Proprietary Versus Open-Source Visual-Inertial Fusion Under GNSS Degradation for Orchard-Scale 3D Fruit Mapping

Thiago T. Santos[1], D. Bharti[1], Luciano Gebler[2] and Andrea de Rossi[2]

[1]Embrapa Agricultura Digital, Campinas, SP, Brazil

[2]Embrapa Uva e Vinho, Vacaria, RS, Brazil

A paper from the Proceedings of the

17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture

13-16 July 2026

Porto Alegre, Brazil

Abstract.

A recent pipeline combining GNSS–visual-inertial odometry with factor-graph refinement of fruit landmarks has reported orchard-level apple counting errors below three percent against harvest totals — a result with few precedents in the agricultural SLAM literature, where GNSS-VIO fusion, landmark-level optimization, and harvest-validated yield estimation have until now appeared mostly in isolation. However, trajectory quality in that pipeline was assessed solely through reprojection error, which can decrease even when camera poses drift, provided poses and landmarks shift together consistently; low reprojection residuals therefore do not guarantee metrically accurate 3D fruit maps. Moreover, the pipeline relies on a proprietary sensor SDK whose internal fusion logic is opaque, making it impossible to attribute trajectory degradation to visual-inertial odometry quality, GNSS weighting, or downstream optimization. This study proposes a systematic evaluation protocol that jointly addresses both limitations. A post-processed kinematic (PPK) GNSS signal is synthetically degraded before entering the fusion stage — through temporal subsampling, additive position noise, systematic drift, intermittent dropout, and a combined low-cost-receiver profile — while the original PPK trajectory is withheld as independent ground truth. Crucially, the degradation experiments are conducted on three parallel VIO-GNSS front-ends processing the same stereo imagery and IMU data from a tractor-mounted platform in a Fuji apple orchard in southern Brazil: the proprietary SDK fusion (ZED-F); an open OpenVINS estimator fused with degraded GNSS through an explicit GTSAM pose graph in which every factor covariance is transparent and controllable (OV-PG); and the same pose graph augmented with visual landmark projection factors (OV-PGL). All three fused trajectories feed the same downstream pipeline — YOLO-based apple detection, point-tracker-based temporal association, and multi-view triangulation — so that differences in trajectory error and counting accuracy are attributable to the fusion stage alone. For each front-end and degradation level, Absolute Trajectory Error and Relative Pose Error are computed against the withheld reference, and reprojection behavior is examined to identify the regime where internal consistency diverges from metric accuracy. The protocol is applied to two independent

The authors are solely responsible for the content of this paper, which is not a refereed publication. Citation of this work should state that it is from the Proceedings of the 17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture. EXAMPLE: Last Name, A. B. & Coauthor, C. D. (2026). Title of paper. In Proceedings of the 17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture (unpaginated, online). Monticello, IL: International Society of Precision Agriculture and Brazilian Congress on Precision and Digital Agriculture.

traversals of the same orchard block collected approximately one year apart, each with its own PPK ground truth. The results disentangle VIO front-end quality from fusion strategy: the open pose graph achieves the lowest absolute error where it completes, the proprietary fusion is the most robust across all conditions, and the landmark-augmented graph — although beneficial for local map rigidity — anchors the trajectory to corrupted GNSS once position noise exceeds a critical threshold, inflating absolute error by up to an order of magnitude while remaining internally consistent. Relative pose error proves insensitive to both architecture and degradation, confirming that internal or relative metrics alone cannot certify georeferenced fruit maps. Multi-view fruit triangulation recall is further shown to be a sensitive proxy for trajectory orientation accuracy, indicating that counting quality under realistic GNSS degradation is governed primarily by orientation consistency rather than by positional accuracy or GNSS signal quality. Together, these findings establish GNSS quality requirements for operational deployment of SLAM-based fruit mapping and contribute toward the reproducible multi-season evaluation protocols currently absent from the agricultural SLAM literature.

Keywords.
SLAM, visual-inertial odometry, GNSS degradation, factor graphs, fruit mapping, yield estimation, apple orchards

Introduction

Accurate and spatially resolved yield estimation is a central requirement for precision orchard management, affecting harvest logistics, labor planning and storage capacity. Ground-based sensing platforms equipped with cameras, inertial measurement units (IMU) and GNSS receivers enable data acquisition at orchard scale from within the crop rows, where individual fruits are visible at high resolution (Santos et al. 2025; Lei et al. 2025). A recent pipeline combining GNSS-visual-inertial odometry (VIO) with incremental factor-graph refinement of fruit landmarks reported orchard-level apple counting errors below three percent against harvest totals (Bharti et al. 2026) — a result with few precedents in the agricultural SLAM literature, where GNSS-VIO fusion, landmark-level optimization, and harvest-validated yield estimation have until now appeared mostly in isolation.

Two limitations, however, remain open in that line of work. First, trajectory quality was assessed solely through reprojection error. Reprojection error measures internal consistency between estimated camera poses and estimated 3D landmarks; because poses and landmarks are jointly optimized, the residuals can decrease even when the camera trajectory drifts, provided poses and landmarks shift together consistently (Zhang and Scaramuzza 2018). Low reprojection residuals therefore do not guarantee metrically accurate, georeferenced 3D fruit maps. Second, the pipeline relies on the proprietary fusion module bundled with the stereo camera SDK. Its internal fusion logic — sensor weighting, outlier rejection, GNSS-VIO calibration — is opaque, making it impossible to attribute trajectory degradation to the VIO front-end, to the GNSS weighting, or to the downstream optimization.

At the same time, the GNSS quality available in production settings spans a wide range, from PPK/RTK solutions with centimeter-level accuracy to low-cost single-frequency receivers subject to meter-level noise, reduced update rates and signal loss under dense canopy. Establishing which GNSS quality is actually required for a target mapping accuracy — and which fusion architecture degrades most gracefully when that quality is not met — is a prerequisite for operational deployment of SLAM-based fruit mapping.

This study addresses both limitations through a systematic evaluation protocol. The PPK-corrected GNSS signal was synthetically degraded before entering the fusion stage — through temporal subsampling, additive position noise, systematic drift, intermittent dropout and a combined low-cost-receiver profile — while the original PPK trajectory was withheld as independent ground truth. Degradation experiments were conducted on three parallel VIO-GNSS fusion front-ends processing the same stereo imagery and IMU data: the proprietary SDK fusion, an open OpenVINS-based estimator (Geneva et al. 2020) fused with GNSS in an explicit GTSAM factor graph (Dellaert and Kaess 2017; Kaess et al. 2012), and the same factor graph augmented with visual landmark projection factors. For each front-end and degradation level, absolute trajectory error (ATE) and relative pose error (RPE) were computed against the withheld reference (Sturm et al. 2012).

The protocol was applied to two independent traversals of the same orchard block collected approximately one year apart (March 2025 and February 2026), each with its own PPK ground truth. The specific contributions are: (i) a reproducible GNSS degradation protocol for evaluating VIO-GNSS fusion in orchards; (ii) a head-to-head comparison of proprietary and open fusion architectures on identical sensor data across two seasons; and (iii) an analysis of the GNSS quality regimes in which each architecture remains within sub-meter trajectory accuracy. It is demonstrated that multi-view fruit triangulation recall is a sensitive proxy for the rotational accuracy of the underlying camera trajectory, and that fruit counting quality under realistic GNSS degradation is driven primarily by trajectory orientation consistency rather than by positional accuracy or GNSS signal quality.

Materials and Methods

Data Acquisition Platform and Field Sessions

Data were collected with the SEEmear proximal sensing platform (Santos et al. 2025), which integrates two stereo cameras with global-shutter sensors of 1920 × 1200 pixels, wide-angle lenses and built-in IMUs (ZED X, Stereolabs Inc., San Francisco, CA, USA), a rover-base pair of multi-band GNSS receivers (Reach M2 rover and Reach RS3 base, Emlid Tech Kft., Budapest, Hungary), and an embedded computer (Jetson AGX Orin Developer Kit, NVIDIA Corporation, Santa Clara, CA, USA). The cameras face opposite sides of the platform, scanning two tree rows simultaneously; the GNSS rover provides GPS time used to synchronize image timestamps with positioning data (Fig. 1).

Field sessions were performed at Embrapa's Temperate Climate Fruit Growing Experimental Station (EFCT) in Vacaria, Rio Grande do Sul, Brazil (28°30'58.2"S, 50°52'52.2"W). The platform was mounted on a tractor traveling at approximately 5 km/h along the corridors of a Fuji apple block trained to a palmette system, with camera-to-canopy distance between 1 m and 1.5 m. The block consists of four *Fuji* apple rows alternating with four *Gala* rows, the latter already harvested at the time of data collection; each session covered both sides of all four *Fuji* rows. Two sessions of the same block were used: 13 March 2025 and 10 February 2026. Each session recorded complete stereo footage at 30 fps, IMU readings, and raw GNSS observations from rover and base. The 2025 session lasted 377 s, covering an estimated 524 m and yielding 11,320 stereo frame pairs, while the 2026 session lasted 394 s, covering an estimated 547 m and yielding 11,833 stereo frame pairs.

Ground Truth and Reference Frames

GNSS observations were post-processed in kinematic mode (PPK) with RTKLIB (Takasu and Yasuda 2009) against the local base station, whose coordinates were determined by IBGE-PPP (SIRGAS2000, compatible with WGS84 at the accuracy level considered here). The fixed-ambiguity PPK solution, with nominal centimeter-level accuracy, was adopted as the ground-truth trajectory (condition C0) and was withheld from the fusion stage in all degraded conditions. All positions were converted to a local East-North-Up (ENU) frame anchored at the same base-station origin for the 2025 and 2026 sessions, enabling cross-session comparison. All trajectories were exchanged in the TUM format (timestamp, position, orientation quaternion) for evaluation.

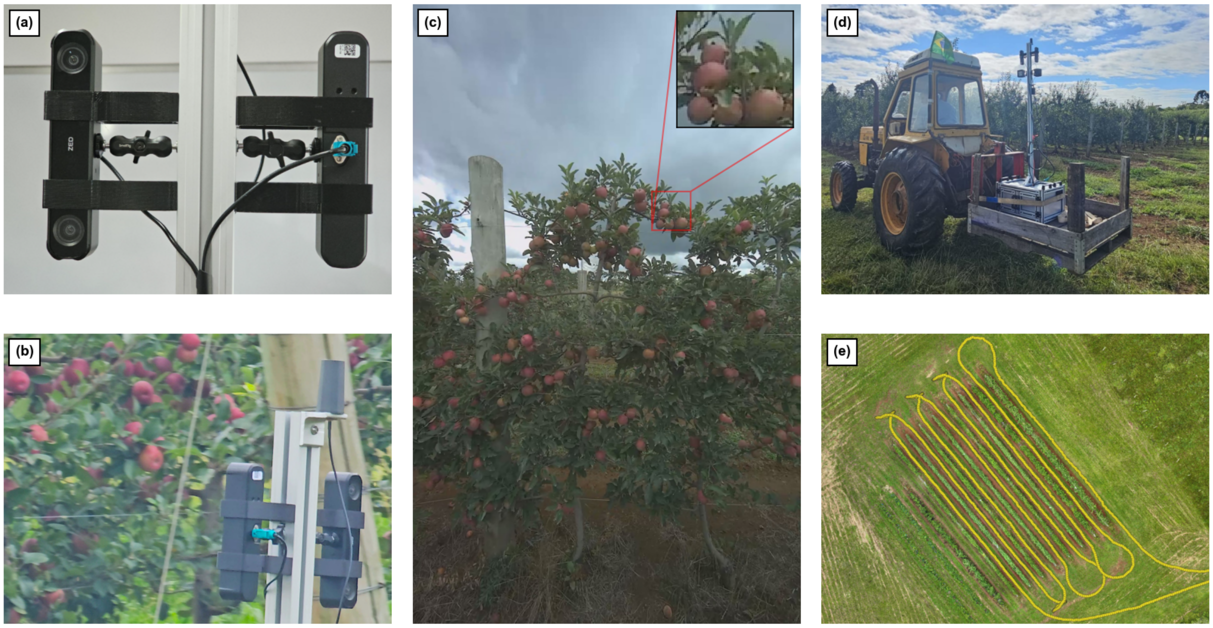

Fig 1. Data acquisition setting as seen in Santos et al. (2025): (a) stereo cameras with integrated IMU, (b) cameras oriented to the apples trees, (c) frame example, the left camera from a stereo pair, (d) the system mounted in a tractor, (e) the path registered by the GNSS system after PPK correction.

Synthetic GNSS Degradation Protocol

Six degradation families were applied to the C0 PPK stream, producing sixteen experimental conditions (Fig 2). The degraded streams — never the original PPK — were the only GNSS input given to the fusion front-ends:

- C1 — temporal subsampling: the PPK rate was reduced to 1.0 Hz and 0.5 Hz, emulating low-rate receivers.
- C2 — additive noise: independent zero-mean Gaussian noise with $\sigma \in \{0.05, 0.10, 0.25, 0.50, 1.00\}$ m was added to the ENU positions; the per-fix covariances delivered to the estimators were inflated accordingly, so each front-end received an accurate description of the degraded covariance.
- C3 — systematic drift: a linear position ramp at 45° in the EN plane, reaching a terminal offset of 0.5, 1.0 or 2.0 m at the end of the sequence, emulating slowly varying biases such as uncorrected atmospheric error.
- C4a — random dropout: each fix was removed with probability 0.3, 0.5 or 0.7, emulating intermittent reception under canopy.
- C4b — structured blackout: all fixes within windows of 3 s or 10 s centered at the headland turns were removed, emulating obstruction at row ends.
- C5 — combined low-cost profile: 1 Hz subsampling, $\sigma = 0.25$ m noise and 30% dropout applied jointly.

VIO-GNSS Fusion Front-Ends

Three fusion front-ends consumed exactly the same stereo image stream, IMU stream and degraded GNSS stream. Differences in output trajectories are therefore attributable to the fusion architecture alone.

ZED-F: Proprietary SDK Fusion

The first front-end is the GNSS-VIO fusion bundled with the camera manufacturer's SDK (ZED SDK Fusion API, Stereolabs Inc.) version 5.2.1. Sequences were replayed from the recorded

SVO2 files with positional tracking in its most recent generation mode with IMU fusion enabled; degraded GNSS fixes were injected through the SDK ingestion interface together with their per-fix covariances and fix-status flags. The SDK performs its own GNSS-VIO calibration (target yaw uncertainty of 7×10^{-3} rad was used) and outputs a fused geodetic pose, converted to the common ENU frame. The internal weighting, outlier rejection and state estimation are not documented and cannot be inspected — this front-end represents an undisclosed, loosely-coupled, proprietary fusion, used here as the black-box baseline in Bharti et al. (2026). Its output is dense (one pose per fused frame, ≈30 Hz).

OV-PG: OpenVINS with GTSAM Pose-Graph Fusion

The second front-end decouples VIO from GNSS fusion. Visual-inertial odometry is computed by OpenVINS (Geneva et al. 2020), an open-source filter-based estimator built on the multi-state constraint Kalman filter (MSCKF; Mourikis and Roumeliotis 2007), configured for the stereo pair with calibrated camera-IMU extrinsics. Its odometry is fused with the degraded GNSS in an explicit factor graph optimized incrementally with iSAM2 (Kaess et al. 2012) in the GTSAM library (Dellaert and Kaess 2017).

The graph structure is shown in Fig 3. Each keyframe k holds a pose X(k) in ENU, a velocity V(k) and an IMU bias B(k). Consecutive states are linked by IMU preintegration factors (`CombinedImuFactor`) built from the raw accelerometer and gyroscope streams, with noise densities taken from the IMU calibration. GNSS fixes are attached as position-only `GPSFactors`, weighted by the (inflated) per-fix standard deviations with a floor of 0.03 m (horizontal) and 0.05 m (vertical). A relative-odometry `BetweenFactor` links consecutive keyframe poses using the OpenVINS relative motion (σ_{rot} = 0.05 rad, σ_{pos} = 0.5 m); this factor constrains the yaw, which IMU plus position-only GNSS leave weakly observable, while leaving the absolute position to the `GPSFactors`.

An initial bootstrap stage aligns the OpenVINS odometry frame to ENU: after the platform has traveled at least 1 m, OpenVINS-GNSS pose pairs are collected over a 5 s window (15 s under C5, whose rate is ≈0.7 Hz) and a least-squares rigid alignment is estimated (Umeyama 1991). The optimized iSAM2 state is periodically written back to the OpenVINS filter to keep the front-end consistent with the fused estimate; the writeback is guarded by two fail-safes — it is blocked when the bootstrap residual RMS exceeds 1 m, and when the optimized gyroscope or accelerometer biases exceed 0.1 rad/s or 0.5 m/s², respectively — preventing a degraded graph solution from destabilizing the filter. The output trajectory is sparse (one pose per keyframe, ≈5 Hz).

OV-PGL: Pose Graph with Visual Landmark Factors

The third front-end extends OV-PG with explicit visual landmarks (Fig 4). Image features tracked by OpenVINS are promoted to 3D landmark variables L(j) in the graph, observed through monocular projection factors (`GenericProjectionFactor`) with a 1.5-pixel noise model wrapped in a Huber robust kernel (k = 1.345). Landmarks are seeded by back-projecting the feature at its stereo depth from the current GNSS-corrected keyframe pose and receive a loose 0.5 m position prior that removes the depth rank-deficiency of two-observation monocular landmarks. A landmark enters the graph only after its second observation, is capped at 500 simultaneous landmarks, and is evicted after 20 keyframes without re-observation. Because the projection factors already constrain inter-frame rotation, the relative-odometry rotation sigma is relaxed to 0.1 rad; an A/B sweep showed that the tighter 0.05 rad value over-constrains the graph under sparse or noisy GNSS and causes optimization failures.

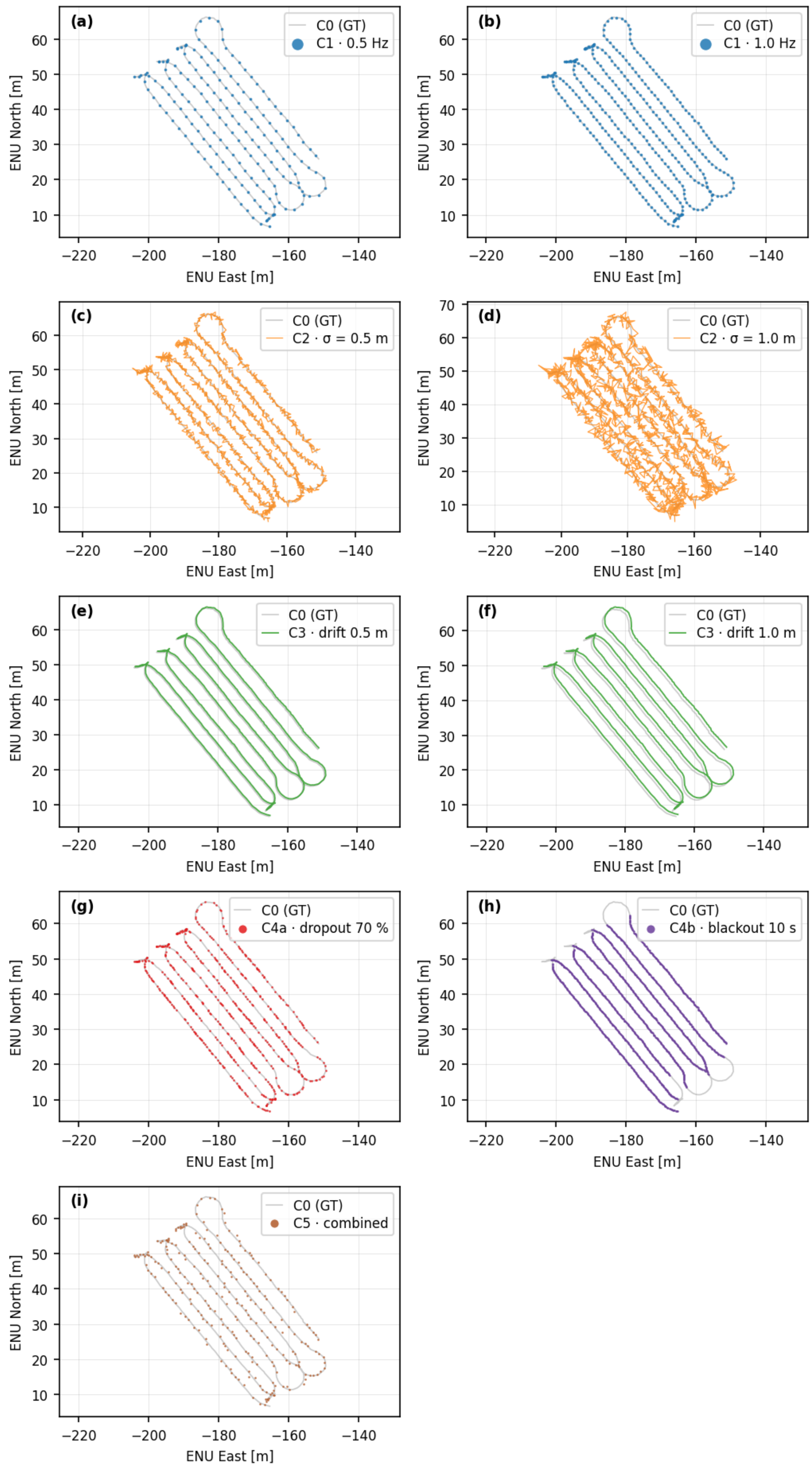

Fig 2. Synthetic GNSS degradation conditions applied to the 2025 PPK trajectory (gray, C0): (a) C1 subsampling at 0.5 Hz, (b) C1 at 1.0 Hz, (c) C2 additive noise σ = 0.5 m, (d) C2 σ = 1.0 m, (e) C3 drift 0.5 m, (f) C3 drift 1.0 m, (g) C4a dropout 70%, (h) C4b blackout 10 s at headland turns, (i) C5 combined profile

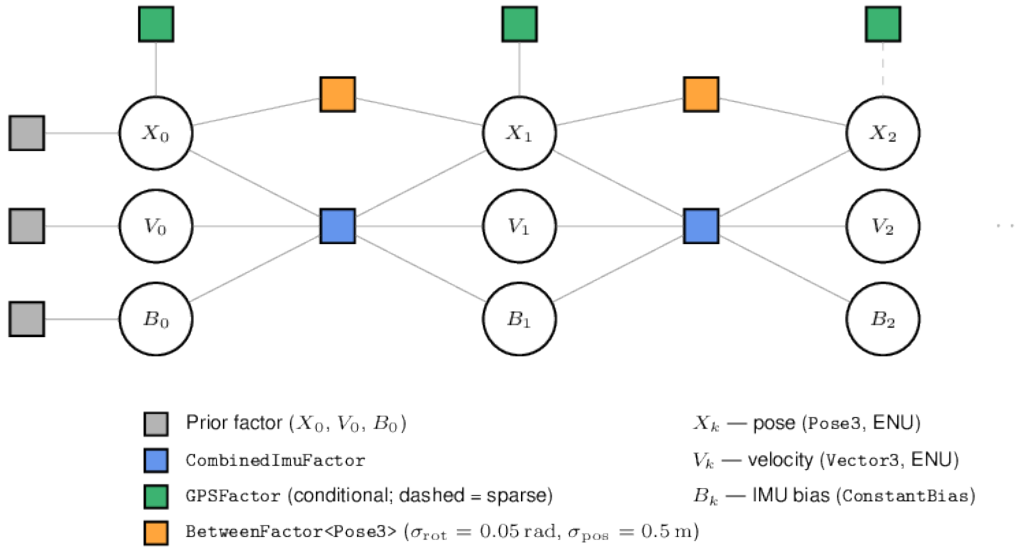

Fig 3. Factor graph of the OV-PG front-end. Circles are variables (X pose, V velocity, B IMU bias); squares are factors: priors (gray), IMU preintegration (blue), position-only GNSS factors (green, dashed when the fix is absent under degradation) and relative-odometry between-factors (orange)

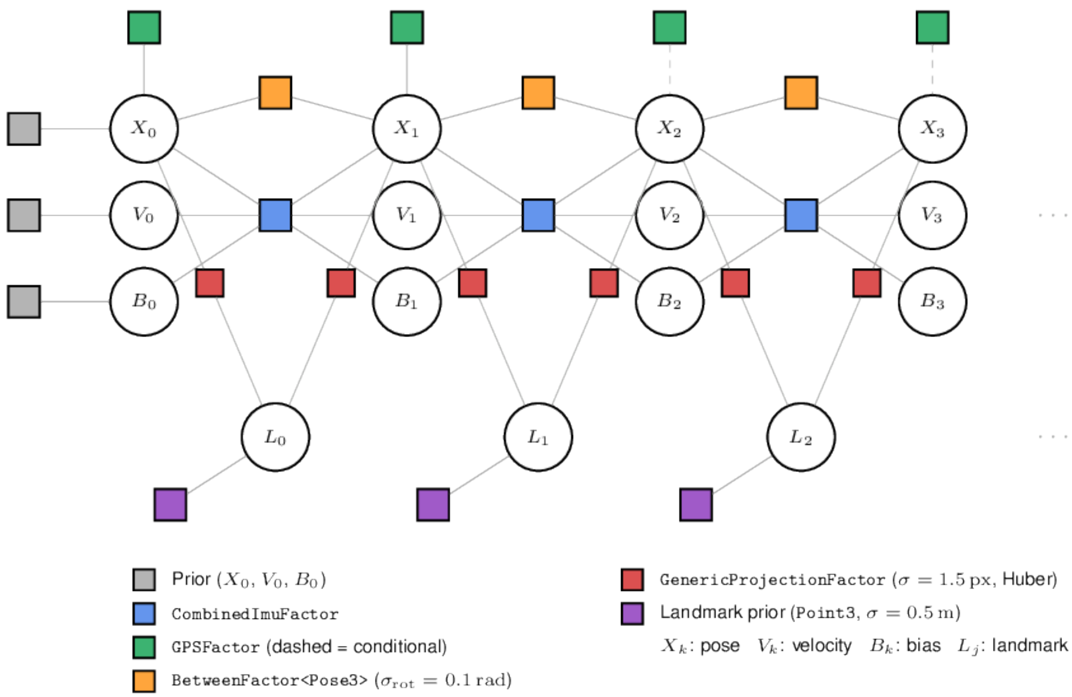

Fig 4. Factor graph of the OV-PGL front-end. In addition to the OV-PG structure, landmark variables L are observed by projection factors (red) with Huber robust kernels and anchored by loose landmark priors (purple)

Trajectory Evaluation

For each front-end and condition, the fused trajectory was compared with the withheld C0 PPK trajectory using the evo package (Grupp 2017). ATE was computed as the translational absolute pose error after a rigid Umeyama alignment without scale correction, with pose association limited to 0.15 s of timestamp difference. RPE was computed on the translational part over sub-trajectory lengths of 1, 5, 10 and 50 m (Sturm et al. 2012; Zhang and Scaramuzza 2018). Note that the three front-ends produce trajectories of different densities (≈30 Hz for ZED-

F, ≈5 Hz keyframes for OV-PG and OV-PGL); metrics are computed over each system's own pose set after timestamp association.

Downstream Fruit Mapping

To connect trajectory accuracy with the final agronomic product, the fused trajectories feed a mapping pipeline similar to Bharti et al. (2026): apples are detected with a YOLO-based model, YOLOv9 (Wang et al. 2024), associated across frames by a point-tracking network, CoTracker3 (Karaev et al. 2024a; Karaev et al. 2024b), and triangulated from multiple views, using the poses from the trajectories (position and orientation). The resulting georeferenced fruit positions are aggregated into orchard-level density maps.

Results and Discussion

The importance of the GNSS signal

Odometry-based estimation—whether derived from visual, inertial, or visual-inertial sensing — integrates relative motion increments over time, and any small per-step error in this integration accumulates into a global pose error that grows unbounded with trajectory length. Fig. 5 illustrates this behavior: although the VIO trajectories (ZED-F, OV-PG, and OV-PGL, all without GNSS factors) closely follow the ground-truth shape over short segments, the accumulated drift causes substantial divergence by the end of each session, with estimated and true trajectories differing by tens of meters in some regions. Correcting this drift requires an external source of global localization information that is not subject to integration error—most commonly GNSS, but potentially also fixed visual landmarks, or pre-surveyed control points—whose absolute position measurements can be incorporated as factors in the estimation graph to anchor the trajectory to a consistent global reference frame.

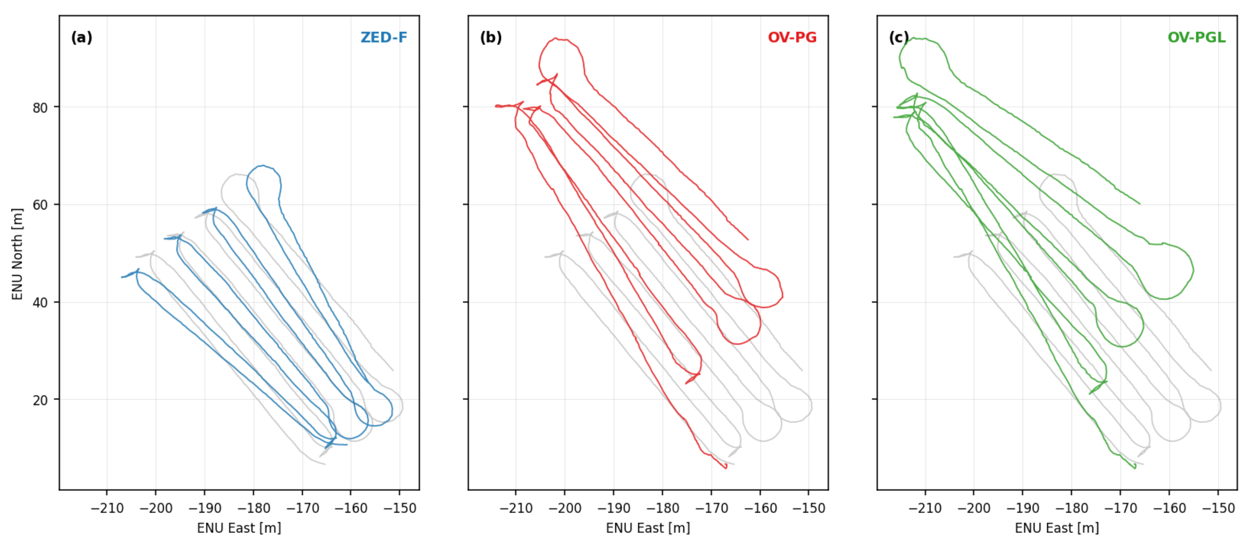

Fig 5. Visual-inertial odometry (VIO) trajectories are locally accurate but accumulate severe drift over time, highlighting the need for a global correction source such as GNSS. Estimated trajectories (colored) are compared against ground truth (C0, gray) for (a) ZED-F (without GNSS fusion), (b) OV-PG, and (c) OV-PGL, none of which include GNSS factors.

Absolute Trajectory Error Under GNSS Degradation

When GNSS fusion was enabled, three distinct patterns emerged, as shown in Table 1 and in the trajectory overlays of Figs 6 and 7. First, the open pose-graph fusion (OV-PG) achieved the lowest absolute errors wherever it completed: 0.08–0.29 m RMSE across the 2025 conditions and 0.12–0.30 m in 2026, including 0.26 m under the combined low-cost profile C5 in 2025 —

five times lower than the proprietary fusion under the same input (1.28 m). With transparent, weighted `GPSFactors`, the graph optimally discounts noisy fixes against the IMU preintegration and relative-odometry constraints.

Second, the proprietary fusion (ZED-F) was the most robust: it completed all sixteen conditions in both seasons, with ATE growing smoothly from 0.14–0.40 m under mild degradation to 1.0–1.3 m under σ = 1.0 m noise and the C5 profile. Its internal model evidently regularizes the trajectory strongly, but the lack of access to its weighting prevents any per-condition tuning, and its error under heavy noise exceeded the open pose graph by a factor of 4–5.

Table 1. ATE RMSE (m) against the withheld PPK ground truth for the three fusion front-ends in the 2025 and 2026 sessions, by degradation condition (C1 subsampling rate, C2 noise σ, C3 drift terminal offset, C4a dropout probability, C4b blackout duration, C5 combined profile). An em-dash (—) denotes a run that failed to converge/complete.

Condition	ZED-F 2025	OV-PG 2025	OV-PGL 2025	ZED-F 2026	OV-PG 2026	OV-PGL 2026
C1 0.5 Hz	0.57	**0.13**	0.66	0.74	—	**0.71**
C1 1.0 Hz	0.46	**0.09**	0.52	**0.35**	—	0.50
C2 0.05 m	0.40	**0.09**	0.40	**0.25**	—	0.38
C2 0.10 m	0.50	**0.10**	0.57	**0.36**	—	0.74
C2 0.25 m	**0.74**	—	1.65	0.58	**0.12**	2.36
C2 0.50 m	**0.99**	—	3.95	0.81	**0.18**	—
C2 1.00 m	**1.20**	—	8.00	0.98	**0.30**	11.69
C3 0.5 m	**0.31**	—	0.34	**0.14**	—	0.25
C3 1.0 m	**0.31**	—	0.34	**0.14**	—	0.25
C3 2.0 m	**0.32**	—	0.35	**0.15**	—	0.26
C4a 30%	0.33	**0.08**	0.37	**0.15**	—	0.30
C4a 50%	0.34	—	0.40	**0.19**	—	—
C4a 70%	0.40	**0.09**	0.52	**0.24**	—	0.41
C4b 3 s	0.41	**0.08**	0.36	**0.15**	—	0.29
C4b 10 s	0.48	**0.29**	0.56	**0.63**	0.18	0.83
C5 combined	1.28	**0.26**	4.68	**1.10**	—	10.52

Third, the landmark-augmented graph (OV-PGL) matched OV-PG under subsampling, drift, dropout and blackout (0.25–0.83 m), but degraded sharply once the GNSS noise reached σ = 0.25 m, with ATE growing roughly linearly with σ up to 8.0 m (2025) and 11.7 m (2026) at σ = 1.0 m, and 4.7–10.5 m under C5. To date, the best explanation that has been identified is structural: landmark seeds are back-projected from GNSS-corrected keyframe poses, and subsequent poses are then anchored — via their priors and projection factors — to a map that was built in the corrupted frame. GNSS noise is thereby propagated into the map, rather than being averaged out, by the same visual factors that confer local rigidity. This is precisely the regime in which internal consistency metrics are found to diverge from metric accuracy: the optimizer is maintained in a self-consistent state — reflected in low reprojection residuals — while the entire pose-landmark assembly is left to drift.

Under systematic drift (C3), all front-ends followed the drifting input as expected — the fusion has no independent absolute reference — and the residual ATE after rigid alignment (0.14–0.35 m) reflects only the non-rigid component of the linear ramp. Detecting such drift therefore requires cross-session or external consistency checks rather than single-session fusion. Under structured blackouts (C4b, 10 s at headland turns) all systems bridged the gaps with visual-inertial odometry, with ATE below 0.84 m in every completed run.

It is worth noting that none of the three systems employ reinitialization: when a significant disparity is detected between GNSS measurements and the current fusion estimate, the system could be reinitialized in an attempt to recover a sane state after divergence. Such a strategy was not employed in this comparison but would be a viable option for a final deployed system.

Relative Pose Error

In contrast with ATE, RPE was nearly invariant across front-ends and degradation conditions (Table 2): 12.3–14.1 m RMSE over 10 m sub-trajectories, and 1.4–2.1 m over 1 m sub-trajectories, with no systematic ordering between systems. Two conclusions follow. First, local trajectory consistency is governed by the visual-inertial front-end — which receives identical imagery and IMU data in all runs — while GNSS quality controls only the absolute placement of the trajectory; degradation therefore manifests almost exclusively in ATE. Second, RPE alone would not have discriminated the fusion architectures at all, reinforcing that internal or relative metrics are insufficient to certify georeferenced fruit maps.

Cross-Season Consistency

The qualitative ranking of the three architectures was preserved across the two sessions, collected eleven months apart with different camera units and field conditions: OV-PG lowest ATE where it completed, ZED-F most robust, OV-PGL noise-sensitive above σ = 0.25 m. The 2026 session was more challenging for the open front-ends — the visual-inertial filter required the bias-guard fail-safe and relative-odometry constraints to complete the runs — while ZED-F errors were, on average, slightly lower than in 2025.

Implications for Fruit Mapping

Fig 8 shows the georeferenced apple map and the corresponding kernel-density estimate produced by the downstream pipeline from the 2025 session (condition C1, 1 Hz). Fruit positions concentrate along the four scanned rows, and the density map exposes within-row production variability at a resolution useful for site-specific management. Given the ATE results, sub-meter fruit geolocation is attainable with any of the three front-ends when GNSS noise stays below σ = 0.25 m; beyond that threshold, only the architectures that decouple the landmark map from the GNSS stream (OV-PG fusion, or ZED-F followed by independent triangulation) preserve map geometry.

Table 2. RPE RMSE (m) over 10 m sub-trajectories for the three fusion front-ends, by degradation condition (codes as in Table 1)

Condition	ZED-F 2025	OV-PG 2025	OV-PGL 2025	ZED-F 2026	OV-PG 2026	OV-PGL 2026
C1 0.5 Hz	**12.77**	14.12	13.22	**12.84**	—	13.40
C1 1.0 Hz	**13.01**	14.08	13.33	**12.64**	—	13.40
C2 0.05 m	**13.16**	13.30	13.30	**12.98**	—	13.39
C2 0.10 m	**13.21**	13.28	13.27	**12.86**	—	13.39
C2 0.25 m	13.26	—	**13.11**	**13.04**	13.54	13.16
C2 0.50 m	13.37	—	**12.84**	**12.90**	13.44	—
C2 1.00 m	**12.95**	—	13.09	12.72	13.44	**12.67**
C3 0.5 m	**12.90**	—	13.34	**12.91**	—	13.40
C3 1.0 m	**12.90**	—	13.34	**12.91**	—	13.40
C3 2.0 m	**12.90**	—	13.34	**12.92**	—	13.40
C4a 30%	**12.90**	13.49	13.32	**12.99**	—	13.38
C4a 50%	**12.91**	—	13.33	**12.69**	—	—
C4a 70%	**12.89**	13.92	13.30	**12.61**	—	13.39
C4b 3 s	**12.96**	13.04	13.35	**12.91**	—	13.40
C4b 10 s	**13.10**	13.29	13.31	**12.57**	13.50	13.34
C5 combined	13.19	14.14	**13.08**	12.82	—	**12.31**

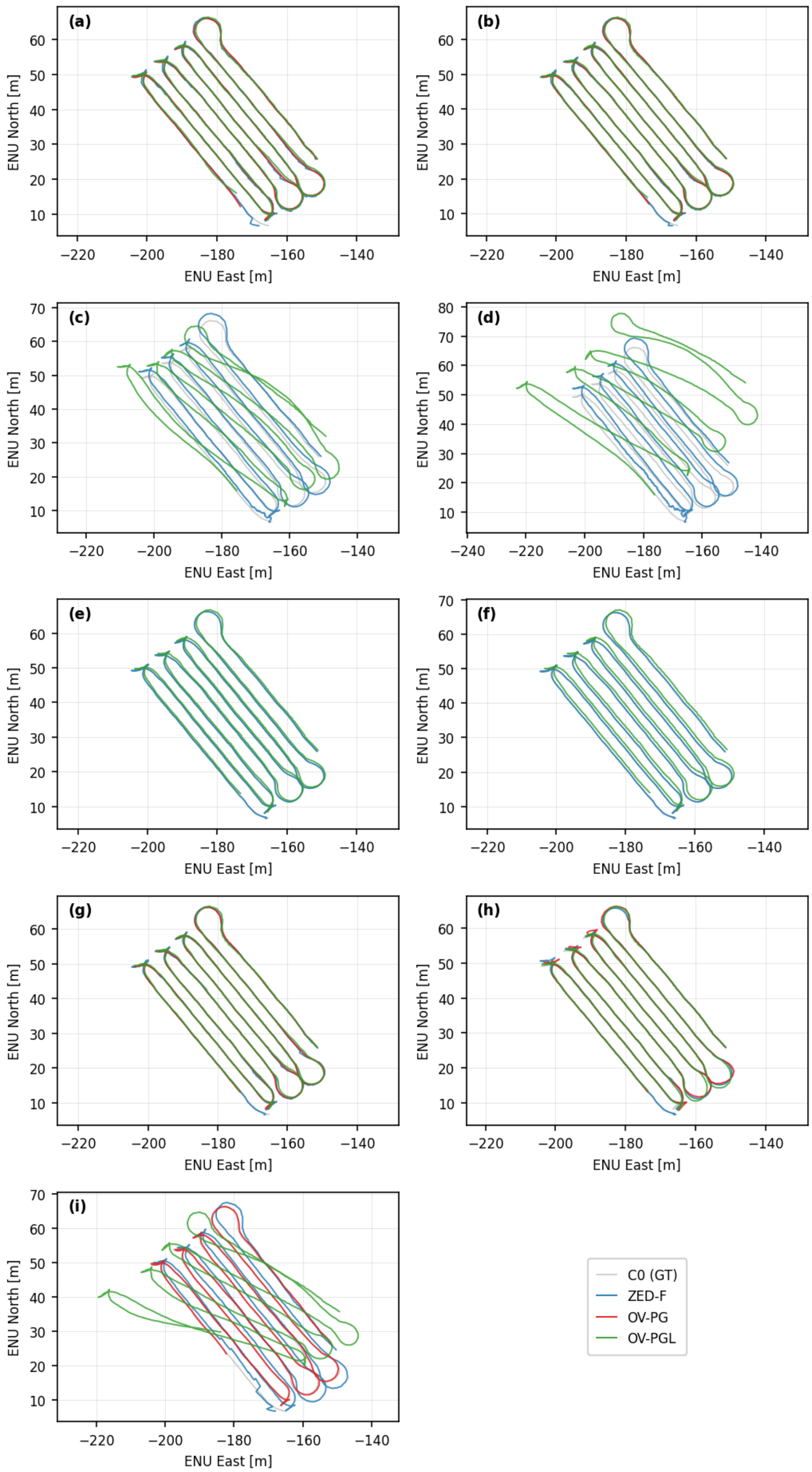

Fig 6. Fused trajectories for the 2025 session against the PPK ground truth (gray): (a) C1 0.5 Hz, (b) C1 1.0 Hz, (c) C2 0.5 m, (d) C2 1.0 m, (e) C3 0.5 m, (f) C3 1.0 m, (g) C4a 70% dropout, (h) C4b 10 s blackout, (i) C5 combined; ZED-F (blue), OV-PG (orange) and OV-PGL (green). Curves of missing runs (failed to converge/complete) are absent.

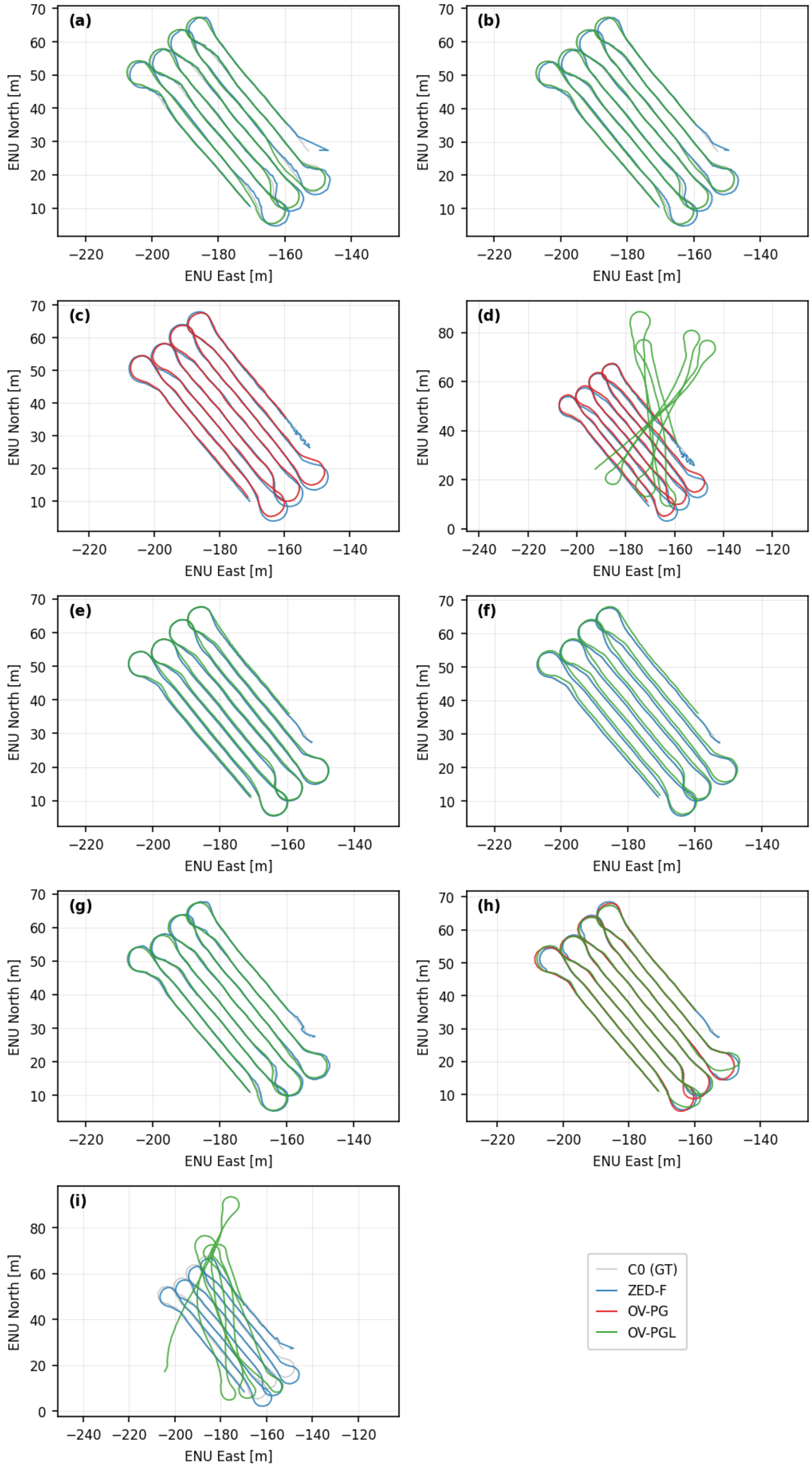

Fig 7. Fused trajectories for the 2026 session against the PPK ground truth (gray), panels as in Fig 6. Curves of missing runs (failed to converge/complete) are absent

Against a ground-truth count of 9,985 fruits across four orchard rows, the apple mapping pipeline based on OV-PGL detected 8,558 fruits under the baseline 1 Hz GNSS condition (C1), yielding an overall recall of 85.7%. Per-row recall was highly uneven: the two central rows achieved 94.6% and 98.2%, while the two edge rows reached only 79.9% and 67.9%. The lower recall on row 1 is largely explained by the system's initialization phase — camera poses are unavailable during the bootstrap window (typically 6–7 s after motion onset), so fruits in the opening segment of the first row are never triangulated. Row 4's lower recall reflects a distinct agronomic factor: that row uses a wider-canopy training system than rows 1–3, increasing fruit occlusion and reducing the number of valid views per track available for triangulation. Moderate GNSS degradation had negligible impact on fruit counts: across all C2 (noise up to 1 m) and C3 (drift up to 2 m) conditions, inter-run variation remained below 30 detections (< 0.4% of total), demonstrating that the triangulation pipeline is robust to the levels of GNSS error tested. Only the most severe condition, C5 (combined degradation at sub-Hz GNSS rate), produced a meaningful drop, with Row 1 losing approximately 300 additional detections relative to baseline, consistent with the trajectory distortion observed in the ATE results for that condition.

The ZED Fusion pipeline achieved a substantially lower overall recall of 67.8% (6,774 fruits), with per-row recalls of 69.4%, 63.5%, 90.2%, and 44.5% for Rows 1–4, respectively. Because both pipelines operate on identical 2D detection tracks, this gap is entirely attributable to differences in the estimated camera poses. Multi-view triangulation reprojection error — the dominant rejection criterion — is primarily sensitive to rotational rather than translational pose errors: a positional offset shifts all reprojections uniformly and is absorbed by the DLT solver, whereas inconsistent orientation estimates across frames rotate the reprojection pattern differently at each view, producing median reprojection errors that exceed the 6 px acceptance threshold. The 18-percentage-point recall gap therefore indicates that the ZED SDK's internal fusion produces less rotationally accurate poses than OV-PGL, whose iSAM2 graph explicitly constrains orientation through odometry `BetweenFactors` and visual projection factors at every keyframe. This interpretation is further supported by ZED's non-monotonic response to GNSS degradation — C0 (PPK-corrected GNSS) yields fewer detections than several C1 and C2 runs — which reveals that fruit count is governed by the SDK's internal rotational accuracy rather than by GNSS input quality, precluding any systematic characterization of its degradation response.

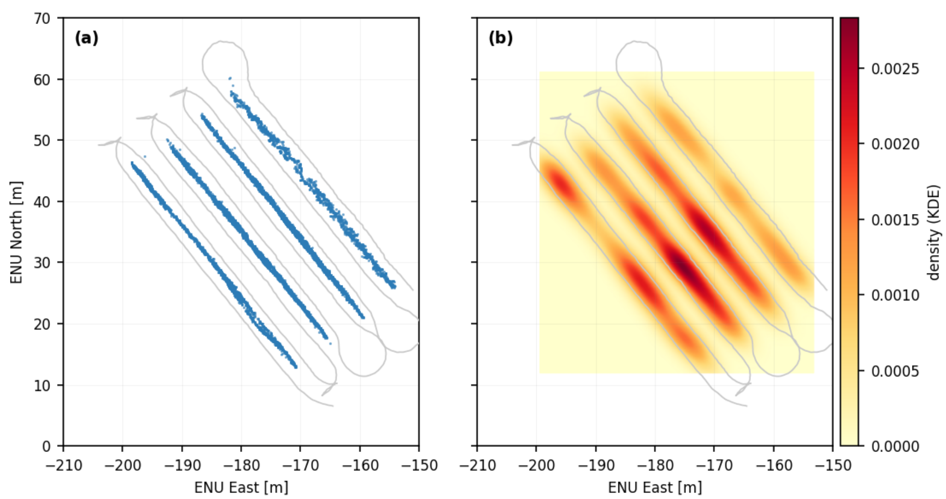

Fig 8. Orchard-scale apple map for the 2025 session: (a) triangulated apple positions (blue) along the scanned rows over the platform trajectory (gray); (b) kernel-density estimate of fruit density (color scale) over the same area

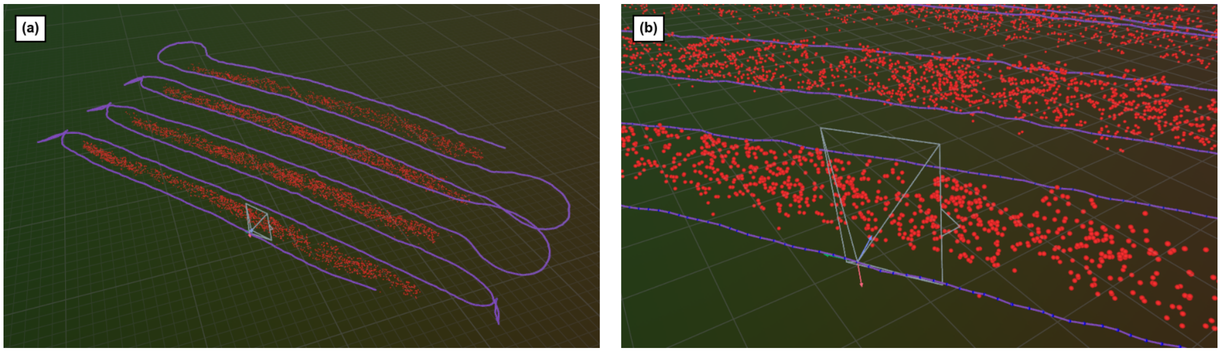

Fig 9. Orchard-scale 3D apple map for the 2025 session: (a) overview showing the estimated camera trajectory (purple line), a single camera frustum for reference, and the reconstructed apples (red spheres); (b) detailed view of the highlighted region in (a), showing individual keyframe poses (blue dots) along the trajectory and the camera frustum at one keyframe.

Fig. 9 shows an example of 3D fruit mapping for the C1 configuration (1 Hz subsampling) using the OV-PGL system. Fig. 9(a) shows the complete estimated trajectory alongside the reconstructed fruit positions (red spheres), while Fig. 9(b) provides a detailed view of a representative region, showing individual keyframe poses and the corresponding camera frustum.

Conclusion

A systematic GNSS degradation protocol was proposed and applied to three VIO-GNSS fusion front-ends processing identical stereo, inertial and degraded GNSS streams from a tractor-mounted platform in a Fuji apple orchard, across two seasons with withheld PPK ground truth. The transparent OpenVINS-GTSAM pose graph delivered the best absolute accuracy, the proprietary SDK fusion the best robustness, and visual landmark factors — beneficial for local map rigidity — were shown to anchor the trajectory to corrupted GNSS once position noise reached a critical threshold, inflating ATE by up to an order of magnitude while remaining internally consistent. Relative pose error was insensitive to both architecture and degradation, confirming that reprojection- or odometry-style metrics cannot certify georeferenced fruit maps. It was further demonstrated that multi-view fruit triangulation recall is a sensitive proxy for trajectory orientation accuracy, and that fruit counting quality under realistic GNSS degradation is driven primarily by orientation consistency rather than by positional accuracy or GNSS signal quality. For operational deployment, GNSS solutions at or below the identified noise threshold — attainable with low-cost RTK — support sub-meter orchard mapping with open components; below that quality, fusion architecture choice dominates the error budget. Future work includes completing the missing runs and the cross-session map-consistency analysis.

Acknowledgments

This research was funded by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), grants 2022/09319-9 and 2024/10267-9.

References

Bharti, D., Faria, L. N., Koenigkan, L. V., Gebler, L., Rossi, A., & Santos, T. T. (2026). Geo-referenced factor-graph SLAM for orchard-scale 3D apple reconstruction and yield estimation. *Agriculture*, 16(7), 764, doi:10.3390/agriculture16070764

Dellaert, F., & Kaess, M. (2017). Factor graphs for robot perception. *Foundations and Trends in Robotics*, 6(1–2), 1–139.

Geneva, P., Eckenhoff, K., Lee, W., Yang, Y., & Huang, G. (2020). OpenVINS: A research platform for visual-inertial estimation. In *Proceedings of the 2020 IEEE International Conference on Robotics and Automation* (pp. 4666–4672). Piscataway, NJ, USA: IEEE.

Grupp, M. (2017). evo: Python package for the evaluation of odometry and SLAM. https://github.com/MichaelGrupp/evo. Accessed 15 June 2026.

Kaess, M., Johannsson, H., Roberts, R., Ila, V., Leonard, J. J., & Dellaert, F. (2012). iSAM2: Incremental smoothing and mapping using the Bayes tree. *The International Journal of Robotics Research*, 31(2), 216–235.

Karaev, N., Rocco, I., Graham, B., Neverova, N., Vedaldi, A., & Rupprecht, C. (2024a). CoTracker: It is better to track together. In *Computer Vision – European Conference on Computer Vision 2024* (pp. 18–35). Cham, Switzerland: Springer.

Karaev, N., Makarov, I., Wang, J., Neverova, N., Vedaldi, A., & Rupprecht, C. (2024b). CoTracker3: Simpler and Better Point Tracking by Pseudo-Labelling Real Videos. arXiv. https://doi.org/10.48550/arxiv.2410.11831

Lei, J., Prabhu, A., Liu, X., Cladera, F., Mortazavi, M., Ehsani et al. (2025). Spatio-temporal metric-semantic mapping for persistent orchard monitoring: Method and dataset. *IEEE Robotics and Automation Letters*, 10(8), 8610–8617.

Mourikis, A. I., & Roumeliotis, S. I. (2007). A multi-state constraint Kalman filter for vision-aided inertial navigation. In *Proceedings of the 2007 IEEE International Conference on Robotics and Automation* (pp. 3565–3572). Piscataway, NJ, USA: IEEE.

Santos, T. T., Koenigkan, L. V., Bharti, D., & Gebler, L. (2025). SEEmear: High-resolution georeferenced 3D mapping methodology for proximal monitoring of permanent crops. In *Anais do XV Congresso Brasileiro de Agroinformática* (pp. 662–674). Online event, Brazil: Associação Brasileira de Agroinformática.

Sturm, J., Engelhard, N., Endres, F., Burgard, W., & Cremers, D. (2012). A benchmark for the evaluation of RGB-D SLAM systems. In *Proceedings of the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 573–580). Piscataway, NJ, USA: IEEE.

Takasu, T., & Yasuda, A. (2009). Development of the low-cost RTK-GPS receiver with an open source program package RTKLIB. In *Proceedings of the International Symposium on GPS/GNSS*. Jeju, Korea: International Convention Center.

Umeyama, S. (1991). Least-squares estimation of transformation parameters between two point patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(4), 376–380.

Wang, C.-Y., Yeh, I.-H., & Liao, H.-Y. M. (2024). YOLOv9: Learning what you want to learn using programmable gradient information. In *Computer Vision – European Conference on Computer Vision 2024* (pp. 1–21). Cham, Switzerland: Springer.

Zhang, Z., & Scaramuzza, D. (2018). A tutorial on quantitative trajectory evaluation for visual(-inertial) odometry. In *Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 7244–7251). Piscataway, NJ, USA: IEEE.