95}(\zeta)}{\max[Q_{0.50}(\zeta), 10^{-6}]}. \tag{16}$$

Here Q_p is the pth quantile and $\mathbb{I}$ is the indicator. The clean references $f_{T,0} = 0.05263$ and

$r_{T,0} = 2.54770$ are fixed from the development sequence. The normalized tail evidence is

$$e_{F,k} = C\left(\frac{f_{T,k} - f_{T,0}}{\max(0.18 - f_{T,0}, 0.05)}\right), \tag{17}$$

$$e_{R,k} = C\left(\frac{r_{T,k}/\max(r_{T,0}, 1) - 1}{2.5}\right), \tag{18}$$

$$e_{T,k} = C[\max(e_{F,k}, 0.75e_{R,k})]. \tag{19}$$

The fraction detects standardized extremes, while the quantile ratio compares the upper tail with its center. Their use after scale removal follows the distinction between residual scale and robust-kernel behavior examined by Das and Gross [26]. These are *heuristic summaries of post-fit residuals*, not statistically identified noise-process parameters or mutually independent fault detectors. A spike can change both dispersion and tail evidence, and the Gaussian pilot can itself be influenced by corruption.

3.3 Frozen robust weighting and second solve

Heavy-tail evidence sets the kernel parameters and normalized pilot magnitude:

$$c_k = 4 - 3e_{T,k}, \qquad p_k = 1 + 1.2e_{T,k}, \qquad u_k = \frac{\|\boldsymbol{\nu}_k\|_2}{\sigma^{\text{eff}}_{D,k}}. \tag{20}$$

Thus (c, p) changes continuously from $(4, 1)$ for clean-like evidence toward $(1, 2.2)$ for strong tails. Following the residual-dependent robust-factor principle [21, 24, 25], the implemented generalized weight and final whitened residual are

$$w_k = \text{clip}_{[0.01,1]}\left[\left(1 + (u_k/c_k)^2\right)^{-p_k}\right], \tag{21}$$

$$\mathbf{r}^A_{D,k} = \frac{\sqrt{w_k}}{\sigma^{\text{eff}}_{D,k}}\left(\hat{\mathbf{v}}^B_{D,k} - \tilde{\mathbf{v}}^B_{D,k}\right). \tag{22}$$

The evidence, scales and weights are computed once from the Gaussian pilot and held fixed during the second solve. The physical prediction inside the residual is recomputed at each candidate state. The second-stage DVL term is $\|\mathbf{r}^A_{D,k}\|_2^2$. Consequently, scale evidence weakens the observation through covariance scaling, while tail evidence adjusts the suppression of large pilot residuals. Adaptation requires no specialist selection or fault classification.

Unavailable DVL observations are omitted in both stages. A missing-fraction and longest-run diagnostic $e_{D,k}$ is recorded, but it does not modify $\sigma^{\text{eff}}_{D,k}, c_k, p_k$ or w_k. Its exact definition is in Supplementary Section S2. The zero-effect dropout ablation is therefore a structural check, rather than validation of a third adaptive channel.

The trailing statistics use only preceding residuals, but the pilot itself is a batch solution. The present implementation is therefore a two-stage batch estimator, and the window construction alone does not establish online causality or real-time execution. The ablation and accuracy results below assess this implemented estimator with its mappings frozen before evaluation.

Algorithm 1 Two-stage pilot-residual-driven estimation

Input: Raw IMU, DVL velocity/validity, calibrated sensor transforms and frozen parameters

Build the common physical graph; omit unavailable DVL factors
Solve the Gaussian graph to obtain $\hat{\mathcal{X}}^{G}$
Compute valid posterior DVL residuals $\boldsymbol{\nu}_k$
for each keyframe k **do**
    Form the trailing 8-s window; initialize $e_{N,k} = e_{T,k} = 0$
    if at least five valid residuals are available **then**
        Compute component medians, MAD scales and $e_{N,k}, e_{T,k}$
    end if
    Set $\sigma_{D,k}^{\text{eff}}, c_k, p_k, w_k$ from the frozen mappings
end for
Solve the graph with fixed adaptive scales/weights and recomputed physical residuals
return Navigation states

4 Experimental setup and accuracy evaluation

4.1 Development, held-out trajectories and external transfer

The primary study uses the public Sonar-Based UUV dataset, recorded with a BlueROV2-class vehicle in an indoor tank, with synchronized IMU/DVL streams and external motion-capture poses [32]. The circle trajectory is reserved for physical-interface auditing, method development and clean residual references. The L-shape and line trajectories are held out for the final evaluation. Imaging sonar is available but excluded, so all estimators use the same IMU–DVL information. Motion capture is an evaluation reference rather than a source of graph factors.

After freezing the primary method, it is transferred to TANK [31]. The released benchmark has eight sequences; this study uses the three Structure sequences because their direct AprilTag/TankGT coverage is much more continuous than that of the HalfTank and WholeTank sequences. The SLAM-augmented `/apriltag_slam/GT_full` stream is excluded. Structure Easy is used only to interpret sensor fields, transforms, DVL sign, specific force, timing and reference-frame conversion. Structure Medium and Hard are held out. Neither clean residual references nor adaptive window, thresholds or kernel mappings are retuned on TANK. Supplementary Section S1 documents the adapter and reference-coverage restriction.

At 2.5 Hz, Medium has 438 keyframes over 174.8 s, with 406 valid real DVL factors; Hard has 182 over 72.4 s, with 153 valid factors. Natural invalidity is preserved, and injected degradations never turn an invalid observation into a valid one. The primary and TANK studies are reported separately: each has two held-out physical sequences, four fault families and 12 seeds, giving 96 matched fault realizations. These counts describe repeated

perturbations, not 96 independent physical missions in either dataset.

4.2 Frozen configuration and matched baselines

Every method shares the state definition, raw-IMU preintegration, calibrated transforms, lever-arm geometry, initialization, missing-factor policy and diagonal physical-factor scales. The graph rate is 2.5 Hz, IMU–DVL lag is zero, and DVL scale is unity. The primary lever arm is $[-0.28, -0.10, 0.12]^{\mathsf{T}}$ m; TANK uses its separately audited calibration. Supplementary Sections S1–S2 give the exact transforms, initial priors and factor scales. Ground-truth orientation and position, pseudo-depth and DVL altitude are not introduced as measurements.

Four variants are compared within the identical physical graph. The Gaussian baseline uses unit DVL weights and $\sigma_{D,0} = 0.07$ m/s. The two *fixed-kernel, pilot-weighted* baselines evaluate the Gaussian-pilot residual norm $u_k^G = \|\boldsymbol{\nu}_k\|_2/\sigma_{D,0}$ and set

$$w_k^H = \mathrm{clip}_{[0.01,1]}\left[\min\left(1, \frac{1.8}{\max(u_k^G, 10^{-8})}\right)\right], \tag{23}$$

$$w_k^{GM} = \mathrm{clip}_{[0.01,1]}\left[\left(1 + (u_k^G/2.5)^2\right)^{-2}\right]. \tag{24}$$

We refer to these as *Huber-style* and *generalized Geman–McClure-type* baselines, respectively. Their weights remain fixed during one subsequent weighted least-squares solve, exactly as the proposed pilot-derived weights do. They are *not* fully iteratively reweighted M-estimator solutions, and our comparisons must not be interpreted as tests against such converged robust optimizers. The proposed estimator instead uses the adaptive scale and tail-shape mappings in Section 3. Ablations disable broad-noise scaling, disable heavy-tail adaptation, or zero the dropout diagnostic, with all remaining settings unchanged. A wide-bias Gaussian variant is reported only in the separate bias stress test and is excluded from the three primary baselines.

The circle development sequence fixes the physical model, residual references, evidence definitions, mappings and evaluation rules before the unseen confirmatory seeds are applied. Solver evaluation limits are numerical termination controls: the primary runner allows 45 evaluations for the Gaussian pilot and 35 for robust solves; TANK uses staged continuation without changing the objective or fixed weights. The continuation schedule and recorded runtimes are in Supplementary Section S4. All 384 external method–case combinations terminate successfully; this is a numerical result, not a guarantee of accurate navigation.

4.3 Controlled degradations and evaluation metrics

Controlled faults are added to recorded real DVL velocities; real IMU, vehicle motion, timing and reference streams are retained. The four primary families are broad Gaussian noise, impulsive spikes, random/burst dropout and mixed degradation. Mixed cases combine these effects with a constant velocity offset; they therefore test an additional limitation outside the validated bias capability. A constant-bias-only stress test is reported separately. Exact generation ranges appear in Supplementary Table S2, and frozen seeds are 20292301–

20292312. Severity is obtained by normalizing the sampled parameters within their ranges and averaging component scores when needed. Scores below 1/3, from 1/3 to below 2/3, and at least 2/3 define mild, moderate and severe subsets. Their sizes depend on the sampled faults and need not be equal.

The primary outcome is strict root-mean-square absolute trajectory error:

$$E = \mathrm{ATE}_{\mathrm{RMSE}} = \sqrt{\frac{1}{N}\sum_{k=1}^{N} \|\tilde{\mathbf{p}}_k - \mathbf{p}_k^{\mathrm{GT}}\|_2^2}. \tag{25}$$

The evaluation resolves only an initial yaw-frame offset using the first 2 s: horizontal velocity correspondence is used when sufficient motion is present, with initial horizontal displacement as fallback. Translation is set by the first position sample. There is no scale fit, whole-trajectory registration or subsequent adjustment using later ground truth. This alignment is applied after estimation, and the same rule is used for every method. Velocity RMSE uses the analogous three-dimensional velocity error. Final position error and horizontal ATE provide supporting summaries.

For each matched case, paired percentage gain over baseline B is $g_B = 100(E_B - E_A)/E_B$, where A is the proposed estimator. Averaging these ratios differs from comparing the two aggregate mean ATE values. Regret uses the retrospectively best fixed baseline:

$$\mathcal{R} = 100\frac{E_A - \min_{B\in\mathcal{B}} E_B}{\min_{B\in\mathcal{B}} E_B}, \qquad \mathcal{B} = \{\mathrm{Gaussian}, \mathrm{Huber}, \mathrm{Geman}\}. \tag{26}$$

This per-case choice is an evaluation oracle unavailable to an online estimator. It measures proximity to the available fixed alternatives rather than deployment performance of an additional method.

4.4 Statistical units and frozen analysis

The primary protocol comprises 96 matched trajectory–fault–seed cases, plus one clean run per held-out trajectory. The reported mean percentage gains are averages of the casewise quantities g_B, whereas the primary two-sided Wilcoxon signed-rank p-values are calculated on the *paired absolute ATE differences* $E_B - E_A$, not on percentage gains. The 5000 case-level bootstrap replicates yield descriptive confidence intervals for the finite fault-perturbation design. Because multiple cases reuse the same two physical trajectories, neither these intervals nor the nominal Wilcoxon p-values establish inference to a population of missions; treating all 96 cases as independent physical observations would overstate generality. Rank-biserial correlation, Cohen's d_z and win fractions provide supporting descriptions. Component ablations use identical cases and are not retuned. An explicitly *post-hoc* sensitivity analysis also aggregates paired percentage gains across both trajectories and all four families within each of the 12 fault-generation seeds. Its exact signed-rank results and seed-bootstrap intervals (Supplementary Table S6) address sensitivity to the sampled perturbations, but do not generate additional independent physical trajectories.

TANK is a subsequent transfer experiment, not an additional tuning set or a pooling

of primary and external cases. Its descriptive means and win fractions use the 96 matched realizations. A seed-cluster sensitivity analysis first averages paired percentage gains over both sequences and all four families within each seed, then applies an exact Wilcoxon test and 5000-replicate bootstrap to the 12 seed-level aggregates. Family-specific sensitivity analyses follow the same principle. Seed-level aggregation reduces the risk of treating fault perturbations sharing a generation seed as independent. It does not account for all dependence induced by reusing the same underlying sensor recordings; the number of independent physical sequences remains two for TANK and two for the primary dataset, and population-level environmental generalization is not established.

AI-assisted manuscript preparation. ChatGPT (OpenAI) assisted with language editing, presentation of existing analyses, structural revisions and journal-specific formatting. No new physical experiments were performed as part of these textual revisions. The authors are responsible for checking all technical statements, source records, statistical interpretations and the final manuscript.

5 Results

5.1 Primary accuracy and paired comparisons

The clean runs show similar nominal accuracy: mean strict ATE is 0.9226 m for the proposed estimator and 0.9235 m for Gaussian, a ratio of 0.9990. Velocity RMSE is 0.1887 versus 0.1899 m/s. The fixed Geman-type baseline gives the lowest clean ATE, 0.9197 m; the proposed difference is about 3 mm. These are two recorded clean trajectories, not a statistical equivalence test. Supplementary Table S3 provides all clean results.

Over the 96 primary fault realizations, the proposed mean strict ATE is 1.1280 m with bootstrap 95% interval $[1.0307, 1.2390]$ m (Table 1). Figure 1 shows the aggregate and family-specific errors. The aggregate proposed–Geman-type difference is small; the comparison should be read alongside paired analyses and ablation results rather than as a meaningful absolute ranking of robust estimators.

Table 1: Aggregate confirmatory performance over the 96 primary fault realizations.

Method	Mean ATE [m]	Bootstrap 95% CI [m]	Median [m]
Gaussian	1.1962	[1.0914, 1.3089]	1.0407
Huber	1.1860	[1.0866, 1.2986]	1.0433
Geman–McClure	1.1319	[1.0344, 1.2432]	1.0120
Proposed	1.1280	[1.0307, 1.2390]	1.0029

Table 2 reports the paired percentage results. Mean casewise percentage gains over Gaussian and the pilot-weighted Huber-style baseline are 4.27% and 3.76%, respectively. The corresponding signed-rank tests on *absolute paired ATE differences* give $p = 0.0110$ and $p = 0.0162$. The Gaussian comparison has rank-biserial correlation 0.299 and Cohen's $d_z = 0.322$; corresponding Huber values are 0.283 and 0.310 (full results in Supplementary Table S4).

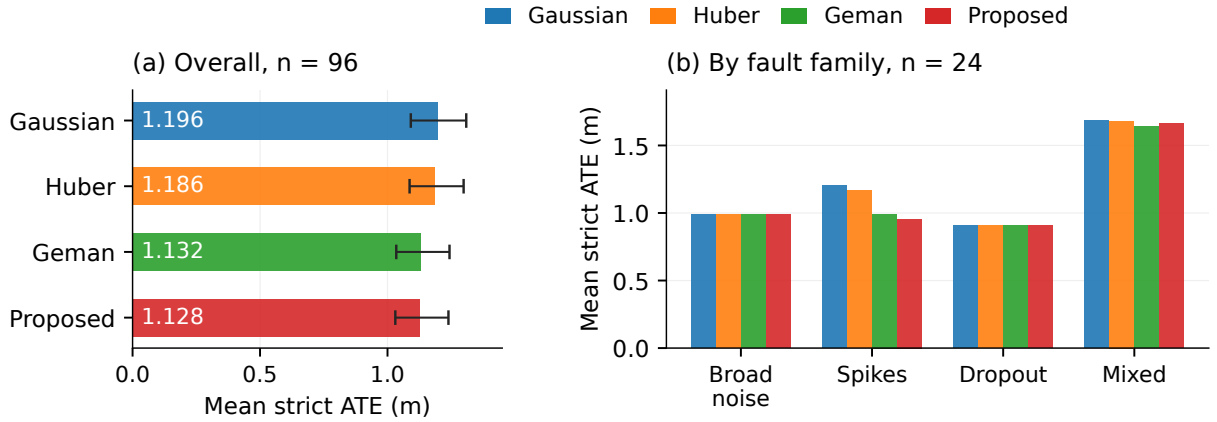


Figure 1: Primary confirmatory performance: (a) mean strict ATE with the original 95% bootstrap intervals over 96 cases; (b) mean ATE by fault family, with 24 matched cases per family. All methods share the physical graph. Geman denotes the pilot-weighted fixed generalized Geman–McClure-type baseline; Huber also uses one fixed pilot-derived weighting step.

Both win fractions are 58.3%. The proposed–pilot-weighted Geman-type difference is not significant on the absolute-error signed-rank test ($p = 0.3402$), and mean casewise percentage gain is -0.62% despite the slightly lower aggregate proposed ATE. This ordering reflects averaging case-normalized ratios rather than absolute errors. The comparisons indicate an advantage over Gaussian and this particular pilot-weighted Huber-style implementation under the chosen perturbations; they do not establish superiority to Geman-type weighting, to converged robust M-estimators or to other adaptive navigation techniques. Neither statistical equivalence to a fixed baseline nor generalization beyond the recorded physical trajectories was tested. A post-hoc primary seed-cluster sensitivity check gives exact two-sided $p = 0.00098$, 0.00244 and 0.67725 for the same three comparisons, respectively (Supplementary Table S6). These tests retain the physical-trajectory limitation and do not replace the pre-specified 96-case statistics.

Table 2: Primary paired comparison across 96 matched fault realizations. Positive gain favors the proposed method; intervals use 5000 case-level bootstrap replicates. Wilcoxon p-values test paired *absolute ATE* differences, not percentage gains. Cases reuse recorded physical paths.

Baseline	Mean gain (%)	95% CI (%)	Wilcoxon p	Win fraction
Gaussian	4.268	[1.578, 7.215]	0.0110	0.583
Huber	3.764	[1.208, 6.578]	0.0162	0.583
Geman–McClure	-0.624	[-3.549, 1.871]	0.3402	0.490

5.2 Fault type and severity

Impulsive spikes produce the clearest benefit. Mean ATE falls from 1.2025 m for Gaussian to 0.9559 m for the proposed method. The mean paired Gaussian gain is 16.45%, with interval $[8.60, 24.65]\%$ and $p = 6.5 \times 10^{-4}$; the Huber gain is 14.74% ($p = 0.0011$). Geman is already strong at 0.9928 m, and the paired difference from it is not significant ($p = 0.178$).

Broad-noise ATE is closely grouped: proposed 0.9866 m versus Gaussian 0.9878 m ($p = 0.922$). Dropout is similarly close, 0.9058 versus 0.9075 m, consistent with shared factor omission. Mixed degradation is hardest: proposed ATE is 1.6636 m versus Gaussian 1.6869, Huber 1.6815 and Geman 1.6395 m, without significant paired differences.

Supplementary Fig. S6 shows the growing spike advantage: under severe spikes, proposed ATE is 0.9855 m, Gaussian 1.4762, Huber 1.3964 and Geman 1.0033 m. Severity subsets contain different sampled realizations, so the curves need not be monotonic. The remaining severity results in Supplementary Section S3 retain contrary outcomes: severe broad noise favors Gaussian (0.8294 versus proposed 0.8717 m), and severe mixed degradation gives proposed 2.9573 m versus Gaussian 2.8463 and Geman 2.7332 m. These cases were retained under the frozen mapping.

5.3 Evidence separation, ablation and time-resolved controls

The scale and shape statistics activate differently (Fig. 2a). Broad-noise cases give intended noise evidence 0.781 versus largest competing activation 0.086; spike cases give tail evidence 0.712 versus 0.298. The dropout diagnostic likewise separates missingness (0.741 versus 0.295). Separation identifies a measurement pattern; it does not establish improved navigation.

The ablation makes that distinction explicit (Fig. 2b; Supplementary Table S5). Removing heavy-tail adaptation increases mean ATE from 1.1280 to 1.1795 m, a 4.57% degradation. Removing broad-noise scaling slightly improves it to 1.1202 m (−0.69% degradation). Zeroing dropout evidence gives exactly zero mean and maximum absolute ATE change across all 96 cases. Heavy-tail adaptation is therefore the strongest independently supported component; broad-noise covariance scaling has no positive aggregate ablation benefit, and dropout evidence has no numerical weighting effect.

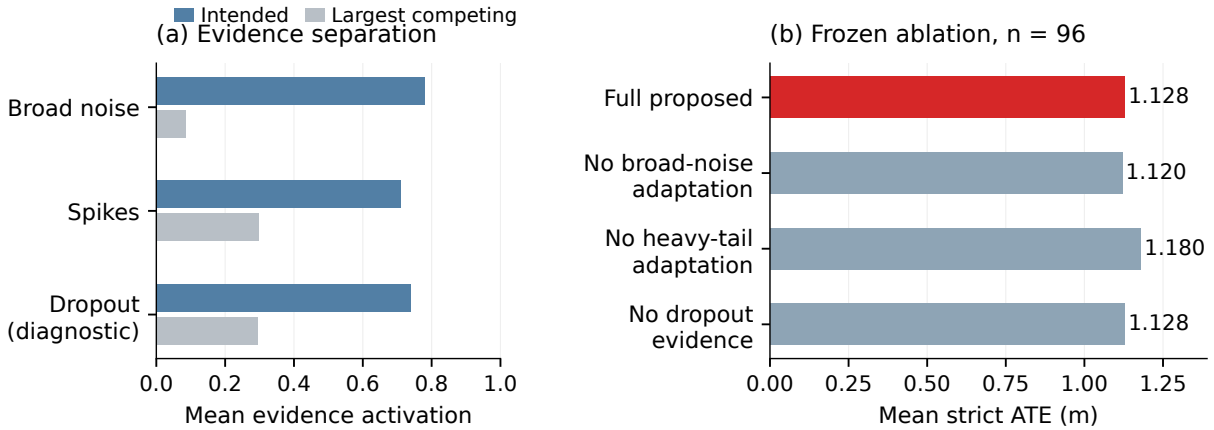


Figure 2: Primary mechanism assessment: (a) mean intended evidence and largest competing activation for each fault pattern; dropout is diagnostic only; (b) mean ATE with individual channels disabled. The no-dropout-evidence variant is identical to the full estimator.

Supplementary Fig. S7 shows the implemented controls on saved cases. Broad noise mainly activates e_N; spikes activate e_T with fluctuating e_N. The robust weight and inverse covariance scale have separate effects. These traces illustrate the frozen mappings, rather than provide new trials or calibrated observation uncertainty.

5.4 Representative trajectories and adverse cases

Figure 3 exposes both a benefit and an operating limitation. The spike example is closest to the median paired Gaussian gain among the 12 L-shape spike cases. The mixed example is closest to median regret among the two severe line mixed cases. Supplementary Section S3 states the complete deterministic selection rules and cache audit. These illustrations supplement, rather than replace, the matched statistics.

In the spike example, proposed ATE is 0.706 m versus Gaussian 0.843 m. In the severe mixed example, proposed ATE is 2.900 m versus Gaussian 2.788 m, and proposed final position error reaches 5.358 m. All four estimators develop substantial drift. The saved states are finite and the solver terminates successfully, but this is a failure of navigation accuracy; successful optimization termination must not be confused with successful tracking. No alignment over the complete trajectory or post-confirmatory tuning was introduced to conceal it.

Trajectory stratification is also informative (Supplementary Fig. S2). Proposed mean ATE is 1.0492 m on L-shape and 1.2067 m on line. Geman is slightly better on L-shape and worse on line, reinforcing the absence of uniform superiority. Across the primary cases, 91.7% of proposed results lie within 10% of the retrospective best fixed baseline; 95th-percentile regret is 18.18% (Supplementary Fig. S1). The bias-only stress condition remains excluded from the 96 primary cases and the novelty claim: a persistent offset is not reliably rejected by the tightly constrained nuisance state.

5.5 Independent TANK external validation

Without adaptive retuning, the proposed estimator gives the lowest external fault mean ATE, 2.8367 m, compared with Gaussian 2.9315, Huber 2.9270 and Geman 2.8806 m (Table 3). Clean methods are closely grouped on Hard, while the proposed method has the lowest clean ATE on Medium. Supplementary Fig. S3 gives the clean comparison.

Table 3: Independent TANK external validation. Clean columns contain one unmodified run per held-out physical sequence; the final two columns summarize the 96 controlled fault realizations across Structure Medium and Structure Hard. Lower ATE is better.

Method	Clean Medium ATE [m]	Clean Hard ATE [m]	Fault mean ATE [m]	Fault median ATE [m]
Gaussian	0.3853	1.0249	2.9315	1.6967
Huber	0.3850	1.0249	2.9270	1.6845
Geman–McClure	0.3851	1.0260	2.8806	1.5666
Proposed	0.3742	1.0248	2.8367	1.5292

Mean paired gains are 3.43%, 3.30% and 1.82% over Gaussian, Huber and Geman, respectively (Table 4). The 12 seed-aggregated bootstrap intervals are positive. Each exact Wilcoxon p-value is 4.88×10^{-4} because all 12 seed-aggregate percentage gains are positive in the respective comparison, giving $2/2^{12}$. This is a sensitivity analysis over the injected fault seeds, not a confirmatory test across independent physical missions; its common sensor backgrounds and potential cross-seed dependencies constrain interpretation.

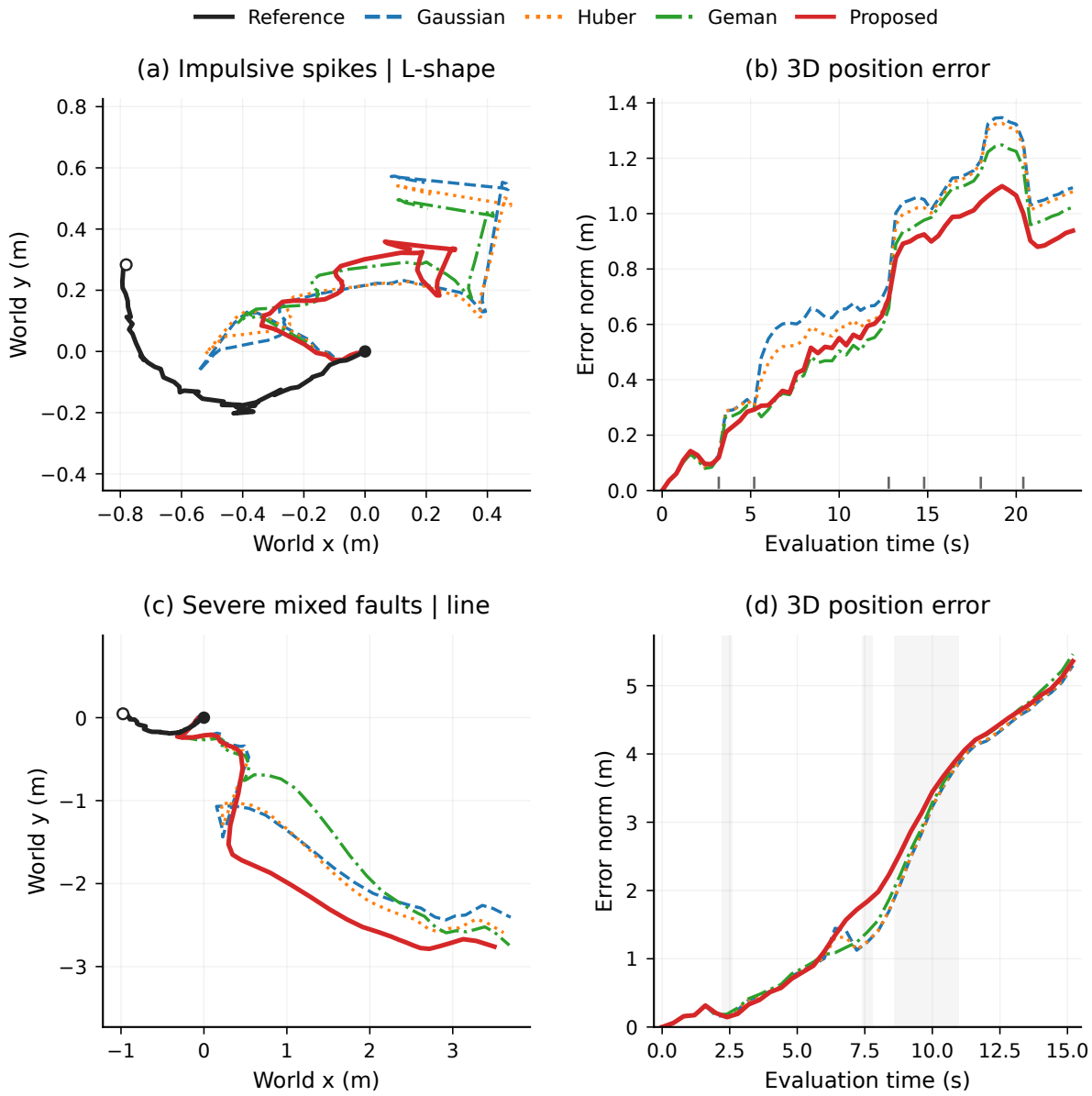


Figure 3: Representative spikes (a,b; L-shape, seed 20292306) and severe mixed degradation (c,d; line, seed 20292302). Left: horizontal trajectory projections; filled/open circles mark the first/last evaluated reference positions. Right: three-dimensional position-error norm. Grey ticks mark spikes; shaded spans mark unavailable DVL factors. Initial-prefix-only yaw/translation alignment is retained. The mixed case shows substantial drift.

Table 4: Paired TANK external-validation summary over 96 matched fault realizations. Confidence intervals and p-values are based on 12 seed-level clusters, each aggregating the two held-out sequences and four fault families. Exact Wilcoxon p-values concern seed-aggregate *percentage gains*; they coincide because all 12 values are positive in each comparison. These seeds do not constitute 12 independently recorded physical missions.

Baseline	Mean paired gain [%]	Seed-cluster 95% CI [%]	Wilcoxon p	Win fraction
Gaussian	3.430	[2.354, 4.717]	4.88×10^{-4}	0.760
Huber	3.295	[2.281, 4.501]	4.88×10^{-4}	0.750
Geman–McClure	1.819	[1.314, 2.318]	4.88×10^{-4}	0.740

The strongest external gain again occurs under spikes (Fig. 4): 11.97% over Gaussian, 11.37% over Huber and 6.09% over Geman. Corresponding spike win fractions are 95.8%, 95.8% and 91.7%. Broad-noise gains range from 0.28% to 0.33%, and dropout gains from 0.54% to 0.74%. Mixed gains are 0.73%, 0.75% and 0.33%; the mixed–Geman seed-level comparison is not significant ($p = 0.092$).

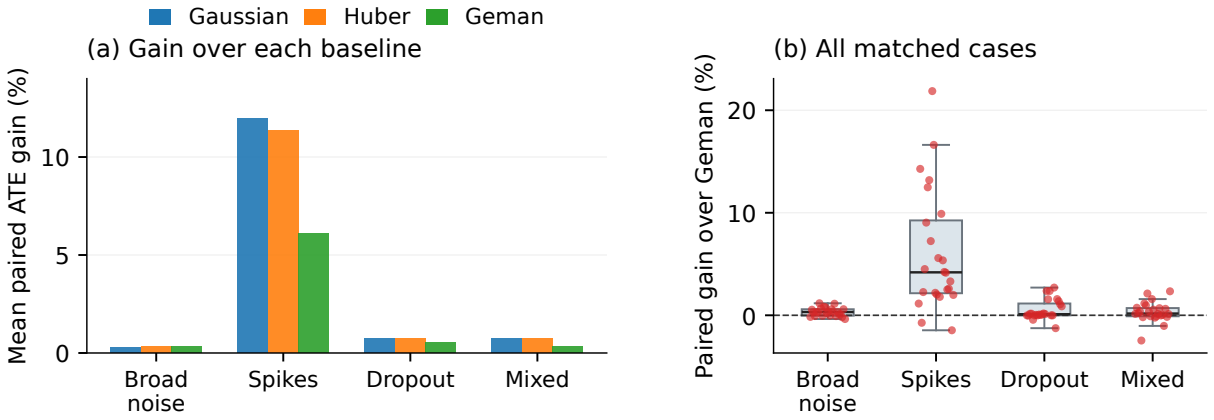


Figure 4: TANK transfer without adaptive retuning: (a) mean paired ATE gain over each fixed baseline; (b) case-wise gains over Geman. Positive gain favors the proposed estimator. Each family has 24 matched realizations (2 held-out sequences ×12 seeds). Boxes show median and interquartile range; points show all cases.

Geman-relative gains are positive on both Medium (1.26%) and Hard (2.38%), as shown in Supplementary Fig. S4. All 384 external solves terminate successfully under the fixed continuation protocol. Recorded runtimes in Supplementary Fig. S5 characterize batch solver calls and do not establish a separately measured end-to-end pilot-plus-adaptive latency or real-time onboard throughput.

6 Discussion

6.1 Supported mechanism and comparison with prior work

The principal supported mechanism is selective suppression of impulsive DVL errors after separating residual scale from standardized tail shape. Sparse large errors can coexist with an otherwise useful velocity stream; increasing its covariance uniformly weakens every observation. The proposed tail statistics instead alter the robust kernel according to the distribution remaining after local scale removal. The strongest fault-family advantage, heavy-tail ablation loss and repetition of the spike benefit on TANK are consistent with this interpretation.

This is an *incremental methodological contribution* in an already active area, rather than the introduction of robust factor graphs, residual feedback, median absolute deviation or adaptive loss functions themselves. Adaptive entropy factors [21, 23], virtual-velocity correction [22] and delayed-marginalization robust graphs [24] already address degraded underwater observations. Our more specific design choice is to use *scale-normalized window-tail evidence* to modulate the exponent and threshold of a DVL weight while a separate

window-scale statistic modifies factor scale. This separation makes the control mechanisms explicit and permits independent ablation, but the two channels are not probabilistically independent and their functional mappings are heuristic, frozen engineering choices rather than optimal or derived uncertainty estimators.

The evidence for the added complexity is also asymmetric. Tail adaptation has a substantive ablation effect; broad-noise scaling does not show a positive aggregate ablation effect. Consequently, the full two-channel architecture should be regarded as an evaluated candidate formulation, not as proof that both adaptation mechanisms are necessary. This limitation narrows the claim of innovation to the *implemented separation and empirically supported tail-control mechanism*.

The comparison is intentionally controlled: all graphs have identical physical factors, and the Huber-style and Geman-type alternatives use one pilot-derived reweighting step. Such fairness in graph structure does not make these alternatives equivalent to converged iteratively reweighted M-estimators or to the latest adaptive FGO methods. No direct comparative evidence is available against those algorithms; neither comparative superiority nor universal optimality is claimed.

The pilot-weighted fixed Geman-type baseline remains competitive on the primary data, and its paired difference from the proposed estimator is not statistically significant. TANK adds a positive transfer result, especially under spikes, but its smaller mixed-fault gains limit any claim of uniform dominance. The retrospective regret statistic further characterizes proximity to the best fixed choice; that choice remains unavailable during deployment. Clean accuracy is preserved descriptively, rather than proven equivalent by an equivalence test.

The broad-noise finding is not a positive validation of the proposed covariance-scaling channel: its evidence identifies dispersion changes, but removing that channel improves mean primary ATE from 1.1280 to 1.1202 m. Increasing factor scale could reduce the influence of useful as well as corrupted DVL measurements; this is a plausible interpretation, not a causally isolated diagnosis. We retain the full frozen mapping to report the tested method transparently rather than selecting a post-hoc winner. Thus, neither a two-channel necessity claim nor reliable noise-covariance calibration follows from these ablations. Dropout handling comes exclusively from factor omission and inertial propagation shared by all methods; the missingness diagnostic does not improve the optimizer.

6.2 Accuracy, uncertainty and physical consistency

Accuracy is measured against external reference trajectories in metres and metres per second after initial-prefix-only alignment. Shared physical factors and matched degradations isolate the statistical DVL treatment. Frame direction, mounting geometry and timing remain essential: residual weighting cannot resolve an incorrectly modeled sensor point or coordinate system. The lever-arm term belongs in the predicted DVL-point velocity, while body-origin compensation is initialization-only. Calibration and explicit reference-center modeling in related measurement studies [28, 30] support this modeling requirement.

The uncertainty quantities have different meanings. Bootstrap intervals describe variation in aggregate errors or paired gains over the specified perturbations. MAD measures posterior

pilot-residual scatter; covariance scaling and robust weights regulate optimization influence. Because the pilot already uses DVL observations, its residual scale is not an unbiased measurement-error covariance estimate. Neither these statistics nor the reported accuracy establish calibrated uncertainty coverage for individual observations or navigation states.

Reference-system and extrinsic-calibration uncertainties are not propagated into a complete uncertainty budget. The IMU factors use conservative diagonal scales, and a full preintegration covariance with suitable Jacobians and residual-consistency checks remains an important extension [14]. The adverse broad-noise ablation result provides a concrete reason to examine uncertainty-consistent scaling before adding more heuristic evidence channels.

6.3 Operating limits and validation scope

Persistent DVL offset is not a validated capability. Sensor-only residuals can confound sustained offsets with initialization errors, model mismatch and inertial drift; restricted motion also limits observability [33]. The tightly regularized nuisance state does not solve that ambiguity. Severe mixed corruption combines offset, missing information and abnormal dispersion/tails, and the retained trajectory example shows substantial navigation drift despite successful optimizer termination. Robust weighting limits a corrupted observation's influence; it cannot create an independent velocity or position constraint when the remaining information is inadequate.

Both datasets contain real motion and sensor backgrounds, but the repeated fault episodes are injected. Their distributions cannot reproduce all operational beam failures, acoustic effects or environmental dependencies. The confirmatory freeze and external transfer reduce tuning bias; they do not expand physical diversity beyond the recorded sequences. TANK's Structure restriction prioritizes direct reference coverage and avoids SLAM-augmented reference intervals, while excluding five released sequences from the navigation evaluation. Seed-cluster sensitivity quantifies repeated fault variability on Medium and Hard, not population-level performance across vehicles, depths or seafloor conditions. Naturally occurring failures over more independent missions are needed for broader validation.

The solver is a CPU-oriented two-stage batch implementation. A trailing evidence window does not make a batch pilot causal, and continuation increases numerical effort without establishing real-time execution. The recorded runtime fields do not isolate complete pilot-plus-adaptive latency, and the available files lack enough hardware detail for a normalized timing benchmark. Incremental or fixed-lag implementations require separate accuracy and timing evaluation. Full cached-state reconstruction checks the consistency of the plotted diagnostics with the stored experiment; it is not an additional physical validation.

The resulting research priorities are focused: validate natural faults across more missions; propagate and assess uncertainty consistently; evaluate a causal incremental implementation; and introduce independently calibrated exteroceptive factors where persistent offsets or long outages need additional observability. Sonar is a possible extension, but its exclusion here preserves the interpretability of the IMU–DVL comparison.

7 Conclusions

A pilot-residual-driven IMU–DVL factor graph was developed with separate controls for residual dispersion and standardized tail shape. It retains a calibrated SO(3) physical model, predicts velocity at the DVL point and omits unavailable factors without reconstructing measurements or classifying faults.

Under the fixed pilot-weighted comparison protocol, the primary study gives mean strict ATE 1.128 m and mean paired percentage gains of 4.27% over Gaussian and 3.76% over the Huber-style baseline, with no significant absolute paired-error difference from the Geman-type baseline. Ablation supports standardized-tail adaptation as the strongest individual component, while broad-noise scaling has no demonstrated aggregate benefit. With its parameters unchanged, the method yields mean ATE 2.837 m on the separate TANK perturbation set and a 6.09% paired gain over the pilot-weighted Geman-type baseline in its impulsive-error subset.

The most defensible novelty is therefore a reproducibly specified pilot-residual scale–shape weighting design with an empirically supported tail-sensitive component, evaluated across two distinct recorded datasets without adaptive retuning. This is not evidence of a generally optimal robust estimator, superior performance to fully iterative or recently published adaptive methods, an independently beneficial broad-noise channel, adaptive dropout weighting, persistent-bias rejection, real-time execution or calibrated state uncertainty. Severe mixed drift and the small number of independent physical trajectories remain substantive limits. Supplementary Information provides physical-model details, fixed settings, supporting results and reconstruction checks.

Author contributions

S.T.: Conceptualization, Methodology, Supervision, Writing–review and editing. S.M.M.: Conceptualization, Software, Validation, Data curation, Visualization, Writing–original draft. Both authors are responsible for verifying and approving the final version before submission.

Funding

This research received no specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability

The primary Sonar-Based UUV dataset is publicly available through the cited project repository [32]. The original TANK dataset [31] is distributed by Sense Robotics Lab subject to its access request and licence terms; its raw recordings are not redistributed here. Supplementary data contain derived primary and TANK evaluation tables, paired statistics, severity and ablation summaries, convergence records, and selected primary traces. These

are derived evaluation products, not raw third-party recordings. Further availability of author-generated files may be discussed with the corresponding author, subject to applicable third-party conditions.

Code availability

The author project contains the cached-primary-state reconstruction script and its accompanying verification outputs. Implementation and reproduction requests may be directed to the corresponding author; access to any third-party input streams remains governed by the original dataset providers' terms.

Competing interests

The authors declare no competing interests.

References

[1] Zhang, B., Ji, D., Liu, S., Zhu, X. & Xu, W. Autonomous underwater vehicle navigation: A review. *Ocean Eng.* **273**, 113861 (2023).

[2] Liu, S., Zhang, T., Zhang, J. & Zhu, Y. A new coupled method of SINS/DVL integrated navigation based on improved dual adaptive factors. *IEEE Trans. Instrum. Meas.* **70**, 1–11 (2021).

[3] Qian, L., Qin, F., Li, A., Li, K. & Zhu, J. An INS/DVL integrated navigation filtering method against complex underwater environment. *Ocean Eng.* **278**, 114398 (2023).

[4] Du, S., Zhu, F., Wang, Z., Huang, Y. & Zhang, Y. A novel lie group framework-based student's t robust filter and its application to INS/DVL tightly integrated navigation. *IEEE Trans. Instrum. Meas.* **73**, 1–21 (2024).

[5] Kang, L., He, K., Zhao, J., Wang, X. & Tan, P. A hybrid-kernel-based adaptive robust kalman filter for INS/DVL integrated underwater navigation. *Ocean Eng.* **350**, 124269 (2026).

[6] Davari, N., Maurelli, F. & Aguiar, A. P. A robust navigation system for underwater vehicles using an adaptive H_∞ filter based on a variational bayesian approach: Development and experimental evaluation. *Ocean Eng.* **354**, 124719 (2026).

[7] Miao, Y. *et al.* Physics-guided adaptive ukf for robust auv integrated navigation under degraded underwater observations. *Ocean Eng.* **362**, 126542 (2026).

[8] Davari, N. & Aguiar, A. P. Real-time outlier detection applied to a DVL sensor based on hybrid autoencoder and RNN. *IEEE J. Ocean. Eng.* **46**, 1288–1301 (2021).

[9] Klein, I., Gutnik, Y. & Lipman, Y. Estimating DVL velocity in complete beam measurement outage scenarios. *IEEE Sens. J.* **22**, 20730–20737 (2022).

[10] Cohen, N. & Klein, I. Beamsnet: A data-driven approach enhancing doppler velocity log measurements for AUV navigation. *Eng. Appl. Artif. Intell.* **114**, 105216 (2022).

[11] Topini, E. *et al.* An experimental comparison of deep learning strategies for AUV navigation in DVL-denied environments. *Ocean Eng.* **274**, 114034 (2023).

[12] Jin, K., Chai, H., Su, C. & Xiang, M. A performance-enhanced DVL/SINS integrated navigation system based on data-driven approach. *Meas. Sci. Technol.* **34**, 095120 (2023).

[13] Yan, X. *et al.* Tcn-ftiakf: A fault-tolerant tightly coupled improved adaptive kalman filtering algorithm for sins/dvl integrated navigation. *Measurement* **268**, 120577 (2026).

[14] Forster, C., Carlone, L., Dellaert, F. & Scaramuzza, D. On-manifold preintegration for real-time visual–inertial odometry. *IEEE Trans. Robot.* **33**, 1–21 (2017).

[15] Wen, W., Pfeifer, T., Bai, X. & Hsu, L.-T. Factor graph optimization for GNSS/INS integration: A comparison with EKF. *NAVIGATION* **68**, 315–331 (2021).

[16] Taylor, C. & Gross, J. N. Factor graphs for navigation applications: A tutorial. *NAVIGATION: Journal of the Institute of Navigation* **71**, navi.653 (2024).

[17] Qin, H. *et al.* A novel INS/USBL/DVL integrated navigation scheme against complex underwater environment. *Ocean Eng.* **286**, 115485 (2023).

[18] Dong, X., Hu, G., Gao, B., Zhong, Y. & Ruan, W. Windowing-based factor graph optimization with anomaly detection using mahalanobis distance for underwater INS/DVL/USBL integration. *IEEE Trans. Instrum. Meas.* **73**, 1–13 (2024).

[19] Du, X., Pang, X., Guan, F., Hu, J. & Zhang, W. A novel multi-source navigation algorithm based on factor graph in complex underwater environments of polar regions. *Ocean Eng.* **301**, 117516 (2024).

[20] Zhang, L., Zhang, T. & Wei, H. A factor graph optimization-based in-motion alignment method for INS/DVL integration. *IEEE Trans. Veh. Technol.* **73**, 18452–18459 (2024).

[21] Cheng, S., Wang, Y., Zhao, Q., Zhu, H. & Qu, X. A robust INS/USBL/DVL integrated navigation method based on adaptive correlation entropy factor graph optimization. *Ocean Eng.* **356**, 125234 (2026).

[22] Liu, R. *et al.* VVFGO: A robust IMU/AHRS/USBL integrated navigation framework with adaptive virtual velocity factor correction for deep-sea docking. *Ocean Eng.* **358**, 125753 (2026).

[23] Cheng, S., Wang, Y., Sun, B., Zhu, H. & Qu, X. Deep learning-enhanced adaptive correntropy factor graph optimization for robust UUV navigation. *Ocean Eng.* **363**, 126690 (2026).

[24] Guo, S. *et al.* An adaptive and robust factor graph optimization with delayed marginalization for ins/dvl/ps integrated navigation. *Ocean Eng.* **366**, 127594 (2026).

[25] Ramezani, M., Mattamala, M. & Fallon, M. AEROS: Adaptive robust least-squares for graph-based SLAM. *Front. Robot. AI* **9**, 789444 (2022).

[26] Das, S. & Gross, J. N. Analysis of scale-variant robust kernel optimization for nonlinear least-squares problems. *IEEE Trans. Aerosp. Electron. Syst.* **59**, 7529–7538 (2023).

[27] Guo, M. *et al.* Robust INS/GNSS/DVL integrated navigation for MASS based on gradient-adaptive FGO. *Electronics* **15**, 634 (2026).

[28] Cao, S. *et al.* Joint calibration of DVL misalignment and anisotropic scale factors using lie algebra-based adaptive optimization. *Ocean Eng.* **364**, 126969 (2026).

[29] Zhuang, J., Cao, Z., Wang, Q. & Sun, D. Research on sins/dvl integrated navigation method assisted by acoustic doppler delay velocity information. *Ocean Eng.* **342**, 122943 (2025).

[30] Song, Z., Gao, F., Liu, Y., Xu, T. & Zhang, S. Improved AUV navigation algorithm fusing SINS/DVL/PS and considering lever arm errors. *Meas. Sci. Technol.* **36**, 066317 (2025).

[31] Xu, S. *et al.* Tank dataset: An underwater multi-sensor dataset for slam evaluation. *Int. J. Robot. Res.* **45**, 541–551 (2026). First published online 30 August 2025.

[32] Sonar-Based UUV DataSet Project. Sonar-based uuv dataset: Bluerov2 indoor underwater dataset. `https://github.com/hwgao1101/Sonar_Based_UUV_DataSet` (2026). Public dataset repository; accessed September 2026.

[33] Sheng, G. *et al.* The EKF-based SINS/DVL integrated navigation for AUV on lie group under hovering condition. *Ocean Eng.* **325**, 120742 (2025).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SRSupplementaryRevisedv2.pdf