Sensors & Transducers, Vol. 272, Issue 1, April 2026, pp. 23-35

www.sensorsportal.com

Sensors & Transducers

Published by IFSA Publishing, S. L., 2026
http://www.sensorsportal.com

Towards Offline Navigation for Small UAVs Using Lightweight CNNs and Satellite Maps

[1, 2, *] Ludwig KEMPF, [1] Mark UMANSKY, [1] Kevin B. KOCHERSBERGER

[1] Virginia Tech, Uncrewed Systems Laboratory, 160 Inventive Lane, Blacksburg, VA 24061, USA
[2] Technical University of Darmstadt, Institute for Flight Systems and Automatic Control, Otto-Berndt Strasse 2, 64287 Darmstadt, Germany
[*] E-mail: lkempf@vt.edu

Received: 30 Jan. 2026 /Revised: 23 April 2026 /Accepted: 25 April 2026 /Published: 28 April 2026

Abstract: Today there is a wide range of different navigation techniques with additional technologies paving the way for more complex fusion approaches. As the current standard, a combination of Inertial Navigation System (INS) and Global Navigation Satellite System (GNSS) enhanced with state-of-the-art augmentation services provide high accuracy, global coverage, and an acceptable baseline robustness. However, its reliance on signal integrity leaves safety-critical applications vulnerable to spoofing and jamming, while global crises increase the likelihood of such destructive interferences. With self-containment and high deployment volumes in mind, a lightweight vision-based reference module is proposed as a GNSS fallback for seamless integration into small Uncrewed Aerial Vehicles (UAVs). Using offline satellite maps and an ensemble of Convolutional Neural Network (CNN) models, 2D offsets and associated uncertainties are fused through the PX4 autopilot's on-board Extended Kalman Filter (EKF). Software-In-The-Loop (SITL) simulations provide first results on the feasibility and real-time capabilities of the module on a resource-limited platform. Systematic mission parameter variation on terrain data of Virginia Tech's Kentland Farm demonstrates the required minimum flight path stabilization capabilities with an R95 accuracy of 4.5 m to 5.5 m. Even in low-texture environments drift stays bounded, while varying signal loss scenarios reveal current stability limitations. Driven by the integration concept and initial SITL results, the resulting hardware test platform is presented as the foundation for upcoming flight testing.

Keywords: GNSS-denied navigation, Vision-based sensing, Cross-view geolocalization.

1. Introduction

While navigation accuracy used to be a question of technological feasibility, today it is a decision taken based on the performance demands of specific mission objectives. With standardized technology, such as the Global Positioning System (GPS) in conjunction with the Wide Area Augmentation System (WAAS), an R95 accuracy of 1.6 m can easily be achieved compared to over 20 m for a non-augmented best-case scenario with initial GPS deployment in 1995 [1, 2].

Fig. 1 illustrates the R95 metric and puts it into perspective compared to alternatives.

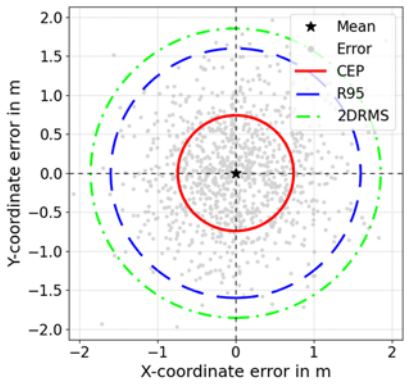


Fig. 1. Common accuracy measures in navigation.

With steady advances in navigation technology – from celestial navigation, to radio-based systems, to the latest intelligent approaches – fusion has always been employed in position estimation. Fig. 2 provides an overview of navigation techniques based on a literature review of methods using passive and active perception modalities. The broadest distinction between navigation techniques can be made by absolute and relative positioning, which are usually combined with each other for complementation, as is the case for a Global Navigation Satellite System (GNSS) with an Inertial Navigation System (INS).

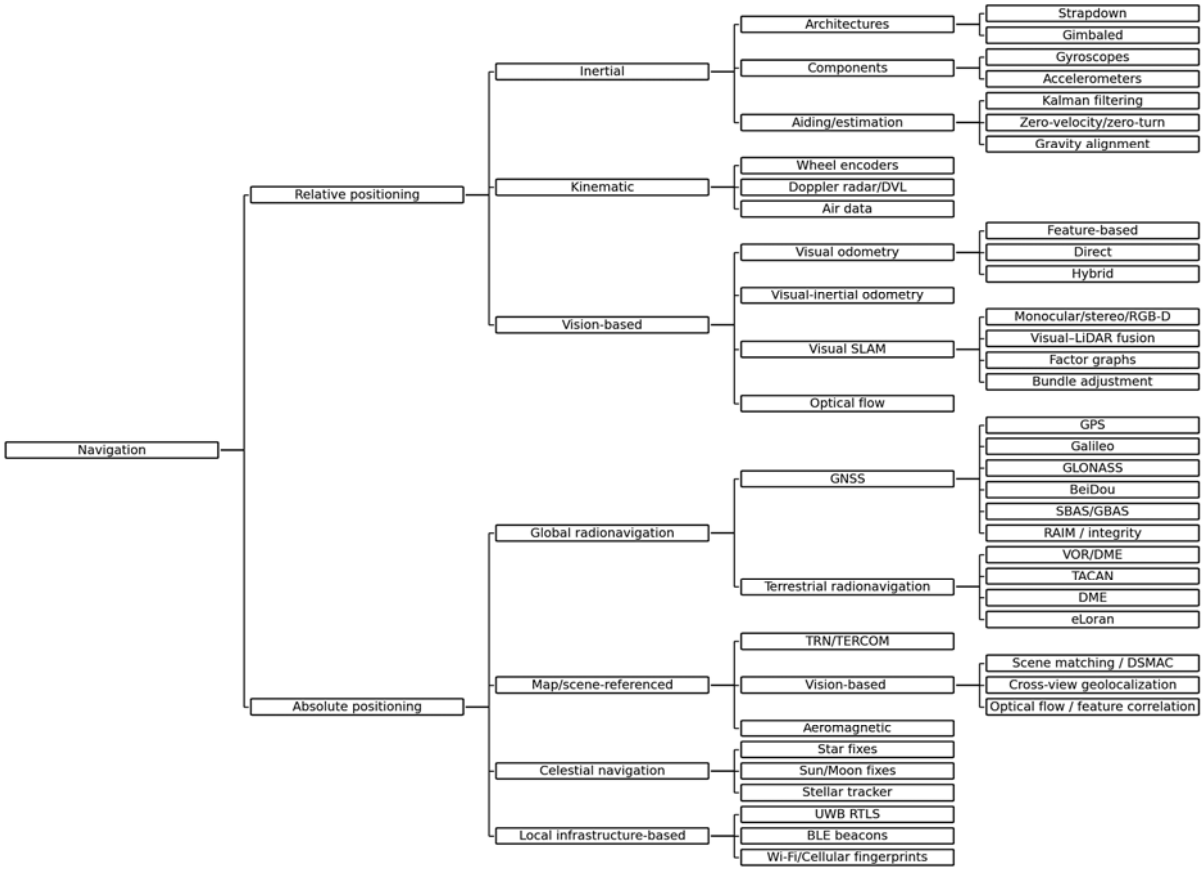


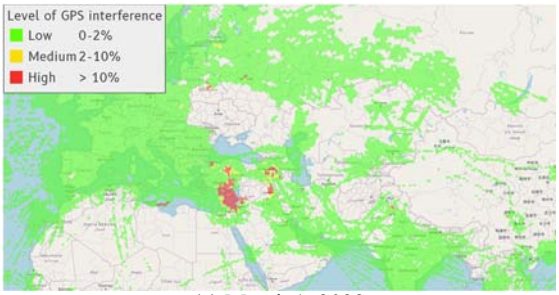
Fig. 2. Overview of navigation techniques.

Despite being self-contained, non-jammable, and non-radiating, standard INS outputs such as those of any stand-alone dead-reckoning system for relative positioning will drift over time due to bias and noise in the sensed signals and their processing. This holds true for high-end sensors and low-end sensors alike, with Fiber-Optic Gyroscopes (FOGs) representing the standard for what is possible in terms of upper price and performance, but they will still drift and require correction information provided by a global position estimator. Cost differences between low- and high-quality systems can reach up to five orders of magnitude, illustrating the economic considerations for high-volume deployment plans [3].

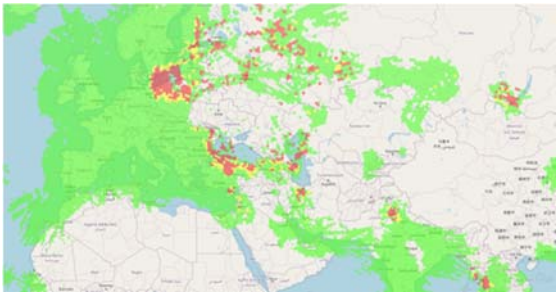
Building upon advanced INS technology, vision-based approaches have augmented and even replaced some conventional systems. Alongside probabilistic techniques, such as Kalman or particle filters, optical and feature-based techniques precede the most recent Artificial Intelligence (AI)-based relative positioning concepts. Nevertheless, even with today's variety of sensors available – such as cameras, barometers, radar, or sonar – each of these imposes its own limitations [4].

In contrast, the spectrum of absolute positioning techniques stretches from the very first celestial navigation approaches to the technological extension of this triangulation concept to global radionavigation, with satellites following the first ground stations and both complementing each other in today's navigation strategies.

However, GNSS depends entirely on signal quality, and Fig. 3 illustrates the increased level of GPS (and other GNSSs) interference today compared to early March 2022. Therefore, with spoofing and jamming becoming major concerns for safety-critical applications, the need for a self-contained GNSS substitute for absolute positioning has grown and scene-referenced navigation has been identified as the most promising class of techniques. Vision-based approaches with lower hardware cost are suitable as a lightweight GNSS alternative for absolute positioning of small Uncrewed Aerial Vehicles (UAVs).



(a) March 1, 2022



(b) February 1, 2026

Fig. 3. Level of GPS (and other GNSSs) interference estimated based on the Navigation Integrity Category (NIC) subscribed to via Automated Dependent Surveillance Broadcast (ADS-B) Exchange API.

1.1. Scene-Referenced Navigation

The very first Terrain Referenced Navigation (TRN) systems, such as Terrain Contour Matching (TERCOM) and Digital Scene Matching Area Correlation (DSMAC), were developed during the Cold War and have been used in GNSS-denied environments ever since, albeit with varying relevance. While TERCOM compares an offline Digital Elevation Model (DEM) to actual radar altimeter readings and is usually used for en-route navigation, DSMAC exploits correlation information of binary images. Consequently, the latter excels on terrain with less variation in ground elevation but structured textures present and is favored for final approach scenarios. Typically, R95 accuracy is approximately 3 times higher for DSMAC [5, 6]. However, reliable operational specifications for TERCOM and DSMAC are not publicly available and reported accuracy values should therefore be treated as indicative estimates rather than strict performance guarantees.

Aeromagnetic navigation is formally not classified as TRN but it is still a field-referenced method with increasing relevance for global self-contained operation due to recent advances in Quantum Scalar Magnetometers (QSMs). These are required to identify minor anomalies in Earth's magnetic field masked by any aircraft's electromagnetic noise. However, when expressed in terms of R95 accuracy, the corresponding radii are more than two orders of magnitude greater than those for DSMAC or legacy single-frequency GPS without any augmentation [7].

Substantial advances in Artificial Intelligence (AI), Machine Learning (ML), and in particular Neural Networks (NNs), have enabled a variety of even more powerful scene-referenced approaches, with the two predominant categories being Template and Feature Matching (TFM) and Semantic Mapping and Recognition (SMR). As a high-level cognitive process, the latter targets content rather than structure alone, exploiting the available data to a greater extent.

As a TFM approach, cross-referencing processed RGB images with a DEM using Simultaneous Localization and Mapping (SLAM) with bundle adjustment on sufficiently high image quality, a converted R95 accuracy of less than 1 m is achievable [8]. Alternatively, using Synthetic Aperture Radar (SAR) images in combination with the range–Doppler algorithm, a converted R95 accuracy of approximately 2.5 m has been demonstrated under ideal conditions, provided that high-resolution SAR reference maps (or known reflectors) are available [9]. Fusing precise TRN, a radar altimeter and LiDAR, converted R95 accuracies between 6.4 m and 17.5 m have been measured for a variety of mission parameters [10].

In the SMR domain, using only a downward-facing monocular camera, an altimeter and a compass to compare measurements against open-source vector maps, a converted R95 accuracy of 10 m has been reported [11]. Employing object-level geolocalization, point-matching Convolutional Neural Networks (CNNs), and pixel-level refinement, a converted R95 accuracy of 0.2 m is achievable under highly favorable conditions, noting that performance depends on the evaluation dataset [12].

Table 1 compares a selection of relevant approaches. Note that, while the accuracy values have been estimated based on reasonable assumptions, they do not correspond to the metrics in which performance was originally reported. More importantly, none of the techniques have been evaluated under identical conditions (and rarely even on the same datasets).

Table 1. Converted R95 accuracies assuming an isotropic Gaussian error with Rayleigh distribution.

Method	R95 accuracy in m
Others [8, 9, 10, 11, 12]	0.2 - 30
GNSS [1, 2]	1.6 - 36
DSMAC [6]	≤ 20
TERCOM [5]	~ 60
Magnetic anomalies [7]	200 - 800

1.2. Design Approach

Guided by the introduced limitations, the proposed concept promotes an easily integrable and deployable module with a high absolute positioning accuracy threshold to bound drift, uncertainty estimation for stable Extended Kalman Filter (EKF) fusion, and real-time capability on a small UAV platform with limited computational resources. The system scales to

different mission scenarios, maintains reasonable robustness, and uses simple, inexpensive sensors, as well as satellite imagery of only moderate resolution.

With nearly 11000 small commercial satellites (≤ 1200 kg) launched between 2015 and 2024, not only has communication technology advanced, but the commercial availability of high-resolution satellite imagery has also increased significantly [13]. Planet Labs PBC alone launched more than 500 satellites during this period, and its SkySat constellation of 15 units can revisit almost any location up to ten times per day [14]. Its Dove fleet provides near-daily coverage of the entire globe [15].

Only through these recent trends providing multiple image updates of the same location every day can the introduced referencing method become a realized concept for future navigation strategies. Therefore, success ultimately depends on providers like Planet Labs PBC or BlackSky (another smaller yet highly relevant provider) supplying new data at high frequencies and reasonable costs, with current trends being very promising [16, 17].

2. Methodology & Modeling

The reference module is designed to directly integrate with the widely accepted and highly compatible PX4 open-source autopilot and its built-in EKF2, which exchanges information with the reference module via ROS2. Fig. 4 shows the methodology blueprint of the reference module. With an actively gimbaled, north-aligned RGB camera providing an observation, the observed scene is then compared to a window of the same extent as the camera frame extracted from a reference satellite map based on the last EKF position estimate. Next, both images are preprocessed and passed to the estimator ensemble, which outputs a mean 2D offset $\boldsymbol{\mu}_{\Delta x}$ alongside its covariance $\boldsymbol{\Sigma}_{\Delta x}$. Using the built-in `VehicleOdometry` topic, the EKF2 on the PX4 fuses the probabilistic estimate and ensures correct allocation in time using the reported timestamps.

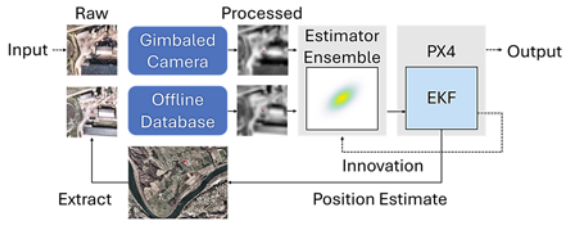


Fig. 4. Methodology blueprint.

2.1. Design Requirements

The module is designed to provide absolute position reference not only when flying across geometrical human-made structures, but also when flying across organic textures. However, any extreme alteration of the terrain or substantially compromised visibility (bright reflections, dark shadows, changed vegetation, clouds, fog, smoke) will put mission success at risk. The employed model ensemble is therefore specialized to the defined mission settings – in particular the type of terrain – and a broader validity range can only be established through ensemble pooling. Since the system is equipped only with an RGB camera, nighttime operations are also currently not supported. Nevertheless, other ranges and resolutions of the electromagnetic spectrum alongside additional filtering techniques are to be explored, potentially leading to extended capabilities in the future.

Since the cruise segment usually dominates the mission profile, it is the most relevant phase for navigation problems. Therefore, the initial reference module explicitly excludes any radical UAV maneuvers and targets a constant altitude Above Ground Level (AGL) of 120 m, providing sufficient spatial context for absolute positioning. The only change in flight conditions that will be further investigated within the scope of this work are wind gusts affecting gimbal stability. A conservative propagation percentage of 50 % is assumed for simulation leading to slight image out-of-plane distortions from changes in pitch and roll that are modeled using a small-angle approximation. These changes are typically below 5°, with occasional rotations up to 7°. All other effects are neglected. This approach ensures a minimum level of robustness under nominal flight conditions, but more variation during training can later enable true adaptability to greater AGL fluctuations or additional effects from a violated cruise flight condition.

Another aspect of the reference module to be closely monitored is the data used for training, currently limited to 9 satellite image sources during this proof-of-concept phase. The chosen sources and the ones thereby excluded determine the strengths and limitations of the employed models. Therefore, it is crucial to acknowledge the inherent bias from data availability and understand its effects on performance with varying terrain data during testing. Additionally, constraints such as overcast weather or satellite constellation limitations during data acquisition are two very immediate factors to be aware of.

Initial testing is conducted in Software-In-The-Loop (SITL) simulation before moving to the hardware stack for flight testing under real-world conditions. These first results provide the baseline for the following compatibility aware hardware adjustments but should at the same time be treated as highly preliminary – in particular as regards latency evaluation. EKF2 accepts signals as old as 1000 ms, allowing for seamless SITL testing without optimizing, but a translation from Python, used for fast prototyping, to another more efficient language, such as C++, may be necessary to meet the 10 Hz target data rate for final deployment.

2.2. Data Sampling

A total of six satellite images of Virginia Tech's Kentland Farm were selected based on its visual correspondence to the test mission from a pool of 9 images (Fig. 5) provided by BlackSky [18]. While the reference satellite map is selected to reflect the latest changes in terrain (1), additional sources are identified with training convergence in mind but still allowing for anticipated fluctuation in lighting conditions (2-4). At the same time, robustness is maintained by also including more significant vegetation and foliage changes (5, 6).

Overall, this yields a 200000-sample dataset of paired satellite windows and surrogate camera frames, downsampled from 1 m to 3.8 m per pixel for low-pass filtering and converted to greyscale. Within the defined sampling domain – enabling a later cross-area evaluation – window centers and associated offsets are randomly chosen. Offsets are bound to ±10 m (8.3 % of the frame size) on both Cartesian axes to optimally use the limited complexity of the CNN.

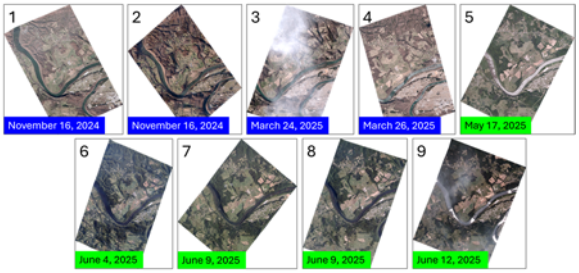


Fig. 5. Raw satellite images sorted by date. Blue labels show low vegetation coverage, and green labels show high vegetation coverage corresponding to seasonal change.

2.3. Gaussian Model

The individual models are trained following a standard supervised curriculum learning approach, with the first 10 epochs training only the means using the mean model, and the next 30 epochs extending the output to uncertainties using the Gaussian model and a custom Negative Log Likelihood (NLL) loss. Fig. 6 shows the overall high-level architecture.

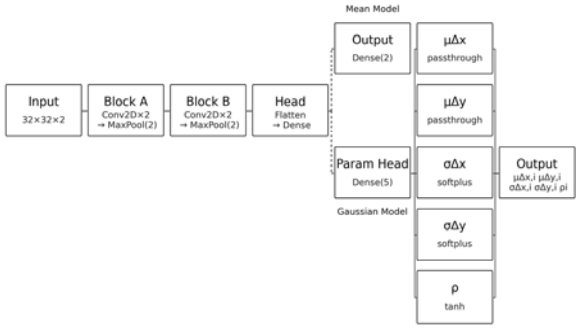


Fig. 6. CNN architecture.

The architecture's overall simplicity allows a straightforward interpretation of performance, representing a solid baseline for the future implementation of more advanced structures and optimization of estimate quality.

An additional calibration term enforces the average squared Mahalanobis distance m^2 to remain close to 2 with the objective of well-scaled uncertainty coverage and Fig. 7 shows the Cumulative Distribution Function (CDF) of m^2, representing the probability of the value being less than or equal to a given threshold. The empirical curve computed from the validation data generally follows the theoretical distribution, indicating that the estimated uncertainties have the correct order of magnitude. Generalization of calibration must later be assessed on the test data.

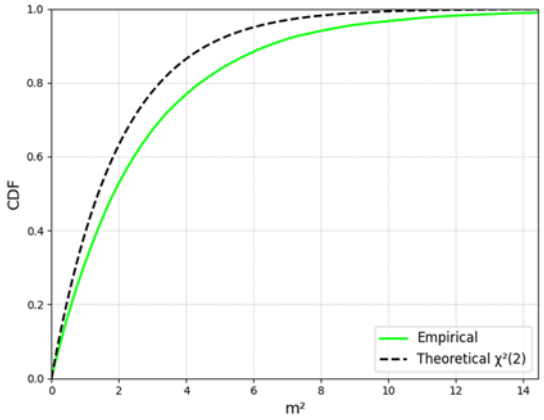


Fig. 7. Calibration curve of the Mahalanobis distance squared m^2 of a typical training run.

2.4. Estimator Ensemble

To constitute the estimator ensemble, 3 separate CNN models are trained on the same dataset but with different sample shuffling, so that the variance across their predictions $\boldsymbol{\Sigma}_{\mu_{\Delta x}}$ provide an inexpensive estimate of systematic uncertainty from model disagreement. This component can then be combined with the average of the stochastic uncertainty $\overline{\boldsymbol{\Sigma}}_{\Delta x}$ of all models M to form the total effective covariance:

$$\boldsymbol{\Sigma}_{eff,\Delta x} = \overline{\boldsymbol{\Sigma}}_{\Delta x} + \boldsymbol{\Sigma}_{\mu_{\Delta x}}, \quad (1)$$

with

$$\overline{\boldsymbol{\Sigma}}_{\Delta x} = \frac{1}{M}\sum_{i=1}^{M} \boldsymbol{\Sigma}_{\Delta x,i}, \quad (2)$$

and

$$\boldsymbol{\Sigma}_{\mu_{\Delta x}} = \frac{1}{M}\sum_{i=1}^{M}\left(\boldsymbol{\mu}_{\Delta x,i} - \overline{\boldsymbol{\mu}}_{\Delta x}\right)\left(\boldsymbol{\mu}_{\Delta x,i} - \overline{\boldsymbol{\mu}}_{\Delta x}\right)^{T} \quad (3)$$

The diagonal elements of the covariance ultimately guide EKF updates and signal fusion on the PX4.

Although systematic uncertainty is already modeled, there remains the risk that all models in an ensemble may share the same bias. This results in an unreasonable estimate that, in turn, produces large innovations within the EKF2. To safeguard against such scenarios and maintain filter consistency, an innovation-based covariance inflation is applied, with the normalized innovation test ratios τ_x and τ_y fed back into the reference module. The ratios per axis are calculated from the innovation ν, describing the disagreement between EKF motion prediction and the estimator ensemble output, and the innovation gate γ. The latter is an integer defining the acceptable multiples of σ (eq. (4)) and is set to 5 – the PX4 default value.

$$\tau_i = \frac{\nu_i^2}{\gamma^2 \sigma_i^2} \tag{4}$$

When these ratios become excessively high, $\boldsymbol{\Sigma}_{eff,\Delta x}$ is inflated by a factor λ, as defined in eq. (5). This adaptive mechanism enhances numerical stability and preserves statistical consistency. Ideally, these ratios remain below 0.5. Values exceeding 1 indicate potential overconfidence of the estimator and lead to the rejection of the measurement.

$$\lambda = 1 + \max(0, \tau_{max} - 0.5), \tag{5}$$

with

$$\tau_{max} = \max(\tau_x, \tau_y) \tag{6}$$

This logic has an intuitive implication: Within a single update step, substantial disagreement is tolerated only when uncertainty is high. Nevertheless, it does not prevent destabilization when motion prediction and estimator output are close to each other but both wrong.

The flowchart in Fig. 8 summarizes the estimator ensemble subprocess, outlining the most important steps.

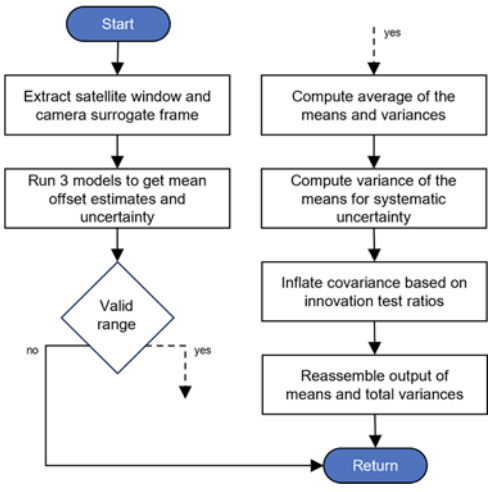


Fig. 8. Flowchart of the estimator ensemble subprocess.

3. Simulation Results

For the systematic evaluation of the reference module, real-time SITL simulation with the X500 Quadrotor model in Gazebo Harmonic is performed on the Seeed Technology reComputer J4012, which is based on the NVIDIA Jetson Orin NX 16 GB module [19]. A dedicated geo-mosaic of Kentland Farm together with its corresponding DEM for relief modeling (virtual object displacement) constitutes the terrain data for simulation. This allows for a simulated camera view for a variety of mission scenarios (Fig. 9). The raw data was collected using a SenseFly eBee fixed-wing mapping drone.

(a) Geo-mosaic image (b) DEM

Fig. 9. Geo-mosaic and associated DEM for camera simulation during SITL testing.

SITL testing starts with an open-loop reference mission, assessing the reference module's performance compared to ground truth but still using GNSS for positioning. The reference is then modified step by step to explore the following modifications:

1. Loop closure (GNSS disabled);
2. Mission location (cross-area);
3. Linear flight path geometry;
4. Wind gust exposure;
5. Signal loss:
   a. Linear w/ wind;
   b. Linear w/o wind;
   c. Circular w/o wind.

This process identifies relevant trends and examines robustness of the proposed reference module by better understanding its strength and limitations. Fig. 10 shows the in-area reference mission, with at least 75 % of the terrain covered by pastures and farmland. The Ground Control Points (GCPs) indicate the centers of the 200 m sampling radii for the employed dataset. As for all subsequent variations, the ground speed is set at 5 m/s.

The statistics in Fig. 11a demonstrate a cumulative R95 accuracy of 4.7 m along the course of the reference mission, including the spike about 400 s into the flight. Given the final resolution of 3.8 m, the mean bias of 1.2 m stays in the sub-pixel range. At the same time, the coverage statistics reveal the reference module's constant overconfidence on unseen data. There is a significant under-coverage compared to the expected percentages (hereafter denoted *exp*) for all three bands and Fig. 11b shows the sudden disagreement between the EKF2's prediction and the external position estimate after a series of lower

innovation test ratios. However, test ratios stay within a healthy regime throughout the mission.

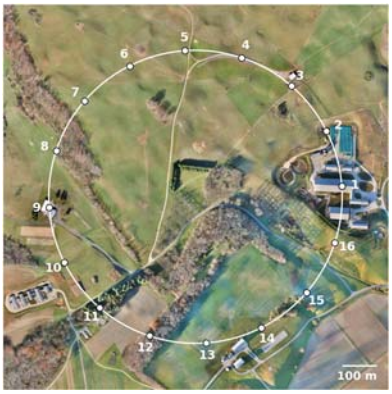
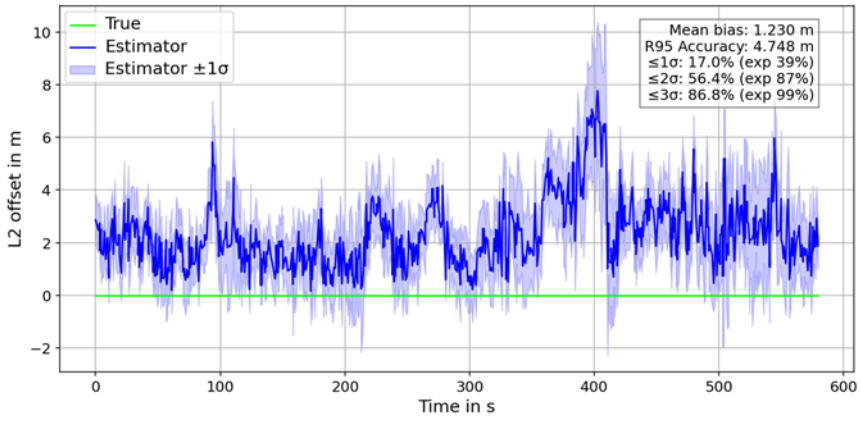


Fig. 10. Route of the reference mission. GCPs indicate the centers of the 200 m sampling radii for the employed training dataset.

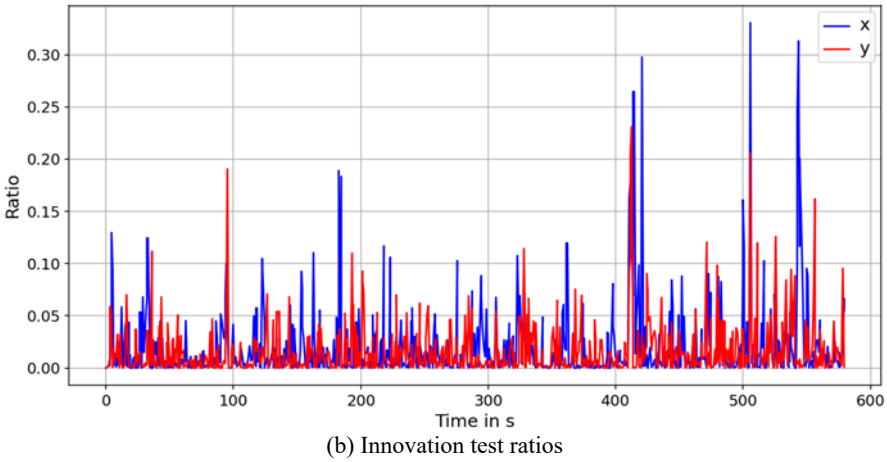
3.1. Loop Closure

When disabling GNSS and thereby closing the control loop, R95 accuracy declines, with a value of 5.1 m for the selected run. This represents an average scenario, but results vary by approximately 0.5 m across different runs.

As anticipated, being a function of the ground textures, coverage is almost identical to the reference mission. Although the noisy estimator outputs induce frequent deviations from the target, the position estimate remains stable and within the indicated accuracy. Fig. 12 shows a typical flight-path anomaly caused by noisy estimates. The blue band indicates the standard deviation obtained by projecting the effective covariance $\boldsymbol{\Sigma}_{eff,\Delta x}$ onto the unit vector perpendicular to the trajectory. It illustrates that while the estimator tracks the overall trend, the 1σ band does not provide sufficient uncertainty coverage.



(a) L2 offset of the estimated position with respect to the truth. The blue band around the estimator represents the radial standard deviation obtained by projecting the effective covariance $\boldsymbol{\Sigma}_{eff,\Delta x}$ onto the offset direction unit vector



(b) Innovation test ratios

Fig. 11. Time metrics for the reference mission.

Computing the relevant ground truth statistics, Fig. 13a uncovers the cause of drift at GCP 12 (400 s into the flight) for the reference and the loop-closure mission: there is a significant bias in the negative x-direction from training-to-truth inconsistency.

Although Fig. 11a earlier revealed significant under-coverage, Fig. 13b now shows that the squared Mahalanobis distance m^2 has the correct order of magnitude in most regions, which is a very promising tendency in the data.

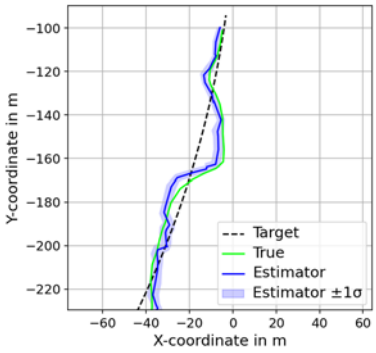
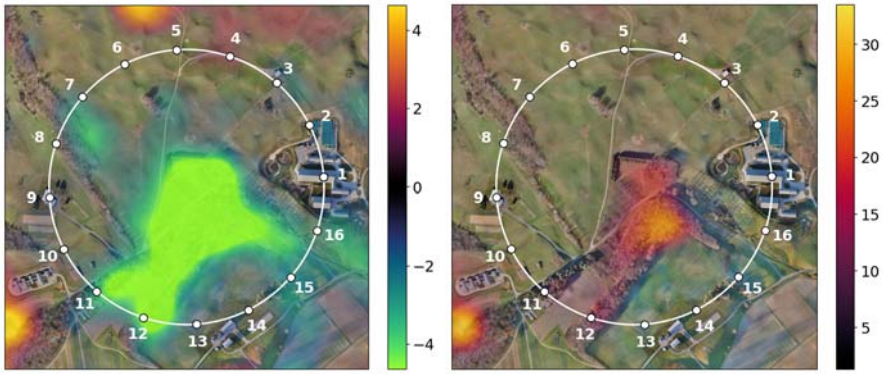


Fig. 12. Flight path anomalies for the loop-closure mission.

3.2. Mission Location

While all simulation is conducted as an out-of-sample test on unseen data, the alternate location additionally represents a cross-area evaluation on a region that contains mostly sections the employed models have not been trained on. In contrast to the reference mission, this mission leads across forests and has a tree coverage of 20 %, with the most challenging part being the continuous nature of the relevant passage. GNSS is disabled (as for all missions except for the reference mission).



(a) X mean error in m (b) Mean squared Mahalanobis m^2 (exp 2)

Fig. 13. Ground truth statistics for the reference mission, generated by performing repeated Monte Carlo simulations of the estimator ensemble on discrete grid cells across the geo-mosaic. Grid resolution is 60×60 m, with 30 runs per cell and a maximum offset of 5 m for regularization.

Fig. 14 illustrates the UAV's behavior at the alternate mission location. After a stable flight, only showing the typical anomalies, a significant and persistent disagreement between the truth and the estimate begins to manifest up to a destabilization.

With the L2 offset Fig. 15 illustrates the gradual unfolding of the destabilization in the time domain.

The 1σ-band indicates a generally larger reported uncertainty over the forest region, resulting in lower innovation test ratios and smaller updates of the motion-predicted position. Consequently, the EKF2 leans more heavily on these predictions and becomes overconfident, with the destructive divergence appearing around 370 s into the flight. During this phase, the motion model and estimator ensemble output agree with each other while both being wrong.

While the models still perform well – even in previously unseen areas – if similarity is provided, a significant inconsistency for an extended period will break the EKF logic and therefore induce a high risk of destabilization (Fig. 16).

3.3. Linear Flight Path

Changing the flight path geometry of the mission could be nontrivial, since a straight line requires less control effort than a continuous circle. However, with the PX4 already being a highly optimized autopilot with a sophisticated cascade control architecture, the flight path geometry doesn't affect the performance of the reference module for a continuous position signal [20]. Later investigation shows that it becomes relevant for a signal loss scenario.

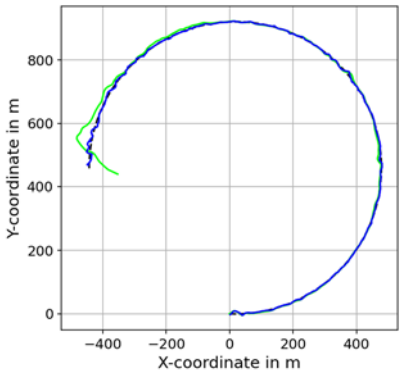


Fig. 14. Destabilization across forest.

3.4. Wind Gust Exposure

For a robust system, wind should only have a minor influence on its functionality. Wind gusts

represent one of the most challenging scenarios as they can have a significant impact on the flight state – particularly on the pitch and roll of the quadrotor model. These dynamic attitude changes must not compromise the output quality of the estimator ensemble under nominal flight conditions.

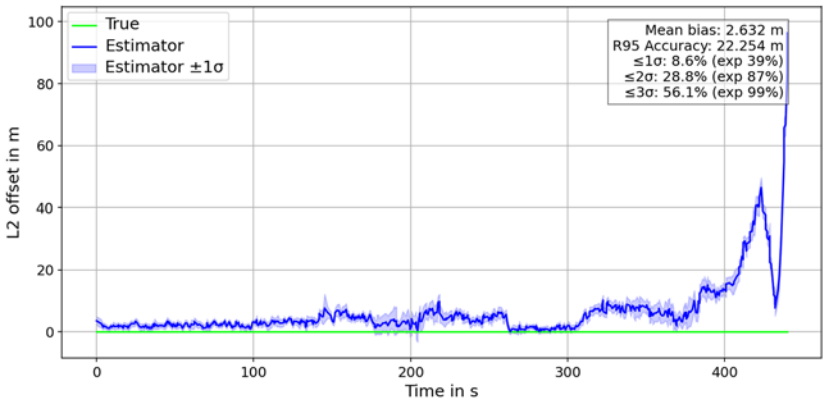


Fig. 15. L2 offset of the estimator relative to the truth for the alternate mission location.

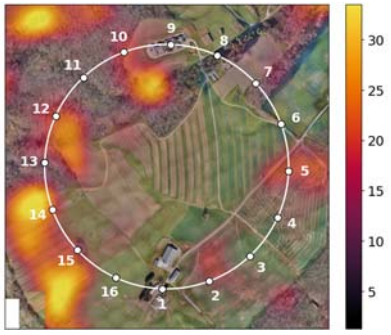


Fig. 16. Mean squared Mahalanobis m^2 for the alternate mission location (exp 2).

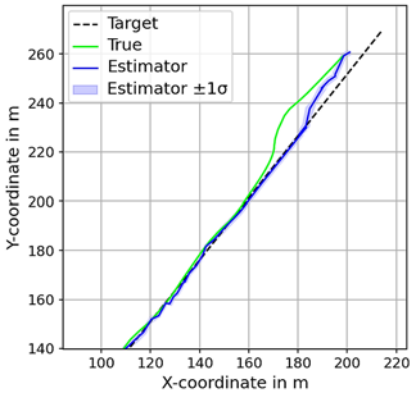
Despite the actively controlled gimbal being part of the system design, strong gusts may still impose a real risk. Therefore, a conservative propagation percentage of 50 % is assumed for simulation and slight image distortions from changes in pitch and roll are modeled using a small-angle approximation. These changes are typically below 5°, with occasional rotations up to 7°. Yaw is assumed constant.

Gazebo Harmonic has limited aerodynamic modeling capabilities compared to Classic (end of life in 2025), so the included `WindEffects` plugin is empirically tuned in a way that simulates a realistic wind gust scenario for a small UAV. Starting from a 7 m/s constant wind baseline, strong gusts are induced.

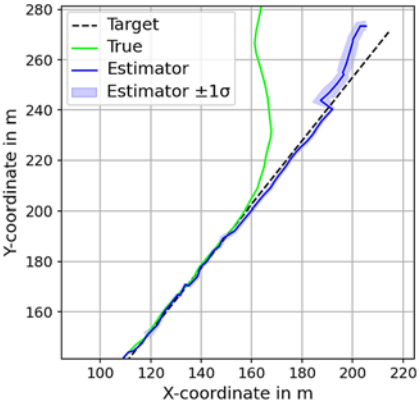
It shows that high-frequency fluctuations of the position observation caused by wind gusts are absorbed by the noise of the estimator ensemble itself and therefore do not compromise the reference module's estimate quality.

3.5. Signal Loss

Undetected signal loss or skipped measurements over extended periods pose potentially catastrophic consequences for flight stability and mission success. Although PX4 will initiate an auto-land once a missing external position estimate is detected, it will continue flying if the signal appears valid. This scenario is simulated with a 10 s blackout and a corresponding estimate freeze. Fig. 17 shows the behavior without wind and with additional wind gust exposure.



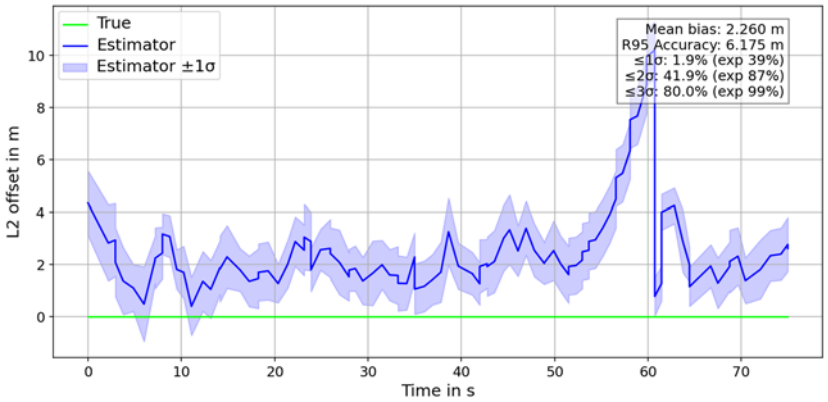
(a) Recovery without wind



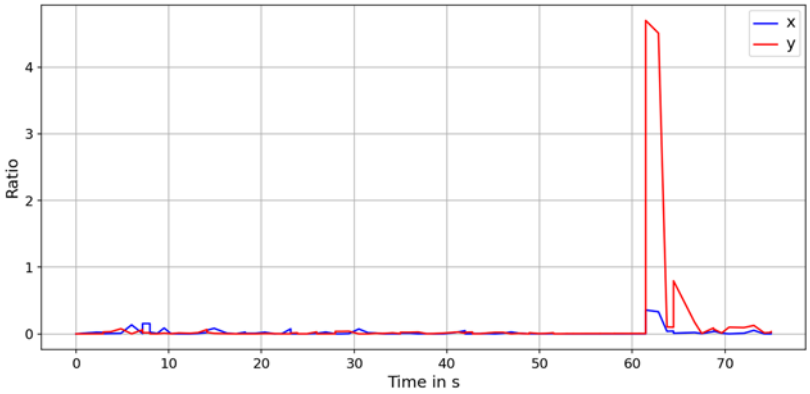
(b) Destabilization with wind

Fig. 17. Behavior during a 10 s signal loss Initiated around x = 150 m.

While the system generally destabilizes under wind-gusts conditions, even on a linear flight path, it has a high likelihood of recovering when no gusts are present. On a circular flight path without wind, the UAV is less likely to recover after a signal loss, although it is possible as captured in Fig. 18.



(a) L2 offset of the estimator relative to the truth



(b) Innovation test ratios

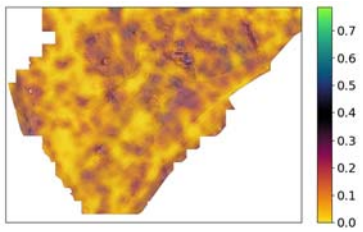
Fig. 18. Time metrics for signal loss after 50 s on the circular path without wind.

The notable spike in the y innovation test ratio indicates the estimator ensemble having regained a valid input signal. However, the EKF still rejects it due to its magnitude. Once innovation-based covariance inflation takes effect, the filter accepts the estimator outputs again, and PX4 regains control.

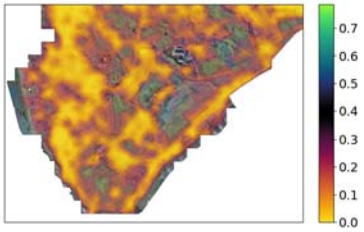
4. Future Test Design

Overall R95 accuracy for the different test cases ranges from 4.5 m to 5.5 m, with flights across distinct human-made structures typically improving accuracy by about 50 %. At the same time, ground truth evaluation suggests that despite performance variation by terrain, the estimator ensemble's baseline uncertainty has the correct order of magnitude to provide long-term EKF consistency.

While additional ablation studies are still to be conducted, it shows that using an ensemble instead of a single model is a promising approach, especially for distinct textures (Fig. 19). However, extended forest coverage imposes a high risk of destabilization without further measures taken. Prior automated terrain classification and specialized CNN models could improve coverage in more complex regions.



(a) 1σ coverage for a single model (exp 39 %)



(b) 1σ coverage for the ensemble (exp 39 %)

Fig. 19. Coverage overview for a single model and the estimator ensemble.

For future testing, instead of looking at a realistic mission scenario with heterogeneous terrain, individual ground textures should be evaluated by simulating test missions on homogenous terrain, with each simulation using a model ensemble solely trained on representative samples. Together with stronger out-of-distribution testing, this would allow for a systematic performance evaluation by terrain, therefore providing the foundation for purposefully increased model and model ensemble complexity.

Additionally, the reference module's current design does not provide any global search – as opposed to GNSS allowing full localization at any point in time. With the proposed basic CNN architecture, results are easier to interpret but the module has a very limited receptive field of ±10 m, providing a drift-bounding reference for absolute positioning only when a rough initial guess is available for reference. Therefore, future test design should evaluate a cascade approach with sequential focusing of the receptive field, thereby refining the search. Visual Transformers (ViTs) could be used for an upstream global search instead of traditional CNNs.

Although closed-loop flight tests had not been performed at the time of article submission, the Tarot T960 hexacopter platform was selected as the base for integrating the rest of the hardware system for Hardware-in-the-Loop (HITL) simulation and future flight tests, as shown in Fig. 20. It provides a payload capacity of about 4 kg and flexible mounting options on the frame are available for adding additional devices such as computers, radios, and sensors.

Fig. 20. The integrated Tarot T960 with camera payload and computing platform.

The Tarot's six brushless motors are controlled by a PixHawk Cube Black (Fig. 21) running the latest stable PX4 version 1.16.0, the exact same version used for SITL evaluation. It connects to the Electronic Speed Controllers (ESCs) for commanding the motors based on its control inputs or preplanned autonomous mission.

With an observed loop rate of 5.7 Hz on average (min 3 Hz, max 10 Hz), the target data rate of 10 Hz is not currently achievable. Since model inference alone takes about 25 ms per cycle, a transition to the Graphics Processing Unit (GPU) and from Python to embedded C++ cannot be the only modification. The NVIDIA Jetson Orin AGX 32GB is selected to replace the NVIDIA Jetson Orin NX 16 GB as an enduring main computing platform. The Orin AGX consumes up to 40W while delivering up to 200 TOPS of processing power, compared to 157 TOPS available in the Orin NX 16GB [29]. The Orin AGX is connected to an Auvidea X230 carrier board, providing the required mechanical, electrical, and communications interfaces for integrating into the Tarot (Fig. 22).

Fig. 21. PixHawk Cube Black [21].

Fig. 22. Auvidea X230 carrier board with Orin AGX [22].

The carrier board connects to the PixHawk Cube Black over a standard serial connection and to the Tarot's main camera sensor over gigabit ethernet. An additional code and language optimization might still be necessary to minimize the overhead from statistical computations of the model ensemble.

Instead of just using an RGB camera, Infrared (IR) could be used for nighttime missions and polarization information could be exploited to improve vision in foggy scenes [23, 24]. The latter will be implemented in future testing, with the primary sensing input being a LUCID Vision Phoenix PHX050S1-QC polarimetric camera. The camera contains a 5MP Sony IMX264MYR CMOS sensor, which is a color sensor that has an array of micro-lenses with linearly polarizing filters on each 4-pixel super-pixel (Fig. 23). It is powered either over its gigabit ethernet connector via PoE or directly via DC voltage.

A custom camera driver has been written using the LUCID Vision Software Development Kit (SDK) to connect to the camera, configure its acquisition parameters, and produce a standard ROS2 image interface.

While SITL simulation shows that gust-induced fluctuations in pitch and roll of less than 7° are tolerated (max 3 % pixel shift), in-plane camera variation from yaw imposes a real challenge. The ensemble is not trained on off-axis distortions. Therefore, to provide a stable observation, the camera is mounted to a 3-DOF HDAir Studio MR-S2 gimbal keeping the sensor north-aligned and pointed downwards in the nadir direction during typical flight maneuvers. The gimbal is preconfigured which removes any need for active external command and control of the gimbal.

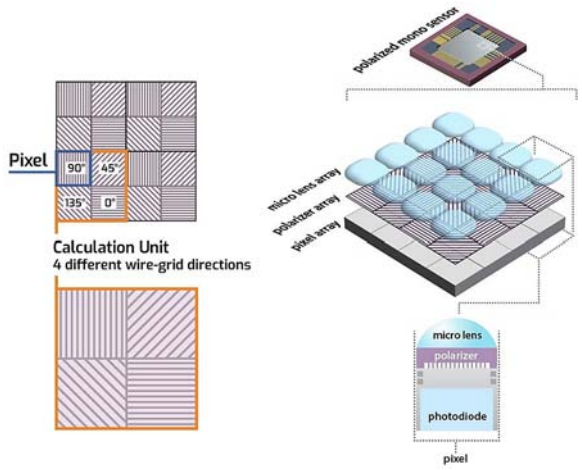


Fig. 23. Sony IMX278 Polarization Sensor [25].

Alternatively, an orthorectification algorithm could be a cost-effective gimbal replacement for a future flight stack and has been proven feasible in past research [26]. Nevertheless, to ensure robustness against long-term magnetometer inaccuracies, an additional yaw Degree Of Freedom (DOF) must be included in the training of the models themselves.

Like in SITL simulation, all communication between the PX4 and any external nodes is established through the micro Unified eXtensible Resource Coordination Entity Data Distribution Service (uXRCE-DDS) as illustrated in Fig. 24.

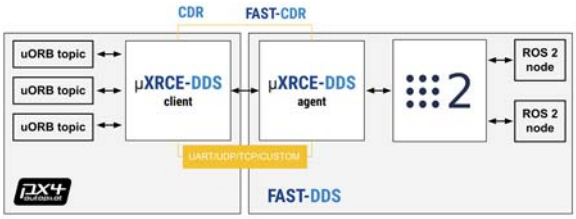


Fig. 24. PX4/ROS2 link over uXRCE-DDS [27].

5. Conclusion & Outlook

While dynamically changing terrain such as tree covered regions is identified as a major challenge, the proposed reference module can still effectively bound drift when flying across low-texture farmland and pastures. The overall R95 accuracy stays between 4.5 m and 5.5 m and the squared Mahalanobis distance has the correct order of magnitude for most of the test terrains. Features persisting across training and truth are the fundamental precondition. However, a systematic classification of terrain-performance relationship is required to push the reference module towards its full potential and boost robustness of uncertainty coverage on unseen data. Additionally, only an upstream global search would transform the current module to a full GNSS fallback. Constrained computational efficiency and sensitivity towards yaw must be addressed as well. Comprehensive benchmarking against alternative techniques, additional ablation studies, and larger datasets for training and testing are required to further substantiate the results and identify potential improvements.

Initial integration tests of the Tarot T960 base flight platform, Orin computing system, camera sensor and gimbal, camera driver, and PX4 communications have been completed, with short test hop flights successfully executed to verify the overall stability of the mechanical and electrical integration of the payloads. The next integration step is to close the loop with the reference module providing absolute position data in real time to the PX4's EKF2. This will enable iterative experiments with actual flight and sensor data, as well as evaluation and extension of the results from SITL simulations completed thus far.

Acknowledgements

The presented research is based on the first author's thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Mechanical Engineering at Virginia Tech (USA) and Technical University of Darmstadt (Germany). The thesis is currently under embargo.

Fundamental findings have already been presented at the *2nd International Conference on Drones and Unmanned Systems (DAUS' 2026)*, with its associated conference proceedings [28]. This article extends the proceedings as follows:

1. Additional background on state-of-the-art navigation techniques and recent motivation is provided;
2. The integration of adaptive covariance inflation is explained and the full model ensemble process illustrated in a flow chart;
3. All scenarios simulated so far are discussed, now including a GNSS-enabled reference mission, a variation in flight path geometry, and wind-gust scenarios with and without signal loss;
4. Future steps in SITL simulation are explicitly derived based on the limitations revealed from the initial results;
5. A specific hardware setup has been developed to provide a test platform for HITL simulation and flight tests in the field.

The research is funded by the National Science Foundation Center for Autonomous Air Mobility and Sensing (CAAMS) and supported by a fellowship of the German Academic Exchange Service (DAAD).

References

[1]. Global Positioning System precise positioning service performance standard, *United States Department of Defense*, 2007.

[2]. Global Positioning System Wide Area Augmentation System (WAAS) performance standard, *Federal Aviation Administration*, 2008.
[3]. Advanced Navigation, MEMS vs FOG: What inertial system should you choose?, https://www.advancednavigation.com/tech-articles/mems-vs-fog-what-inertial-system-should-you-choose/
[4]. I. Jarraya, A. B. Mohammed, H. N. Noura, O. Salman, et al., GNSS-denied unmanned aerial vehicle navigation: Analyzing computational complexity, sensor fusion, and localization methodologies, *Satellite Navigation*, Vol. 6, Issue 1, 2025, 9.
[5]. Worldwide equipment guide, Volume 2: Airspace and air defense systems, *Department of the Army*, 2011.
[6]. J. C. O'Halloran (Ed.), Jane's Weapons: Strategic 2015-2016, *IHS Global Inc., Coulsdon*, 2015.
[7]. M. Muradoglu, M. T. Johnsson, N. Wilson, T. Navickas, et al., Quantum-assured magnetic navigation achieves positioning accuracy better than a strategic-grade INS in airborne and ground-based field trials, *arXiv*, 2025, arXiv:2504.08167.
[8]. X. Wan, M. Zhang, Y. Wang, B. Li, Terrain aided planetary UAV localization based on geo-referencing, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 60, 2022, pp. 1-18.
[9]. C. Lindstrom, J. L. P. Barker, T. E. Humphreys, GPS-denied navigation aided by synthetic aperture radar using the range-Doppler algorithm, *NAVIGATION: Journal of the Institute of Navigation*, Vol. 69, Issue 3, 2022, 533.
[10]. J. Lee, C.-K. Sung, J. Oh, K. Han, et al., A pragmatic approach to the design of advanced precision terrain-aided navigation for UAVs and its verification, *Remote Sensing*, Vol. 12, Issue 9, 2020, 1396.
[11]. C. Ouyang, Y. Wang, J. Chen, X. Wang, et al., A semantic vector map-based approach for aircraft positioning in GNSS/GPS denied large-scale environment, *Defence Technology*, Vol. 34, 2024, pp. 1-10.
[12]. Q. Ye, J. Luo, Y. Lin, A coarse-to-fine visual geo-localization method for GNSS-denied UAV with oblique-view imagery, *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 212, 2024, pp. 306-322.
[13]. Smallsats by the numbers 2025, BryceTech, https://brycetech.com/reports
[14]. SkySat documentation, https://docs.planet.com/data/imagery/skysat/
[15]. PlanetScope documentation, https://docs.planet.com/data/imagery/planetscope/
[16]. Form 10-K, fiscal year ended December 31, 2024, *BlackSky Technology Inc*., 2025.
[17]. Form 10-K, fiscal year ended January 31, *Planet Labs PBC*, 2025.
[18]. BlackSky imagery and analytics: Satellite images of Kentland Farm (Virginia Tech), *BlackSky Technology Inc*., 2025.
[19]. Seeed Technology, NVIDIA Jetson Orin NX 16GB AI device – reComputer J4012, https://www.seeedstudio.com/reComputer-J4012-p-5586.html
[20]. Controller diagrams, https://docs.px4.io/main/en/flight_stack/controller_diagrams
[21]. Hex Cube Black flight controller, https://docs.px4.io/main/en/flight_controller/pixhawk-2
[22]. Auvidea X230, https://auvidea.eu/product/x230-70419/
[23]. J. Fan, E. Zheng, Y. He, J. Yang, A cross-view geo-localization algorithm using UAV image and satellite image, *Sensors*, Vol. 24, Issue 12, 2024, 3719.
[24]. M. Garcia, T. Davis, S. Blair, N. Cui, et al., Bioinspired polarization imager with high dynamic range, *Optica*, Vol. 5, Issue 10, 2018, pp. 1240-1246.
[25]. LUCID Vision Labs, Polarization explained, https://thinklucid.com/tech-briefs/polarization-explained-sony-polarized-sensor/
[26]. J. Kinnari, R. Renzulli, F. Verdoja, V. Kyrki, LSVL: Large-scale season-invariant visual localization for UAVs, *Robotics and Autonomous Systems*, Vol. 168, 2023, 104497.
[27]. PX4 ROS 2 user guide, https://docs.px4.io/main/en/ros2/user_guide
[28]. L. Kempf, K. B. Kochersberger, A lightweight vision-based GNSS substitute for small UAVs, in *Proceedings of the 2nd International Conference on Drones and Unmanned Systems (DAUS'26)*, 2026, pp. 173-178.
[29]. NVIDIA Jetson Orin, https://www.nvidia.com/en-us/autonomous-machines/embedded-systems/jetson-orin/