Supplementary Information

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

Sobhan Toulabi and Seyyede Marzieh Mousavi*

*Correspondence: mmousavi9429@aut.ac.ir

Title: Supplementary Information

Description: Physical models, frozen estimator settings, supporting navigation results, diagnostic reconstruction and external-validation solver records for pilot-residual-driven IMU–DVL fusion.

The reported comparisons retain the frozen estimator; diagnostic reconstruction introduces no new optimization or tuning.

The Huber-style ($c = 1.8$) and generalized Geman–McClure-type ($c = 2.5$, $p = 2.0$) comparators use weights calculated once from the same Gaussian-pilot posterior DVL residuals and then held fixed during a single subsequent solve. They are not converged iteratively reweighted M-estimators. The proposed estimator also freezes its pilot-derived weights, with additional scale- and tail-dependent mappings. The principal source of comparison here is the shared physical graph and fixed-pilot-weighting protocol; comparisons with fully iterative or other recently proposed adaptive solvers were not performed.

S1 Physical model and sensor interfaces

S1.1 Frame conventions and inertial dynamics

World, body, IMU and DVL frames are W, B, M, D. The homogeneous transform $\mathbf{T}_{AB} = [\mathbf{R}_{AB}, \mathbf{t}_{AB}; \mathbf{0}^{\mathsf{T}}, 1]$ maps coordinates from B to A: $\mathbf{p}^A = \mathbf{R}_{AB}\mathbf{p}^B + \mathbf{t}_{AB}$. Attitude $\mathbf{R}_{WB}$ maps body vectors to world coordinates; its inverse is $\mathbf{R}_{BW} = \mathbf{R}_{WB}^{\mathsf{T}}$. The state uses body-origin position and velocity, attitude, accelerometer and gyroscope biases, and a tightly constrained DVL-offset nuisance state. The implemented 18-parameter vector uses three attitude rotation parameters. Exponential and logarithmic maps are on SO(3) [1, 2].

After rotating IMU observations into the body frame, the gyroscope measures angular rate plus bias/noise and the accelerometer measures specific force. The continuous dynamics

are

$$\dot{\mathbf{p}}^W = \mathbf{v}^W, \tag{S1}$$
$$\dot{\mathbf{v}}^W = \mathbf{R}_{WB}(\tilde{\mathbf{f}} - \mathbf{b}_a - \mathbf{n}_a) + \mathbf{g}^W, \tag{S2}$$
$$\dot{\mathbf{R}}_{WB} = \mathbf{R}_{WB}[\tilde{\boldsymbol{\omega}} - \mathbf{b}_g - \mathbf{n}_g]_\times, \tag{S3}$$
$$\dot{\mathbf{b}}_a = \mathbf{n}_{wa}, \qquad \dot{\mathbf{b}}_g = \mathbf{n}_{wg}. \tag{S4}$$

Here $[\mathbf{a}]_\times \mathbf{b} = \mathbf{a} \times \mathbf{b}$; $\mathbf{g}^W$ is the frozen gravity convention, and the $\mathbf{n}$ terms denote measurement and bias-process noise. Frame and specific-force interpretations are audited before validation, rather than inferred from faulted held-out data.

S1.2 Preintegration and graph residuals

For raw sub-interval Δt_n between keyframes i, j, the basic preintegration recursions are [1]

$$\Delta\mathbf{R}_{n+1} = \Delta\mathbf{R}_n \operatorname{Exp}[(\tilde{\boldsymbol{\omega}}_n - \mathbf{b}_{g,i})\Delta t_n], \tag{S5}$$
$$\Delta\mathbf{v}_{n+1} = \Delta\mathbf{v}_n + \Delta\mathbf{R}_n(\tilde{\mathbf{f}}_n - \mathbf{b}_{a,i})\Delta t_n, \tag{S6}$$
$$\Delta\mathbf{p}_{n+1} = \Delta\mathbf{p}_n + \Delta\mathbf{v}_n\Delta t_n + \tfrac{1}{2}\Delta\mathbf{R}_n(\tilde{\mathbf{f}}_n - \mathbf{b}_{a,i})\Delta t_n^2, \tag{S7}$$

initialized with $\Delta\mathbf{R}_i = \mathbf{I}$ and $\Delta\mathbf{v}_i = \Delta\mathbf{p}_i = \mathbf{0}$. In the actual implementation, angular rate and specific force are midpoint-averaged on each sub-interval, and the force uses the half-step attitude. The raw sub-intervals are retained and increments are recomputed at each candidate initial bias during nonlinear residual evaluation. There is no frozen-bias first-order approximation around one preintegration point.

For $\Delta t_{ij} = t_j - t_i$, the IMU factor stacks

$$\mathbf{r}_{R,ij} = \operatorname{Log}(\Delta\mathbf{R}_{ij}^\mathsf{T}\mathbf{R}_{WB,i}^\mathsf{T}\mathbf{R}_{WB,j}), \tag{S8}$$
$$\mathbf{r}_{v,ij} = \mathbf{R}_{WB,i}^\mathsf{T}(\mathbf{v}_j^W - \mathbf{v}_i^W - \mathbf{g}^W\Delta t_{ij}) - \Delta\mathbf{v}_{ij}, \tag{S9}$$
$$\mathbf{r}_{p,ij} = \mathbf{R}_{WB,i}^\mathsf{T}(\mathbf{p}_j^W - \mathbf{p}_i^W - \mathbf{v}_i^W\Delta t_{ij} - \tfrac{1}{2}\mathbf{g}^W\Delta t_{ij}^2) - \Delta\mathbf{p}_{ij}. \tag{S10}$$

Adjacent bias residuals are $\mathbf{b}_{a,j} - \mathbf{b}_{a,i}$ and $\mathbf{b}_{g,j} - \mathbf{b}_{g,i}$, whitened by $\sigma_{b_a,rw}\sqrt{\max(\Delta t_{ij}, 10^{-6}\text{ s})}$ and $\sigma_{b_g,rw}\sqrt{\max(\Delta t_{ij}, 10^{-6}\text{ s})}$, respectively. Initial priors, these preintegration terms and IMU-bias evolution form the common physical objective. The DVL nuisance terms use the fixed zero-centered and adjacent-difference whitenings specified in the main paper. They enter as squared whitened residuals. Table S1 gives the diagonal factor scales; full preintegration covariance is not propagated.

S1.3 Primary sensor transforms and initialization

The official primary dataset transforms [3] are

$$\mathbf{T}_{BM} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \qquad \mathbf{T}_{BD} = \begin{bmatrix} -1 & 0 & 0 & -0.28 \\ 0 & -1 & 0 & -0.10 \\ 0 & 0 & 1 & 0.12 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{S11}$$

The body-frame DVL lever arm is $\mathbf{r}_{BD}^{B} = [-0.28, -0.10, 0.12]^{\mathsf{T}}$ m. The factor measurement is only rotated: $\tilde{\mathbf{v}}_D^B = \mathbf{R}_{BD}\tilde{\mathbf{v}}_D^D$. No rotational lever-arm contribution is subtracted from it. The prediction adds $(\tilde{\boldsymbol{\omega}} - \mathbf{b}_g) \times \mathbf{r}_{BD}^{B}$ to $\mathbf{R}_{BW}\mathbf{v}^W$, with the nuisance offset.

For initialization only, an approximate body-origin velocity is

$$\mathbf{v}_{B,\text{org}}^{\text{init}} = \tilde{\mathbf{v}}_D^B - \boldsymbol{\omega}^B \times \mathbf{r}_{BD}^{B}. \tag{S12}$$

This quantity is not inserted as the factor observation; measured and predicted factor velocities remain at the DVL point. Scale is unity, lag is zero, and graph rate is 2.5 Hz. Raw IMU samples are retained between keyframes. DVL altitude is not treated as vehicle depth, and no motion-capture or pseudo-depth factor is introduced. Onboard quaternions support initialization only.

S1.4 TANK physical adapter and reference restriction

The TANK body origin is the IMU origin. Raw calibration matrices are interpreted as Camera→IMU and Camera→DVL. With $\mathbf{T}_{CI}$ denoting IMU→Camera and $\mathbf{T}_{CD}$ DVL→Camera, the adapter uses

$$\mathbf{T}_{ID} = \mathbf{T}_{CI}^{-1}\mathbf{T}_{CD}, \qquad \mathbf{r}_{ID} = [-0.075797, 0.140476, -0.091004]^{\mathsf{T}} \text{ m}, \tag{S13}$$

with $\|\mathbf{r}_{ID}\|_2 = 0.18374$ m. DVL sign is +1, accelerometer readings are specific force, lag is zero, and natural invalid samples remain invalid. The raw reference pose is Camera→Global, converted to the IMU/body origin for evaluation only. Structure Easy audits these conventions without retuning any adaptive reference or mapping.

The benchmark [4] releases Structure Easy/Medium/Hard, HalfTank Easy/Medium/Hard and WholeTank Medium/Hard. Reported direct AprilTag/TankGT coverage for Structure is 99.37%, 99.67% and 84.65%, versus approximately 18.97%–60.73% for HalfTank and WholeTank. The Structure restriction is based on reference continuity, not estimator ranking. The partially referenced trajectories and `/apriltag_slam/GT_full` are not used: the latter supplements TankGT using AQUA-SLAM rather than providing an entirely direct reference. This protects the reference interpretation while limiting physical diversity.

S2 Frozen settings and fault generation

S2.1 Physical factors and adaptive references

The settings in Table S1 apply to the primary study and TANK transfer after the dataset-specific physical adapter. They are frozen on primary circle development data. Structure Easy audits only the TANK interface; Medium and Hard do not tune either physical or adaptive settings. Solver budgets are termination controls, documented in Section S4.

Table S1: Frozen estimator parameters shared by the primary confirmatory study and the TANK external transfer after dataset-specific physical-frame adaptation. Numerical nonlinear-solver stopping budgets are reported separately in the text because they are implementation-level termination controls rather than estimator parameters.

Parameter	Value	Unit
Keyframe rate	2.5	Hz
IMU–DVL lag	0	s
DVL scale	1.0	–
Initial position prior σ_{p_0}	0.02	m
Initial velocity prior σ_{v_0}	0.15	m/s
Initial attitude prior σ_{q_0}	0.06	rad
Initial accelerometer-bias prior $\sigma_{b_{a,0}}$	0.45	$\mathrm{m/s^2}$
Initial gyroscope-bias prior $\sigma_{b_{g,0}}$	0.04	rad/s
IMU position-factor scale	0.10	m
IMU velocity-factor scale	0.20	m/s
IMU rotation-factor scale	0.055	rad
Accelerometer-bias random walk	0.08	$\mathrm{m/s^2}/\sqrt{\mathrm{s}}$
Gyroscope-bias random walk	0.010	$\mathrm{rad/s}/\sqrt{\mathrm{s}}$
Nominal DVL factor scale σ_{DVL}	0.07	m/s
DVL-offset zero prior $\sigma_{b_D,0}$	0.010	m/s
DVL-offset random walk $\sigma_{b_D,rw}$	0.004	m/s

The evidence window is 8 s, with at least five valid residuals and a component MAD floor 10^{-4} m/s. Clean reference values used by the code, before presentation rounding, are

$$s_0 = 0.0070746275889730405 \text{ m/s}, \tag{S14}$$

$$f_{T,0} = 0.05263157894736842, \tag{S15}$$

$$r_{T,0} = 2.5476985724752854. \tag{S16}$$

The standardized-tail threshold is 3.6; the quantile-ratio denominator floor is 10^{-6}. The main paper gives the complete mappings: noise normalization $\ln(4.5)$, multiplier $1 + 3e_N$ clipped to $[1, 4]$, tail fraction target 0.18 with denominator floor 0.05, tail ratio gain 2.5 and relative coefficient 0.75, kernel $c = 4 - 3e_T$, $p = 1 + 1.2e_T$, and weight clipping to $[0.01, 1]$. Evidence is zero before the minimum valid window is reached. The final solve holds pilot-derived scales and weights fixed while recomputing the physical prediction.

Let m_k be the missing fraction of graph keyframes in the current window and ℓ_k the

longest missing run divided by its keyframe count. The diagnostic is

$$e_{D,k} = \mathrm{clip}_{[0,1]} \left[\max \left(m_k/0.40, \ell_k/0.30 \right) \right]. \tag{S17}$$

It does not enter the effective covariance, kernel parameters or robust weights. Missing factors are omitted in every estimator. The nuisance offset remains tightly constrained at the same fixed scales; there is no adaptive persistent-bias detector.

S2.2 Recorded-stream perturbations

Synthetic faults modify real recorded DVL velocities, leaving IMU, reference poses and timing unchanged. Table S2 gives all ranges. Amplitudes refer to the nominal DVL factor scale 0.07 m/s.

Table S2: Ranges used to generate controlled DVL degradations from the recorded real DVL measurements. All multiplicative amplitude factors are relative to $\sigma_{\mathrm{DVL}} = 0.07$ m/s.

Scenario	Noise factor	Spike process	Dropout process	Persistent offset
Broad noise	0.8–4.5	–	–	–
Spikes	–	rate 0.015–0.15, amplitude 4–20	–	–
Dropout	–	–	random fraction 0.05–0.45, burst 0.4–4.0 s	–
Mixed	0.4–3.0	rate 0.01–0.12, amplitude 4–16	random fraction 0.03–0.30, burst 0.4–2.5 s	0.3–3.5
Bias stress test	–	–	–	0.7–5.0

Broad noise adds zero-mean Gaussian perturbations to each velocity component. Spikes contaminate randomly selected observations with zero-mean impulsive disturbances. Dropout combines random loss and a contiguous burst. No missing sample is interpolated or reconstructed. Mixed cases combine noise, spikes, random/burst loss and an unmodeled constant offset; the separate bias-only condition is outside the primary novelty claim.

Each sampled parameter is linearly normalized within its range. Single-fault severity averages the relevant normalized rate and amplitude or duration terms when two define a fault; mixed severity averages its normalized components. Scores in $[0, 1/3)$, $[1/3, 2/3)$ and $[2/3, 1]$ define mild, moderate and severe. Subset sizes are determined by sampled realizations. Confirmatory seeds 20292301–20292312 are unseen during development. For each dataset, 2 held-out sequences $\times 4$ families $\times 12$ seeds produce 96 matched cases. Clean runs and bias-only stress tests are separate. TANK preserves natural invalidity before adding any controlled fault.

S3 Supporting primary results and diagnostic reconstruction

S3.1 Clean accuracy, complete paired statistics and ablation

Table S3 averages one unmodified run on each held-out primary trajectory. Table S4 contains the full paired statistics over the 96 primary fault cases, including effect sizes; Table S5 reports frozen channel removals. Positive paired improvement favors the proposed method; positive ablation degradation indicates worse ATE than the full method. These conditional results do not constitute equivalence tests or independent-mission inference.

Table S3: Clean-data performance on the held-out validation trajectories.

Method	Strict ATE RMSE [m]	Velocity RMSE [m/s]
Gaussian	0.9235	0.1899
Huber	0.9234	0.1899
Geman–McClure	**0.9197**	0.1899
Proposed	0.9226	**0.1887**

Table S4: Paired confirmatory comparison of the proposed estimator with fixed baselines over all 96 primary fault realizations. Positive improvement indicates lower ATE for the proposed estimator.

Baseline	Mean paired improvement [%]	Bootstrap 95% CI [%]	Wilcoxon p	Rank-biserial	Cohen's d_z	Win fraction
Gaussian	4.268	[1.578, 7.215]	0.0110	0.299	0.322	0.583
Huber	3.764	[1.208, 6.578]	0.0162	0.283	0.310	0.583
Geman–McClure	-0.624	[-3.549, 1.871]	0.3402	0.112	0.033	0.490

Table S5: Confirmatory ablation results. Positive degradation indicates worse performance than the full proposed method.

Variant	Mean ATE [m]	Degradation [%]
Proposed	1.1280	0.000
No broad-noise adaptation	**1.1202**	-0.690
No heavy-tail adaptation	1.1795	4.567
No dropout evidence	1.1280	0.000

S3.2 Severity, trajectory stratification and regret

For mild spikes, proposed mean ATE is 1.0109 m, versus Gaussian 1.0571 and Geman 1.0464 m. Under moderate spikes it is 0.9178 m, versus Gaussian 1.0686, Huber 1.0505 and Geman 0.9679 m. Severe-spike values are 0.9855, 1.4762, 1.3964 and 1.0033 m for proposed, Gaussian, Huber and Geman, respectively. Different severity groups contain different sampled

cases; non-monotonicity does not describe the response of one fixed trajectory to increasing amplitude.

Under moderate broad noise, proposed ATE is 1.1146 m versus approximately 1.14 m for fixed baselines; under severe broad noise it is 0.8717 versus Gaussian 0.8294 m. Dropout remains closely grouped. Proposed mixed ATE is 1.2574 and 1.4418 m for mild and moderate subsets, respectively, but rises to 2.9573 m for severe mixed degradation, versus Gaussian 2.8463 and Geman 2.7332 m. The supplementary data archive contains all severity-stratified estimates and subset sizes.

Figure S1 shows per-case regret relative to the retrospectively best of Gaussian, Huber and Geman. The proposed estimator lies within 10% of that fixed-baseline oracle in 91.7% of the primary cases, with 95th-percentile regret 18.18%. The oracle choice is not an available online method.

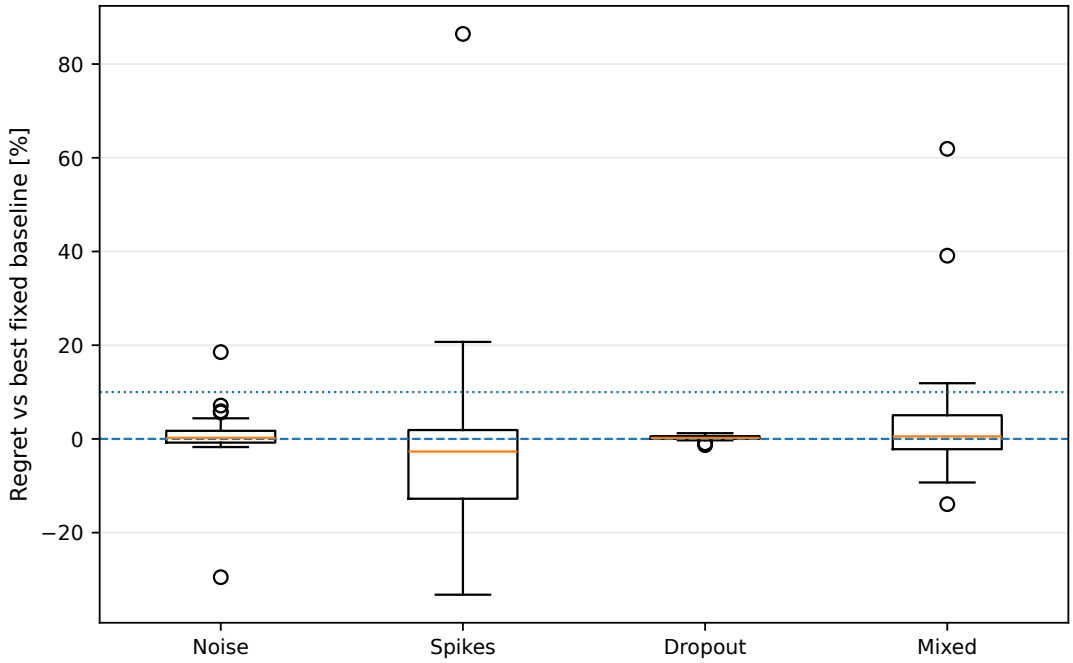


Figure S1: Primary case-wise regret relative to the retrospective best fixed baseline. Negative values favor the proposed estimator. Boxes show median and interquartile range, whiskers extend to observations within 1.5 interquartile ranges, and circles show observations beyond the whiskers. Dashed zero and dotted 10% lines are reference levels. Each family has 24 matched realizations.

On L-shape, mean ATE is proposed 1.0492, Gaussian 1.1288, Huber 1.1148 and Geman 1.0423 m. On line it is proposed 1.2067, Gaussian 1.2635, Huber 1.2573 and Geman 1.2215 m. Figure S2 retains the stratification. The proposed–Geman ordering reverses between trajectories, consistent with the conservative primary paired conclusion.

The bias-only stress condition is not included in the primary 96 cases. Development did not establish reliable persistent-offset detection from the available sensor residuals. The nuisance state therefore remains tightly regularized. Bias-only output records are supplied for transparent stress characterization, not as a validated rejection capability.

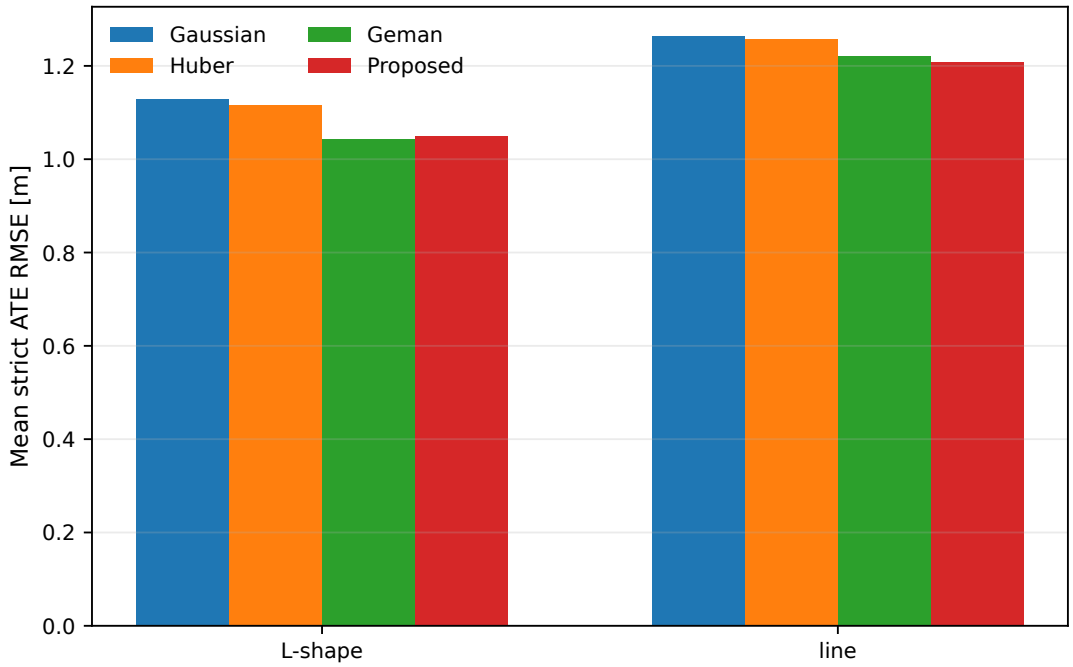


Figure S2: Primary trajectory-stratified mean strict ATE. Each bar aggregates four fault families and 12 seeds per family on a held-out sequence, with one common frozen estimator configuration.

S3.3 Selection rules and cached-state audit

Main Figure 3 and Supplementary Figure S7 use saved states rather than rerun experiments. The spike example is L-shape seed 20292306, chosen closest to the median paired Gaussian percentage gain among 12 L-shape spike cases. The selected gain is 16.25% and the subset median is 18.99%. The severe mixed example is line seed 20292302, chosen closest to median regret among the two severe line mixed cases; selected regret is 4.03% and the median is 4.45%. The broad-noise diagnostic is L-shape seed 20292301, the median-severity case among three severe L-shape noise cases. Ties use seed order.

The spike trace contains 59 evaluated keyframes over 23.2 s, with proposed ATE 0.705999 m and final error 0.938275 m; Gaussian ATE is 0.843031 m and final error 1.093707 m. The mixed trace contains 39 keyframes over 15.2 s, with proposed ATE 2.900208 m and final error 5.358250 m; Gaussian ATE is 2.787749 m and final error 5.296600 m. This mixed example represents substantial navigation drift. Finite states and successful solver termination do not establish acceptable accuracy.

The portable reconstruction matched all 806 cached primary state vectors to their original result rows and recomputed eight position/velocity error summaries. The maximum absolute discrepancy was below 5.37×10^{-14}. Six adaptive-summary fields for proposed cases matched below 10^{-15}. The original sensor transformations, fault generation, keyframe timing, reference preprocessing and initial-prefix alignment were retained. No optimizer was run and no parameters were tuned. This audit verifies internal reproducibility of the stored experiment, not an independent trial. The supplementary archive contains derived tables and selected time traces; the editable author project contains the original source, cached-input ZIP, reconstruction script and complete audit outputs.

S4 External TANK validation and solver records

S4.1 Clean and sequence-specific transfer

The adaptive settings remain those frozen on primary development data. Structure Easy audits only the physical interface; Medium and Hard are held out. Figures S3 and S4 complement the main-paper aggregate results. The 96 fault realizations are repeated perturbations of two physical sequences. No TANK sequence retunes the residual references or mappings.

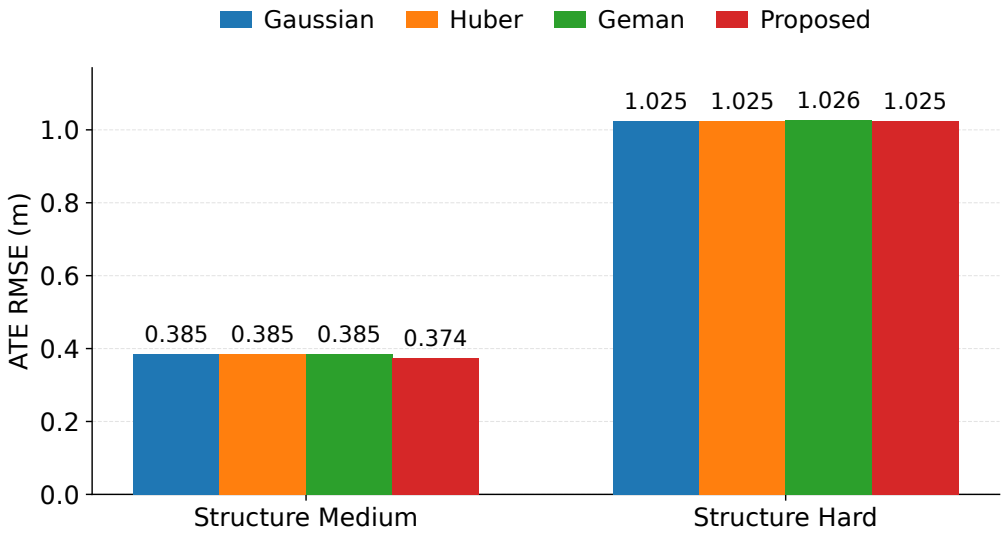


Figure S3: Clean TANK ATE on the two held-out Structure sequences. Proposed ATE is lowest on Medium; differences on Hard are small. Each method has one unmodified run per sequence.

Relative to Geman, proposed paired gains are 1.26% on Medium and 2.38% on Hard; Gaussian-relative gains are 2.52% and 4.34%. The positive Geman-relative transfer occurs on both held-out sequences. Family-specific gains remain uneven, with spikes strongest and the seed-level mixed–Geman comparison non-significant ($p = 0.092$).

S4.2 Continuation, numerical termination and recorded timing

The frozen primary runner allows 45 nonlinear evaluations for Gaussian and 35 for subsequent robust solves. The recorded primary Gaussian runs use at most 18 evaluations, and proposed runs at most 17; all recorded primary method solutions terminate successfully below their respective caps. External TANK continuation uses caps $45 \to 120 \to 220$ for Gaussian and $35 \to 120 \to 220$ for Huber, Geman and proposed robust solves. When a stage reaches its cap, the next starts from the returned state with the same objective, factor weights, factor scales, physical model and adaptive mappings. Continuation increases numerical work only; it does not redefine a method or retune the comparison.

All 384 method–case combinations terminate successfully in the supplied fault records. The recorded runtime summary is shown in Fig. S5. These fields characterize solver calls in the batch CPU implementation; they do not isolate Gaussian-pilot cost or provide a separately

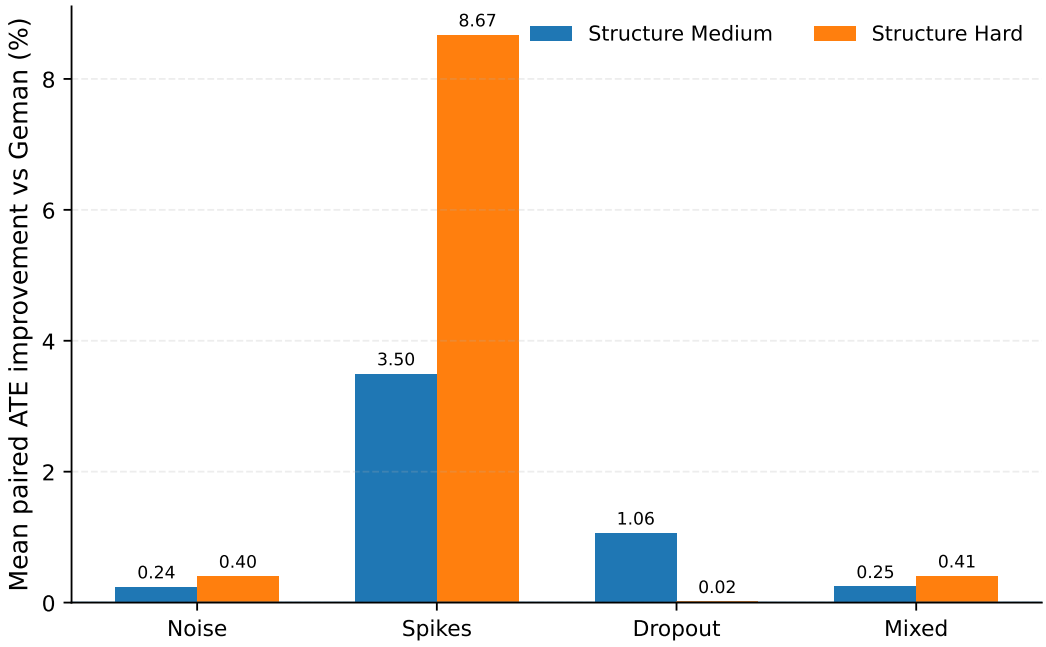


Figure S4: Sequence-specific mean paired ATE gain relative to fixed Geman on held-out TANK Medium and Hard. Positive values favor the proposed method.

measured complete pilot-plus-adaptive latency. The available records lack sufficient hardware detail for a platform-normalized benchmark. Successful termination likewise does not imply small ATE or real-time onboard execution.

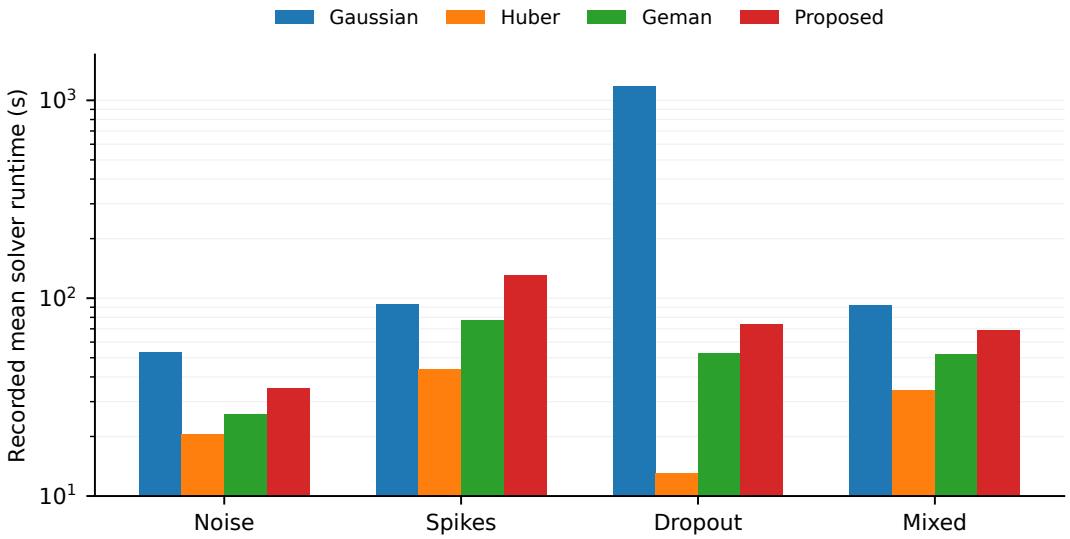


Figure S5: Recorded TANK solver runtimes by controlled fault family. Values are the supplied solver-runtime fields, not online processing rates or separately measured end-to-end pilot-plus-adaptive latency.

The data archive provides the original four derived TANK CSVs: clean results, all completed fault results, paired-gain summary and convergence summary. Blank optional fault-metadata fields remain blank; generation ranges are documented in Section S2 rather than reconstructed from missing fields. The archive contains derived evaluation output, not

raw TANK sensor streams.

S5 Additional primary sensitivity and time-resolved results

S5.1 Post-hoc primary seed-cluster sensitivity

To examine dependence induced by repeated perturbations, we averaged the proposed-versus-baseline paired percentage gain across the two primary held-out trajectories and four fault families for each of the 12 injected-fault seeds. Each seed then contributed one mean difference to an exact two-sided Wilcoxon signed-rank test. The 95% percentile intervals in Table S6 were computed by 5,000 seed-level bootstrap resamples using a fixed analysis seed (20261026). This is a *post-hoc sensitivity analysis*, not an additional experiment or a revision of the frozen primary protocol. The two physical trajectories are unchanged and are not independent mission replicates. The primary 96-case Wilcoxon tests in the main article are performed on paired absolute ATE differences, whereas the seed-aggregate sensitivity tests here use paired percentage gains. The estimands and aggregation units must not be conflated.

Table S6: Post-hoc primary 12-seed sensitivity for mean paired ATE gains. Each seed aggregates two trajectories and four injected-fault families. Positive gain favors the proposed estimator; these inferences remain conditional on the recorded physical sequences.

Baseline	Mean gain [%]	Seed-bootstrap 95% CI [%]	Exact p	Positive seeds
Gaussian	4.268	[2.869, 5.648]	0.00098	11/12
Huber	3.764	[2.441, 5.072]	0.00244	10/12
Geman–McClure	-0.624	[-4.365, 2.067]	0.67725	7/12

S5.2 Supplementary spike-severity and temporal diagnostics

The following two plots were moved from the main manuscript to satisfy the eight-display-item limit without changing the underlying data.

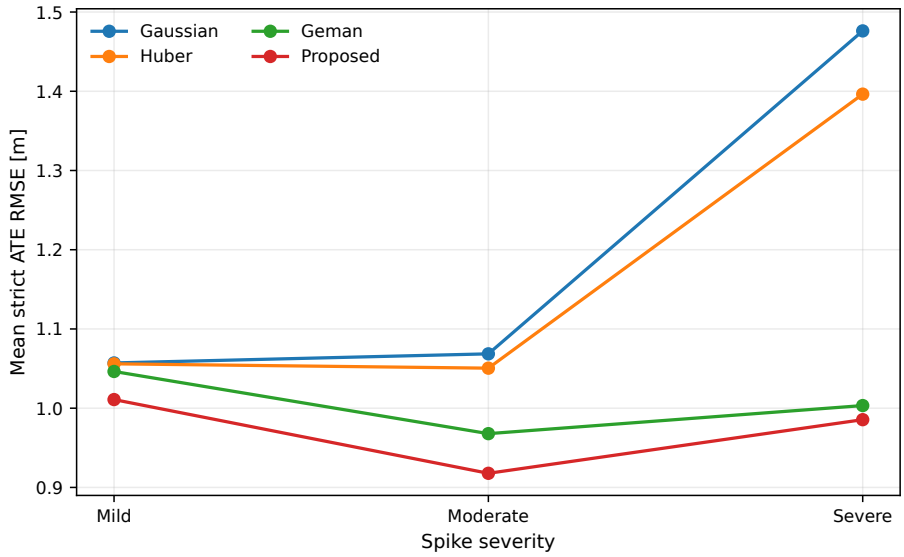


Figure S6: Mean strict ATE by sampled spike severity. The strongest advantage over Gaussian and Huber occurs in the severe subset; Geman remains competitive. Severity subsets are groups of different realizations, not a single trajectory with an increasing fault amplitude.

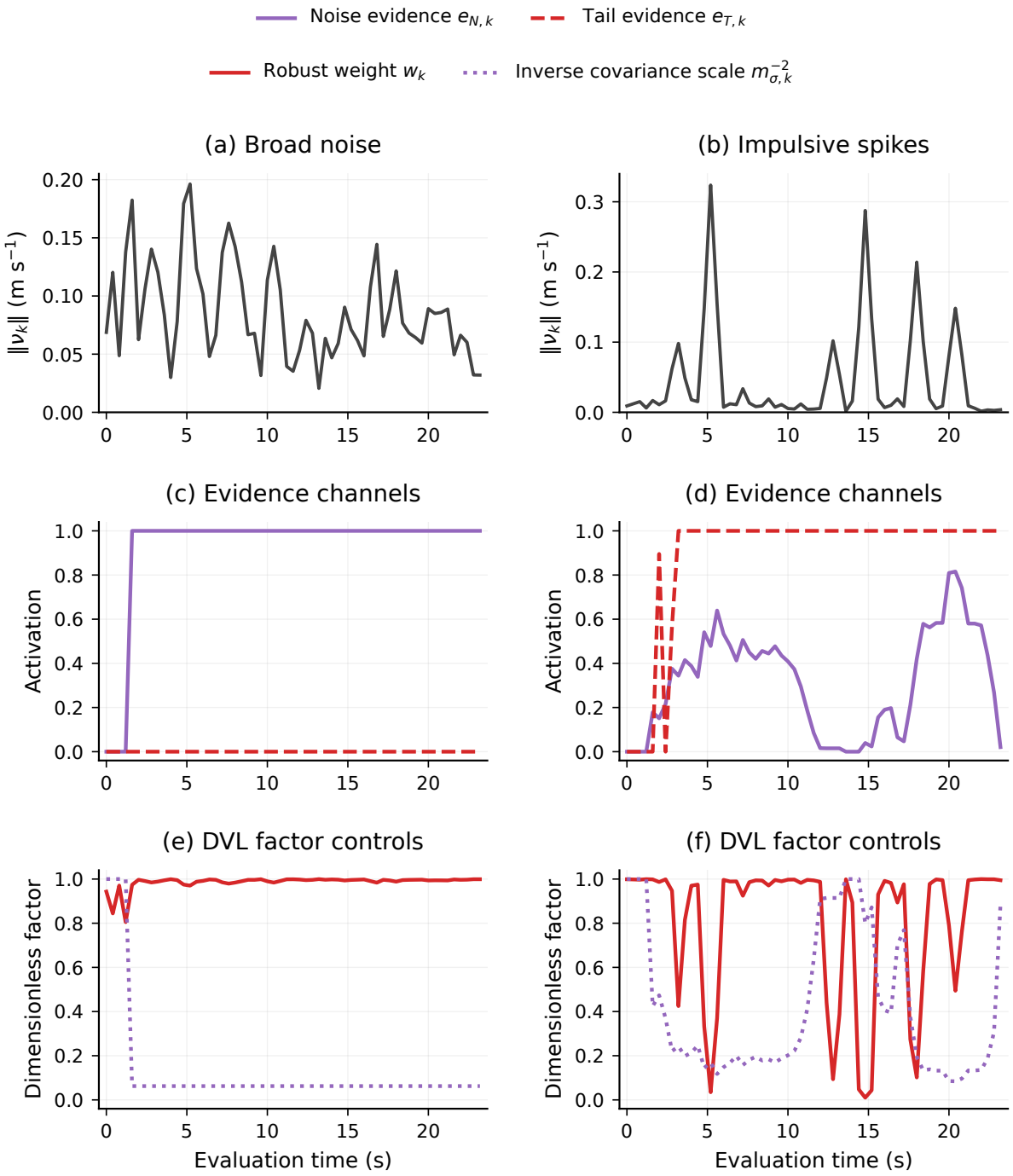


Figure S7: Time-resolved behavior on held-out L-shape data: noise seed 20292301 (left) and spike seed 20292306 (right). Panels show pilot DVL residual norm (a,b), evidence (c,d), and robust weight w_k with inverse covariance scale $m_{\sigma,k}^{-2}$ (e,f). Traces use the original cached states and mappings. Time starts at the first evaluated keyframe.

References

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

[2] 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).

[3] 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.

[4] 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.