

Kalman Filtering for Multirotor State Estimation

Dev Kumar

▶ **To cite this version:**

Dev Kumar. Kalman Filtering for Multirotor State Estimation. Independent research; original academic project at École Nationale Supérieure d'Arts et Métiers. 2026, pp.1-38. ⟨hal-05873382⟩

HAL Id: hal-05873382

https://hal.science/view/index/docid/5873382

Submitted on 20 Nov 2026

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Copyright - All rights reserved

Kalman Filtering for Multirotor State Estimation

Nonlinear filtering benchmarks, inertial estimation on SO(3)
and the geometry of a gimballed camera

Dev KUMAR

École Nationale Supérieure d'Arts et Métiers

Paris, France

Research report
November 2026 · Revision 2

Abstract

This study develops and tests Kalman estimators for simulated multirotor state estimation. The first experiment derives the exact discrete flow and Jacobian of a two-state nonlinear model from an archived MATLAB study. A 256-run paired comparison evaluates a fixed linear Kalman filter, an extended filter, an unscented filter and an offline extended RTS smoother. Under the large input excursions, the extended filter reduces x_1 RMSE from 0.338 to 0.036 with matched parameters. A deliberate 20% error in the nonlinear coefficient reverses that ordering for x_1 and exposes significant covariance miscalibration. The second experiment uses a 15-dimensional error-state formulation on SO(3) with course-independent heading, GNSS, inertial and barometric data. Over 64 clean trials, horizontal and orientation RMSE are 0.337 m and 1.108°. In a 15 s GNSS outage the gated filter retains 0.588° yaw RMSE with independent heading, while horizontal RMSE rises to 10.82 m. An additional experiment shows how uncertainty in camera orientation enters ground-plane projection. A separate replay uses native sensors from three recorded Gazebo flights. Horizontal RMSE is 0.058 m nominally and 1.135 m during a 15 s GNSS blackout. With 200 ms GNSS delay, acquisition-time replay reduces it from 0.066 to 0.060 m over the flight. The numerical and recorded-simulation experiments are reproducible in Python.

Keywords State estimation, nonlinear filtering, error-state Kalman filter, SO(3), multirotor, observability, camera geometry, Monte Carlo.

Résumé

Cette étude développe des estimateurs de Kalman pour un multirotor simulé. La première expérience dérive le flot discret exact et la jacobienne d’un modèle non linéaire à deux états issu d’une étude MATLAB. Sur 256 essais appariés, le filtre de Kalman étendu réduit la RMSE de x_1 de 0,338 à 0,036 en forte excitation avec le bon modèle. Une erreur de 20 % sur un coefficient inverse ce classement pour x_1 et dégrade le calibrage des incertitudes. La deuxième expérience développe un filtre à état d’erreur de dimension 15 sur SO(3) fusionnant mesures inertielles, GNSS, baromètre et cap absolu. Sur 64 essais nominaux, les RMSE de position horizontale et d’orientation valent 0,337 m et 1,108°. Lors d’une coupure GNSS de 15 s, la présence d’un cap indépendant maintient une RMSE de lacet de 0,588°, tandis que la RMSE horizontale atteint 10,82 m. Une troisième expérience propage l’incertitude de l’orientation de la caméra jusqu’à l’intersection d’un rayon avec le sol. Un rejeu sur trois vols Gazebo enregistrés donne une RMSE horizontale de 0,058 m en nominal et de 1,135 m pendant la coupure GNSS. Avec 200 ms de retard GNSS, le rejeu à la date d’acquisition réduit la RMSE de 0,066 à 0,060 m sur le vol. Les expériences sont reproductibles en Python.

Mots-clés Estimation d’état, filtre de Kalman étendu, multirotor, observabilité, Monte-Carlo, caméra.

Nomenclature

Symbol	Meaning
x, u, z	State, control input and observation
F, H, K	Transition or Jacobian, observation Jacobian and gain
P, Q, R_m	Error covariance, process covariance, measurement covariance
ν, S	Innovation and its predicted covariance
NIS, NEES	Normalised innovation and estimation error squared

Symbol	Meaning
p, v, R	Position, velocity and body-to-world rotation
b_a, b_g	Accelerometer and gyro biases
δθ, $[\cdot]_\times$	Right attitude error and skew matrix
φ, θ, ψ	Roll, pitch, yaw where specified
Δ, T, f	Sample time, total thrust and specific force

Contents

1. Research questions and evidence 3
   1.1 Problem formulation 3
   1.2 What is observed in each source 3
   1.3 Comparison and inference 4
2. Bayesian state estimation 4
   2.1 Prediction and correction 4
   2.2 Discretisation and units of noise 4
   2.3 Detectability and the Riccati equation 5
   2.4 Consistency of uncertainty 5
3. Exact sampled model of the MATLAB example 6
   3.1 Equilibrium, local model and observability 6
   3.2 Nonlinear flow under zero-order hold 6
   3.3 Archival comparison and controlled inputs 7
4. Nonlinear filtering and smoothing results 7
   4.1 Extended and unscented filter construction 7
   4.2 Paired Monte Carlo comparison 8
   4.3 Misspecification and covariance tuning 10
5. Multirotor model and archived geometry 11
   5.1 Vehicle and sensor conventions 11
   5.2 Mechanics and a controlled reference trajectory 11
   5.3 Measurements and placement 12
6. Error-state Kalman filtering on SO(3) 12
   6.1 Nominal state and local errors 12
   6.2 Nominal propagation 12
   6.3 Error dynamics 13
   6.4 Discretising the stochastic dynamics 13
   6.5 Injection and covariance reset 13
7. Measurement updates and observability 14
   7.1 Position, altitude and heading 14
   7.2 Static information and dynamic excitation 14
   7.3 Outlier rejection and consistency 15
   7.4 Multirate timing and delayed observations 15
8. Multirotor estimation benchmark 15
   8.1 Reproducible design 15
   8.2 Definition of errors 16
   8.3 Matched clean case 16
   8.4 GNSS loss and residual gating 17
   8.5 Timestamps and latency 19
9. Camera geometry and the archived position filter 20
   9.1 What the archive implements 20
   9.2 Projection and linear sensitivity 20
   9.3 Monte Carlo on a static calibration point 21
10. ESKF replay on recorded Gazebo flights 23
   10.1 Acquisition and data checks 23
   10.2 Model and sensor initialisation 24
   10.3 Causal handling of recorded and delayed packets 25
   10.4 Nominal replay and the vertical datum 26
   10.5 Position and heading through the GNSS outage 27
   10.6 Effect of acquisition-time replay 28
   10.7 Reprojection on the recorded checkerboard 29
11. Interpretation and validation protocol 30
   11.1 What changes the conclusions 30
   11.2 Scope of the Gazebo replay 31
   11.3 From idealised to measured sensors 31
   11.4 Conclusion 31
References 32
Appendix A. Original estimation results 33
   A.1 MATLAB parameters and figure provenance 33
   A.2 Archived linear model 33
   A.3 Archived linear filter on nonlinear data 34
Appendix B. Reproduction and checks 36
   B.1 Recreating the results 36
   B.2 Verification tests and data audit 36
   B.3 Event sequence for the inertial filter 36
   B.4 Reading a paired result 37
   B.5 Recorded Gazebo replay and exports 37

1. Research questions and evidence

1.1 Problem formulation

Can an estimator provide both an accurate state and a covariance that describes how uncertain it is after a manoeuvre, a sensor outage or a timestamp error? For a multirotor the question is coupled to orientation. A small roll or pitch error rotates the measured specific force and creates a persistent horizontal acceleration error. Heading can remain well estimated while the horizontal position becomes uncertain. A camera adds another dependency because the estimated rotation determines where an observed ray intersects the scene.

Four experiments isolate these mechanisms. E1 uses the two-state model from the 2026 estimation dossier to compare local linearisation, unscented moment propagation and batch smoothing. Its exact discrete flow separates discretisation effects from estimation errors. E2 uses the Gazebo project to establish coordinate, airframe and sensor conventions, then evaluates a multirate 15-state inertial error-state Kalman filter. E3 follows the archived camera projection and measures how an orientation or pixel error affects recovery of the position of a static calibration point on a plane. E4 evaluates the 15-state estimator, sensor timestamps and optical geometry on recorded Gazebo streams [12].

Table 1.1 – Status of the material used in this work.

Evidence	Available material	Use in the report
2026 estimation dossier [1]	Word equations and 16 distinct MATLAB plots. Printed initial state, measurement variance and step size.	Derive the exact transition, reconstruct a likely step and run a fully specified new benchmark.
ROS/Gazebo archive [2]	SDF, bridge configuration, source code and RGB recordings.	Recover dynamics and frame conventions, document the archived pose and position estimators.
Python experiments	Controlled truth, stored seeds, run-level errors, sensor innovations and numerical tests.	Quantify algorithm gains, sensor outages and timestamp sensitivity in simulation.
Gazebo acquisition [12]	Four native sensor and reference recordings, timed availability, instrumented SDF and checkerboard frames.	Test the same error-state structure on logged Gazebo signals and compare replay scheduling.

1.2 What is observed in each source

The source estimation report gives nonlinear dynamics, an equilibrium, sampled inputs for initialisation and measurement covariance, and archived MATLAB plots [1]. The Word version preserves the equations and original chart images. The supplied versions reach the linear filter and include a linear filter run against nonlinear truth. Their EKF and Monte Carlo sections consist of headings. Appendix A preserves the MATLAB record. The newly computed benchmark specifies its own process covariance, random initialisation, input and seeds.

The archived Gazebo simulation uses a velocity-control plugin and obtains the vehicle pose from Gazebo's model state [2]. It also contains a separate four-state Kalman estimator of a simulated person's position using GPS fixes and a ground-plane camera projection. The vehicle pose used by the control and projection code is supplied by the simulator. The IMU, aircraft GNSS, barometer and independent heading in E2 are modelled sensor channels. The archived images and videos document the visual scenario. E1–E3 use controlled numerical studies. E4 replays native sensor streams from the instrumented simulator and scores against a separately recorded reference.

1.3 Comparison and inference

Each comparison inside a scenario uses common truth and noise draws. Estimators differ only in the intended mechanism, such as the transition model, attitude observations or rejection gate. E1 tests whether a local approximation captures the experimental model near equilibrium and after a large change in input. E2 evaluates the roles of correction sources and timestamp assumptions around the same trajectory. The E1 and E2 statistical comparisons use independent Monte Carlo trials as sampling units. The three E4 flights are a small set of noise realisations on the same command sequence. The absolute performance numbers depend on the specified excitation, noise and outage schedule.

2. Bayesian state estimation

2.1 Prediction and correction

Let x_k be a state at sample k and z_k the observation. In a Markov state model with conditionally independent measurements, knowledge from the previous sample is expressed by the filtering posterior. Prediction and correction give the Chapman–Kolmogorov integral and Bayes update [3,4].

$$p(x_k \mid z_{1:k-1}) = \int p(x_k \mid x_{k-1})\, p(x_{k-1} \mid z_{1:k-1})\, \mathrm{d}x_{k-1} \tag{2.1}$$

$$p(x_k \mid z_{1:k}) \propto p(z_k \mid x_k)\, p(x_k \mid z_{1:k-1}). \tag{2.2}$$

Known inputs in the transition have been suppressed in Eqs. (2.1)–(2.2). For a linear Gaussian setting let F_k map k−1 to k, B_k be the input matrix and H_k the observation matrix. Independent state and measurement noise with zero mean have covariances Q_k and $R_{m,k}$.

$$\begin{aligned} x_k &= F_k x_{k-1} + B_k u_{k-1} + w_k, \\ z_k &= H_k x_k + v_k. \end{aligned} \tag{2.3}$$

For Gaussian priors the prediction is entirely specified by propagating the mean and covariance.

$$\begin{aligned} \hat{x}_k^- &= F_k \hat{x}_{k-1}^+ + B_k u_{k-1}, \\ P_k^- &= F_k P_{k-1}^+ F_k^\top + Q_k. \end{aligned} \tag{2.4}$$

The superscripts refer to estimates before and after conditioning on z_k. Subtracting the expected measurement yields the innovation $\nu = z - H\hat{x}^-$. Its cross covariance with the prior error is $P^- H^\top$, and its covariance is $S = H P^- H^\top + R_m$. Conditioning the joint Gaussian gives

$$\begin{aligned} K &= P^- H^\top S^{-1}, \\ \hat{x}^+ &= \hat{x}^- + K\nu. \end{aligned} \tag{2.5}$$

The same gain is the linear minimum mean-square-error choice under the stated second moments. For a generic gain the posterior error is $e^+ = (I - KH)e^- - K\nu$. Its covariance is

$$P^+ = (I - KH)P^-(I - KH)^\top + KR_mK^\top. \tag{2.6}$$

Expanding gives a quadratic in the gain. Subtracting its value at $K_\star = P^- H^\top S^{-1}$ leaves the positive semidefinite term $(K - K_\star)S(K - K_\star)^\top$. This establishes optimality and separates the role of the gain from the numerical implementation. The Joseph form in Eq. (2.6) uses two positive semidefinite terms. The code solves the system involving S and symmetrises the posterior. Covariance symmetry is a numerical check. Squared innovations and squared estimation errors test whether the underlying statistical assumptions are credible.

2.2 Discretisation and units of noise

For a continuous linear error process use $\dot{e} = Ae + Lw_c$, with the formal white-noise density defined by $E[w_c(t)w_c(s)^\top] = Q_c\delta(t - s)$. Over a constant interval Δ define $W = LQ_cL^\top$ and $\Phi = \exp(A\Delta)$. Integrating through the transition gives

$$Q_d = \int_0^\Delta \exp(A\tau) W \exp(A^\top \tau)\, \mathrm{d}\tau. \tag{2.7}$$

The Van Loan block exponential [5] obtains both quantities from

$$\begin{aligned}\exp(\Delta M) &= \begin{bmatrix}\Phi & \Gamma\\ 0 & \Phi^{-\top}\end{bmatrix},\\ M &= \begin{bmatrix}A & W\\ 0 & -A^{\top}\end{bmatrix},\\ Q_d &= \Gamma\Phi^{\top}.\end{aligned} \tag{2.8}$$

For a scalar position–velocity pair driven by white acceleration with spectral density q_c, this produces a position variance proportional to Δ^3. An independent, held acceleration error with variance σ_a^2 produces a position variance proportional to Δ^4. Both models are used in practice and apply to different time correlations.

$$\begin{aligned}Q_{\text{white}} &= q_c \begin{bmatrix}\Delta^3/3 & \Delta^2/2\\ \Delta^2/2 & \Delta\end{bmatrix},\\ Q_{\text{held}} &= \sigma_a^2 \begin{bmatrix}\Delta^4/4 & \Delta^3/2\\ \Delta^3/2 & \Delta^2\end{bmatrix}.\end{aligned} \tag{2.9}$$

The archived four-state position filter uses the held model. E1 fixes a per-step Q for a 0.1 s sample. E2 derives the corresponding continuous covariance from the inertial noise amplitudes and changes it with the elapsed integration time. A variation of sample rate requires stating which physical model is held constant.

2.3 Detectability and the Riccati equation

For a time-invariant linear plant, the observability matrix stacks $H, HF, \ldots, HF^{n-1}$. It has full column rank when the noise-free output sequence recovers all initial directions. Detectability requires the unobservable modes to be stable. When the standard stabilisability and detectability conditions hold, the discrete steady prior covariance P_∞^- solves an algebraic Riccati equation.

$$\begin{aligned}P_\infty^- = F(P_\infty^- - K_\infty S_\infty K_\infty^{\top})\\ F^{\top} + Q.\end{aligned} \tag{2.10}$$

Here K_∞, S_∞ follow Eq. (2.5) with P_∞^-. The steady solution is a useful analytical reference when the operating point, sample rate and tuning are fixed. When data disappear or arrive on several clocks, the correction sequence changes the covariance. A state-space model can already distinguish those cases through the time-varying sequence H_k, F_k.

2.4 Consistency of uncertainty

With a synthetic truth and correctly specified Gaussian model, the squared Mahalanobis state error (NEES) has mean n. The squared Mahalanobis innovation (NIS) has mean m, the number of independent measurement coordinates.

$$\begin{aligned}\varepsilon_k &= e_k^{\top} P_k^{-1} e_k,\\ \eta_k &= \nu_k^{\top} S_k^{-1} \nu_k.\end{aligned} \tag{2.11}$$

For M independent runs at the same fixed time, $M\bar{\varepsilon}_k$ is compared with χ^2_{Mn}, and $M\bar{\eta}_k$ with χ^2_{Mm}. These are pointwise reference bands. The trials are independent by construction, and the samples inside a trajectory share history. Averaging the statistics across time is useful descriptively and does not multiply the number of independent trials. A nominal marginal interval uses $1.96\sqrt{(P_{ii})}$. A correct marginal hit rate for one coordinate does not certify the coupled covariance or the other state components.

For nonlinearity, rejection and misspecification the chi-square law is a reference approximation. Biases show up as elevated NEES, non-white innovations, poor coverage or a shift in the residual mean. If truth is unavailable then consistency can be probed from innovation magnitude and autocorrelation, taking care to use all pre-gate innovations when testing whether the innovation distribution matches the assumed model.

3. Exact sampled model of the MATLAB example

3.1 Equilibrium, local model and observability

The dossier models two states x_1,x_2, input u, and an observation of the first state [1]. Write $s=x_2$, $\alpha=4,\beta=1$.

$$\begin{aligned} \dot{x}_1 &= \beta - us, \\ \dot{s} &= us - \alpha s^2, \\ z &= x_1 + v. \end{aligned} \tag{3.1}$$

At equilibrium the first equation gives $u_e s_e=\beta$. A minus sign on an intermediate line in the source conflicts with its numerical equilibrium and final matrices. With $\beta>0$, $s_e=0$ is excluded. The second equation then gives $u_e=\alpha\, s_e$, so the signs of s_e,u_e are paired. The positive branch is (0.5,2); $x_{1,e}$ can be chosen freely. The negative branch has a locally unstable s mode, with derivative $u_e-2\alpha\, s_e=2$. The positive branch has the derivative −2 and a neutral x_1 equilibrium direction. The positive branch is used here.

For centred deviations $x_\ell=x-x_e$, $u_\ell=u-u_e$, let $x_e=(0,0.5)$. The $\partial f/\partial x$ and $\partial f/\partial u$ derivatives at that point give

$$\begin{aligned} A &= \begin{bmatrix} 0 & -2 \\ 0 & -2 \end{bmatrix}, \\ B &= \begin{bmatrix} -0.5 \\ 0.5 \end{bmatrix}, \\ C &= \begin{bmatrix} 1 & 0 \end{bmatrix}. \end{aligned} \tag{3.2}$$

For input held one sample, write $d=\exp(-2\Delta)$. Direct integration gives

$$\begin{aligned} F &= \begin{bmatrix} 1 & d-1 \\ 0 & d \end{bmatrix}, \\ B_d &= \begin{bmatrix} -\Delta + (1-d)/4 \\ (1-d)/4 \end{bmatrix}. \end{aligned} \tag{3.3}$$

For example, $\Delta=0.1$ s gives $F_{12}=-0.181269$ and $B_{d,1}=-0.054683$. The matrix formed by stacking C and CF has determinant d−1, so the second state is locally observable through the influence it has on x_1. At this sample rate its two singular values are 1.420 and 0.128. For a shorter interval $d-1=-2\Delta+O(\Delta^2)$, so an adjacent pair of samples is less informative about the hidden coordinate at the same per-sample measurement variance. Multiple samples over a fixed duration have to be considered for a sampling-rate design.

3.2 Nonlinear flow under zero-order hold

For a constant u over Δ, the second equation is logistic. Transforming by $r=1/s$ gives the linear differential equation $\dot{r}=-ur+\alpha$ on the positive branch. Define $E=\exp(u\Delta)$ and $\varphi=(E-1)/u$, with continuous limit $\varphi=\Delta$ at zero input. Then $D=1+\alpha s\varphi$. The solution for s is sE/D. Since $\dot{D}/D=\alpha\, s(t)$ within the sample, the integral of s(t) is $\log(D)/\alpha$. Both states therefore admit a closed form.

$$\begin{aligned} g_1(x,u) &= x_1 + \beta\Delta - \frac{u}{\alpha}\log D, \\ g_2(x,u) &= \frac{sE}{D}. \end{aligned} \tag{3.4}$$

The partial derivatives at the current estimate are

$$F(x,u) = \begin{bmatrix} 1 & -u\phi/D \\ 0 & E/D^2 \end{bmatrix}. \tag{3.5}$$

Using expm1(uΔ) gives a stable calculation near zero input. Positive s and α keep $D>0$ since $\varphi>0$ for $\Delta>0$. The experiment checks domain validity for states and sigma points. Process noise is an additive Gaussian perturbation of the newly evaluated step. This discrete stochastic model differs from imposing continuous white noise inside the nonlinear ODE. Its declared covariance makes the Monte Carlo comparison self-contained.

When u=0, x_1 increases at rate β independently of s. Local sensitivity F_{12} is then zero and s cannot be identified from that input segment alone. In the nonzero-input experiments, observations of x_1 constrain s through the accumulated dynamics. Exact discretisation is verified against an independent Dormand–Prince integration. The Jacobian is checked with centred differences.

3.3 Archival comparison and controlled inputs

The plotted x_2 in the archived step changes from 0.5 to about 0.75 after 4 s. With $s_e=u/\alpha$, that final value suggests u=3 after the step. The figure order and trend also show a later mismatch in x_1 when the fixed linear model sees nonlinear truth. Figure 3.1 reconstructs a candidate input with that change. This run uses deterministic nonlinear truth and added measurement noise. The source's exact command sequence, Q, initial covariance and random seed are unavailable.

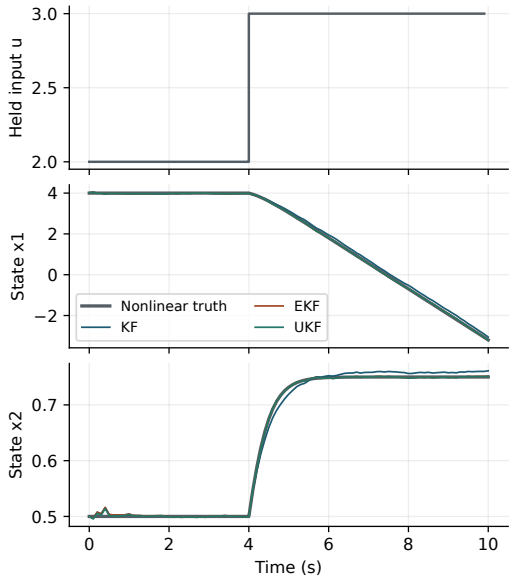

Figure 3.1 – Reconstruction of an inferred u=2 to u=3 step at 4 s. Nonlinear truth is exact under held input. Estimates use a new noisy measurement stream and the covariances of Table 3.1.

One mechanism is visible without simulation. At the new equilibrium in s, the nonlinear slope of x_1 is $1-u^2/4$. It is −1.25 at u=3. With a fixed local linearisation at u=2, the new s deviation is 0.25 and the corresponding x_1 slope is $-2(0.25)-0.5(1)=-1.0$. The difference is the second-order contribution $-(\delta u)^2/4$. It accumulates over time in the neutral x_1 direction. The filter can correct this offset through measurements, with a correction weight set by its gain and R_m.

Table 3.1 – Numerical conditions for the new two-state simulations. Noise units follow the two-state model.

Quantity	Value or protocol
Sampling period	0.1 s (both archived and new)
x prior and covariance	(4, 0.5), diag(0.12², 0.07²)
Nominal α, β	4, 1
Q per step	diag(0.012², 0.008²)
R per observation	0.01 (printed in the dossier)
Near input	2 + 0.20 sin(2πt/7)
Manoeuvre input	2, 3.70, 1.05, 4.10, 1.65
Mismatched truth	Same input as manoeuvre, α = 4.8

The near-equilibrium case uses a smooth input around u=2. The large-excursion case uses the five piecewise constant inputs in Table 3.1, changing at 4, 8, 12 and 16 s. In the misspecified case the truth changes α to 4.8 and the filters keep 4. All filters within a scenario process the same 201 timestamps on each run, including the measurement at t=0. The initial truth is sampled from the prior and the noise is independent across trials. The reconstructed step is a separate illustrative ten-second run.

4. Nonlinear filtering and smoothing results

4.1 Extended and unscented filter construction

The fixed KF uses the centred model in Eq. (3.3) for every input. An extended Kalman filter instead predicts its mean with Eq. (3.4) and propagates covariance through the current Jacobian in Eq. (3.5). Both have the same linear observation matrix C. With $f_k=g(\cdot,u_{k-1})$, the EKF propagation is

$$\begin{aligned} \hat{x}_k^- &= f_k(\hat{x}_{k-1}^+), \\ P_k^- &= F_k P_{k-1}^+ F_k^\top + Q, \\ F_k &= Df_k(\hat{x}_{k-1}^+). \end{aligned} \tag{4.1}$$

The unscented filter propagates deterministic sigma points through the exact map and reconstructs its mean and covariance [6]. In dimension n=2, let $\lambda=a^2(n+\kappa)-n$ and let the columns of L be from the Cholesky factor of $(n+\lambda)P^+$. The five points are the state mean and the offsets $\pm L_{:j}$. Here

a=0.85, κ=0, with Gaussian shape parameter b=2.

$$W_0^{(m)} = \frac{\lambda}{n+\lambda},$$
$$W_0^{(c)} = W_0^{(m)} + 1 - a^2 + b, \tag{4.2}$$
$$W_i^{(m)} = W_i^{(c)} = \frac{1}{2(n+\lambda)}.$$

The weighted predicted mean uses $W^{(m)}$. Its covariance uses $W^{(c)}$, centred at that weighted mean, plus Q. For this affine observation the same Joseph measurement update applies to the approximate predicted moments. UKF benefits depend on the predicted curvature and spread. The exact discrete flow at small sample intervals has mild curvature on the sampled positive branch.

To test the value of observations after k, an offline extended Rauch–Tung–Striebel smoother runs backwards over the EKF history [4,7]. Let P_{k+1}^{-} be the predicted covariance and F_{k+1} the Jacobian used from k to k+1. A local Gaussian conditional gives

$$J_k = P_k^{+} F_{k+1}^{\top} (P_{k+1}^{-})^{-1}, \tag{4.3}$$

$$\hat{x}_k^s = \hat{x}_k^{+} + J_k(\hat{x}_{k+1}^s - \hat{x}_{k+1}^{-}),$$

$$P_k^s = P_k^{+} + J_k(P_{k+1}^s - P_{k+1}^{-})J_k^{\top}. \tag{4.4}$$

The terminal smoothed estimate is its filtered counterpart. The implementation checks that smoothing contracts the covariance to numerical tolerance. The method estimates past states using present observations and is evaluated for offline replay.

4.2 Paired Monte Carlo comparison

The sampled time span for scalar scores is 1–20 s. Each table RMSE pools all included time samples and all trials. For an interval on a between-method difference, compute the time RMSE within each trial first and form its paired change. Resampling 256 complete pairs with replacement 5,000 times gives a percentile 95% interval for the mean change. This targets a mean of per-trial RMSE differences. Its value can differ from subtracting two pooled table RMSEs.

Table 4.1 – Two-state results over 256 paired trials and samples from 1 s to 20 s. EKS is the offline extended RTS smoother. Coverage is the x_2 marginal 95% interval hit rate.

Input/model	Method	RMSE x_1	RMSE x_2	ANEES	ANIS	Coverage x_2 (%)
Near	KF	0.0353	0.0138	1.98	1.00	95.1
Near	EKF	0.0352	0.0138	1.98	1.00	95.1
Near	UKF	0.0352	0.0138	1.97	1.00	95.1
Near	EKS	0.0255	0.0132	1.98	—	95.1
Manoeuvre	KF	0.3381	0.0367	104.74	13.98	65.9
Manoeuvre	EKF	0.0359	0.0136	1.97	1.00	95.2
Manoeuvre	UKF	0.0359	0.0136	1.96	1.00	95.2
Manoeuvre	EKS	0.0258	0.0127	1.98	—	95.1
Mismatch	KF	0.1266	0.1190	94.51	2.70	2.8
Mismatch	EKF	0.2368	0.1086	115.35	7.28	4.7
Mismatch	UKF	0.2364	0.1083	114.93	7.26	5.0
Mismatch	EKS	0.0770	0.0773	57.33	—	14.1

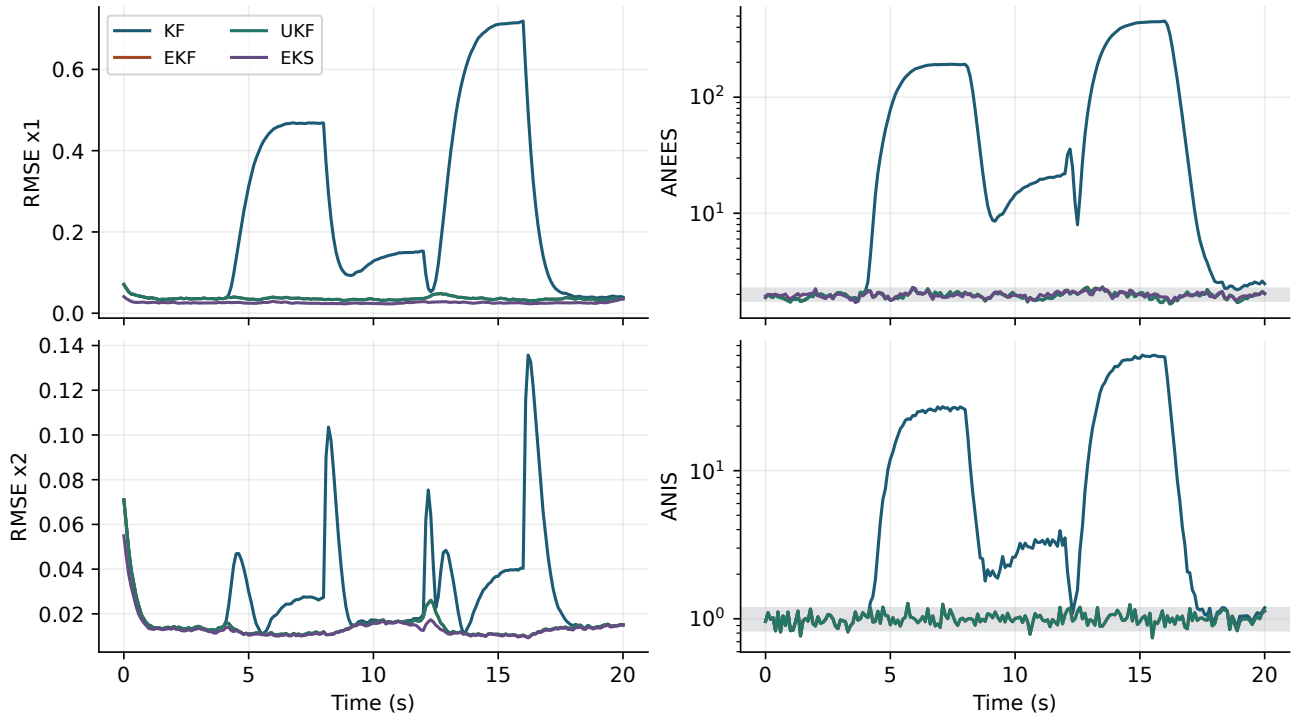

Figure 4.1 – Time-resolved errors and consistency for the strong input changes. Each point averages the 256 trials at one time. Shaded bands are 95% chi-square reference intervals at the indicated dimensions.

Near equilibrium, the KF, EKF and UKF have x_2 RMSE close to 0.0138. The EKF marginal hit rate is 95.1% and average NEES is 1.98 for two coordinates. The lag-one sample correlation of the standardised EKF innovation, computed across the retained run-time pairs, is approximately 0.00035. These are examples of the near-equilibrium conditions in which a fixed linear model can be effective.

The equilibrium Riccati calculation gives $K_\infty=(0.1261,-0.00873)^\top$ for the new Q and archived R_m. Figure 4.2 shows convergence near this reference over the small-excursion input. A negative second gain means that a positive innovation reduces the estimate of x_2. This is consistent with the negative effect of x_2 on the next observed x_1. The gain scale depends on the process and measurement covariances. Accordingly it cannot be read directly from the archived figure as a recovered numerical Q.

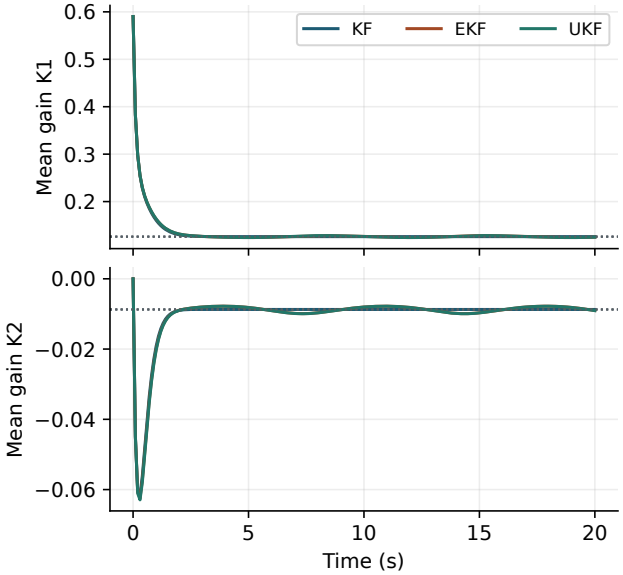

Figure 4.2 – Mean gain components over the near-equilibrium ensemble. The dotted lines give the fixed equilibrium Riccati solution for the new process covariance.

On the stronger manoeuvre, the KF's x_1 error increases from 0.0353 in the near case to 0.3381, with average NEES 104.7 and x_2 coverage 65.9%. The EKF stays near 0.0359 for x_1, with average NEES 1.97 and a 95.2% x_2 hit rate. The mean paired change in x_2 RMSE for EKF minus KF is −0.02310, with interval [−0.02333,−0.02287]. The

UKF minus EKF change is approximately −0.000010, with interval [−0.000025,+0.000006]. Resolving a practical improvement between these two nonlinear filters would require a case with stronger one-step curvature, broader uncertainty or a different observation structure.

In Fig. 4.1 the error separation starts after input changes break the local drift approximation. For the EKF and UKF, smaller x_2 uncertainty after the high-input segments also reflects a rapidly contracting logistic mode. The extended smoother improves the matched-motion x_1 RMSE from 0.0359 to 0.0258, a reduction of about 28%. The x_2 reduction is from 0.0136 to 0.0127. Future x_1 observations constrain the hidden state via the same temporal coupling found in the observability calculation.

4.3 Misspecification and covariance tuning

With α=4.8 in truth, all the tested filters understate uncertainty. Coverage for x_2 falls below 6% for all three causal filters. In this experiment the fixed KF has lower pooled x_1 RMSE than the nonlinear filters, at 0.1266 against 0.2368 for the EKF. At high positive u, x_1 slope is $1-u^2/\alpha$. Increasing the true divisor reduces the magnitude of its downward drift. The exact model at α=4 systematically predicts a slope that is too negative. At the same high inputs, the slope error from the KF's local approximation can partly cancel that parameter error. The table shows why a comparison with matched parameters alone can overstate the value of nonlinear filtering in an application.

For tuning sensitivity, hold all near-equilibrium data fixed and multiply only the assumed Q by scale c_Q. At c_Q=0.1, x_2 RMSE stays within 4.1×10^{-5} of its value at c_Q=1, yet NEES grows from 1.98 to 14.42. The filter reports very different confidence for similar accuracy in that coordinate. Increasing Q by a factor of ten gives ANEES 0.75 and increasing x_1 RMSE. The best value on this grid is c_Q=1 for x_1 with the known generator. Figure 4.3 displays both accuracy and calibration.

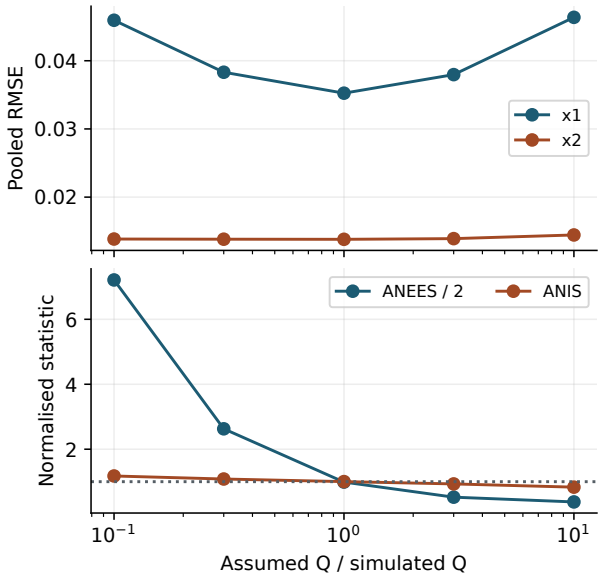

Figure 4.3 – Near-equilibrium tracking and consistency when the assumed process covariance is scaled. NEES is divided by two. The expected value of either normalised statistic is one in the matched linear-Gaussian limit.

A parameter-estimation extension could augment the state with ℓ=logα to enforce positivity. Chain rule gives a Jacobian column with components $\alpha\,\partial g_i/\partial\alpha$. The exact derivatives for the held model are

$$\frac{\partial g_1}{\partial \alpha} = \frac{u \log D}{\alpha^2} - \frac{us\phi}{\alpha D}, \tag{4.5}$$

$$\frac{\partial g_2}{\partial \alpha} = -\frac{s^2 \phi E}{D^2}.$$

Identification should use persistently informative inputs and held-out segments to separate parameter bias from process noise. The results above measure specified fixed-parameter cases.

5. Multirotor model and archived geometry

5.1 Vehicle and sensor conventions

The source airframe is defined in the world file **empty_field.sdf** [2]. The origin of its base link is used as a body-frame reference. The model includes four rotor links and one camera link. Their stated masses sum to 1.415 kg with the base. The base diagonal inertia and arm coordinates are listed in Table 5.1. The assembled inertia would also require transforming component inertias to a common centre of mass using the parallel-axis theorem. The benchmark uses the mass only to illustrate the scale of prescribed thrust.

Table 5.1 – Vehicle properties directly read from the Gazebo archive, with derived mass stated separately.

Property	Value
Base-link mass	1.35 kg
Four rotor links	0.010 kg each
Camera link	0.025 kg
Summed link mass	1.415 kg
Base inertia diagonal	0.025, 0.025, 0.045 kg·m²
Rotor centres body XY	±0.24 m in X and Y
Camera position at mount	(+0.22, 0, −0.065) m
Camera SDF stream	640 × 360 at 15 Hz, 1.25 rad HFOV
Gimbal Y limits	−0.30 to +1.50 rad

The Python world frame is a right-handed local East–North–Up frame. The body axes are forward, left, up, and R maps a body vector to world coordinates. When writing a quaternion, use the Hamilton product and denote its scalar first as $q=(q_w,q_v)$ in the formulas. ROS messages store the four components as x, y, z, w. The bearing of the projected body X axis is used as yaw and wraps at ±π. These choices follow the frame conventions in REP 103 [10]. Gazebo camera rays in the archive use X forward and Y left. A ROS optical frame uses Z forward, X right, Y down. That fixed axis conversion must precede use of optical-frame camera models.

5.2 Mechanics and a controlled reference trajectory

For total thrust T along the positive body Z axis, body angular velocity ω, inertia J, and torque τ, the ideal Newton–Euler model is

$$\begin{aligned} \dot{p} &= v, \\ m\dot{v} &= mg_w + TRe_3 + f_{\text{ext}}, \\ g_w &= (0, 0, -g)^\top, \end{aligned} \tag{5.1}$$

$$\begin{aligned} \dot{R} &= R[\omega]_\times, \\ J\dot{\omega} &= \tau - \omega \times (J\omega). \end{aligned} \tag{5.2}$$

If rotor i acts at body lever arm r_i, an ideal force and moment model adds $f_i\, e_3$ and $r_i \times f_i\, e_3$, with signed reaction torque around e_3. The archive's velocity controller supplies simulated motor speeds inside Gazebo. The offline benchmark prescribes a trajectory with acceleration and orientation consistent with the translational thrust equation, using $f_{ext}=0$. It does not use the archive's feedback controller or model actuator response.

Let $a_w=\ddot{p}_{ref}$ from a twice-differentiable reference position. Choose $z_b=(a_w-g_w)/\|a_w-g_w\|$, and a non-collinear horizontal heading reference $c=(\cos\psi_r,\sin\psi_r,0)^\top$. Then

$$\begin{aligned} y_b &= \frac{z_b \times c}{\|z_b \times c\|}, \\ x_b &= y_b \times z_b, \\ R &= (x_b,\ y_b,\ z_b). \end{aligned} \tag{5.3}$$

This construction gives a proper rotation and thrust $T=m\|a_w-g_w\|$. The actual heading measurement later uses x_b, which is the projection of c orthogonal to z_b. Near zero tilt the headings agree. Angular rate between discrete poses is obtained from the log of $R_k^\top R_{k+1}$. This generates a known attitude reference and a consistent aggregate thrust axis without replaying archived motor commands.

The E2 path combines three coordinate sinusoids with a compact manoeuvre in the east coordinate, added from 19 to 24 s. The bump, its first derivative and its second derivative vanish at both boundaries.

Its form is $2.8[4s(1-s)]^3$ metres for $s=(t-19)/5$ within that interval. The reference approaches 3.19 m/s² horizontal acceleration and 17.93° tilt, with theoretical total thrust reaching 14.64 N under the summed mass. These values describe the estimator test. Its speeds and accelerations can exceed the particular velocity-controller settings in the Gazebo archive.

5.3 Measurements and placement

A body-fixed accelerometer at the reference point measures specific force expressed in its axes. A gyroscope measures body angular velocity. Including slowly varying bias,

$$\begin{aligned} a_m &= R^\top(\dot{v} - g_w) + b_a + n_a, \\ \omega_m &= \omega + b_g + n_g. \end{aligned} \tag{5.4}$$

An accelerometer under thrust measures nonzero specific force even in hover. Rotating that measurement and adding gravity gives the inertial acceleration. A gravity-only tilt update assumes that translational acceleration is small. For example, a horizontal acceleration of 3.19 m/s² would contribute a direction deviation of order $\arctan(3.19/9.81)$, about 18°. The benchmark therefore uses acceleration as part of the propagation model and relies on independent reference observations.

Example sensor readings are generated at the body origin. For a known lever arm l^b, a GNSS antenna position would be $p+Rl^b$, with attitude Jacobian $-R[l^b]_\times$ in the convention below. The camera is mounted at a nonzero offset and has a pitch joint. A rotation error then affects camera orientation and the transformed lever arm. Section 9 derives the camera geometry directly from the recorded conventions and treats it as a separate experiment.

6. Error-state Kalman filtering on SO(3)

6.1 Nominal state and local errors

Represent R on the rotation group, with $R^\top R=I$ and $\det R=1$. The nominal state uses p,v,R,b_a,b_g. Express estimation errors by Euclidean increments for position, velocity and biases, and a right rotation increment. The error vector has 15 components.

$$\begin{aligned} R &= \hat{R}\mathrm{Exp}(\delta\theta), \\ \delta x &= (\delta p, \delta v, \delta\theta, \delta b_a, \delta b_g). \end{aligned} \tag{6.1}$$

Here $\mathrm{Exp}(\xi)=\exp([\xi]_\times)$, with $[a]_\times b=a\times b$. For an angle–axis vector ξ and $r=\|\xi\|$, Rodrigues' formula is

$$\begin{aligned} \mathrm{Exp}(\xi) &= I \\ &+ \frac{\sin r}{r}[\xi]_\times \\ &+ \frac{1-\cos r}{r^2}[\xi]_\times^2. \end{aligned} \tag{6.2}$$

The continuous limits are used near zero. Continuous quaternion propagation is equivalent to $\dot{q}=\tfrac{1}{2}\, q\otimes(0,\omega)$. The local covariance describes the three components of a small angular displacement in the body tangent frame [8,9]. Normalised quaternions q and −q give the same rotation. Angle-error comparisons therefore use the geodesic norm of the log of the relative rotation, which also removes yaw wrap artefacts from the orientation comparison.

6.2 Nominal propagation

For one interval, subtract estimated biases to form $f=a_m-\hat{b}_a$ and $w=\omega_m-\hat{b}_g$. With corrected samples held and bias means kept constant through propagation, compute

$$\begin{aligned} R_m^* &= \hat{R}_k\mathrm{Exp}(w\Delta/2), \\ a^* &= R_m^* f + g_w, \\ \hat{R}_{k+1}^- &= \hat{R}_k\mathrm{Exp}(w\Delta). \end{aligned} \tag{6.3}$$

The mid-interval orientation represents the averaged force direction to second order in short, smooth intervals. For constant a^* over that update, set

$$\begin{aligned} \hat{p}_{k+1}^- &= \hat{p}_k + \hat{v}_k\Delta + \frac{a^*\Delta^2}{2}, \\ \hat{v}_{k+1}^- &= \hat{v}_k + a^*\Delta. \end{aligned} \tag{6.4}$$

The gyroscope sample is applied over the IMU interval just completed. The reference attitude increments and mid-interval specific force in E2 create a consistent step input at the selected rate. White rate noise is drawn with per-sample standard deviation $\sigma/\sqrt{(\Delta)}$ and bias random-walk increments with $\sigma_b\sqrt{(\Delta)}$. This defines the time-domain covariance convention of the experiment.

6.3 Error dynamics

To first order, $R\,f=(\hat{R}+\hat{R}[\delta\theta]_\times)f$. The change of rotated force is $-\hat{R}[f]_\times\delta\theta$. The accelerometer bias and white measurement terms enter with signs $-\hat{R}\delta b_a$ and $-\hat{R}\,n_a$. Differentiating the right rotation error under $\dot{R}=R[\omega]_\times$ gives the $-[w]_\times$ term from the commutator.

$$\begin{aligned}\delta\dot{p} &= \delta v,\\ \delta\dot{v} &= -\hat{R}[f]_\times\delta\theta\\ &\quad -\hat{R}(\delta b_a + n_a),\end{aligned} \tag{6.5}$$

$$\delta\dot{\theta} = -[w]_\times\delta\theta - \delta b_g - n_g,$$

$$\delta\dot{b}_a = n_{ba},\quad \delta\dot{b}_g = n_{bg}. \tag{6.6}$$

These equations specify the F_c blocks of the local continuous linear system. The bias drives are independent white noises. Independent isotropic acceleration noise stays isotropic after rotation. Thus W in Eq. (2.7) has position block zero and velocity, rotation, accelerometer bias and gyro bias blocks $\sigma_a^2\,I_3$,$\sigma_g^2\,I_3$,$\sigma_{ba}^2\,I_3$,$\sigma_{bg}^2\,I_3$. A calibrated non-isotropic accelerometer covariance would instead use $R\,Q_a\,R^\top$ on velocity. The derivation uses a small local error and short update interval. Large uncorrected rotations place the global error outside that approximation.

6.4 Discretising the stochastic dynamics

At 100 Hz, use the frozen midpoint F_c and $A(s)=I+F_c s+F_c^2\,s^2/2$. The transition is $A(\Delta)$, and a two-point Gauss–Legendre approximation to the noise integral is

$$s_\pm = \Delta\left(\frac{1}{2}\pm\frac{\sqrt{3}}{6}\right),$$

$$Q_d \simeq \frac{\Delta}{2}\sum_{\pm} A(s_\pm)WA(s_\pm)^\top. \tag{6.7}$$

Every contribution has nonnegative quadrature weight and positive semidefinite covariance. This also captures position–velocity coupling from the integration of velocity noise. A frozen-dynamics reference using a Van Loan exponential at 0.01 s gives 1.41×10^{-9} relative Frobenius error in Q_d and 8.21×10^{-7} in the transition for the verification case. Comparing the approximate transition to numerical differentiation of the nonlinear midpoint step gives Frobenius errors 2.71×10^{-4},3.39×10^{-5},4.23×10^{-6},5.29×10^{-7} at sample intervals 0.04,0.02,0.01,0.005 s. The reduction by about eight per halving is consistent with a third-order one-step error in that case. These checks verify implementation of the chosen model locally.

6.5 Injection and covariance reset

The measurement update gives a local correction $\delta\,\hat{x}=Kv$. Add position, velocity and bias increments to their nominal values and update the nominal rotation by $\hat{R}^+=\hat{R}^-\mathrm{Exp}(\delta\hat{\theta})$. The local error around the new nominal need not have the same coordinates as before injection. In the attitude block its change follows from

$$\begin{aligned}&\mathrm{Exp}(\delta\theta^+)\\ &= \mathrm{Exp}(-\delta\hat{\theta})\\ &\quad \mathrm{Exp}(\delta\theta^-).\end{aligned} \tag{6.8}$$

At $\delta\theta^-=\delta\hat{\theta}$, the tangent derivative is the right Jacobian at the injected angle. For $D=[\delta\hat{\theta}]_\times$, $r=\|\delta\hat{\theta}\|$,

$$\begin{aligned}J_r &= I - \frac{1-\cos r}{r^2}D\\ &\quad +\frac{r-\sin r}{r^3}D^2.\end{aligned} \tag{6.9}$$

The covariance is first updated in the old tangent frame by the Joseph equation, then transformed by $G_r P_J G_r^\top$, where G_r is identity except for J_r in the attitude block. This transforms cross covariances as

well as angle variances. In the small-angle limit $J_r=I-D/2+O(r^2)$. Central-difference verification gives a maximum entry error 1.41×10^{-12} for the reset Jacobian at the sampled nonzero correction. The code compares rotation-matrix orthogonality and samples covariance eigenvalues through each run.

7. Measurement updates and observability

7.1 Position, altitude and heading

For a GNSS fix in the local world frame, the measurement is $z_{GPS}=p+v_{GPS}$, with position identity block and zeros elsewhere in H. A calibrated barometric height supplies $z_b=e_3^\top p+v_b$. The simulated height is already relative to the same datum and has zero bias. A real barometer may need a pressure-to-altitude model, a bias state and a stable reference datum.

For an independent heading observation, define $r=Re_1=(r_x,r_y,r_z)^\top$ and $h_\psi(R)=\operatorname{atan2}(r_y,r_x)$. The innovation is wrapped to $[-\pi,\pi)$. This is an artificial absolute heading stream with independent declared noise, not the heading of the measured velocity vector. Differentiating r in the right tangent error gives $\delta r=-R[e_1]_\times\delta\theta$. The resulting attitude row is

$$H_{\psi,\theta} = \frac{(-r_y, r_x, 0)}{r_x^2+r_y^2}(-R[e_1]_\times). \tag{7.1}$$

When $r_x^2+r_y^2$ approaches zero the projected azimuth is ill-conditioned. At the test trajectory tilt this singularity is remote. A finite-difference check with wrapped residuals gives 1.34×10^{-11} maximum error in this row at the verification rotation. A vehicle rotating through the branch cut should use a wrapped innovation rather than differencing raw azimuths. The archived panorama logic likewise wraps successive simulator azimuths before accumulating turn angle [2]. This is a related use of angular increments, with different measurement provenance.

7.2 Static information and dynamic excitation

The stationary test freezes $R=I,\omega=0,f=ge_3$. A discrete, noise-weighted observability stack for samples t_i uses $R_{m,i}^{-1/2}H_i\exp(F_ct_i)D_0$. Here D_0 scales the columns by physical prior standard deviations, so the singular values are comparable across state units. For example, at rest the horizontal accelerometer bias can compensate a roll or pitch error in the velocity equation:

$$\delta b_a = -[ge_3]_\times\delta\theta \implies \delta\dot v = 0. \tag{7.2}$$

If the perturbation begins with $\delta p=\delta v=0$ and $\delta b_g=0$, the resulting constant tilt–bias combination is invisible to position. In the upright static linearisation, a yaw error and yaw-axis gyro bias are two other position-only null directions. Heading makes those two directions visible. Horizontal translational motion and attitude variation can separate more of these modes over time through $R[f]_\times$. Excitation of particular directions and timing of measurements matter.

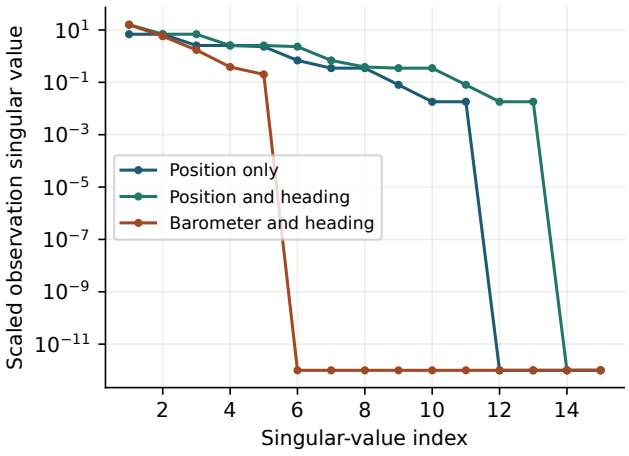

Figure 7.1 – Singular values of a scaled stationary observation stack at eleven equally spaced samples over 2 s. Small values are floored for plotting. Ranks at a relative 10^{-9} threshold are 11, 13 and 5 for the three cases.

The combined GNSS-position and heading stack has rank 13, with the two horizontal tilt–bias combinations remaining. With barometer and

heading only, the five observed coordinates are altitude, vertical velocity, vertical accelerometer bias, yaw and yaw-axis gyro bias at upright rest. This is a finite-interval local test, not a rank claim for a manoeuvring quadrotor or a guarantee that every nominally visible mode is well conditioned. It explains why a well-behaved heading estimate alone provides little evidence that horizontal position is accurate.

7.3 Outlier rejection and consistency

For each new sensor message, evaluate NIS using the prior innovation and covariance. The gated variant accepts when $\eta \le \chi^2_m(0.9973)$, with m matching that sensor's observation dimension. A rejected message leaves the predicted state and covariance as they were before that correction. Clean E2 uses no gate to assess the nominal innovation distribution. Gating can cause run-to-run differences after the same event because state evolution and prior uncertainty depend on accepted observations.

A threshold at the stated percentile has nominal rejection probability approximately 0.27% per independent Gaussian innovation under its model. A per-message nominal false-rejection proportion does not describe the chance of any rejection in a run, which depends on the number of updates and their dependence. The benchmark logs the pre-gate NIS and uses explicit injection tags to calculate false rejection and detection rates. Persistent bias, moderate errors within the gate, and repetitive interference require tests beyond one independent residual.

7.4 Multirate timing and delayed observations

The Python filter propagates at IMU time and updates only at the scheduled sensor events. The order for simultaneous events is position, height, then heading. Under independent sensors with linear models this is equivalent to one stacked correction. With relinearisation and rejection, the order can affect the results. Sensor timestamp is the instant at which the measurement applies. Arrival time can differ because of computation, bandwidth or transport.

For an observation $z=h(x(t-\tau))+v$ treated as current, a locally smooth trajectory gives a systematic innovation component $-H\dot{x}\,\tau$. For position the lag is approximately $-v\tau$, and for heading it is $-\dot{\psi}\,\tau$ followed by wrapping. Queued observations can therefore generate apparently large innovations during motion, even with the correct sensor variance.

A fixed-lag design can retain posterior states, covariances and IMU increments. When a late measurement arrives with a known acquisition time, restore or interpolate the buffered state at that time, apply the update and replay subsequent input increments and dependent corrections in timestamp order. Retaining just a current pose and an old image gives a different measurement model. The numerical latency case below compares chronological zero-delay updates with an explicit delayed-as-current misuse of the same packets. This provides a controlled estimate of the timing penalty for the chosen trajectory.

8. Multirotor estimation benchmark

8.1 Reproducible design

The duration is 56 s. The 64 paired realisations use the same noise-free flight and independent noise arrays within each run. True biases have Gaussian offsets and random walks, both known for scoring. At initialisation, perturb nominal position, velocity and attitude around truth using prior standard deviations. Position uses (0.65,0.65,0.40) m, velocity uses 0.30 m/s on each axis, and the small rotation errors use (4,4,12)°. Initial bias estimates are zero with standard deviations 0.11 m/s² and 0.46°/s per axis. The Gaussian errors are sampled independently across initial coordinates. This is an informed initial prior for study of drift after convergence.

Table 8.1 – Synthetic sensor model. IMU parameters are continuous covariance amplitudes. RW denotes bias random walk.

Signal	Rate	Noise or specification
Accelerometer	100 Hz	0.052 m/s²√s, RW 0.008 m/s²/√s
Gyroscope	100 Hz	0.19 °/√s, RW 0.018 (°/s)/√s
GNSS position	5 Hz	0.45 m (x, y), 0.75 m (z)
Barometric height	10 Hz	0.30 m, zero assumed offset
Absolute heading	5 Hz	2.5°, independent synthetic source

Reference process noise and the simulated noise amplitudes are matched in E2. GNSS fixes and absolute headings arrive at 5 Hz, heights at 10 Hz. Stress applies a 15 s GNSS blackout on 24≤ t<39 s, position errors of roughly 7 m in norm to individual fixes at 20, 42 and 49 s, isolated 3 m barometric errors at 21 and 47 s, and 32° and −27° heading offsets at 18 and 43 s. The matched runs use the same underlying sensor draws. Auxiliary channels continue through the GNSS outage. The no-heading run omits all absolute heading updates, while the inertial-only run applies none of the correction channels.

The clean latency case delays GNSS and absolute heading by 0.20 s and barometer by 0.10 s, then applies the stale values as current. The chronology-correct, zero-delay condition uses the original times. The trigger windows and noise parameters are prescribed, so the conclusions are conditional on this disturbance design.

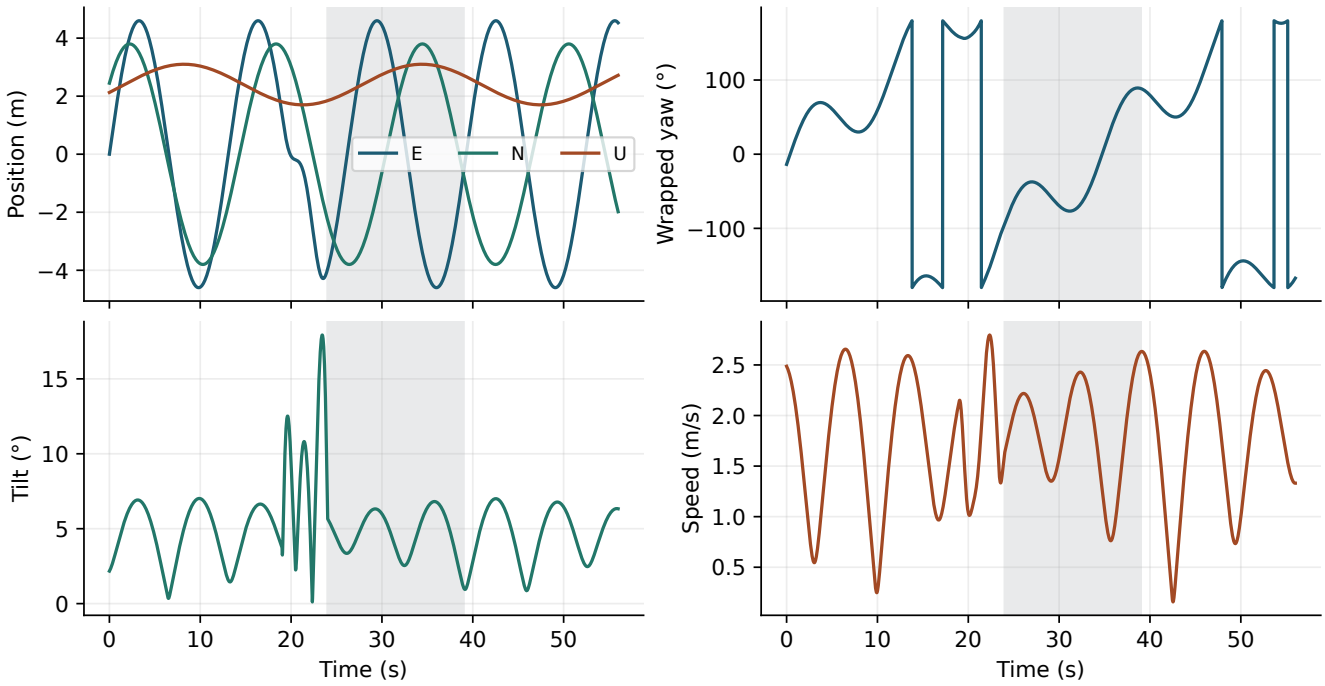

Figure 8.1 – Synthetic pose, yaw, tilt and speed. The shaded interval is the GNSS outage of the stress case. The heading wraps in the upper-right panel, while the physical rotation remains continuous.

8.2 Definition of errors

Horizontal error uses the Euclidean norm of the east and north errors. Vertical error uses the local up coordinate. The main orientation statistic is the geodesic angle $\|\mathrm{Log}(\hat{R}^\top R)\|$, equivalent to rotation distance for angles at most π. The yaw statistic uses the circular difference of the two projected azimuths and is reported separately. All angle RMSE values are converted from radians to degrees. For pooled metrics, the mean of squared errors is taken jointly over time and runs. Confidence intervals use paired trial-level RMSE differences as in Section 4.

8.3 Matched clean case

With uninterrupted clean updates, horizontal RMSE is 0.337 m and vertical RMSE 0.101 m after

5 s. Orientation and yaw RMSE are 1.108° and 0.611°. ANEES is 14.91 against a 15-dimensional reference, averaged over runs and time. Mean pre-update NIS values are 3.020, 0.983 and 1.008 for position, barometric height and heading. Marginal position hit rates are 94.6%, 95.2% and 94.9% across the three axes. The metrics agree with the model and uncertainty assumptions of this matched simulation, with correlated temporal excursions from the pointwise bands in Fig. 8.2.

Table 8.2 – Drone simulation over 64 trials and times t ≥ 5 s. Mean NEES is over 15 local error coordinates. The inertial-only case loses the small-error assumption during the run.

Case	Horiz. (m)	Vert. (m)	SO(3) (°)	Yaw (°)	ANEES
Clean full	0.337	0.101	1.108	0.611	14.91
Stress ungated	6.189	0.115	1.416	0.784	20.84
Stress gated	5.876	0.104	1.280	0.624	15.29
Stress no heading	6.311	0.104	7.381	7.278	16.02
Inertial only	906.960	123.741	21.832	20.173	60.49
Clean delayed as fresh	0.606	0.102	2.990	2.773	58.57

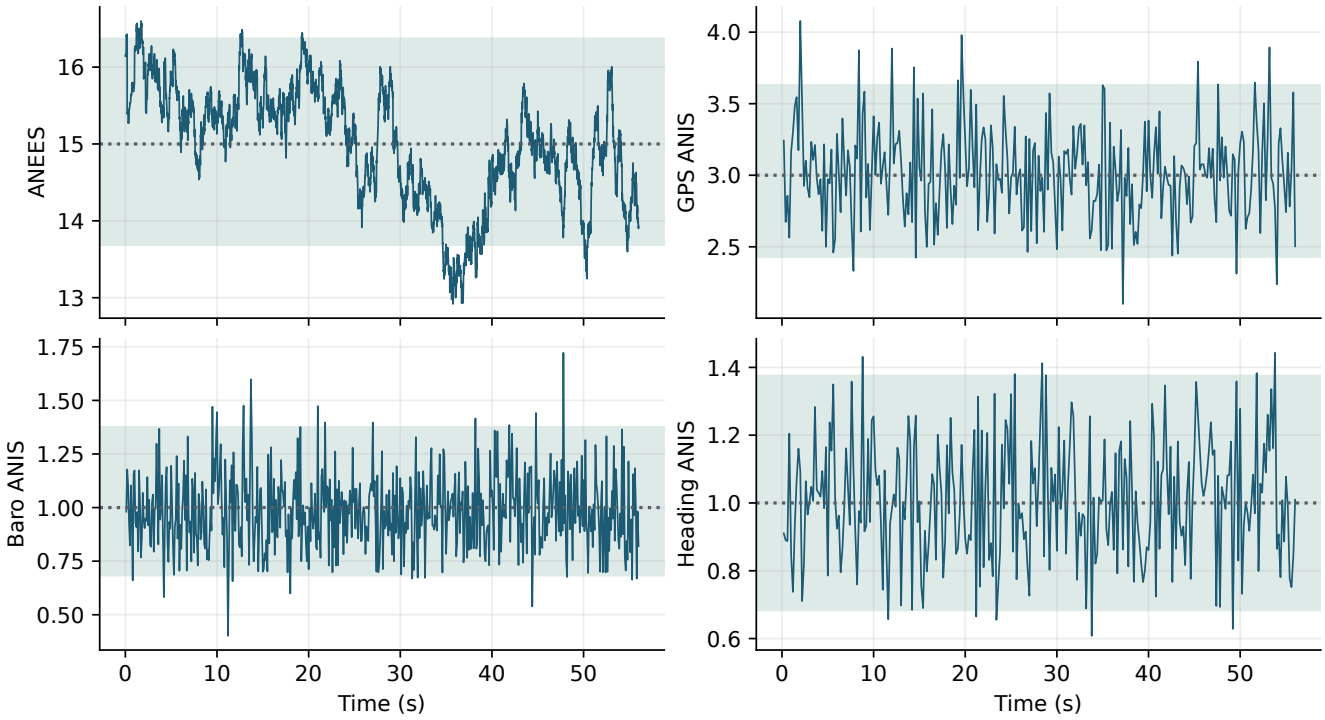

Figure 8.2 – Fixed-time ensemble means for the clean case. The shaded ranges give pointwise 95% chi-square reference intervals averaged over 64 independent trials. The lines at 15, 3 or 1 are the respective dimensional means.

For the clean run, the smallest sampled covariance eigenvalue remains about 1.11×10^{-6}, and the maximum orthogonality deviation in the rotation matrices is 4.30×10^{-14} in the maximum-entry norm. These checks establish numerical symmetry and rotation integrity in the simulated case. The rotation orthogonality check complements the geodesic and circular angle errors, which detect attitude differences even when both matrices are proper rotations.

8.4 GNSS loss and residual gating

Absolute heading strongly reduces yaw error during the GNSS outage. With it, the gated variant has 0.588° gap yaw RMSE. Without it, gap yaw RMSE is 8.220°. The paired difference in per-run

yaw RMSE from adding heading is −5.94°, 95% interval [−7.24,−4.70]°. Despite retained orientation information, the aided filter reaches 10.82 m horizontal RMSE during the gap. At 38.99 s its horizontal median error is 20.36 m and 95th percentile is 35.53 m. Over the first five seconds after fixes return, its pooled horizontal RMSE falls to 0.420 m.

Table 8.3 – Stress-test metrics for the 24–39 s GNSS outage and the first five seconds after GNSS return. Pre-return median and P95 use the sample at 38.99 s.

ESKF variant	Gap XY (m)	Gap yaw (°)	Pre-return XY median / P95 (m)	Recovery XY (m)
Ungated	11.38	0.62	21.46 / 38.12	0.81
Gated	10.82	0.59	20.36 / 35.53	0.42
Gated, no heading	11.62	8.22	20.05 / 37.73	0.44

The mechanism is exposed by an acceleration-bias calculation. Horizontal error can accumulate as roughly ½δ a t^2 when a persistent acceleration error remains uncorrected. A 0.10 m/s² constant error for 15 s yields 11.25 m of displacement from that term alone. The measured gap metrics also include initial velocity error and uncertain orientation and gyro bias. The clean and stress curves make the effect of removing positional updates visible without inferring any fault in the integrated motion equations.

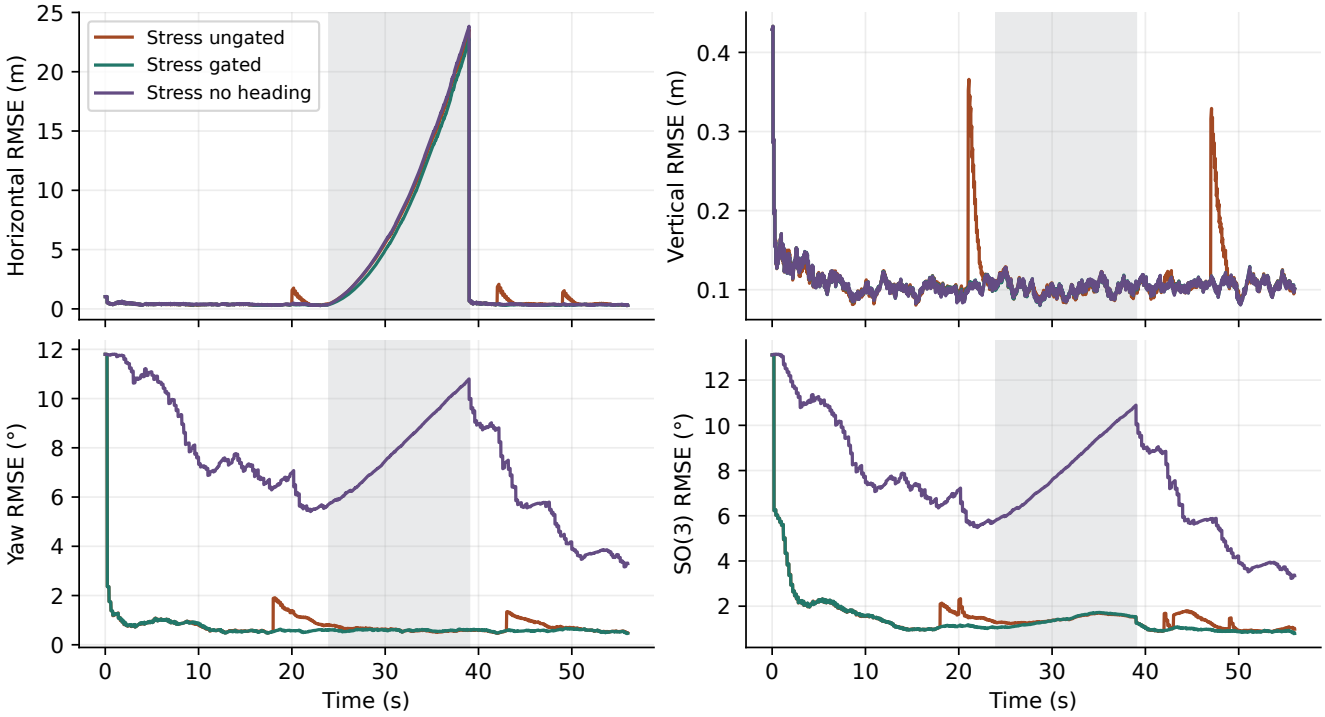

Figure 8.3 – Time-resolved RMSE across 64 stress trials. All curves share the dropout and injected measurements. The no-heading ablation leaves the barometer and available GNSS corrections in place.

Table 8.4 – Pre-update gate decisions for the full stress filter. Counts include all trials.

Sensor	Injected rejected	Nominal rejected	Rate (%)
GPS	192 / 192	41 / 12928	0.32
Baro	128 / 128	84 / 35712	0.24
Heading	128 / 128	48 / 17792	0.27

Gating rejects all 448 of the deliberately large injected sensor events in the matched stress runs. It rejects 173 nominal samples among 66,432 nominal updates, 0.26%. For horizontal RMSE over the complete scored run, the paired mean change from gating is −0.354 m with 95% interval [−0.697, +0.003] m. The data therefore support separation of the specified injected readings from typical residuals. Evidence of a global position improvement is weaker because the uncorrected outage dominates the horizontal error and the paired interval reaches zero. Post-return horizontal RMSE is 0.420 m with the gate and 0.807 m without it in these runs.

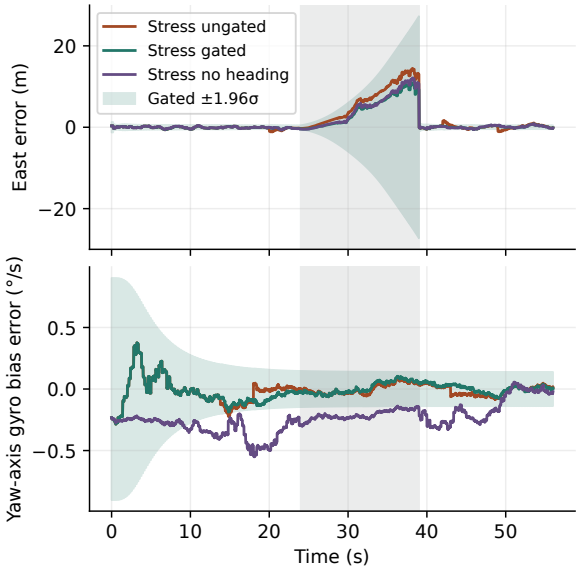

Figure 8.4 – East error and yaw-axis gyro bias error for the first fixed trial. The band is the gated variant's marginal uncertainty and the shaded interval is the GNSS outage.

The lone inertial trajectory baseline has very large errors. Across its runs the total horizontal RMSE is 907 m over the scored times. This baseline quantifies integration of sampled bias and orientation error without correction under the assumed initial conditions. In that regime the local Gaussian error approximation is stretched, so its NEES also loses its nominal interpretation.

8.5 Timestamps and latency

The delayed-as-current case gives 0.606 m horizontal RMSE and 2.773° yaw RMSE, compared with 0.337 m and 0.611° in the clean chronology. The mean paired increase in per-run horizontal RMSE is 0.268 m, interval [0.260, 0.276] m. Position coverage on the east axis falls from 94.6% to 57.2%, and heading ANIS grows from 1.01 to 2.81. Mean ANEES increases from 14.91 to 58.57. Measurement latency therefore changes both estimator accuracy and covariance credibility when the delayed state is interpreted as a present observation.

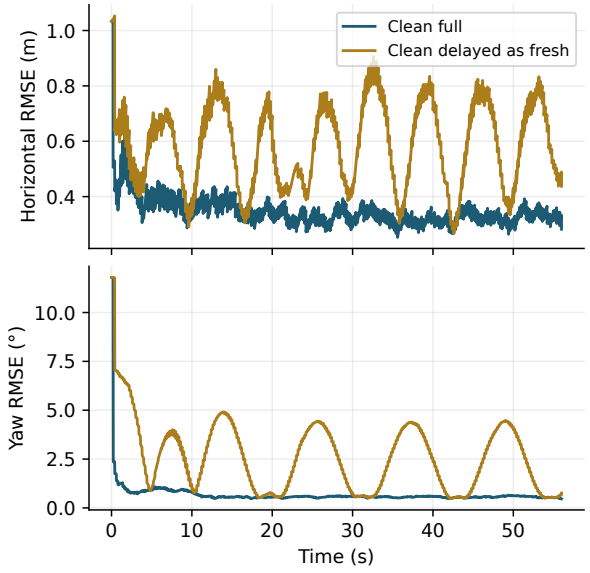

Figure 8.5 – Matched runs with chronological updates and the same sensor packets treated as fresh on delayed reception. The different time dependence of the position and yaw errors reflects the motions driving their residuals.

This is a fixed-lag input error, distinct from sensor dropout. Retimestamping estimates or broadening model covariance alone would obscure the physical source of the residual. Section 10 tests acquisition-time replay on logged Gazebo data, with the delay applied to GNSS alone.

9. Camera geometry and the archived position filter

9.1 What the archive implements

The ROS node named **human_sensor_fusion.py** tracks a planar position with state (x,y,v_x,v_y) [2]. Its two updates accept a synthetic GPS fix and a point inferred by intersecting a camera ray with a horizontal plane. The GPS feed is formed in **simulation_state.py** by adding 1.5 m standard-deviation noise to the verified simulator position before encoding latitude and longitude, nominally once per second in wall time. The filter converts that fix back to an approximate local tangent plane. Rotation of the ray uses the simulator-derived camera orientation, and the image path estimates the ground point from the lower centre of a detection box. A speed direction computed from filtered velocity drives the reported bearing, with a low-speed fallback. That bearing describes the tracked planar motion.



Figure 9.1 – Original simulated camera still [2], **photo_163539_007258.jpg**. It documents the RGB scenario and viewing angle. The archive supplies no synchronised state-error series for this image.



Figure 9.2 – Original simulated camera still [2], **photo_162232_511053.jpg**. The apparent scale illustrates why pixel statistics alone cannot specify metric position uncertainty.

The stills use the Gazebo actor model credited to Open Source Robotics Foundation, © 2012, Apache License 2.0. The same provenance is documented in the simulation archive [2]. The archived transition is constant velocity. Each axis uses $(p,v)\mapsto(p+\Delta v,v)$. The reported process covariance uses the held-acceleration form in Eq. (2.9) with an assumed 0.8 m/s² standard deviation along each axis. The camera update constructs range-dependent, anisotropic radial and lateral variances around the projected point. The position filter applies the Joseph covariance update. Its first update keeps the diagonal position variances, and later updates use the supplied two-by-two noise matrix. Internally the elapsed time is based on monotonic processing time and clamped between 0.001 s and 1 s. The publishing timer reports the latest updated state. Thus the current implementation is a small asynchronous position estimator operating on simulated data and pose. A time-indexed replay would distinguish capture time, update arrival time and the time at which the estimate is reported.

9.2 Projection and linear sensitivity

For an image with width W, horizontal field of view γ, and centre point (c_x,c_y), the focal length under square pixels is $f_x=f_y=W/[2\tan(\gamma/2)]$ [11]. The archived SDF values give f=443.53 px to two decimal places. The camera axes X forward, Y left and Z up give the unnormalised ray

$$d_c = \left(1, \frac{c_x - u}{f_x}, \frac{c_y - v}{f_y}\right)^\top, \qquad d_w = R_{wc} d_c. \tag{9.1}$$

With origin o and plane $\{r : n^\top r = d\}$, the intersection is $o + \lambda\, d_w$. Its parameter is

$$\lambda = \frac{d - n^\top o}{n^\top d_w}, \qquad r = o + \lambda d_w. \tag{9.2}$$

Forward intersection requires $\lambda > 0$. If the ray approaches parallel with the plane, the denominator approaches zero and the projected covariance grows. Differentiation at fixed plane gives

$$\delta r = \left(I - \frac{d_w n^\top}{n^\top d_w}\right)(\delta o + \lambda\, \delta d_w). \tag{9.3}$$

For an independent rotation error in camera coordinates, $\delta\, d_w = -R_{wc}[d_c]_\times \delta\theta_c$. Translation and rotation of the camera origin are generally coupled to the vehicle estimate through its lever arm. In the archive the camera world transform composes the vehicle pose with the gimbal pose. If working with detected pixels instead of an already projected position, the motion estimate and projection must use poses evaluated at matching times. Even small tilt or gimbal errors drive large metric errors at shallow depression.

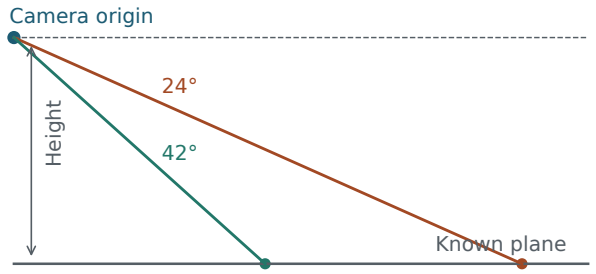

Figure 9.3 – Side-view sensitivity of a principal ray to pitch. The drawing illustrates the denominator in Eq. (9.2). Forward distance increases rapidly as the ray approaches the horizontal.

9.3 Monte Carlo on a static calibration point

E3 uses a stationary plane, a camera origin 2.35 m above it and a static calibration point centred in the image at each nominal angle. In this configuration the ground range is $\rho = H\cot\theta$, with depression $\theta > 0$ and camera-to-plane height H. Its derivative with respect to pitch is $-H\csc^2\theta$. To first order the vertical pixel error near the principal point acts like additional depression δ v/f. The range derivative therefore scales with inverse squared sine.

On each angle use 30,000 independent perturbations of pixel U and V (2 px), pitch (0.45°), yaw (0.40°), horizontal origin (0.07 m per axis), and height (0.10 m), each declared as a separate standard deviation. The same perturbations are reused across angles. Take J as the Jacobian mapping these inputs to range and lateral error. The linear covariance is $J\Sigma\, J^\top$. A central difference at 13° verifies the full seven-input Jacobian to 3.79×10^{-9} maximum absolute entry discrepancy. The nonlinear experiment computes the ray intersection sample by sample and reports the half-width between the 15.9th and 84.1st percentiles. It counts two acceptance definitions, forward direction away from a numerical horizon and the archived validity checks. The archive's geometry accepts a downward ray slope below −0.03 in the unnormalised coordinates, positive height above its 0.05 m plane and a horizontal range from 1 to 45 m.

Table 9.1 – Ray–plane sensitivity with a 2.35 m camera-to-plane height. The MC spread uses forward intersections. Archive validity additionally filters depth, ray slope and range.

Depression (°)	Range (m)	Range σ (m)	MC half-width (m)	Lateral σ (m)	Valid (%)
3	44.84	8.00	8.33	0.38	50.8
5	26.86	3.03	3.08	0.23	100.0
10	13.33	0.91	0.91	0.13	100.0
20	6.46	0.34	0.34	0.09	100.0
40	2.80	0.15	0.15	0.07	100.0

At 20° the nominal range is 6.46 m and the range spread is approximately 0.34 m. At 5° those values are 26.86 m and about 3.08 m. At 3°, the nominal range is 44.84 m and the acceptance share under the archived criteria is about 50.8%, even though essentially all sampled rays intersect the forward plane. The acceptance operation truncates the sample distribution near the range limit. The half-width in Table 9.1 is computed before that outer validity cut, among the forward intersections. It is descriptive of the projected uncertainty.

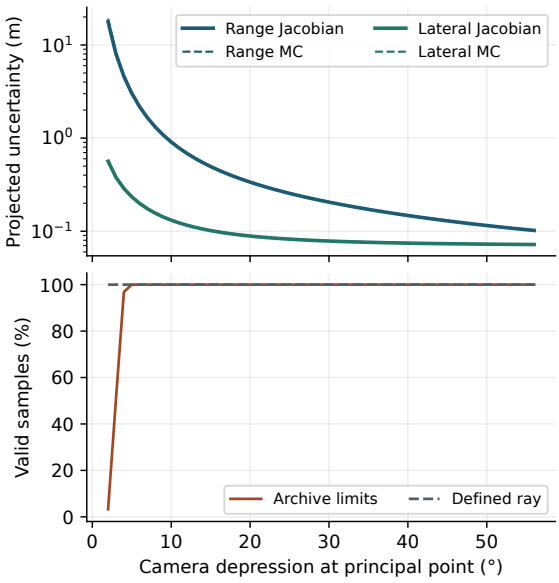

Figure 9.4 – Linearised standard deviation versus nonlinear Monte Carlo central spread. The second panel shows how projecting a shallower ray can fail the archive's geometric limits. Samples are for a fixed calibration point model at each depression.

If a camera projection is formed using the drone's own uncertain pose and then fed back into the same filter as if it were independent, correlations enter the innovation. For linear residual ν=Hε+ν with C_e=Cov(ε,ν),

$$S = HPH^{\top} + R_m + HC_e + C_e^{\top} H^{\top}, \qquad K = (PH^{\top} + C_e)S^{-1}. \tag{9.4}$$

In a more complete visual-inertial model, the raw pixel can be expressed as a function of state, camera calibration and a landmark. This preserves the dependence directly. Synchronised images, a representative calibration scene and an independently measured camera trajectory would be necessary to test a coupled visual update. E3 measures projection sensitivity at fixed inputs, with independent pose and pixel perturbations as stated above.

10. ESKF replay on recorded Gazebo flights

10.1 Acquisition and data checks

E4 replays the estimator on recorded ROS 2 Jazzy and Gazebo Harmonic streams [12]. The acquisition used an instrumented copy of the airframe with a 2 ms physics step and native IMU, NavSat, pressure and magnetic sensors. Three 56 s flights repeat the same velocity-command sequence using simulator seeds 4102–4104. Each flight begins with hover, translates, crosses the wrapped yaw boundary, climbs, translates laterally, descends and returns to hover. A separate 25 s hover is retained for sensor diagnostics and nominal replay. The archived camera scenes in Sec. 9 and this new instrumented scene are separate trials of the same project family.

The validation copy centres the combined mass on **base_link**. If m_b is the base mass and m_j,r_j are the attached masses and offsets in that frame, the inertial origin of the base is set to $r_b = -\sum_j m_j r_j / m_b$. With the masses in Table 5.1, the correction is −4.074 mm along X and +0.611 mm along Z. This preparation affects the Gazebo plant through its mass distribution. The archival acquisition records the original and instrumented SDF and their provenance. The controller operates against Gazebo's internal state and the recorded reference topics are isolated from estimator feedback.

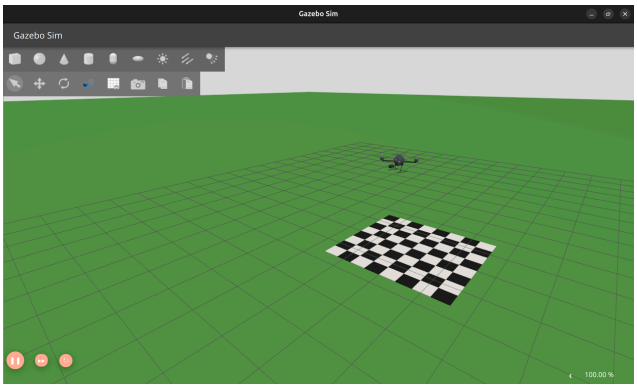

Figure 10.1 – Gazebo capture from flight 4102 during the ascent. The checkerboard is a fixed landmark in the same recorded scene [12].

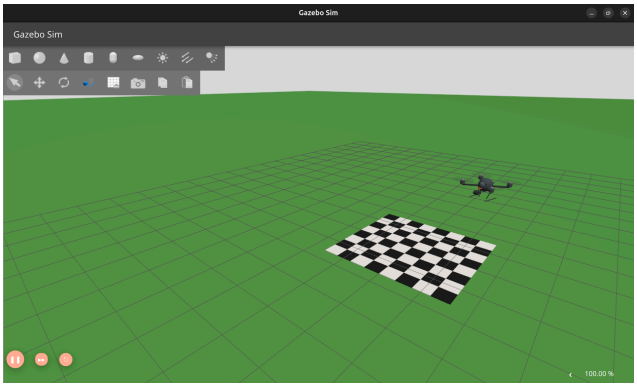

Figure 10.2 – Gazebo capture from flight 4102 during the final hover [12]. Capture images illustrate the scene; native timestamps are used for scoring.

Table 10.1 – Native observations in the instrumented Gazebo trials. The yaw wrap count uses reference orientation. GNSS position and velocity each have the stated count.

Run	Overlap (s)	IMU samples	Truth samples	GNSS fixes	Yaw wraps
flight 4102	55.992	14000	7000	560	1
flight 4103	55.992	13999	6999	560	1
flight 4104	55.992	14000	7000	560	1
hover 4101	24.992	6250	3125	250	0

The supplied MCAP bags and CSV exports were cross-checked offline. All 331 recorded file hashes agree with the archive manifest. The four bags contain 260,294 messages. Direct CDR decoding verifies 119,657 payloads for the quantities used in the replay and their original logging times [13]. This includes timestamps, the body accelerometer and gyro, reference pose and world velocity, GNSS position and velocity, pressure and the magnetic field. Camera geometry uses the separately supplied native camera exports. The source acquisition also records a successful ROS bag playback check.

The IMU has 4 ms sample intervals except for one 8 ms interval in flight 4103. The overlap between IMU and the reference spans 55.992 s in each flight and 24.992 s in hover. The maximum difference between the acquisition stamp and the recorded clock in the checked estimation streams is 6 ms. The world and body reference velocity exports agree to within 2.3×10^{-16} m/s after rotation using the logged orientation. In the separate hover after 5 s, the sample standard deviations range from 0.02978 to 0.03008 m/s² for the accelerometer and 0.00199 to 0.00203 rad/s for the gyroscope. These values agree with the configured 0.030 m/s² and 0.002 rad/s noise at the recorded rate. They characterise this simulator configuration.

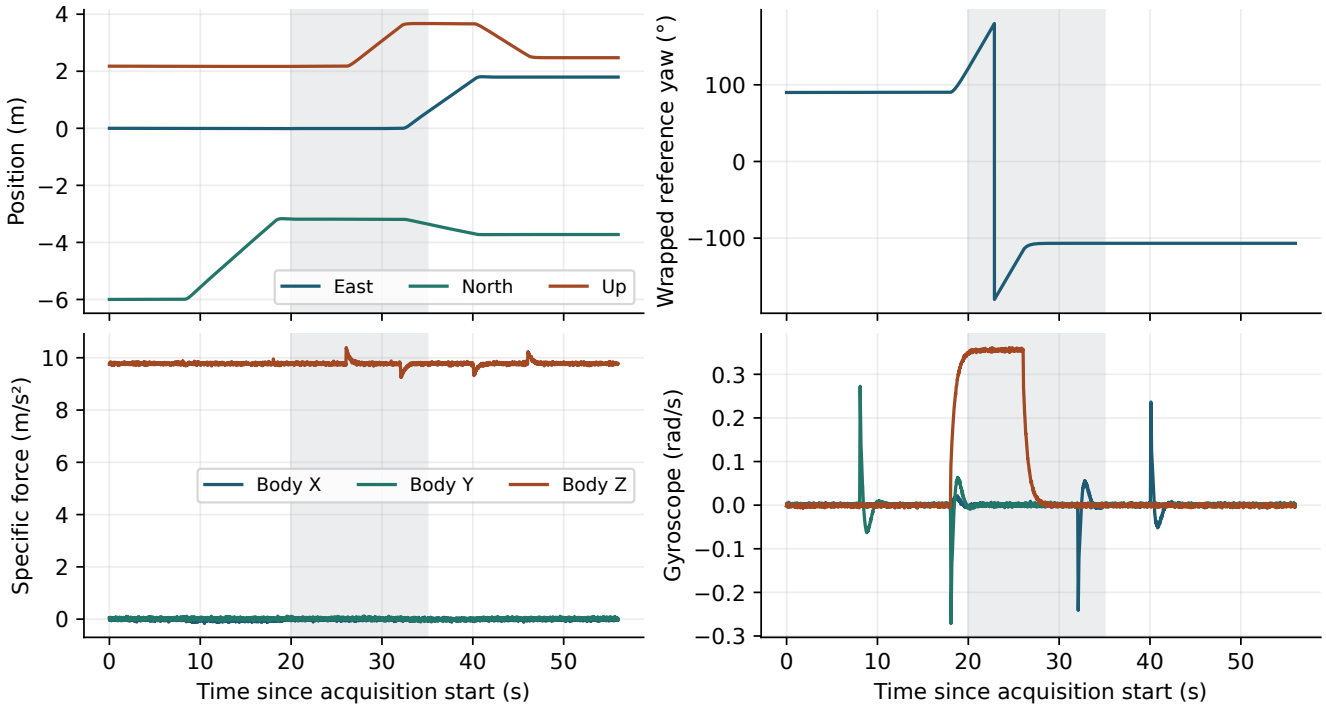

Figure 10.3 – Recorded ENU position, wrapped yaw, specific force and gyro for flight 4102. The shaded interval locates the artificial GNSS blackout applied only in the replay.

10.2 Model and sensor initialisation

The nominal and error states retain the convention in Sec. 6. The rotation maps the forward–left–up body frame into the east–north–up world frame. Raw gyroscope and accelerometer drive propagation, with the IMU's ideal orientation field excluded. Gravity is (0,0,−9.8) m/s² in this world. For observed time increment Δt_k, the nominal step uses the newly available sample and the midpoint rotation. The white process power over the recorded accelerometer and gyro channels uses $\sigma_{sample}^2 \Delta t_0$, with nominal Δt_0=0.004 s. The two-point discretisation from Eq. (6.7) uses the actual step, including the 8 ms interval. Constant bias states match the configured static bias model.

Measurements over the first two seconds initialise all cases identically. The averaged specific force levels the initial orientation. The horizontal projection of the averaged magnetic field fixes yaw relative to the local magnetic model. Mean GNSS position and velocity initialise the translation, and the stationary gyroscope average initialises its bias. The accelerometer bias starts at zero with prior standard deviation 0.016 m/s² per axis. Reference pose and velocity enter scoring after estimation. Position and velocity reference interpolation is linear, and orientation interpolation uses spherical interpolation of quaternions.

GNSS position and velocity

Raw NavSat fixes are converted from WGS84 geodetic coordinates to ECEF, then rotated into local ENU using the configured origin at 48.8566° N, 2.3522° E and 35 m elevation. Version 8.2.2

applies the archived angular perturbation in the latitude and longitude values [12,14]. At the origin and to first order, an equal angular standard deviation s_γ in both coordinates gives horizontal metric deviations

$$\sigma_E = (N+h)\cos(\varphi)s_\gamma, \quad \sigma_N = (M+h)s_\gamma. \tag{10.1}$$

Here N and M are local WGS84 curvature radii. The replay uses (0.330,0.500,0.800) m for GNSS position and (0.05,0.05,0.08) m/s for the native ENU velocity. Reported zero or unknown covariances are interpreted using the configured noise model. The velocity observable is the Gazebo NavSat velocity field with its own noise, not the true velocity or a finite difference of the position fix. The nominal case fuses both GNSS fields. The position-only ablation retains the same calibration and omits subsequent velocity updates.

Pressure altitude and relative datum

Let $\bar{P}$ and $\bar{z}_G$ be the mean initial pressure and the mean initial GNSS height. The measurement channel uses a relative pressure conversion local to that datum.

$$z_P = \bar{z}_G + 44330\left[1-\left(\frac{P}{\bar{P}}\right)^{1/5.255}\right]. \tag{10.2}$$

The derivative near $\bar{P}$ converts 1 Pa of configured noise to 0.0833 m of height noise. The reference $\bar{z}_G$ comes from 20 noisy fixes in each trial and is common to every barometer update in that trial. Its standard error from the GNSS noise model is about 0.179 m. The 15-state covariance is conditional on the selected datum. Repeated pressure samples can determine relative vertical changes precisely while that datum contributes a common absolute-height error. This distinction is tested below.

Three-axis magnetic field

The raw magnetic vector provides a body-frame observation. Gazebo's geographic implementation at the configured origin gives declination 0.70566°, inclination 64.09182° and field strength 47.8922 μT from the versioned upstream tables [14]. In ENU the resulting local reference vector is approximately (0.258,20.924,−43.079) μT. The units and axes are enabled explicitly in the experiment. This field is held constant over the few metres of the recorded path. With m^w the local field and $\delta\theta$ the right attitude error, the observation and its attitude Jacobian are

$$z_m = R^\top m^w + n_m, \quad H_{m,\theta} = [\hat{R}^\top m^w]_\times. \tag{10.3}$$

To obtain the sign, write $R^\top m^w = \mathrm{Exp}(-\delta\theta)\hat{R}^\top m^w$ and expand to first order. Central differences check the resulting Jacobian to better than 10^{-7} in μT per radian. The measurement standard deviation is 0.2 μT per axis, and the NIS gate uses dimension three. The exported scalar magnetic heading is useful as a diagnostic. During manoeuvres its accelerometer-based tilt compensation would reuse part of the information already driving the inertial prediction. The replay instead corrects with the raw magnetic vector under the specified observation model.

10.3 Causal handling of recorded and delayed packets

The 15 supplied input schedules specify message retention and planned receipt. The outage removes both GNSS position and velocity from 20 to 35 s after recording start. In the no-magnetic-aiding case the magnetic updates are removed after the common initialisation. The delay adds 200 ms only to GNSS reception. Other sensor stamps and values are held fixed. The combined schedule applies both changes.

For a sensor packet with acquisition stamp t_a and latest logged simulation clock t_c, the base delivery time is $\max(t_a,t_c)$. This rule accommodates callbacks recorded just before receipt of the next clock tick. A packet is eligible for assimilation at the first IMU epoch at or after its acquisition; the quantisation is at most 8 ms in these data. For delayed receipt the estimator first processes the packet only once its scheduled delivery is reached. A raw 200 ms delay and asynchronous event alignment produce a measured rollback bound of 208 ms.

In the reception-time comparator, the received value is treated as a measurement of the present state. In the acquisition-time policy, the filter restores the preceding checkpoint at the measurement epoch, applies the observation in chronological order and repeats all subsequent available updates and inertial steps through the current tick. The code retains the history for offline replay. An already released estimate is frozen before any later packet can arrive. Scores use this causal sequence of published states. The comparison therefore includes the interval during which delayed observations were still unavailable.
Acquisition-time replay can reduce lag bias, and updates still wait for transmission. On a delayed flight, replacing the reference position, velocity and rotation left the estimates bitwise identical. Perturbing a GNSS packet acquired at 29.096 s left all 6,824 released post-calibration outputs before its 29.296 s receipt unchanged.

10.4 Nominal replay and the vertical datum

Table 10.2 – Nominal ESKF accuracy on the common recorded interval from 5 s. Position, velocity and rotation use the Gazebo reference only for scoring.

Run	XY (m)	Z (m)	Velocity (m/s)	SO(3) (°)	Yaw (°)
flight 4102	0.057	0.023	0.020	0.169	0.151
flight 4103	0.057	0.146	0.018	0.105	0.095
flight 4104	0.059	0.170	0.016	0.133	0.118
Three-flight aggregate	0.058	0.130	0.018	0.138	0.124
Separate hover	0.065	0.071	0.023	0.292	0.268

The cross-flight aggregate is the square root of the mean of the three squared per-flight RMSEs. Each run is scored from 5 s to its final common reference interval. The aggregate nominal horizontal RMSE is 0.058 m and the rotation geodesic RMSE is 0.138°. Raw GNSS position fixes in the same time window have about 0.601 m horizontal RMSE at their own sample times. The output and raw baseline use different sample rates. The comparison shows the benefit of the recorded measurements together and does not isolate a single update channel.

The absolute vertical RMSE spans 0.023 to 0.170 m across the three nominal flights. Mean signed true-minus-estimate vertical errors are −0.021, −0.146 and −0.169 m. The pressure updates preserve a nearly constant offset established during calibration. The conditional vertical covariance does not include this datum uncertainty. In flights 4103 and 4104, the nominal 95% marginal intervals at the estimator's output fail to cover the reference throughout the scored period in Z. This is evidence of a missing uncertainty contribution despite small pressure residuals. A future implementation could estimate the pressure datum as a nuisance parameter or use independently calibrated vertical aiding.

Table 10.3 – Nominal pre-gate NIS. Counts are distinct assimilated packets after initialisation across the three flights.

Channel	Dim.	Packets	Mean NIS	Rejected
GNSS position	3	1620	3.068	3
GNSS velocity	3	1618	3.136	4
Pressure height	1	8100	0.971	23
Magnetic vector	3	8100	3.037	14

Mean pre-gate NIS values remain close to their measurement dimensions for the nominal runs. The counts in Table 10.3 include all distinct sensor packets assimilated after initialisation, with NIS evaluated at the last processing of each packet when replay occurs. They are descriptive. Repeated sensor innovations and state errors share information over time. The nominal barometric NIS is 0.971 even with substantial incorrect vertical coverage in two runs, illustrating that the pressure residual alone cannot diagnose a common datum error. Actual sensor bias draws are unavailable as truth, so a full 15-coordinate NEES is reported in the synthetic study E2 only.

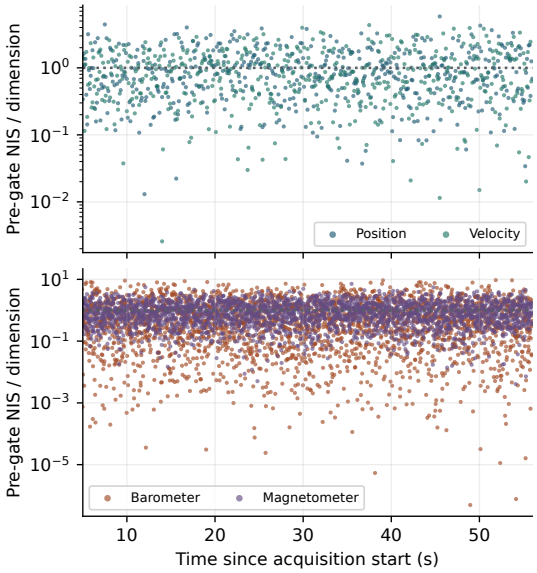

Figure 10.4 – Pre-gate normalised innovation squared divided by measurement dimension in nominal flight 4102. The reference line gives the ideal mean. Each plotted sample belongs to a single sensor update.

10.5 Position and heading through the GNSS outage

Table 10.4 – GNSS outage scores. M and –M denote magnetic aiding present and removed after initialisation. End XY is the horizontal error at the last common sample before 35 s.

Flight	XY M (m)	XY –M (m)	Yaw M (°)	Yaw –M (°)	End M (m)	End –M (m)
flight 4102	1.227	1.469	0.158	0.145	2.836	3.408
flight 4103	0.948	2.044	0.156	0.520	1.991	4.611
flight 4104	1.208	1.618	0.194	0.043	2.720	3.713
Aggregate	1.135	1.727	0.170	0.313	2.543	3.944

During the 15 s outage the aggregate horizontal RMSE increases to 1.135 m with magnetic aiding, from 0.060 m on the same time window in the nominal runs. Just before GNSS returns the horizontal errors are 2.84, 1.99 and 2.72 m. The outage yaw RMSE is 0.170° with the magnetic update, and 0.313° when it is removed after initialisation. Per-flight yaw results differ. Flight 4103 shows the largest magnetic benefit, while the gyro-based short-horizon yaw in 4104 is already accurate on this gap. Horizontal error rises in all three flights when magnetic aiding is removed. The initial stationary calibration, native GNSS velocity when available, gentle manoeuvre and noise settings differ substantially from E2. The 10.82 m horizontal result in E2 and 1.135 m here therefore describe different test conditions.

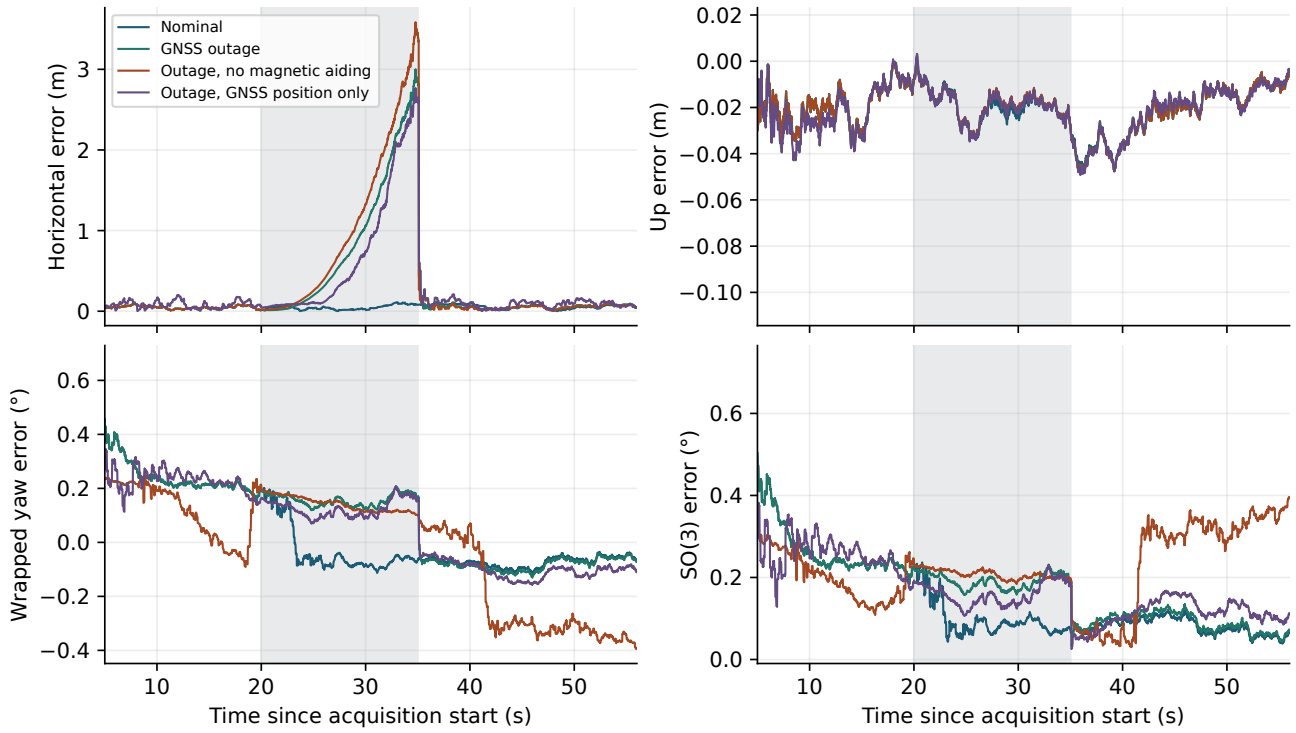

Figure 10.5 – Nominal and outage comparison for flight 4102. The horizontal norm, signed altitude and wrapped yaw errors and the rotation geodesic error are shown from 5 s onward.

The position-only ablation yields 1.019 m aggregate gap RMSE with magnetic aiding. Removing velocity from after the calibration helps two runs and degrades one. During the blackout neither method receives GNSS. Any subsequent drift difference comes from its state and covariance when the blackout begins, with possible interaction with the remaining pressure and magnetic updates. Three realisations of the same trajectory do not establish a general ranking of the two GNSS configurations.

10.6 Effect of acquisition-time replay

Table 10.5 – Causal timing comparison. The last two columns refer to the same time windows in every case. Values aggregate three flights.

Receipt and handling	Overall XY (m)	20–35 s XY (m)	35–40 s XY (m)
Nominal delivery	0.058	0.060	0.078
Delay, as current	0.066	0.065	0.101
Delay, acquisition-time replay	0.060	0.062	0.080
Outage only	0.628	1.135	0.376
Outage + delay, as current	0.736	1.285	0.740
Outage + delay, acquisition-time replay	0.649	1.135	0.641

A 200 ms GNSS reception delay without an outage changes the horizontal RMSE from 0.058 m in the nominal case to 0.066 m when treated as current and 0.060 m with acquisition-time replay. The latter improves all three recorded runs relative to the reception-time comparator. For the combined outage and delay, the first five seconds after the nominal GNSS return have horizontal RMSE 0.740 m when measurements are treated as current and 0.641 m with timestamp-aware replay. The corresponding value with the outage alone is 0.376 m. The original 200 ms delay still holds back

the first usable packet and delays correction. The two delayed cases differ because their state updates treat the same data differently.

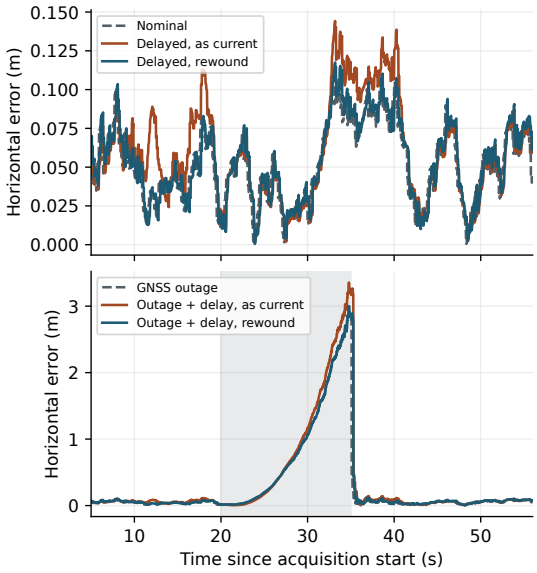

Figure 10.6 – Causal horizontal errors for flight 4102. Top compares a delayed GNSS stream with the nominal case. Bottom repeats the comparison under the 15 s loss of GNSS. State estimates are scored on their initial release, before any future rewind.

The largest observed rewind is 208 ms. Depending on packet arrival grouping, repeated prediction accounts for about 40,700 to 47,700 extra integration steps per 56 s flight in the delayed case. This is a computational cost of an unoptimised checkpoint replay, distinct from the imposed measurement delay. The implementation tests the algorithmic timing effect offline, and these step counts are not a measured embedded compute budget.

10.7 Reprojection on the recorded checkerboard

The archived world contains a fixed checkerboard with 49 inner corners at known world positions. For each RGB frame the native calibration and acquisition stamps are used, with the transform chain world–body–gimbal–camera–optical. The optical pose is where the pinhole formula applies. The focal lengths from **CameraInfo** are 443.530 px in both axes to three decimals, with principal point (320,180) px and zero declared distortion. In camera coordinates $X_c = R_{wc}^{\top}(X_w - p_{wc})$, the pixel prediction is

$$\begin{aligned} u &= f_x X_{c,1}/X_{c,3} + c_x, \\ v &= f_y X_{c,2}/X_{c,3} + c_y. \end{aligned} \tag{10.4}$$

Candidate chessboard corners are detected independently from the negative determinant of a smoothed image Hessian within the achromatic board. A one-to-one minimum-distance assignment pairs candidate image saddles to native geometric predictions within 4 px. This tests the calibration and transform chain, with geometry providing the correspondence. All 5,733 expected corner observations match across the 117 exported RGB frames. The native reprojection RMSE is 0.199 px and the largest matched discrepancy is 0.389 px. The images come from the initial stationary interval with repeated views of the same board. Rendered frame gaps are kept, and the count is descriptive of coverage in this dataset.

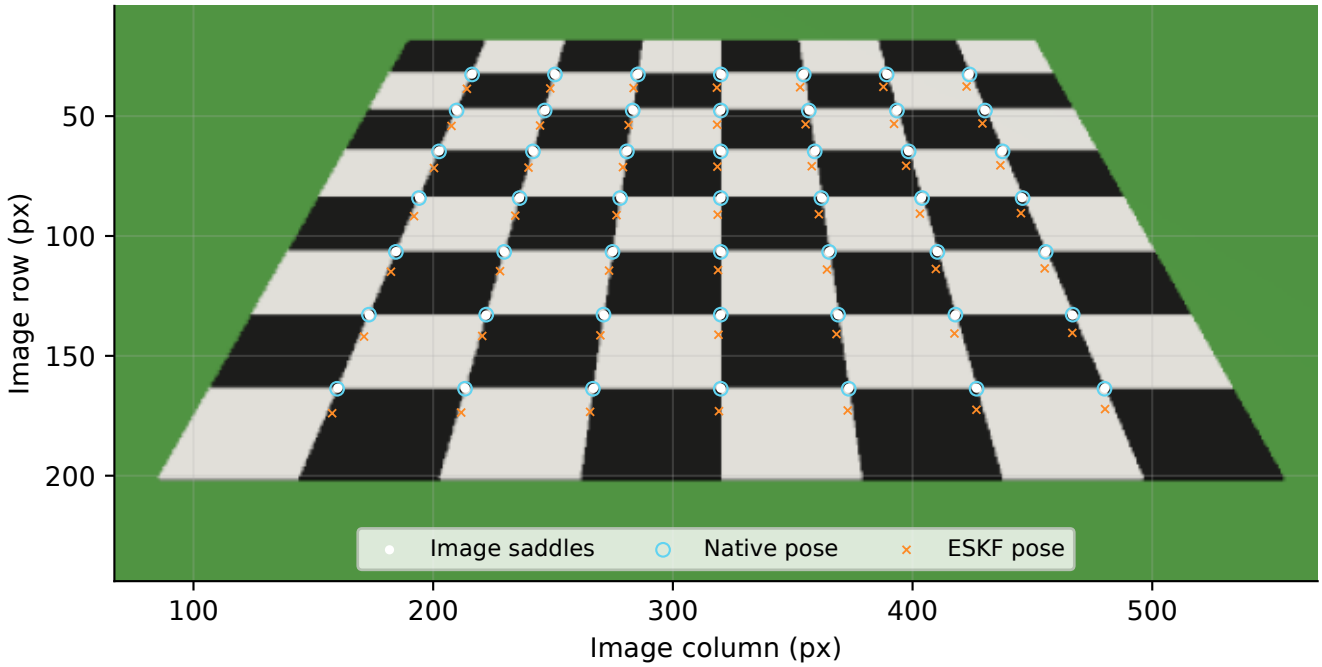

Figure 10.7 – Reprojection on a flight 4102 frame acquired at 3.092 s. White marks are image saddle detections, blue circles use the native body pose, and orange marks use the causal ESKF body pose. Both projections use the same camera calibration and relative gimbal geometry.

Table 10.6 – Reprojection diagnostic on the same 3,773 matched corners of 77 frames. Hybrids use one or more native reference quantities for scoring.

Body pose used	Pixel RMSE
Native pose	0.200
Causal ESKF pose	15.17
ESKF with reference height	7.64
ESKF position and native rotation	14.88
Native position and ESKF rotation	1.56

For 77 frames after initialisation, the most recent released filter pose at the 10 Hz archive cadence gives matched-pixel RMSE 15.17 px. Its maximum age relative to image acquisition is 98 ms. The native projection on exactly those frames gives 0.200 px. Substituting only reference height for the ESKF height in this scoring diagnostic reduces the error to 7.64 px. With reference position and estimated attitude the error is 1.56 px. These substitutions use the reference only to diagnose error sources. They show that the vertical datum and position accuracy matter to visual reprojection even when the orientation error is small.

11. Interpretation and validation protocol

11.1 What changes the conclusions

Algorithm choice and parameter accuracy have distinct effects. In E1 fixed linearisation incurs a deterministic drift error under larger input steps. Moving to the exact flow and updating the covariance Jacobian substantially improves the matched problem, but cannot repair a wrong nonlinear drift coefficient. A covariance scale of one happens to match the generator by design. Calibration against a measured case instead requires independent validation data and physically plausible Q and R_m.

E2 indicates why different diagnostics belong together. An independent heading source can bound yaw error even while horizontal position drifts under missing GNSS. A residual gate removes the

designed impulsive outliers with few nominal rejections. A constant timestamp error instead introduces motion-dependent bias and impaired consistency even in the absence of injected transients. These mechanisms produce different innovation patterns and are associated with different physical corrections.

E3 links pose quality to image-derived quantities. For a shallow ground ray, the projection is strongly sensitive to pitch and vertical pixel error. An estimator with a scalar camera covariance cannot express all of that anisotropy. A linearised covariance is useful well away from tangency. The reported empirical spread and acceptance at shallower angles quantify the limits of representing the distribution as one untruncated Gaussian.

11.2 Scope of the Gazebo replay

E4 extends the estimator to logged data from a configured simulator with physical dynamics, native synthetic sensors and distinct noise seeds. The world experiment adds instrumentation and a commensurate initialisation procedure. The analysis verifies timestamps and frame conventions, then reserves native pose and velocity for scoring. Three runs cover one scripted motion and three noise realisations. The numbers describe that repeatable experiment. The much greater horizontal drift in E2 also reflects its different motion, IMU noise, initial errors and sensor updates.

A camera reprojection using simulator transforms checks the optical model against rendered data. Using the released nominal ESKF pose for the same corners compounds its position and attitude error with that transform. Since the sampled images cover the initial stationary interval, the check does not establish timing accuracy during motion or a visual-inertial filter. A dynamic camera follow-up can use multiple calibrated viewing directions and timestamped images throughout the manoeuvre.

11.3 From idealised to measured sensors

Noise amplitudes, scale factors and bias dynamics in E2 are assigned research inputs. In a measured study, a static IMU run, temperature-controlled drift test, and independent attitude reference would separate white noise, random walk and systematic calibration error. GNSS error can be coloured by multipath and may cease to look Gaussian. Barometer samples may carry drift and motion-correlated pressure bias. A physical absolute heading sensor can suffer magnetic error or require separate calibration. Each of these changes the modelled likelihood and the interpretation of NIS. E4 provides synchronised simulator truth and the native sensor streams. The original project videos in Sec. 9 illustrate the camera scenario.

The numerical priority is to separate faults in the model from faults in the sensor schedule. A recorded residual should contain the acquisition stamp, elapsed inertial integration interval and innovation before gating. Comparing fixed-trajectory runs with alternative timing, sensor outages or validity conditions can then identify which mechanism changes performance. Finally, a position or attitude error estimate is meaningful only with a stated frame, time and data source.

11.4 Conclusion

The exact sampled solution of the two-state model provides a reference against which local linearisation, EKF, UKF and smoothing can be compared. The matched high-excitation trials show a large advantage for the EKF over a fixed KF in x_1, while the mismatched trials demonstrate how model bias can reverse that result. The aircraft benchmark extends the same predict–correct framework to a rotation-aware 15-state filter with gyro and accelerometer biases. The 15 s outage and latency tests make observability and timing effects measurable. The camera calculation shows how these orientation and timing problems propagate into geometric estimates. E4 tests that estimator on three synchronised Gazebo flights using native simulated sensor streams. Acquisition-time replay reduced the penalty of stale GNSS updates, while the blackout and pressure datum constrained performance. The supplied data, code and uncertainty summaries give a reproducible reference for later measured sensor experiments.

References

[1] D. Kumar. *Estimation d'état par filtre de Kalman.* Estimation theory report, École nationale supérieure d'arts et métiers, Paris, 2026. Supplied PDF and Word versions. Equations and MATLAB figures in Appendix A.

[2] D. Kumar. *Multirotor simulation project.* Local ROS 2 Jazzy and Gazebo Harmonic source archive, 18 October 2026 snapshot. SDF geometry, bridge configuration, state publisher, camera and position fusion code, and simulated recordings.

[3] R. E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82(1), pp. 35–45, 1960. 10.1115/1.3662552.

[4] S. Särkkä and L. Svensson. *Bayesian Filtering and Smoothing.* 2nd ed., Cambridge University Press, 2023. Publisher record.

[5] C. F. Van Loan. Computing Integrals Involving the Matrix Exponential. *IEEE Transactions on Automatic Control*, 23(3), pp. 395–404, 1978. 10.1109/TAC.1978.1101743.

[6] S. J. Julier and J. K. Uhlmann. Unscented Filtering and Nonlinear Estimation. *Proceedings of the IEEE*, 92(3), pp. 401–422, 2004. 10.1109/JPROC.2003.823141.

[7] H. E. Rauch, F. Tung and C. T. Striebel. Maximum Likelihood Estimates of Linear Dynamic Systems. *AIAA Journal*, 3(8), pp. 1445–1450, 1965. 10.2514/3.3166.

[8] J. Solà. *Quaternion Kinematics for the Error-State Kalman Filter*. arXiv:1711.02508, 2017. Preprint.

[9] F. L. Markley. Attitude Error Representations for Kalman Filtering. *Journal of Guidance, Control, and Dynamics*, 26(2), pp. 311–317, 2003. 10.2514/2.5048.

[10] T. Foote and M. Purvis. *Standard Units of Measure and Coordinate Conventions.* ROS Enhancement Proposal 103, 2010. REP 103 source.

[11] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision.* 2nd ed., Cambridge University Press, 2004. Authors' companion site.

[12] D. Kumar. *Validation Kalman Drone.* Recorded ROS 2 Jazzy and Gazebo Harmonic experiment archive, November 2026. Four MCAP recordings, CSV exports, instrumented SDF, acquisition scripts and source manifest. Supplied with this study.

[13] Foxglove. *MCAP Format Specification.* File, chunk, channel and message record definitions. Format specification.

[14] Open Source Robotics Foundation and Gazebo contributors. *Gazebo Sim 8.15.0 and Gazebo Sensors 8.2.2.* Magnetometer and NavSat implementations. Exact tagged source files and Apache-2.0 license in archive [12]. Gazebo Sim and Gazebo Sensors.

Appendix A. Original estimation results

A.1 MATLAB parameters and figure provenance

The original report identifies α=4,β=1, $x_0=(4,x_{2,e})$, sample period 0.1 s and observation variance 10^{-2} [1, pp. 4–6]. The local-model Jacobians are the same as in Eq. (3.2). The short MATLAB transcription below preserves the posted numerical model and uses ASCII names for the variables. It documents the stated input parameters. The subsequent figure atlas preserves the 16 chart images extracted directly from the Word file, including both gain variants in the archived sequence. Axes and legends remain as exported.

```
alpha = 4; beta = 1;
x2e = sqrt(beta/alpha);
ue = beta/x2e;
A = [0, -ue; 0, -2*alpha*x2e + ue];
B = [-x2e; x2e]; C = [1, 0];
X0 = [4; x2e]; X0lin = X0 - [0; x2e];
Te = 0.1; R = 1e-2;
```

A.2 Archived linear model

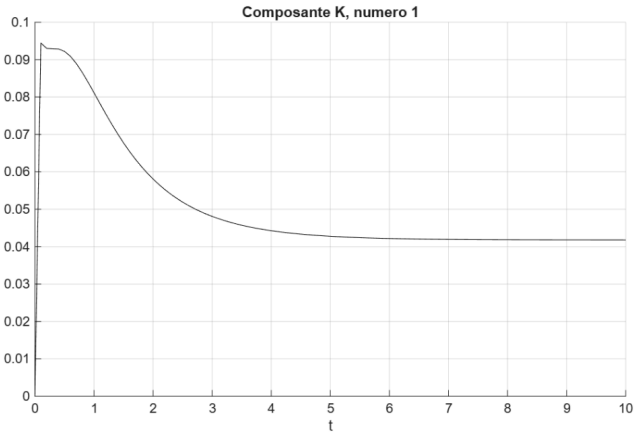

Figure A.1 – Original MATLAB first gain component [1, p. 7].

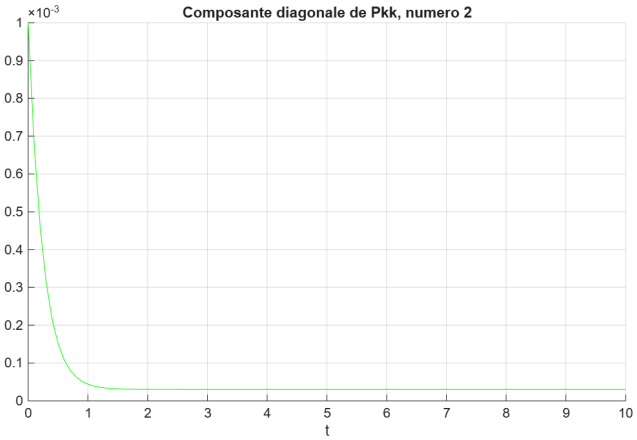

Figure A.2 – Original second covariance diagonal [1, p. 7].

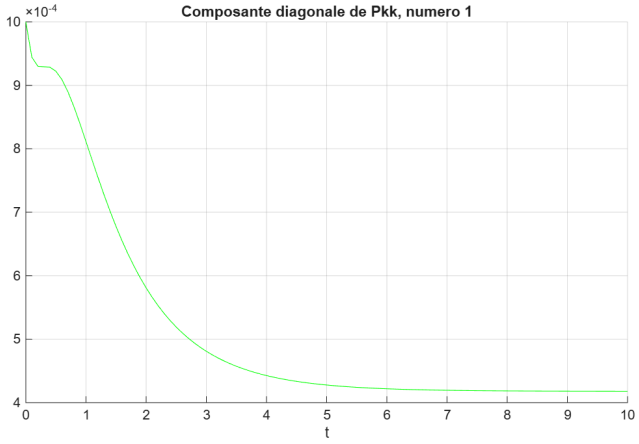

Figure A.3 – Original first covariance diagonal [1, p. 7].

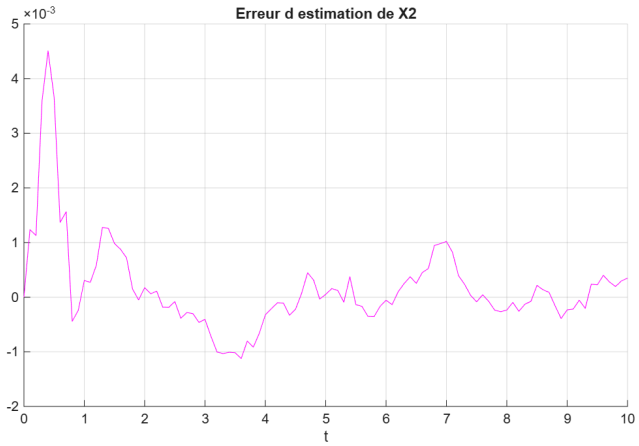

Figure A.4 – Original second-state estimation error [1, p. 8].

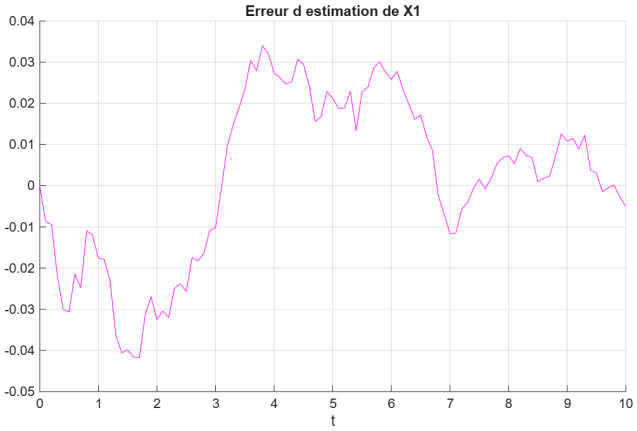

Figure A.5 – Original first-state estimation error [1, p. 8].

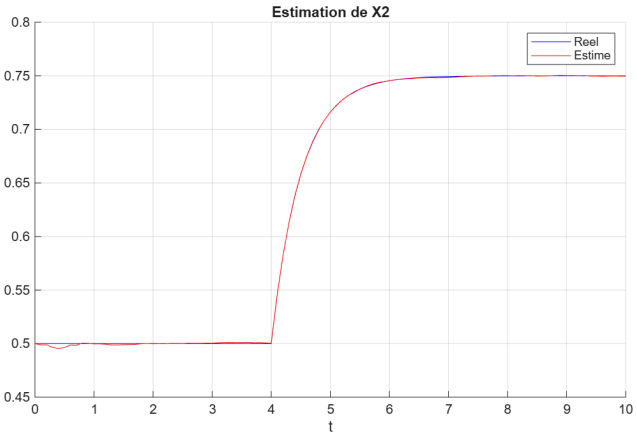

Figure A.6 – Original second-state reference and estimate [1, p. 8].

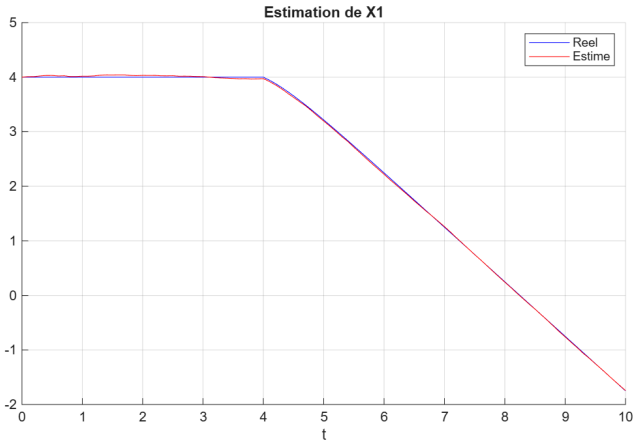

Figure A.7 – Original first-state reference and estimate [1, p. 9].

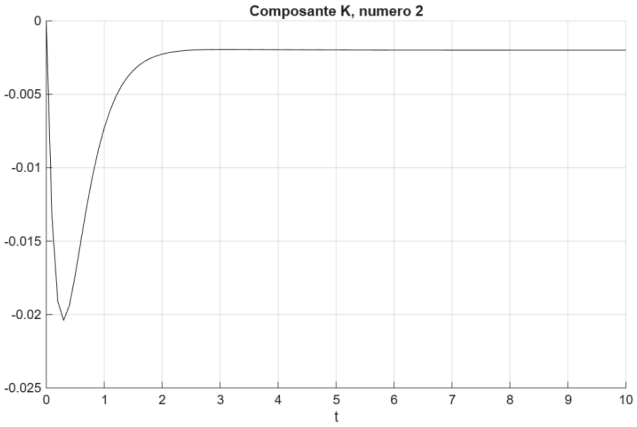

Figure A.8 – Original MATLAB second gain component [1, p. 9].

A.3 Archived linear filter on nonlinear data

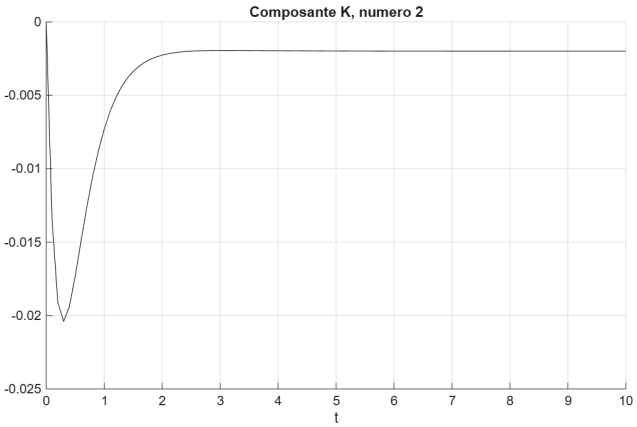

Figure A.9 – Original second gain component for the nonlinear-data comparison [1, p. 10].

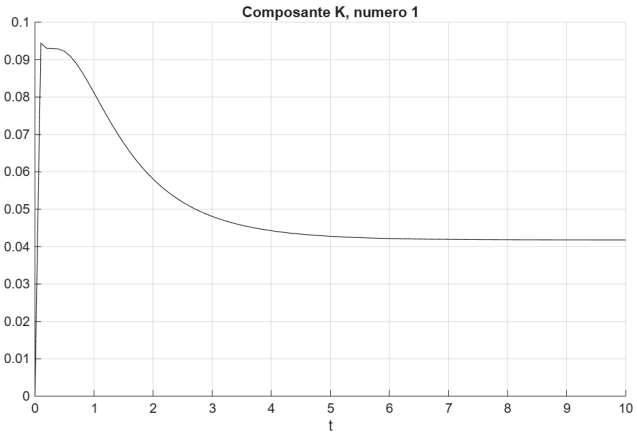

Figure A.10 – Original first gain component for the nonlinear-data comparison [1, p. 10].

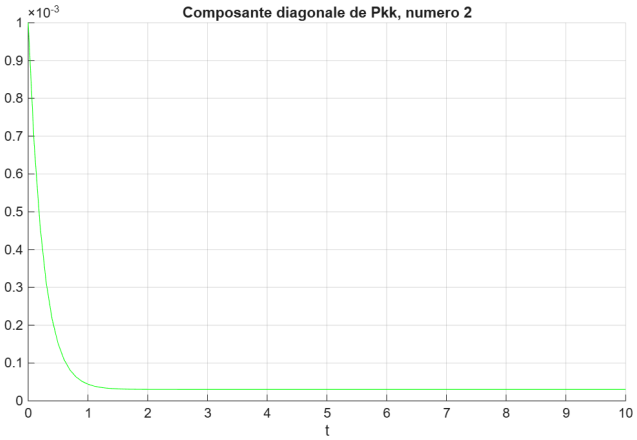

Figure A.11 – Original second covariance diagonal [1, p. 11].

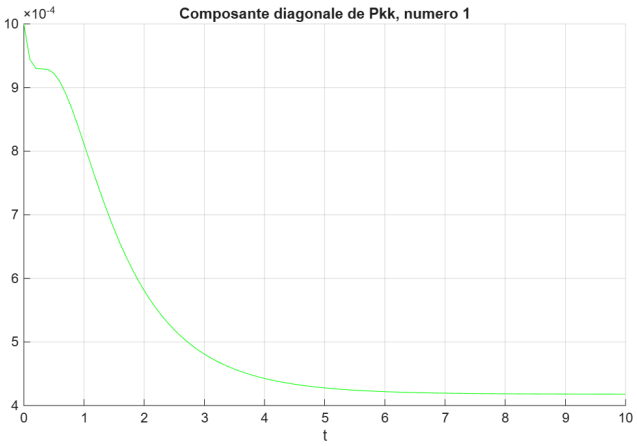

Figure A.12 – Original first covariance diagonal [1, p. 11].

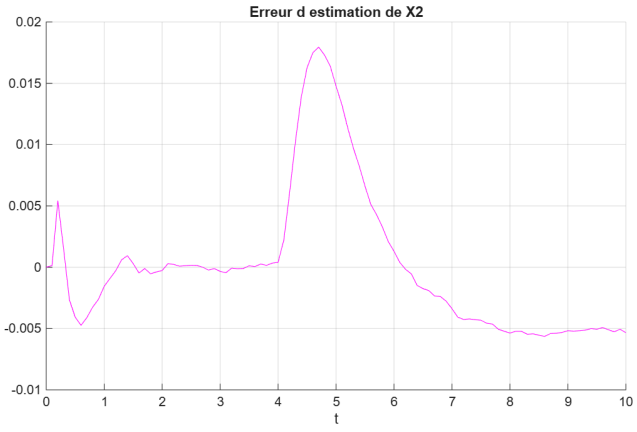

Figure A.13 – Original nonlinear-data second-state estimation error [1, p. 12].

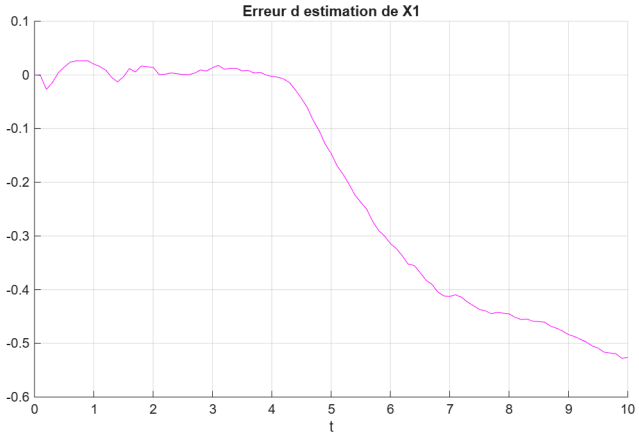

Figure A.14 – Original nonlinear-data first-state estimation error [1, p. 12].

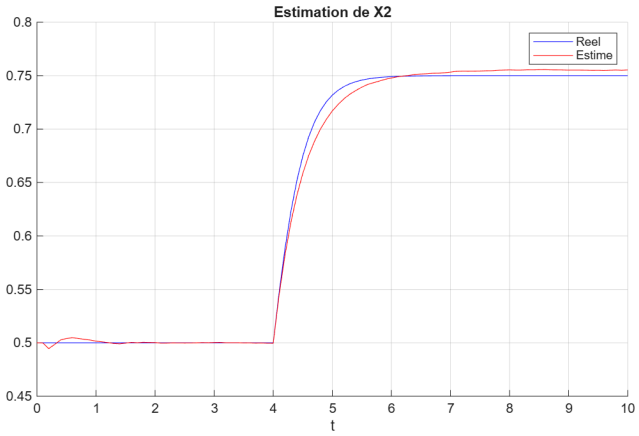

Figure A.15 – Original nonlinear-data second-state reference and estimate [1, p. 13].

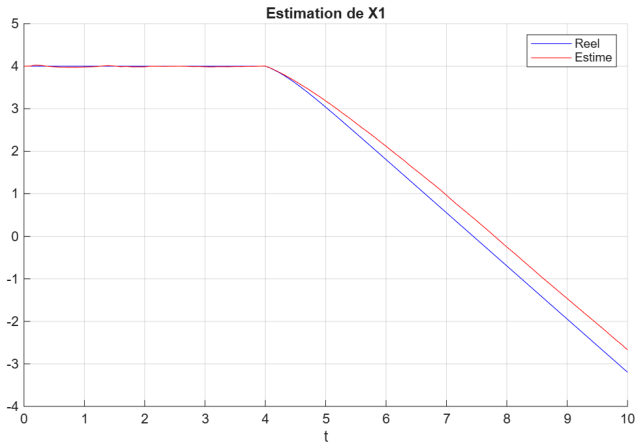

Figure A.16 – Original nonlinear-data first-state reference and estimate [1, p. 13].

Appendix B. Reproduction and checks

B.1 Recreating the results

The PDF includes the attachment **Kalman_Drone_Reproduction.zip**. It contains code, numerical outputs, vector figures and provenance for E1–E3. A second attachment, **Gazebo_Rejeu_ESKF.zip**, provides the E4 replay. No MATLAB, ROS or Gazebo runtime is needed for the numerical experiments. Unzip the attachment from a PDF reader and run the following from the extracted directory.

```
python -m pip install -r requirements.txt
python run_study.py --out reproduced
```

The declared environment is Python 3.12.14, NumPy 2.3.5, SciPy 1.17.0 and Matplotlib 3.10.8. E1 uses generator seed 3161960. E2 uses 21192026, E3 uses 40892552 and bootstrap intervals use 66541. Archived chart images are historical inputs. The scripts recreate the new figures independently from the model and noise assumptions.

Table B.1 – Executable files and principal output arrays in the attached archive.

Name	Reconstruction
run_study.py	Runs E1, E2 and E3. Emits figures and JSON.
lab_kalman.py	Exact model flow, fixed KF, EKF, UKF and extended RTS.
drone_eskf.py	Thrust-consistent trajectory, 15-state ESKF, Monte Carlo and event windows.
camera_geometry.py	Projective camera model and geometric uncertainty propagation.
lab_per_trial.npz	Trials × states per method. Compare per-run RMSE.
drone_per_trial.npz	Four metrics × 64 runs per condition.
drone_first_run.npz	Time, truth ENU position and downsampled state errors.

B.2 Verification tests and data audit

For E1, validation of the analytic transition against an independent integration yields maximum absolute error 3.52×10^{-14} in the selected case. A centred finite difference checks the Jacobian to 4.55×10^{-12}. The analytic held-input matrix matches the augmented block exponential to 6.94×10^{-18}. The constant-velocity white process covariance matches the Van Loan evaluation to numerical precision. These numbers are also in **lab_results.json**. The state is checked to stay on its chosen domain for replay.

E2 checks the rotation-error convention and both the yaw and reset Jacobians by finite differences separately. It compares the frozen-dynamics transition and process covariance to a block exponential and compares the approximate propagation Jacobian to the numerical derivative of the actual nonlinear step. The recorded minimum eigenvalues sample full matrices every two seconds. Orthogonality is measured at those same points. E3 checks the projection Jacobian against nonlinear finite differences. The Monte Carlo numbers, timing conditions, and file hashes in the archive allow those tests to be audited.

B.3 Event sequence for the inertial filter

Both the no-heading and gated variants use the following update order. The benchmark logs innovation statistics before any decision to reject.

```
for each IMU interval in acquisition order:
    propagate position, velocity and rotation with bias-corrected IMU
    propagate the local error covariance from the midpoint dynamics
    for each scheduled measurement at the new timestamp:
        form the predicted measurement, innovation and local Jacobian
        wrap the yaw residual when the observation is heading
        evaluate and record the pre-gate normalised innovation
        if accepted, apply the Joseph update and inject the correction
        transform covariance to the tangent frame of the new rotation
    evaluate errors against truth when building the simulated benchmark
```

The corresponding calculations live in **nominal_step**, **covariance_step**, **observation**, **update** and **filter_batch**. An active sensor update may have more than one component. Using its measurement dimension in the NIS reference is independent of the 15-dimensional error state. A stable Cholesky or square-root formulation is a natural extension for more severely conditioned or lower precision systems.

B.4 Reading a paired result

The run-level arrays preserve the pairs used for comparisons. The method order for each E1 array is KF, EKF, UKF, then extended smoother. The columns of each E2 condition align with the same truth and sensor realisation. For example, the first run-level metric in E2 is horizontal RMSE from 5 s to the end in metres. The following code recomputes the mean paired change when gating is applied.

```
from numpy import load
runs = load("results/drone_per_trial.npz")
ungated = runs["Stress ungated"][0]
gated = runs["Stress gated"][0]
delta = gated - ungated
print(delta.mean())
```

The expected mean change is approximately −0.354 m. The comparison in Section 8 uses complete run pairs to estimate uncertainty. The downsampled trace in **drone_first_run.npz** gives an independent way to plot the simulated truth and error, with state ordering in the attached readme.

B.5 Recorded Gazebo replay and exports

Extract **Gazebo_Rejeu_ESKF.zip** and use the dependency versions in its requirements file. Its compact **recorded_inputs** directory contains the retained columns and frames needed to repeat E4. The full user-provided **Validation_Kalman_Drone.zip** is also a compatible source for the replay. It is required for the independent MCAP audit.

```
python gazebo_replay.py recorded_inputs --out replay
python gazebo_raw_baseline.py recorded_inputs --out replay/raw_gnss.json
python gazebo_camera.py recorded_inputs --out replay
python audit_gazebo.py /path/to/Validation_Kalman_Drone --out audit.json
```

Each replay trace stores causal position, ENU velocity, ROS-order quaternion and the two estimated bias vectors. Covariance snapshots export the upper triangular 120 entries of the 15 by 15 causal covariance at every 25th IMU sample. The diagonals and signed error traces are supplied separately. Per-packet CSV files include acquisition time, scheduled receipt, assimilation time, innovations, innovation-covariance diagonals and gate decisions after the last processing of each input. The README gives array order, units and retained columns. The available reference poses and velocities are used for scoring, with camera reprojection evaluated separately from the ESKF update.