Machine Learning versus Analytical Ground Effect Estimation over Irregular Terrain

Muwon Lee[1*]

[1*]Chungbuk Science High School, Gyoyugwon-ro, Gadeok-myeon, Cheongju-si, 28189, Chungcheongbuk-do, Republic of Korea.

Corresponding author(s). E-mail(s): lmw.hpc@gmail.com;

Abstract

Ground-effect (GE) compensation can improve multirotor altitude tracking, but forecast accuracy may rank estimators differently from closed-loop performance. We tested analytical and machine-learning GE estimates in a simulator with procedurally generated terrain, corrected vehicle dynamics, and disjoint terrain splits. On 106,920 common test targets at 100 ms lead, a long short-term memory network achieved RMSE 0.007417 ($\boldsymbol{R^2 = 0.920}$), versus 0.024537 ($\boldsymbol{R^2 = 0.124}$) for the analytical baseline. Across 30 unseen terrain seeds, above-ground-level tracking RMS averaged 0.060436 m with analytical feedforward, 0.061283 m with the deployed hybrid multilayer perceptron, and 0.060578 m with the current-ground-truth oracle. The oracle improved on analytical feedforward by 1.466 mm under an alternative aerodynamic model. Positive prediction bias reduced clearance margins and increased low-clearance exposure. Active-controller pulses showed transient coupling, with no demonstrated sustained error growth. A validation-selected asymmetric clamp narrowed the hybrid tracking gap. The gap remained small and positive across nine slope and curvature settings (1,620 runs). Linearized poles were stable without added delay, while a nominal 100 ms delay yielded a negative phase margin. Bench tests separately recorded thrust at fixed PWM and the response of a range sensor to a moving target. Their scope was limited by imperfect RPM matching and the choice of physical timing reference. The closed-loop results show that lower offline forecast error alone was insufficient to improve tracking with the tested collective-thrust controller.

Keywords: ground effect, multirotor UAV, machine learning, closed-loop control, distribution shift, feedforward control

1 Introduction

When a multirotor unmanned aerial vehicle (UAV) operates close to the ground, interactions between rotor-induced downwash and the underlying surface can augment thrust relative to free-air conditions. This phenomenon, known as ground effect (GE), was characterized analytically for a rotor operating above a flat, infinite ground plane by Cheeseman and Bennett [1], who derived a closed-form thrust augmentation factor as a function of rotor radius and ground clearance. Because the magnitude of GE depends on instantaneous clearance and may also be influenced by local terrain geometry, uncompensated aerodynamic disturbances can contribute to altitude and attitude oscillations during low-altitude flight and landing. We refer to such disturbances and their associated vehicle responses as GE-induced motion irregularity. However, the precise influence of terrain slope and curvature on multirotor ground effect remains dependent on the aerodynamic configuration and cannot be fully established using simplified analytical assumptions alone.

The Cheeseman–Bennett model and its direct derivatives rely on idealized flat-ground and isolated-rotor assumptions. Sanchez-Cuevas et al. [2] experimentally demonstrated that ground-effect behavior in multirotor configurations differs from predictions based on simplified single-rotor formulations, partly because of aerodynamic interactions among rotors. These limitations motivate per-rotor representations of GE and raise the question of whether data-driven estimation using onboard-observable measurements can improve upon conventional analytical approximations over irregular terrain. Nevertheless, improved estimation accuracy does not necessarily imply improved closed-loop control, particularly when the estimated aerodynamic correction is applied through a collective-thrust command rather than independently to each rotor.

A potential approach, motivated by learning-based aerodynamic compensation methods such as Neural-Lander [3] and Neural-Swarm2 [4], is to estimate GE using onboard-observable signals, including rangefinder altitude, gyroscope angular rates, motor RPM and current, and commanded thrust. These measurements are representative of signals that may be available to an onboard flight controller without requiring explicit terrain reconstruction. Machine learning (ML) models can approximate nonlinear relationships between these measurements and the simulated GE factor, potentially providing predictions at a specified forecast horizon. However, their practical value depends not only on prediction accuracy but also on prediction timing, actuator response, controller dynamics, and the consequences of estimation errors when the predictions actively influence subsequent vehicle motion. Accordingly, offline predictive performance alone is insufficient to establish the effectiveness of an ML-based feedforward compensation strategy.

This paper makes two principal contributions. First, using a physics-based simulator with procedurally generated irregular terrain and heuristic aerodynamic corrections, we systematically compare ML-based GE estimators with analytical baselines under corrected and consistent flight dynamics. The evaluation employs disjoint terrain-seed partitions, common held-out samples, multiple prediction horizons, and terrain-dependent error analyses to distinguish predictive performance from potential evaluation artifacts. Second, we investigate whether improved offline GE prediction

translates into better closed-loop altitude tracking through matched 30-seed evaluations of analytical, learned, hybrid, and oracle feedforward controllers. Controlled prediction-bias interventions, active-controller pulse experiments, aerodynamic-model sensitivity tests, and linearized delay analysis are used to examine the relationship between estimation errors, clearance margins, and control performance. The results show that substantially improved offline prediction does not necessarily produce superior closed-loop tracking under the investigated collective-thrust feedforward architecture. Although prediction errors influence vehicle trajectories and clearance-related metrics, the experiments do not establish a sustained self-reinforcing prediction-error mechanism. Furthermore, the relative performance of analytical and oracle compensation depends on the assumed aerodynamic model, limiting conclusions about the attainable performance of analytical feedforward. We present these findings as a simulation-based assessment of the relationship between predictive accuracy and control effectiveness rather than as validation of a directly deployable compensation system.

An informal indoor hover observation with a DJI Air 3 motivated this study. The supplementary file `Pilot_observation.pdf` presents the original setup photographs, hover trajectories in pixel coordinates, and limitations of the observation.

2 Related Work

Cheeseman and Bennett [1] provided a foundational analytical formulation of rotor-in-ground-effect thrust augmentation, which serves as the basis for the flat-ground analytical reference considered in this study. However, the underlying assumptions of an isolated rotor operating above a flat ground plane limit its direct applicability to multirotor vehicles and irregular terrain. Sanchez-Cuevas et al. [2] investigated ground-effect behavior in multirotor configurations and demonstrated the importance of aerodynamic interactions that are not fully represented by simplified single-rotor formulations. Their findings motivate the per-rotor representation of GE adopted in our simulator, although the slope- and curvature-dependent corrections used here remain heuristic and have not been independently validated against aerodynamic measurements. Hooi et al. [5] investigated ground-effect-aware rotorcraft control using real-time aerodynamic estimation, illustrating an alternative to relying exclusively on fixed analytical corrections. Matus-Vargas et al. [6] examined altitude-estimation challenges associated with near-ground flight and investigated visual-sensor fusion as a means of improving state estimation. These studies highlight the relevance of aerodynamic modeling and onboard sensing to near-ground control. In contrast, the present work focuses specifically on predicting a simulated GE factor from onboard-observable measurements and evaluating whether the resulting prediction improves collective-thrust compensation. The inclusion of a rangefinder-only learned baseline further allows the contribution of additional sensor channels to be examined separately from nonlinear approximation.

Learning-based aerodynamic compensation has been investigated as a means of addressing disturbances that are difficult to represent using simplified physical models. Shi et al. [3] demonstrated learning-based compensation of aerodynamic effects during

quadrotor landing in Neural-Lander, while Shi et al. [4] extended learning-based aerodynamic modeling to interactions among multiple aerial vehicles in Neural-Swarm2. These studies establish the broader motivation for combining physical modeling with learned aerodynamic estimates, but do not directly validate the trust-band gating or low-clearance switching rules implemented in the present controller. O'Connell et al. [7] investigated adaptive learning-based approaches for changing aerodynamic conditions, providing a relevant contrast to the offline-trained, fixed estimators evaluated here. Although adaptation may offer a potential means of addressing changes in the operating distribution, the present experiments do not establish that such changes are the primary cause of the observed closed-loop performance gap. Instead, controlled bias and active-controller pulse interventions are used to examine how prediction errors influence tracking and clearance, without assuming a sustained self-reinforcing feedback mechanism.

The relationship between aerodynamic estimation and actuator-level compensation also depends on the accuracy of the propulsion model and the timing of control actions. Santos et al. [8] investigated thrust estimation using feedback-based methods, while Gräfe et al. [9] addressed the identification of propulsion-system characteristics. These studies motivate explicit consideration of the mapping between rotor speed, thrust, and the applied compensation command. In the present work, the corrected actuator mapping accounts for the quadratic dependence of thrust on rotor speed, and oracle experiments distinguish current ground-truth GE compensation from a kinematically projected 100 ms preview. The resulting comparisons are intended to assess the investigated simulation and controller architecture rather than to establish a general performance limit for analytical compensation.

Yang et al. [10] investigated oscillatory behavior during near-ground multirotor operation, providing a relevant context for the motion irregularity motivating this study. However, observations from physical flight experiments cannot independently validate the heuristic terrain-dependent aerodynamic corrections used in our simulator. Zou et al. [11] investigated the constructive exploitation of near-surface aerodynamic interactions through ceiling-effect-assisted perching, illustrating that proximity-induced aerodynamic forces can be relevant not only as disturbances but also as potential control resources. Together, these studies demonstrate the broader importance of modeling and estimating aerodynamic interactions near surfaces. The distinctive focus of the present work is the controlled comparison between offline predictive accuracy and closed-loop control effectiveness under matched simulation conditions, including seed-wise evaluation, predictive and sensor-restricted baselines, ground-truth oracle interventions, aerodynamic-model sensitivity, and delay-dependent stability analysis.

3 Methods

Figure 1 shows where GE compensation enters the collective command and how the controller receives the next sensor sample.

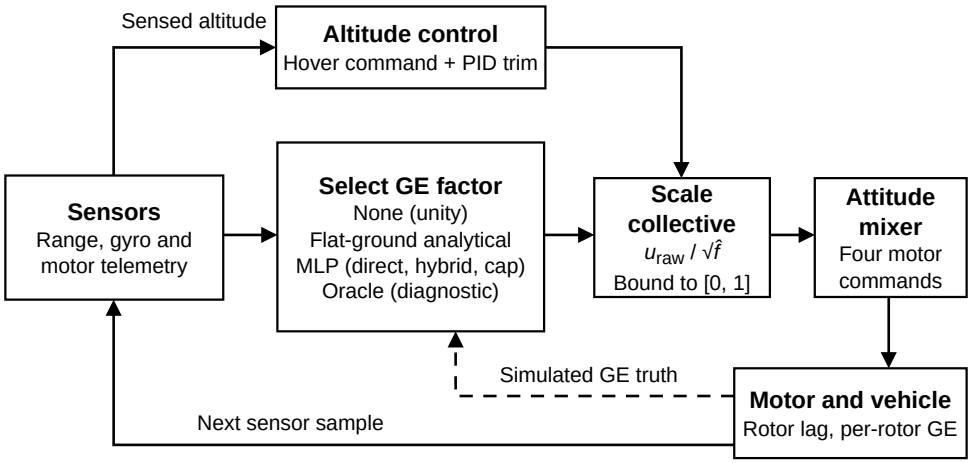


Fig. 1 Architecture of the simulated closed-loop controller. The GE factor $\hat{f}$ scales the collective command u_{raw} before the mixer adds attitude corrections. The dashed arrow represents ground truth supplied to the simulator-only oracle. The solid return arrow represents sensor feedback from the simulated vehicle.

3.1 Simulation environment

We implemented a physics-based multirotor simulator (`Simulator-Development/simulation/`) consisting of a rigid-body dynamics engine, a per-rotor motor model, a procedural terrain generator, and a ground-truth GE model, coupled to a PID-based flight controller and a configurable feedforward correction stage. The reviewer-revision experiments used consistently corrected cruise-speed dynamics and a revised actuator mapping for both dataset generation and closed-loop evaluation.

The simulator computes per-rotor terrain corrections for the dynamics and diagnostic oracle. The analytical estimator uses rangefinder clearance and a flat-ground ratio. The learned estimator uses onboard-observable signals. The selected factor scales the collective command before the attitude mixer.

Vehicle and simulation parameters. The simulated vehicle has a mass of 1.50 kg, an arm length of 0.18 m, a rotor radius of 0.127 m, and principal moments of inertia of 0.0208, 0.0208, and 0.0385 $\text{kg}\,\text{m}^2$. The motor thrust and torque coefficients are 1.472×10^{-7} and 2.940×10^{-9}, respectively, with a maximum rotor speed of 9000 rpm and a motor time constant of 0.030 s. The simulation uses an air density of 1.225 $\text{kg}\,\text{m}^{-3}$ and gravitational acceleration of 9.81 $\text{m}\,\text{s}^{-2}$. Physics and controller updates run at 1000 Hz, with telemetry logged at approximately 60 Hz. The body drag coefficient is 0.30 and reference area is 0.03 m^2. Each experiment lasts 60 s, with performance metrics calculated after excluding the first 10 s. The nominal flight plan specifies a cruise speed of 3.5 $\text{m}\,\text{s}^{-1}$ and a target above-ground-level (AGL) clearance of 0.22 m.

Ground-truth GE model. The baseline term follows the Cheeseman–Bennett flat-ground ratio,

$$f_{\text{flat}} = \frac{1}{1 - (R/4z_c)^2}, \tag{1}$$

where R is rotor radius and $z_c = \max(z, 0.5R)$ regularizes rotor clearance z. For terrain height h, slope $\theta = \arctan \|\nabla h\|$, curvature surrogate $\kappa = h_{xx} + h_{yy}$, slope exponent p, and curvature gain g, the simulator computes

$$a = (f_{\text{flat}} - 1) \max(0, \cos\theta)^p, \tag{2}$$

$$c = \text{clip}(1 + g\kappa R^2, 0.5, 1.8), \tag{3}$$

$$f_{\text{true}} = \text{clip}(1 + ac, 0.85, 1.60). \tag{4}$$

Centered differences use steps of 0.02 m for slope and 0.05 m for curvature. Defaults are $p = 1$ and $g = 0.6$. The Hayden sensitivity uses $q = \max(z_c/(2R), 0.15)$ and $f_{\text{flat}} = \{\min[1, (0.9926 + 0.03794/q^2)^{-1}]\}^{-2/3}$ with the same terrain corrections [12].

The analytical baseline uses the flat-ground expression evaluated from the available rangefinder measurement without access to the simulator's ground-truth terrain slope or curvature. This distinction ensures that privileged terrain geometry is not supplied to the ordinary onboard estimators.

The terrain terms operate on the augmentation above unity. On a horizontal local patch, the slope multiplier is one for every tested exponent. Positive curvature increases the augmentation when the curvature multiplier is below its upper bound, whereas negative curvature reduces it. The clearance floor and the clipping of the final factor play separate roles. The floor regularizes the flat-ground denominator. The final clip bounds the thrust multiplier passed to the dynamics. Because both develop from simulator assumptions, low-clearance estimation errors measure agreement with the specified synthetic target.

Terrain generation. Terrain is generated independently for each episode using a seeded xorshift32 pseudorandom number generator. The generator places 800–1499 non-overlapping bump, depression, and undulation features along the flight track and superimposes high-frequency roughness. Each seed determines a reproducible terrain realization, allowing independent terrain seeds to be reserved for training, validation, testing, and closed-loop evaluation.

Sensor features. The learned models use three-axis gyroscope angular rates, rangefinder altitude, per-rotor motor RPM and current measurements, and normalized commanded thrust. The rangefinder has a maximum measurement range of 3.0 m. The prediction target is the simulated arithmetic mean GE factor across the four rotors. Ground-truth terrain geometry and individual rotor GE factors are retained for diagnostic analysis but are not provided directly to the learned estimators.

The sensor synthesis adds zero-mean Gaussian noise with standard deviations 0.003 rad/s per gyro axis, 0.05 m/s^2 per accelerometer axis, 0.010 m for the rangefinder, 0.05 m for barometric altitude, 20 rpm per rotor, 0.1 A per rotor, and 0.02 V for battery voltage. The accelerometer uses the gravity reference rotated into the body frame. Range is measured vertically, with no range quantization implemented. The altitude

PID uses integral-state and output limits of 0.20 and 0.40. Outer roll and pitch proportional gains are 4.0. Inner roll and pitch PID gains are $(0.060, 0.020, 0.002)$ with derivative filter constant 0.010 s and yaw rate PI gains are $(0.080, 0.010)$.

Actuator mapping. Motor thrust is proportional to the square of rotor speed, $T = K_T n^2$ with speed n in rpm, while the commanded rotor speed is proportional to normalized throttle. Consequently, compensation for an estimated GE thrust multiplier $\hat{f}$ requires

$$u_{\text{comp}} = \frac{u_{\text{nom}}}{\sqrt{\hat{f}}}, \tag{5}$$

rather than scaling the throttle command by $1/\hat{f}$. The latter mapping was used in an earlier implementation and was corrected before regenerating the revision dataset and repeating the principal experiments. The estimated GE factor is constrained to the interval $[0.85, 1.60]$ in the compensation pathway.

3.2 Offline model comparison and seed-wise data splitting

We evaluated eight learned regressors, including linear regression, decision tree, Random Forest, Gradient Boosting, a multilayer perceptron (MLP), a long short-term memory network (LSTM), and a Transformer. The eighth learned regressor was a rangefinder-only MLP. A predictive analytical baseline was evaluated alongside the eight learned models. The primary task was to estimate the GE factor 100 ms ahead from the available current sensor observations.

An earlier evaluation used temporal block-wise splitting, which allowed samples originating from the same terrain realization to appear in different partitions. To eliminate this potential source of leakage, the revision regenerated the complete 300-seed dataset using corrected flight dynamics and introduced a deterministic seed-wise partition comprising 240 training, 30 validation, and 30 test seeds. The partition was generated using random seed 20260825, and disjointness was verified programmatically.

The principal revised model comparison uses 106,920 common held-out target samples. All reported MAE, RMSE, and R^2 values in this comparison are computed on identical target observations, ensuring consistent sample support and metric aggregation. Model selection and training decisions use the validation partition rather than the held-out test partition.

The corrected dataset comprises 1,080,300 logged rows across 300 flights. For a history window ending at row i within a seed, the 100 ms forecasting task uses the truth at $i + 6$. Both the analytical value and the pointwise learned inputs come from row i. The common-support mask requires the history needed by the sequence predictors and the future target to be available within the same seed. The comparison therefore scores the analytical value as a persistence forecast over the same lead as the learned estimates. The fixed split, target definition, and common-support construction are shared across the reported model families.

The rangefinder-only MLP provides a learned baseline without the additional gyroscope and motor measurements, allowing nonlinear approximation and sensor availability to be considered separately. The predictive analytical baseline estimates

range rate from the current and sixth preceding telemetry samples using their timestamp difference. It projects the current range by this rate times the target-minus-source time, then evaluates Eq. 1 with the same clearance floor and GE bounds.

The sequence models use 32 consecutive rows from the same seed. Pointwise models use the last row, and the rangefinder-only MLP uses the 32 range measurements. Neural models use training-only standardization, Adam (10^{-3}), MSE, batch sizes 256/2048 for training/validation, a 25-epoch cap, and the best validation RMSE checkpoint with patience six and minimum improvement 10^{-7}. Both MLPs have ReLU layers of widths 64 and 32. The LSTM uses one 48-unit layer and a 32-unit ReLU head. The Transformer uses width 32, four attention heads, two encoder layers, feedforward width 128, learned positional embeddings, dropout 0.1, and a 32-unit ReLU head. The random forest uses 100 trees, maximum depth 14, minimum leaf size three, and bootstrap fraction 0.5. The decision tree uses depth 10 and minimum leaf size five. Histogram gradient boosting uses depth six and at most 300 iterations. These comparisons are intended to distinguish estimation performance from the downstream consequences of using the estimates in a flight controller.

3.3 Oracle controller and asymmetric clamp

To distinguish estimation limitations from limitations of the collective-thrust compensation architecture, we implemented an oracle feedforward mode using the simulator's ground-truth GE factor. The principal oracle supplies the current GE factor, rather than the 100 ms-ahead target used in the offline prediction task. This distinction is important because access to the current ground-truth factor does not provide exact future-state knowledge or automatically compensate for actuator and controller delays.

Additional oracle experiments compare current-factor compensation with a 100 ms preview obtained through constant-velocity kinematic terrain projection. The preview is an approximation based on projected motion, not the exact future GE factor. The arithmetic mean matches the offline target and corresponds to equal free-air rotor thrusts. The weighted variant uses current free-air weights $K_T n_i^2$ and directly represents the ratio of summed augmented to summed free-air thrust at the measured RPM. The preview translates position by current velocity times 100 ms and holds attitude fixed. The motor time constant is 30 ms.

An asymmetric clamp was implemented to constrain overprediction by the ML feedforward pathway relative to the analytical estimate. Candidate clamp parameters were 0, 0.005, 0.01, 0.02, and 0.04. The final parameter was selected exclusively using controller-validation seeds 301–330 and fixed at 0.0 before the final evaluation. Selection minimized mean validation AGL RMS, with low-clearance duration and smaller δ as successive tie-breakers. Above cutover, the capped mode replaces the trust-band decision with $\min(\hat{f}_{\mathrm{ML}}, f_{\mathrm{analytical}} + \delta)$ for finite predictions and uses unity for invalid predictions.

3.4 Forecast-horizon sweep

To characterize how predictive performance changes with lookahead, we evaluated forecast horizons of 0, 50, 100, and 200 ms. The revised horizon comparison includes LSTM, MLP, Random Forest, and analytical estimators. Each model is evaluated against the GE target corresponding to the specified horizon.

3.5 Terrain-irregularity error decomposition

To evaluate whether prediction performance depends on terrain geometry, we analyzed estimation errors across synthetic terrain classes and clearance intervals. The revised analysis considers flat, bump, dip, and ripple terrain configurations and compares the analytical estimator with the LSTM model.

Particular attention was given to the 0.15–0.25 m clearance interval, where ground-effect compensation is relevant to the nominal low-altitude flight task. Terrain classification and geometry information were used only for post-hoc evaluation, not as privileged inputs to the learned onboard estimators.

3.6 Closed-loop feedforward integration

An MLP with architecture $13 \to 64 \to 32 \to 1$ was exported to a static-array C header and integrated into the flight controller (`firmware/ge_ff.c`). Although the LSTM achieved the best offline prediction accuracy, the MLP was selected for closed-loop deployment because its feedforward computation is straightforward to implement in C without recurrent-state management.

The altitude controller adds a PID trim to the nominal free-air hover command. It uses the rangefinder for altitude feedback, with barometric altitude as the fallback for an unavailable range reading. Before the attitude mixer, the selected feedforward factor scales this collective command. The mixer adds the attitude corrections and bounds the individual motor commands. The motor model relaxes each rotor toward its commanded speed with a first-order lag. For a settled, unsaturated rotor with nominal speed proportional to collective command alone, the actual thrust after compensation is $T_{\mathrm{nom}} f_{\mathrm{true}} / \hat{f}$. An overestimated factor then reduces thrust. This steady relation describes the direction of the commanded correction. Lag, differential mixer terms, and command limits prevent it from being an exact identity throughout a maneuver.

Six controller configurations were evaluated through `GE_FF_MODE`: `none`, `analytical`, `oracle`, `ml`, `ml_direct`, and `ml_asym_clamp`.

The `none` configuration applies no GE feedforward compensation. The `analytical` configuration uses the flat-ground estimate. The `oracle` configuration uses the simulated ground-truth GE factor. The `ml_direct` configuration applies the learned prediction without hybrid switching. The `ml` configuration uses a hybrid strategy, switching to analytical compensation below 0.15 m rangefinder clearance and otherwise accepting ML predictions within a trust band of $\pm 20\%$ around the analytical estimate. The rejected-prediction fallback is unity. The `ml_asym_clamp` configuration additionally applies the selected asymmetric overprediction constraint.

All configurations use the corrected actuator mapping in Eq. 5. Final closed-loop comparisons were conducted on 30 unseen terrain seeds, numbered 331–360, which were disjoint from the offline dataset and the controller-validation seeds.

The primary tracking metric was the RMS AGL error,

$$e_{\mathrm{AGL,RMS}} = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(z_{\mathrm{AGL},k} - 0.22\right)^2}, \tag{6}$$

where $z_{\mathrm{AGL},k}$ is the simulator's true body AGL. Gyroscope RMS is the RMS of the three-axis rate magnitude. The clearance signal is the minimum across the four regularized rotor clearances z_c at each logged sample. Its minimum is therefore bounded by 0.0635 m. Each contiguous segment strictly below 0.15 m counts as one event, including an event present at the start of the analysis interval. Exposure is the number of low samples times the median logging interval. P05 is the fifth percentile of the clearance signal.

3.7 Causal bias intervention and loop-closure analysis

To investigate the directional effects of GE-factor prediction error, we performed controlled bias interventions using the oracle controller. A fixed additive bias was applied to the ground-truth estimate,

$$\hat{f} = f_{\mathrm{true}} + \delta, \qquad \delta \in \{-0.02, -0.01, 0, +0.01, +0.02\}. \tag{7}$$

The resulting closed-loop trajectories were evaluated on matched terrain seeds using AGL tracking RMS, minimum clearance, fifth-percentile clearance, and low-clearance exposure duration.

These experiments isolate the effect of an externally imposed prediction bias on the simulated controller response. However, a causal effect of imposed bias on clearance does not by itself establish that naturally occurring learned-model errors form a persistent positive-feedback mechanism.

To investigate this distinction, additional bias-pulse experiments were conducted with the actively deployed hybrid ML controller. Pulse magnitudes of ± 0.01 and ± 0.02 were applied for approximately one telemetry interval (0.0167 s), 0.1 s, and 0.25 s. Pulses start at 30 s. Post-pulse clearance and natural absolute prediction error are compared with matched zero-bias runs over 0–0.5, 0.5–1, 1–2, 2–5, and 5–10 s windows.

The revised intervention design extends the earlier oracle-pulse experiment, in which a non-actuating shadow predictor was monitored following a temporary oracle bias. Because the shadow predictor did not directly influence actuation, that earlier experiment alone could not establish the behavior of the actively deployed learned controller.

3.8 Robustness to ground-truth correction strength

Because the terrain-dependent GE corrections are heuristic, we examined whether the principal closed-loop comparison depended on their parameterization. A factorial grid

was constructed using slope exponents of 0.5, 1.0, and 2.0 and curvature gains of 0.3, 0.6, and 0.9.

Each of the nine configurations was evaluated using six controller modes and 30 terrain seeds, producing

$$N_{\text{runs}} = 3 \times 3 \times 6 \times 30 = 1620. \tag{8}$$

The default configuration corresponds to slope exponent 1.0 and curvature gain 0.6.

To probe the role of feedback tuning, we repeated the none, analytical, oracle, and hybrid conditions on the same 30 evaluation seeds using common multipliers of 1, 0.5, and 0.25 on the three altitude PID gains. Motor dynamics, limits, and the remaining controller settings were held fixed.

A separate aerodynamic-model sensitivity experiment compared Cheeseman-based and Hayden-based configurations. This analysis was intended to determine whether conclusions concerning the relative performance of analytical and oracle compensation were dependent on the assumed ground-effect formulation.

Neither sensitivity analysis constitutes independent aerodynamic validation, because both are conducted within the synthetic simulation framework.

3.9 Linearized stability and delay analysis

We constructed a local vertical-axis model around level hover, zero vertical velocity, unsaturated actuators, and a constant GE factor. The motor and vertical dynamics were represented as

$$\tau_m \dot{r} + r = u, \tag{9}$$

$$\ddot{z} = K_m r, \tag{10}$$

where $\tau_m = 0.030\,\text{s}$ and K_m is the local acceleration-to-command gain.

The altitude controller was modeled as

$$C(s) = K_p + \frac{K_i}{s} + \frac{K_d s}{1 + \tau_d s}, \tag{11}$$

with $K_p = 0.60$, $K_i = 0.15$, $K_d = 0.25$, and $\tau_d = 0.15\,\text{s}$.

The resulting delay-free fifth-order closed-loop system was represented in companion form,

$$\dot{\mathbf{x}} = A\mathbf{x}, \qquad A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ a_0 & a_1 & a_2 & a_3 & a_4 \end{bmatrix}. \tag{12}$$

At the nominal GE factor of 1.0, the final row of A was

$$\begin{bmatrix} -1177.39998301 & -4886.20992951 & -2668.77329483 & -222.222222222 & -40 \end{bmatrix}. \tag{13}$$

The corresponding numerical poles were

$$\lambda_1 = -35.767037, \tag{14}$$
$$\lambda_{2,3} = -1.09444227 \pm 8.04050374i, \tag{15}$$
$$\lambda_4 = -1.76003841, \tag{16}$$
$$\lambda_5 = -0.28404006. \tag{17}$$

With fixed GE multiplier f, hover occurs at $u_0 = \sqrt{mg/(4K_T n_{\max}^2 f)}$ and $K_m = 2g/u_0$. The characteristic polynomial before monic normalization is $s^3(1+\tau_m s)(1+\tau_d s) + K_m[(K_p\tau_d + K_d)s^2 + (K_p + K_i\tau_d)s + K_i]$. The last companion row contains the negated monic coefficients in ascending order. The frequency response uses $L(s) = \rho C(s) K_m e^{-sD}/[s^2(1+\tau_m s)]$, where ρ is a constant gain multiplier and D is added delay. The steady unsaturated ratio $f_{\text{true}}/\hat{f}$ motivates ρ as a local sensitivity parameter. Mixer, lag, and switching limit the interpretation of that ratio as a bias bound.

Table 1 Operating points and local delay-free poles.

GE	u_0	K_m	Poles (s^{-1})
0.85	0.60248	32.5653	$-35.6001, -1.1669 \pm 7.6743i, -1.7823, -0.2839$
1.0	0.55546	35.3220	$-35.7670, -1.0944 \pm 8.0405i, -1.7600, -0.2840$
1.2	0.50706	38.6933	$-35.9672, -1.0054 \pm 8.4618i, -1.7379, -0.2842$
1.6	0.43913	44.6792	$-36.3122, -0.8477 \pm 9.1502i, -1.7081, -0.2843$

The pole analysis describes only the local delay-free vertical-axis model. It excludes the complete nonlinear dynamics, learned-model feedback, switching, actuator saturation, and terrain-dependent perturbations. Consequently, negative real parts of the reported poles do not establish stability of the complete actively deployed controller.

3.10 Hardware bench measurements and sensing-delay sensitivity

The bench experiments characterize a single-rotor thrust stand and a separate range-sensing setup (Fig. 2). Figure 3 shows the image features used in the high-speed analysis. The thrust stand used an A2212/10T 1400 KV motor, a 4S pack, and a 1045 propeller (nominal radius 127 mm). At each of 25, 20, and 15 cm from the rotor plane to the plate, three runs used a 1400 μs command for 5 s. The 25 cm condition is a finite-clearance reference. The first 25 cm run was acquired on the preceding day. The load-cell measurement was the difference between the mean in the 2–5 s steady window and the pre-run, PWM-off mean. We converted counts to force with a gain of 34,206 counts/N from a 30 g reverse-direction calibration load. Absolute thrust values at the operating point therefore depend on extrapolation of this gain.

A high-speed camera (Photron FASTCAM SA-X2) recorded each video at 1000 frames/s, with an exposure of approximately 0.998 ms. The 30 frames/s exports retained one image per recorded frame in each export interval. The rotor speed estimate used the recurrence of a blue marker on one blade. We computed the dominant marker frequency in 500-frame Hann windows with a 250-frame hop. We selected nine windows from the late part of the longest continuous span with clear periodic motion and skipped the last two windows of that span to reduce coast-down contamination. Stopped, intermittent motion and low-confidence intervals did not enter the estimate. The camera and the Uno were not hardware-synchronized. Steady force and speed windows were selected on their respective clocks, and repeated runs were the units of analysis. A retrospective analysis searched the retained 500-frame windows for RPM matches within 1%. For eligible same-block comparisons, we approximately aligned the end of the optically periodic span to the end of the PWM command and computed the force change as a ratio of load-cell counts above each run's pre-run mean. We varied the camera-to-Uno alignment independently by ± 250 ms as a sensitivity check.

The sensor bench used the output of a Sharp GP2Y0A41SK0F analog infrared range sensor and an Arduino Uno that toggled a visible LED at direction-dependent thresholds. A hand-translated MDF plate changed the sensor target. The camera recorded the right edge of the plate and the LED in the same view. We detected LED transitions from red-minus-green intensity and retained transitions that persisted for at least 20 frames. We measured projected MDF motion from its right edge, using pre- and post-event plateau positions and persistent crossings at 25%, 50%, and 75% of the apparent motion. The time difference between each crossing and the LED edge is a signed bench measurement relative to that image reference.

In a separate simulator sensitivity experiment, we tested added fixed rangefinder delays of 0, 20, and 70 ms with the original 1 ms sensor sampling interval. The two nonzero delays bracket the scale of the midpoint-referenced bench times and serve as hypothetical perturbations. The delayed signal was used by the altitude PID, the analytical and learned observation paths, and hybrid switching logic. Other sensor timing, the motor model, gains, and trained weights remained fixed. The oracle retained direct access to simulator truth. We used all six control modes with the same 30 evaluation seeds at each delay, for 540 runs. The local, exact-delay analysis in Section 3.9 uses a separate linearized model.

3.11 Statistical analysis and reproducibility

Closed-loop controller comparisons used paired terrain seeds to reduce variability associated with terrain realization. Paired t tests use two-sided alternatives. Tables show ordinary 95% paired t intervals. Supplementary tables also give 20,000-resample paired percentile bootstrap intervals (seed 20260825) and paired Cohen's d_z. Holm adjustment covers all pairwise comparisons within each metric and design cell. Pulse-window comparisons are exploratory.

The corrected offline dataset used seeds 1–300, partitioned into 240 training, 30 validation, and 30 test seeds. Controller-validation seeds 301–330 were reserved for clamp

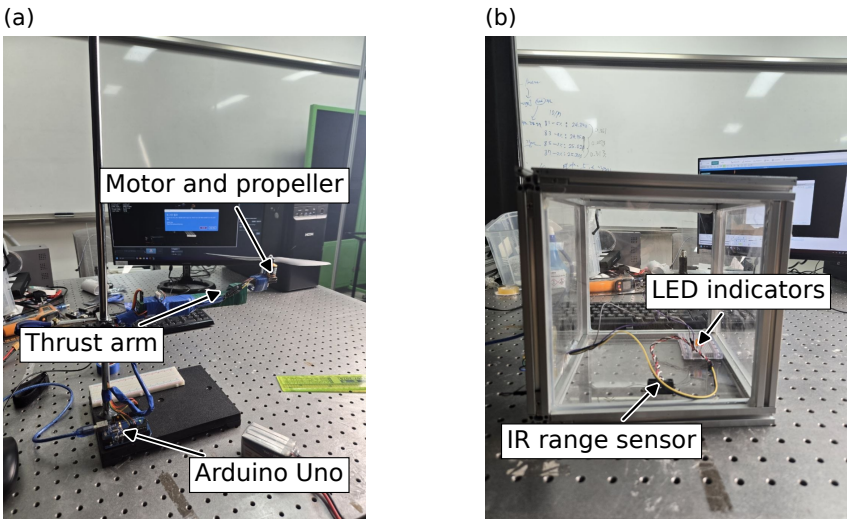


Fig. 2 Experimental bench setups. Panel (a) identifies the motor and propeller, thrust arm, and Arduino Uno on the single-rotor stand. Panel (b) identifies the infrared range sensor and LED indicators used in the moving-target test.

selection, while final controller evaluation used seeds 331–360. The public implementation snapshot is `7e2ea2eb8b8f801d37301d86c83510003b5b36eb`. Run metadata identify the original working-tree run where no commit ID was recorded. Appendix E lists the exact seeds. The supplement gives run manifests and full-precision parameters.

The controller comparison is restricted to the investigated synthetic simulation configurations. The separate bench data describe single-rotor thrust and a moving-target sensing test. The slope and curvature corrections and the complete controller remain to be validated in multirotor flight.

4 Results

4.1 Offline model comparison

On the common held-out targets, the LSTM ranked first, followed by the Transformer (Table 2, Fig. 4). The Transformer had lower RMSE and higher R^2 than the random forest. Both metrics are monotone functions of the same pooled squared error on these identical targets. Appendix A provides the horizon metrics and terrain sample support. The rangefinder-only MLP outperformed the pointwise full-sensor MLP, although the range-only model also receives a longer input history. The predictive analytical baseline performed worse than holding the source-row analytical estimate over the 100 ms lead.

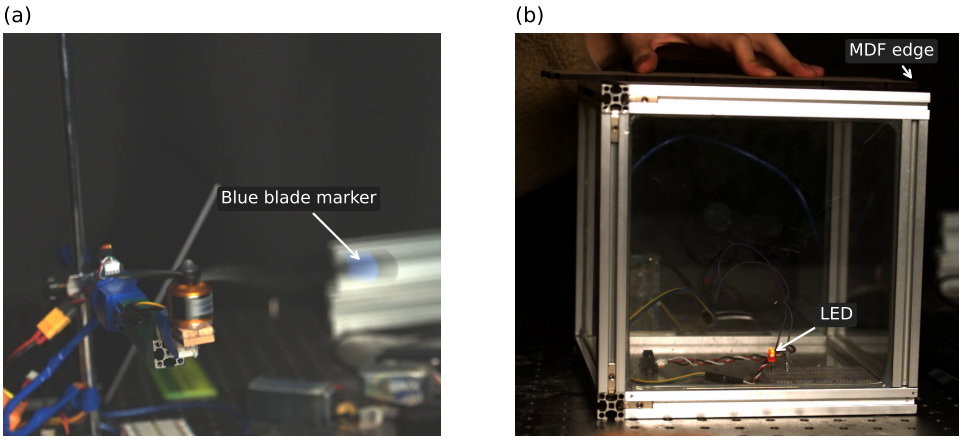


Fig. 3 Representative frames of the two bench measurements, captured at 1000 frames/s. Panel (a) shows the blue blade marker during steady rotation at 20 cm (`20cm-2`, frame 5245). Panel (b) shows the moving MDF edge and the illuminated LED one frame after the first detected rising edge (`tof-delay-video`, frame 1553). Frame indices are zero-based within each export. Arrows identify the image features used in the analyses.

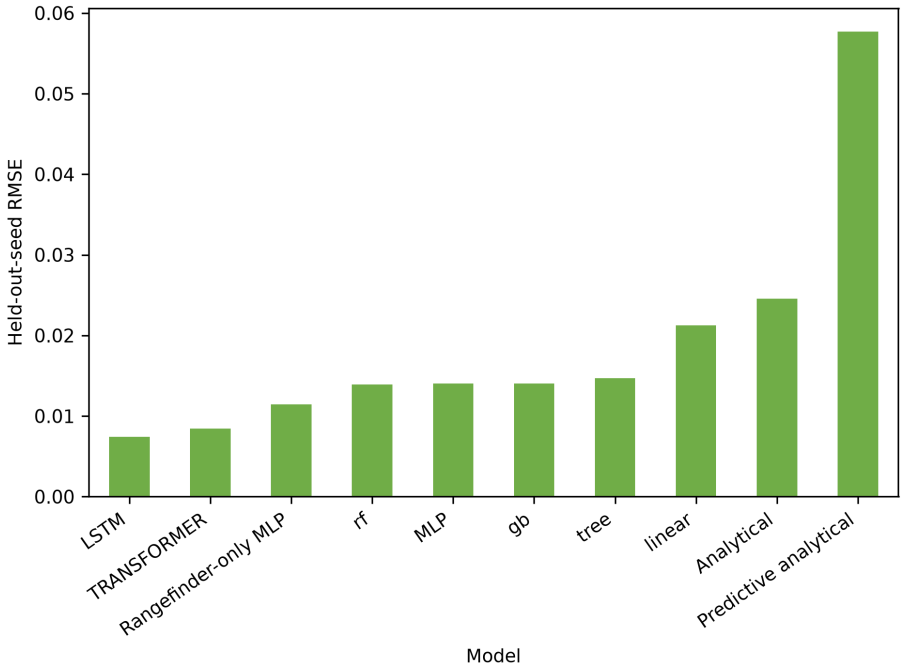


Fig. 4 Pooled held-out prediction RMSE at 100 ms.

The best sequence predictor lowered RMSE from 0.024537 to 0.007417, a reduction of about 70%. The deployed pointwise MLP achieved 0.014010, a reduction of about

43%. The difference between these two improvements matters for interpreting the closed-loop experiment, which actuates the pointwise predictor. Random forest, MLP, and histogram gradient boosting have similar pooled RMSEs but different MAEs. Their metric values reflect different distributions of errors across the same targets. The predictive analytical baseline assumes a constant rate estimated from two noisy range samples. Its high RMSE shows that this particular extrapolation is ineffective on this task. The comparison does not establish how a filtered or model-based range forecast would perform.

Table 2 Pooled errors on 106,920 common targets from 30 held-out seeds at 100 ms lead.

Model	RMSE	MAE	R^2
LSTM	0.007417	0.003333	0.919927
Transformer	0.008458	0.003741	0.895880
Rangefinder-only MLP	0.011421	0.005133	0.810154
Random forest	0.013927	0.006079	0.717709
MLP	0.014010	0.006420	0.714331
Histogram gradient boosting	0.014057	0.006219	0.712400
Decision tree	0.014672	0.006490	0.686682
Linear regression	0.021232	0.012112	0.343891
Analytical	0.024537	0.010253	0.123749
Predictive analytical	0.057712	0.021673	-3.847702

4.2 Forecast horizon

LSTM RMSE increased from 0.001859 at 0 ms to 0.005251, 0.007613, and 0.010013 at 50, 100, and 200 ms, respectively (Fig. 5). The LSTM ranked first at all four horizons. The sweep trained separate models, including at 100 ms. Within each horizon the models shared evaluation targets.

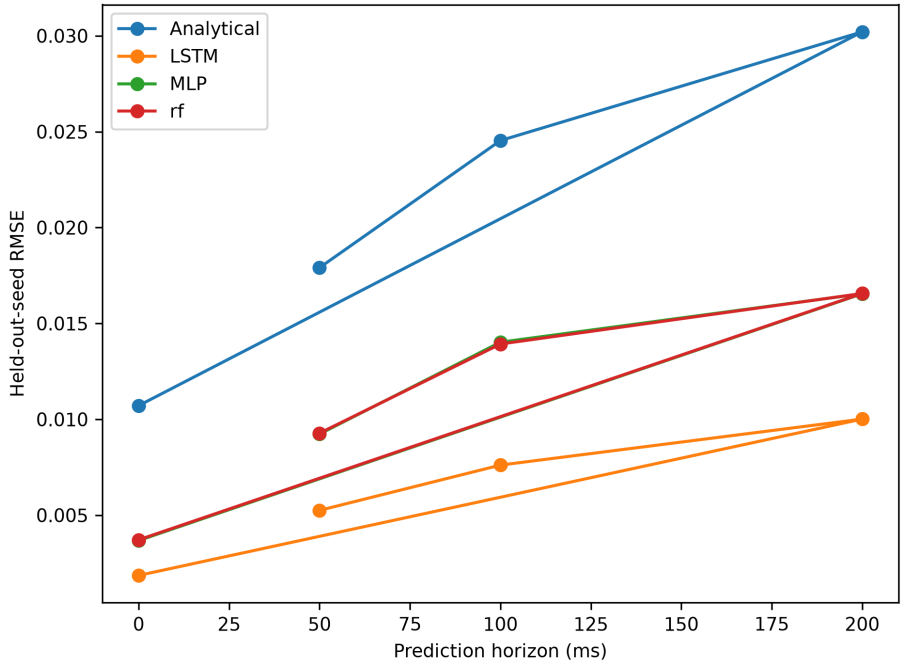


Fig. 5 RMSE across four lead times on the held-out seeds.

At zero lead, the analytical estimator achieved RMSE 0.010714. At 200 ms its RMSE rose to 0.030208, while the pointwise MLP and random forest both reached approximately 0.0166. The decrease in predictability with lead is visible across model families. The later-horizon scores use slightly fewer examples because each flight ends at a fixed time. The sweep measures the ability to forecast a future factor from recorded signals. A longer forecast only helps compensation if the controller applies it at an appropriate time, which motivates the separate oracle timing experiment.

4.3 Terrain-irregularity error decomposition

In the 0.15–0.25 m clearance band, analytical and LSTM seed-mean RMSE were 0.00622 and 0.00200 on flat segments, 0.00953 and 0.00543 on bumps, 0.01634 and 0.00491 on dips, and 0.03928 and 0.01622 on ripples (Fig. 6). These errors use the corrected offline test predictions. Each seed contributes its own subgroup metric before averaging, whereas Table 2 pools all target rows.

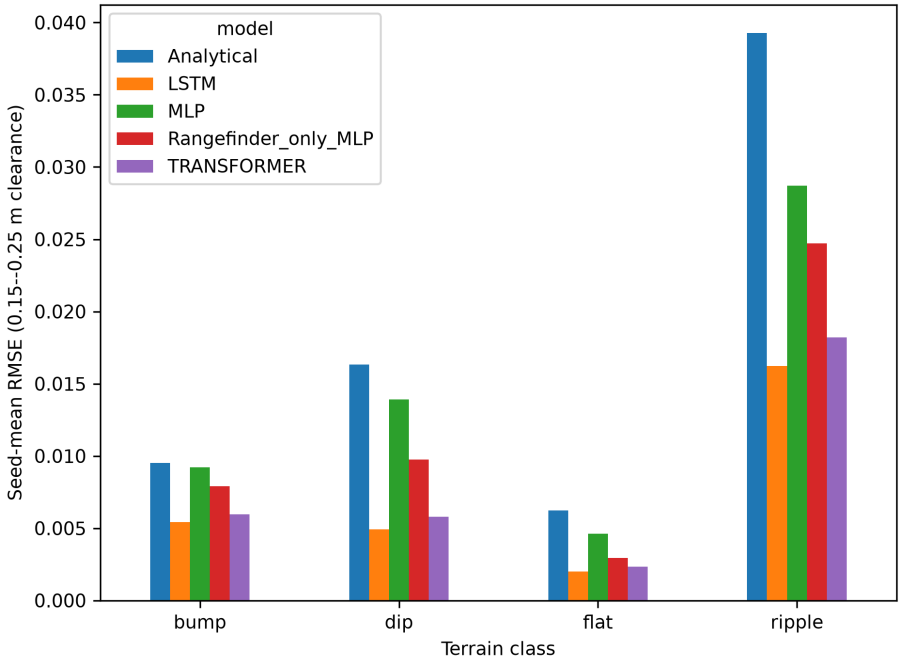


Fig. 6 Seed-mean RMSE by synthetic terrain class within the 0.15–0.25 m clearance band.

4.4 Closed-loop deployment: offline accuracy does not transfer

Analytical and hybrid MLP both improved tracking relative to no feedforward (Tables 3–4, Fig. 7). Hybrid-minus-analytical RMS was 0.847 mm (95% paired t CI 0.681–1.012 mm, Holm-adjusted $p = 1.43 \times 10^{-10}$). Oracle-minus-analytical RMS was 0.142 mm with adjusted $p = 0.0986$. We did not detect an oracle advantage or test statistical equivalence. Table 5 gives selected paired contrasts. Appendix B gives all contrasts against analytical. The supplement includes all 90 pair-by-metric contrasts.

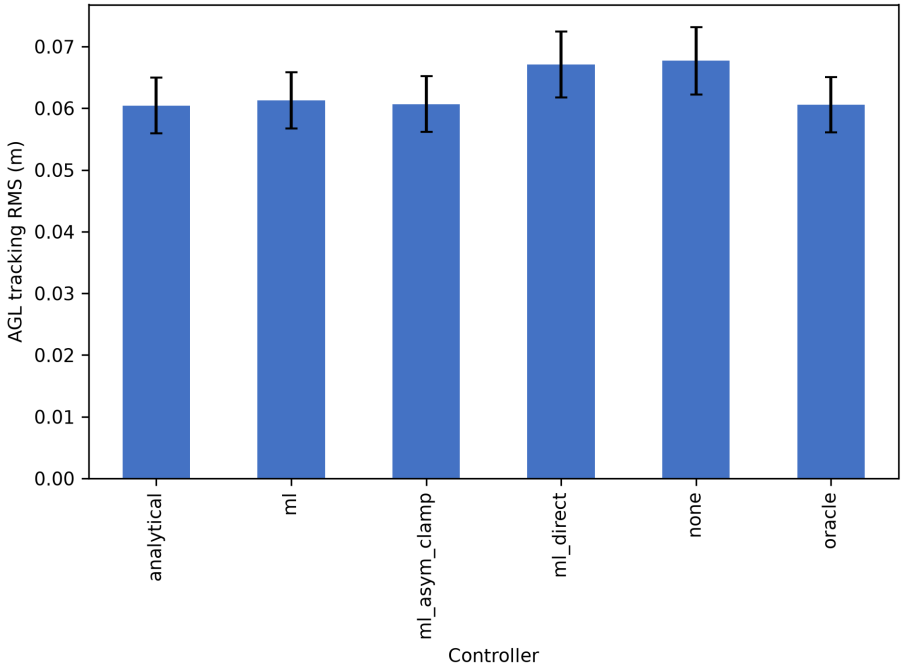


Fig. 7 Tracking RMS on 30 final seeds with marginal 95% t intervals.

Relative to no feedforward, the hybrid reduced RMS by 6.405 mm (95% paired t CI 5.194–7.616 mm), while analytical reduced it by 7.252 mm (6.038–8.465 mm). Direct MLP achieved 67.082 mm, compared with 61.283 mm for the hybrid. This distinction is important because changing from direct prediction to the hybrid changes which estimate is used near the ground and what happens when a prediction is rejected. Mean tracking performance alone cannot attribute the difference to either of these design choices.

The clearance outcomes rank the controllers differently. No feedforward had the highest mean fifth-percentile clearance (137.27 mm) and the shortest mean low-clearance exposure (3.95 s). Relative to this mode, hybrid tracking improved while fifth-percentile clearance decreased by 5.311 mm (95% paired t CI 4.226–6.396 mm lower) and low-clearance exposure increased by 0.550 s (0.407–0.694 s). A reduction in RMS error about the target AGL therefore need not increase the margin above a lower threshold. The minimum-clearance statistic also clusters near the simulator's 63.5 mm floor. Its limited discrimination makes P05 and exposure useful complementary measures within this simulation.

Table 3 Tracking and gyroscope RMS across 30 matched terrain seeds, mean and standard deviation.

Controller	AGL RMS (mm)	Gyro RMS (rad/s)
None	67.69 ± 14.59	0.0565 ± 0.0138
Analytical	60.44 ± 12.12	0.0565 ± 0.0151
Oracle	60.58 ± 11.96	0.0563 ± 0.0150
Direct MLP	67.08 ± 14.27	0.0573 ± 0.0146
Hybrid MLP	61.28 ± 12.13	0.0574 ± 0.0148
Asymmetric clamp	60.68 ± 12.17	0.0566 ± 0.0151

Table 4 Mean and standard deviation across 30 matched terrain seeds. Minimum and P05 clearance summarize the respective per-seed statistics.

Controller	Min (mm)	P05 (mm)	Low time (s)	Events
None	63.50 ± 0.01	137.27 ± 9.75	3.95 ± 1.03	26.87 ± 6.03
Analytical	63.63 ± 0.70	133.85 ± 11.05	4.32 ± 1.21	28.00 ± 6.69
Oracle	63.65 ± 0.82	133.87 ± 11.04	4.31 ± 1.21	27.87 ± 6.65
Direct MLP	63.56 ± 0.34	132.62 ± 10.48	4.49 ± 1.21	29.37 ± 6.19
Hybrid MLP	63.57 ± 0.39	131.96 ± 11.16	4.50 ± 1.24	28.23 ± 6.39
Asymmetric clamp	63.62 ± 0.65	133.71 ± 11.07	4.33 ± 1.21	27.93 ± 6.61

Figure 8 shows the seed-level differences behind the pooled controller summaries. Each point compares the candidate and analytical controller on the same terrain. A positive value indicates poorer tracking for the upper panel and greater fifth-percentile clearance for the lower panel. The asymmetric clamp shifts the differences toward the analytical reference for both metrics, with the direction and magnitude of improvement varying across seeds.

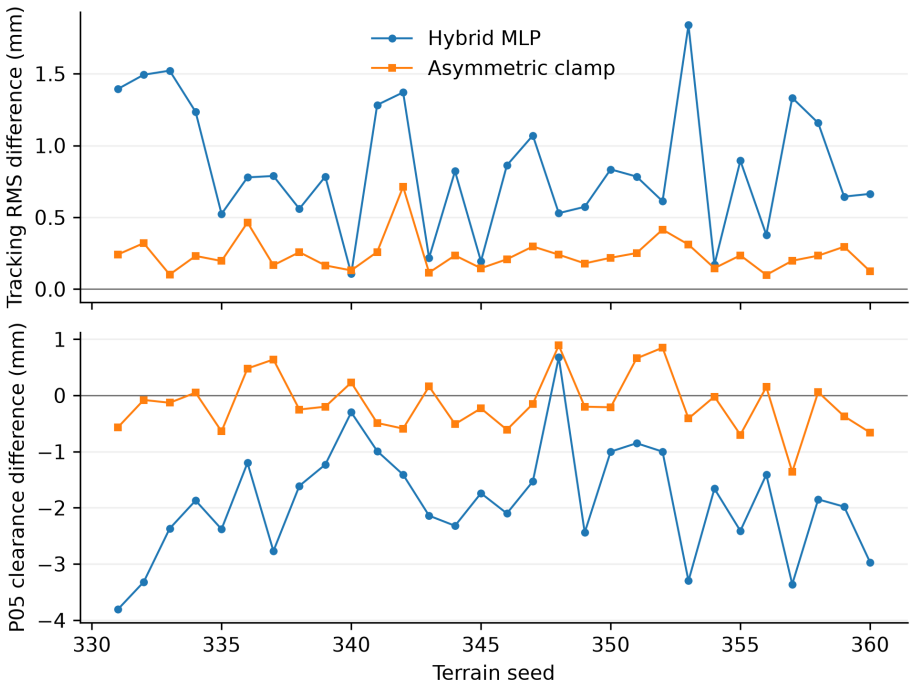


Fig. 8 Seed-level tracking and fifth-percentile clearance contrasts against analytical feedforward. The same 30 terrains contribute to both panels.

Controlled actuator cancellation.

To isolate the effect of motor lag, we ran the firmware mapping and simulator motor model at a fixed nominal command of 0.55, starting at free-air equilibrium. At time zero we stepped the prescribed GE multiplier from unity to 0.85, 1.2, or 1.6, with and without oracle compensation. The motor update interval was 1 ms, with the same 30 ms time constant as the flight simulator. At multiplier 1.6, compensated thrust divided by nominal free-air thrust was 1.600 at onset, 1.208 at 30 ms, 1.004 at 150 ms, and 1.000 to six decimals at 500 ms. Without compensation the ratio stayed at 1.600. All three compensated step tests reached unity within 10^{-6} at 500 ms. The source and traces are supplied in the supplement. This test separates steady cancellation from the initial thrust transient in the motor model. The coupled terrain, mixer, and tracking effects remain those of the closed-loop comparisons.

4.5 Causal bias intervention

Compared with −0.02, the +0.02 bias lowered fifth-percentile clearance by 2.571 mm (95% paired t CI 1.721–3.422 mm lower, adjusted $p = 9.66 \times 10^{-6}$) and added 0.326 s below the clearance threshold (95% CI 0.207–0.444 s, adjusted $p = 3.60 \times 10^{-5}$).

Table 5 Controller minus analytical paired differences, ordinary 95% paired t intervals, and Holm-adjusted p values.

Controller	Metric	Difference	95% CI	Adj. p
None	AGL RMS (mm)	+7.252	[+6.038, +8.465]	6.98e-12
None	P05 (mm)	+3.423	[+2.303, +4.543]	6.4e-06
None	Low time (s)	-0.377	[-0.518, -0.236]	6.17e-05
Oracle	AGL RMS (mm)	+0.142	[+0.000, +0.284]	0.0986
Oracle	P05 (mm)	+0.025	[-0.213, +0.263]	0.829
Oracle	Low time (s)	-0.011	[-0.035, +0.012]	0.991
Direct MLP	AGL RMS (mm)	+6.646	[+5.621, +7.671]	1.15e-12
Direct MLP	P05 (mm)	-1.223	[-1.937, -0.509]	0.00858
Direct MLP	Low time (s)	+0.170	[+0.084, +0.256]	0.00182
Hybrid MLP	AGL RMS (mm)	+0.847	[+0.681, +1.012]	1.43e-10
Hybrid MLP	P05 (mm)	-1.888	[-2.251, -1.525]	2.1e-10
Hybrid MLP	Low time (s)	+0.173	[+0.142, +0.204]	3.67e-11
Asymmetric clamp	AGL RMS (mm)	+0.240	[+0.194, +0.286]	1.1e-10
Asymmetric clamp	P05 (mm)	-0.140	[-0.329, +0.049]	0.419
Asymmetric clamp	Low time (s)	+0.005	[-0.019, +0.029]	1

Mean tracking RMS decreased by 0.351 mm (adjusted $p = 0.155$). Table 6 and Fig. 9 distinguish tracking and clearance outcomes.

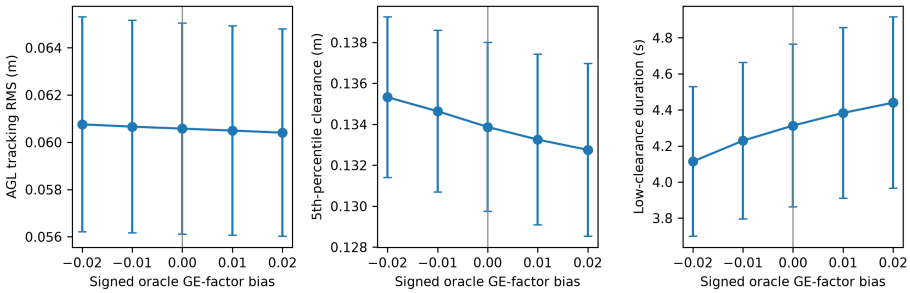


Fig. 9 Tracking RMS, fifth-percentile clearance, and low-clearance exposure with signed oracle bias. Bars show marginal 95% t intervals.

The mean P05 clearance decreased across the five ordered bias levels, from 135.32 to 132.75 mm. Mean exposure increased from 4.114 to 4.440 s. Applying the bias to the oracle isolates an exogenous change in the controller input on matched terrains. The outcomes describe the total effect of that imposed input, including the PID and trajectory response. They do not measure the frequency or duration of such biases in natural hybrid operation. The difference between the tracking and clearance results also motivates reporting adjustment families and safety metrics alongside the primary tracking outcome.

Table 6 Oracle-bias condition means on 30 seeds. AGL RMS intervals are marginal 95% t intervals.

Bias	AGL RMS (m)	95% CI (m)	P05 (m)	Low time (s)
-0.02	0.060758	[0.056211, 0.065305]	0.135324	4.114
-0.01	0.060662	[0.056161, 0.065163]	0.134639	4.229
+0.00	0.060578	[0.056112, 0.065045]	0.133871	4.313
+0.01	0.060497	[0.056070, 0.064923]	0.133253	4.383
+0.02	0.060407	[0.056022, 0.064792]	0.132753	4.440

4.6 Loop-closure and stability analysis

For the 0.25 s, +0.02 bias pulse, mean clearance in the first 0.5 s after pulse termination fell by 2.391 mm relative to matched zero-bias runs. Natural absolute prediction error sometimes increased briefly for either sign. Later windows did not show consistent sustained growth. These exploratory comparisons and the pulse peak responses (Fig. 10) do not establish self-reinforcement.

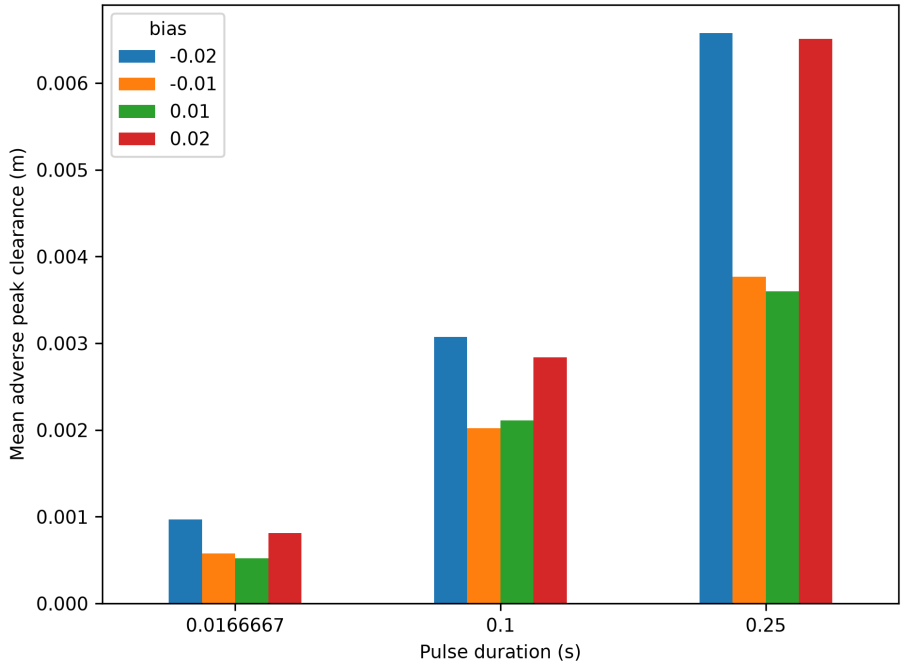


Fig. 10 Sign-aligned peak clearance response to finite bias pulses in the active hybrid controller.

The active pulse design pairs each perturbation with an otherwise matched zero-bias flight. This allows the measured clearance and error to depart from the reference

as the altered actuator command changes the trajectory. The short post-pulse error increase can arise after either sign of the perturbation, which weakens the claim that lower clearance consistently produces the same signed error that initially drove the vehicle down. The analysis follows absolute natural error as an outcome and does not estimate a complete signed feedback gain. A transient return or overshoot from one perturbation cannot establish stability across the switching thresholds at all operating conditions.

Figure 11 shows the post-pulse windows for the two largest signed biases at 0.25 s. The initial clearance changes have opposite signs. The natural absolute-error contrast, which is unsigned at the trajectory level, does not track the sign of the first clearance change. The intervals are exploratory paired summaries for each window. They support a distinction between an immediate actuator effect and evidence of subsequent error growth.

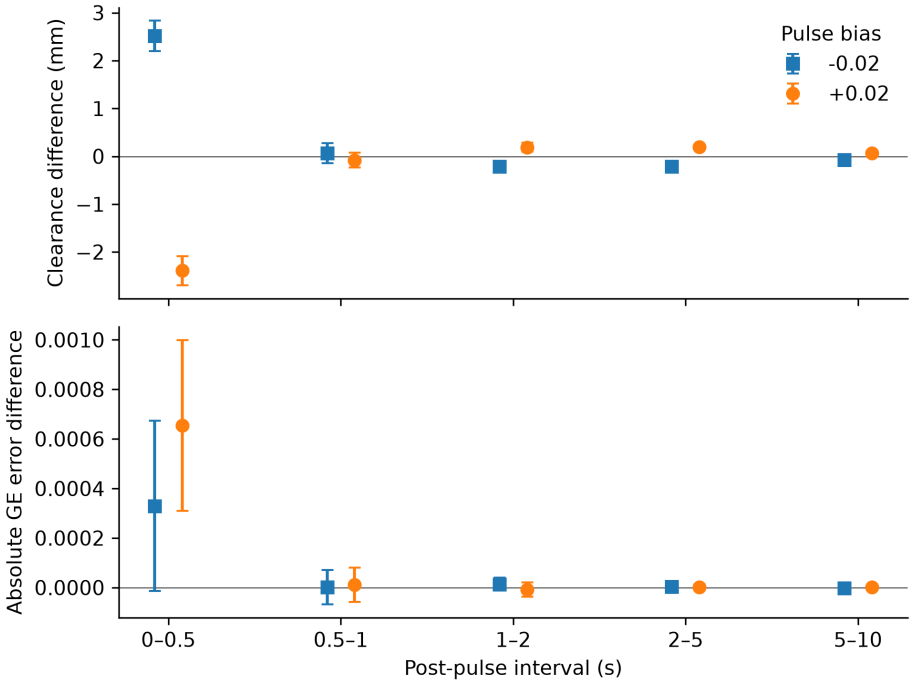


Fig. 11 Differences from the matched zero-bias trajectory following the 0.25 s active-hybrid pulses of −0.02 and +0.02. Bars are ordinary 95% paired t intervals.

Appendix C gives the numerical matrices and delay margins. Appendix D gives the post-pulse window estimates. The delay-free local model had negative pole real parts at all four sampled operating points (Table 1). At unit GE and nominal gains, added delays of 0, 30, and 100 ms gave phase margins of 14.15°, 0.80°, and −30.37° (Fig. 12). The corresponding upper gain ratios were 2.236 and 1.033 at 0 and 30 ms. At 100 ms the nominal linear loop was unstable. These are local margins for a constant

gain perturbation, with hypothetical added pure delays. The forecast horizon does not represent a measured system delay.

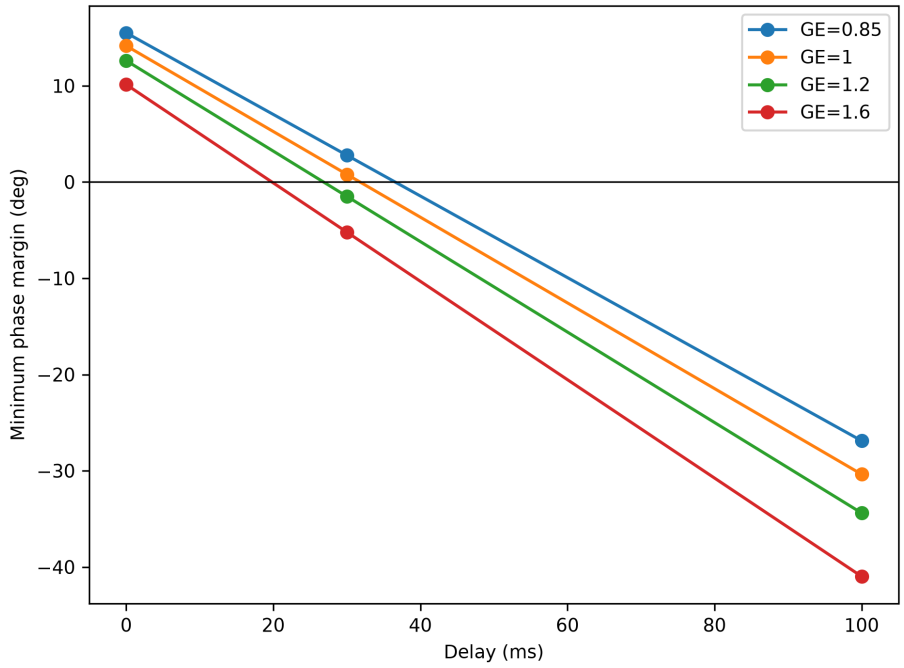


Fig. 12 Local phase margin versus added pure delay at nominal altitude gain.

Figure 13 shows the local frequency response at those delays.

4.7 Mitigation

Relative to standard hybrid, the validation-selected policy reduced tracking RMS by 0.607 mm (95% paired t CI 0.453–0.761 mm), increased fifth-percentile clearance by 1.748 mm (1.454–2.042 mm), and reduced low-clearance exposure by 0.168 s (0.139–0.198 s). Holm-adjusted p values were 3.39×10^{-8}, 9.34×10^{-12}, and 2.07×10^{-11}, respectively. Differences in minimum clearance and event count were inconclusive. Above cutover, the capped policy replaces the symmetric trust-band decision.

On validation terrains, the candidate $\delta = 0$ achieved mean RMS 59.393 mm and mean exposure 4.140 s, while the loosest tested cap ($\delta = 0.04$) achieved 60.049 mm and 4.340 s. The selected threshold was frozen before evaluation on seeds 331–360. On those final terrains, the capped policy recovered approximately 72% of the tracking gap from hybrid to analytical. The remaining capped-minus-analytical tracking difference was 0.240 mm. Relative to analytical, the capped policy's confidence intervals for P05 and low-clearance duration straddled zero, so the safety improvement relative to the standard hybrid should not be generalized to an advantage over analytical feedforward.

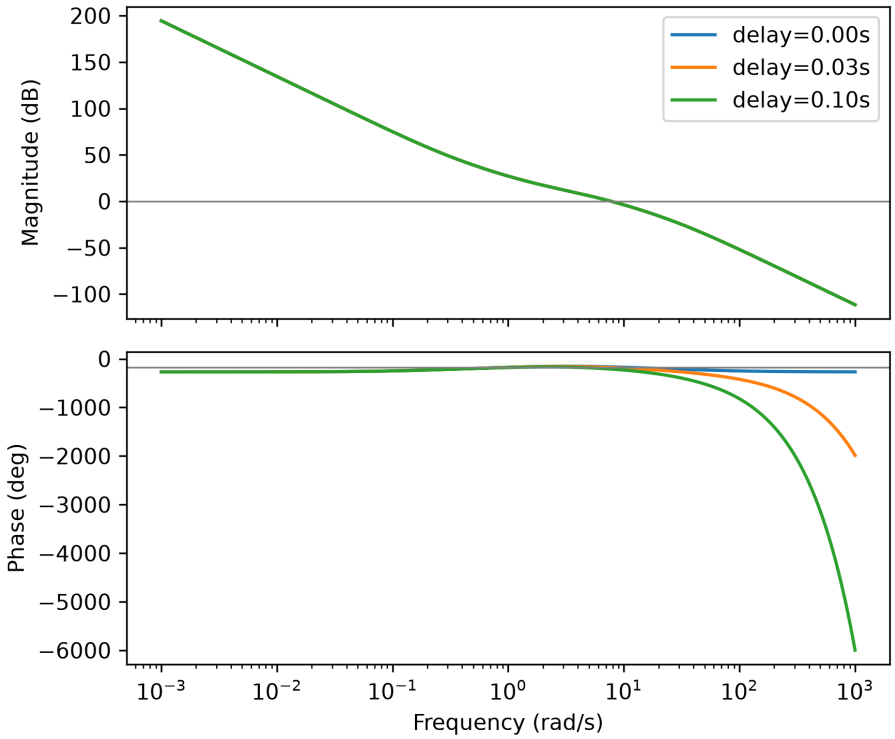


Fig. 13 Frequency response for the local vertical loop with added pure delay.

4.8 Robustness to ground-truth correction strength

The hybrid-minus-analytical RMS difference ranged from 0.845 to 0.868 mm across the nine correction settings (Fig. 14).

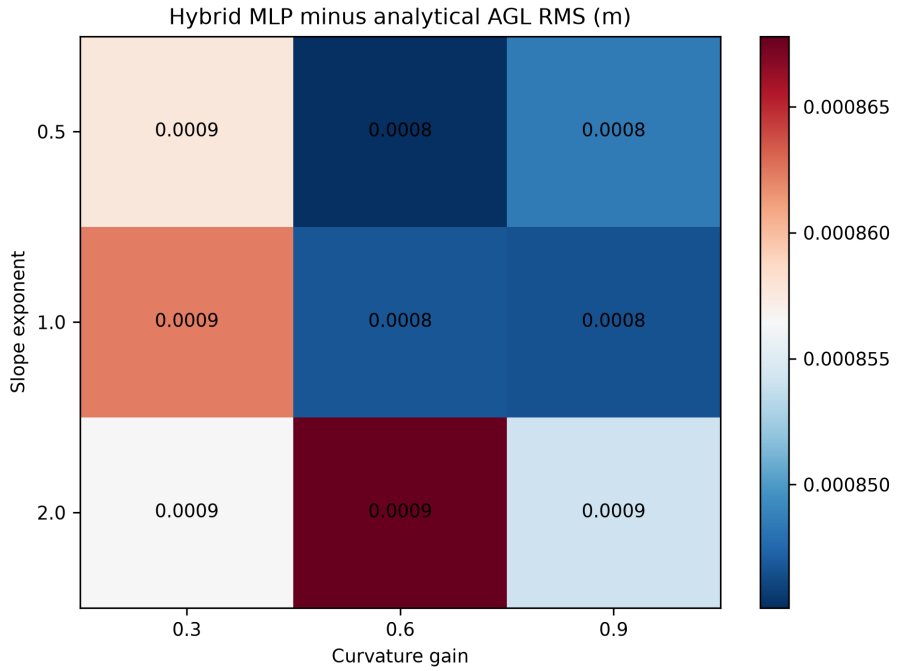


Fig. 14 Hybrid minus analytical tracking RMS over nine correction settings.

In the Hayden configuration, hybrid still trailed analytical while oracle improved upon analytical by 1.466 mm (95% paired t CI 1.185–1.747 mm, adjusted $p = 6.05 \times 10^{-11}$, Fig. 15).

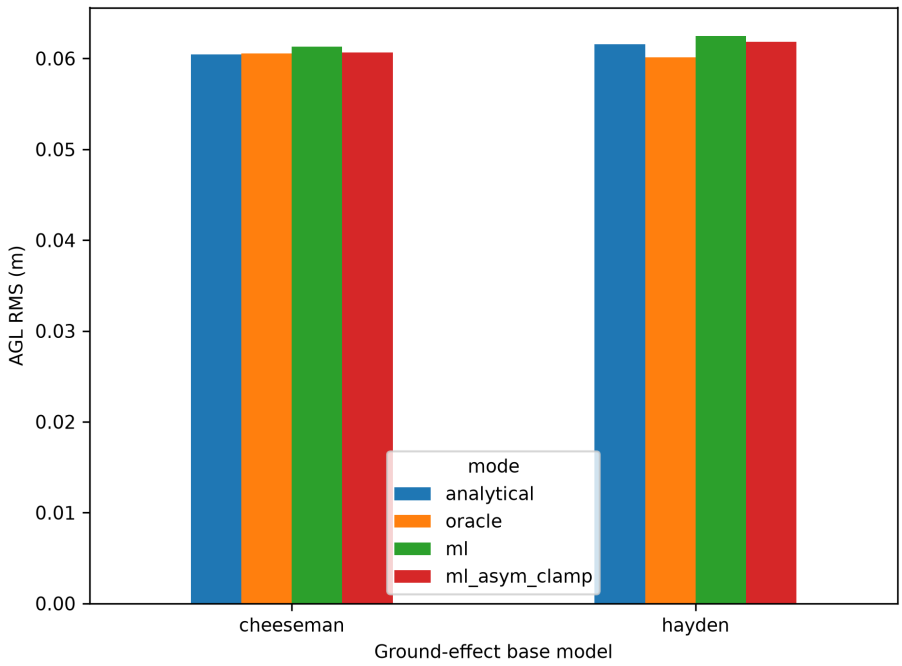


Fig. 15 Tracking RMS under two ground-effect formulations.

Current-factor arithmetic and thrust-weighted oracles achieved 0.060578 and 0.060794 m, respectively. Their 100 ms kinematic-preview variants achieved 0.064186 and 0.064189 m (Fig. 16). This comparison evaluates applying the approximate preview immediately.

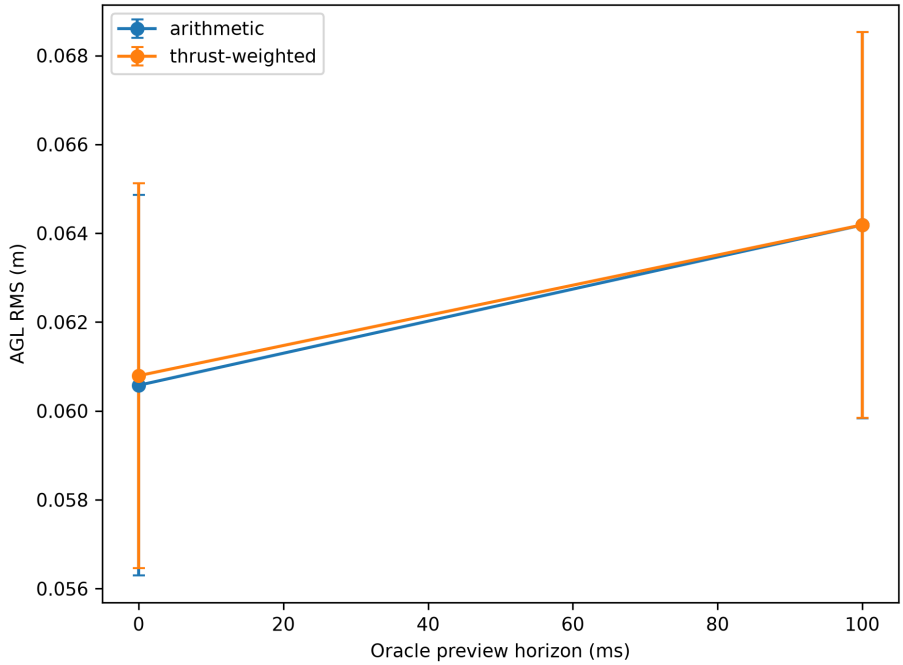


Fig. 16 Tracking RMS for current and previewed oracle factors. Bars show mean plus or minus 1.96 standard errors.

Table 7 reports the feedback sensitivity test. All altitude PID gains were multiplied by the indicated scale with the same seeds, motor dynamics, and flight plan. The oracle mean remained slightly higher than analytical at each scale. This experiment did not reveal a tracking advantage hidden by nominal feedback gains. The scaling changes the feedback response, so the experiment does not isolate the fraction of disturbance rejected by feedback.

At scales 0.5 and 0.25, oracle-minus-analytical RMS was 0.222 mm (95% paired t CI 0.167–0.277 mm) and 0.307 mm (0.239–0.375 mm), respectively. Both analytical and oracle achieved lower absolute tracking RMS at scale 0.5 than at scale 1.0, and both worsened at scale 0.25. These changes show that the tracking metric responds to feedback tuning. The tested gain multipliers are changes to the common PID settings and do not correspond to independently matched closed-loop bandwidths across all hybrid and oracle trajectories.

Figure 17 shows the paired tracking differences over the same gain sweep. Pairing reveals the small separation between the oracle and the analytical controller within each gain setting. The intervals summarize contrasts within each setting and do not test interactions across gain settings.

Table 7 Mean AGL RMS (m) under altitude PID gain scaling on 30 matched seeds.

Gain scale	None	Analytical	Oracle	Hybrid MLP
1	0.067688	0.060436	0.060578	0.061283
0.5	0.062406	0.057084	0.057306	0.057810
0.25	0.075474	0.065994	0.066301	0.067192

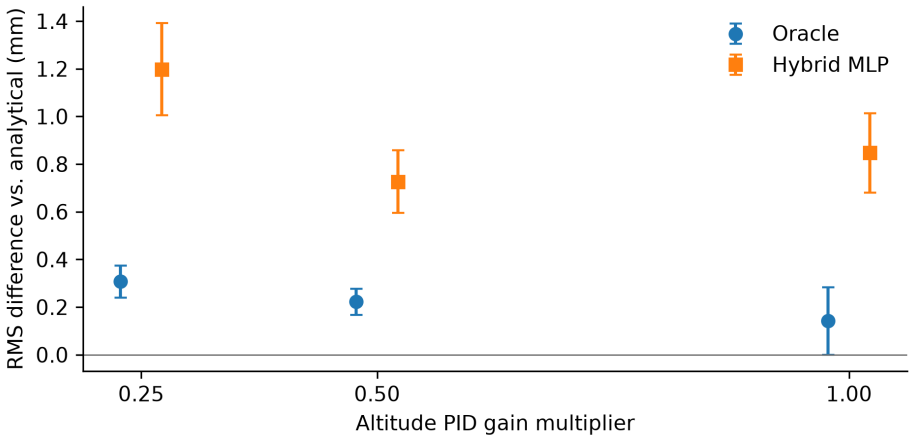


Fig. 17 Oracle and hybrid tracking RMS minus analytical RMS under altitude PID gain scaling. Bars show 95% paired t intervals. The two series are offset horizontally for readability.

4.9 Physical bench observations and rangefinder delay sensitivity

Table 8 and Fig. 18 summarize the nine filmed thrust runs. Estimated mean thrust was 3.003 N at 25 cm and about 3.138 N at both closer heights. Within the three blocks, the six near/reference RPM mismatches ranged from 2.08% to 3.17% in magnitude. None was within 1% or 2%. An exploratory correction dividing the paired thrust ratio by the squared RPM ratio yields mean differences of +2.28% at 20 cm and +3.32% at 15 cm. Those values depend on a local RPM-squared scaling and may reflect acquisition order, battery state, and other bench effects. We retain every run in the supplement. The closest retained subwindows from 25 cm and either closer height still differed by 1.09% in RPM. The 15 and 20 cm conditions, however, had RPM-matched subwindows in all three blocks (0.008–0.302% mismatch). For 15 versus 20 cm, the mean count-derived relative force change at the central alignment was +0.85%, with individual contrasts of −0.45, +1.70, and +1.29%. Independent alignment shifts changed the individual estimates by up to 1.10 percentage points, and one contrast changed sign. These comparisons are exploratory because the clocks were not synchronized and the matching was performed after acquisition.

Table 8 Single-rotor bench runs at 1400 μs. The force values are means and sample standard deviations over three runs per height, using the original calibration gain and each run's measured zero.

Clearance (cm)	Clearance / radius	Mean speed (rpm)	Estimated thrust (N)
25	1.969	4,983	3.003 ± 0.149
20	1.575	5,037	3.138 ± 0.051
15	1.181	5,012	3.139 ± 0.016

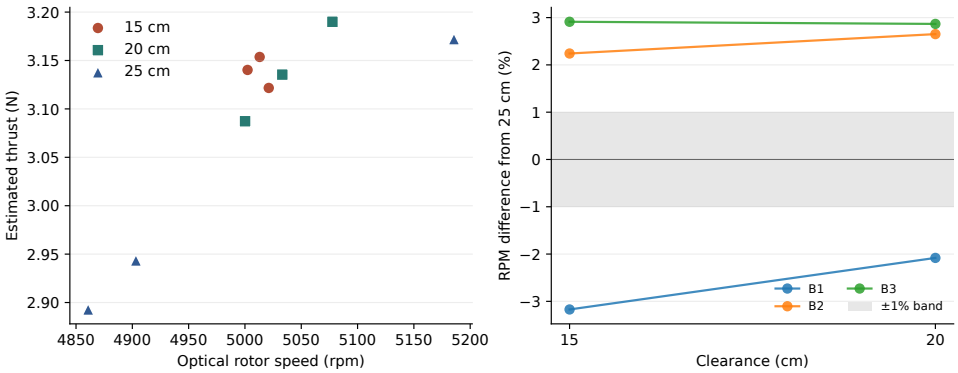


Fig. 18 Estimated thrust and measured speed at the same PWM command. The second panel shows the within-block speed difference from the 25 cm reference. The shaded band is ±1%.

The sensor video contained six sustained rising and six sustained falling LED edges. Six additional 2–4-frame excursions are listed separately in the supplement. Relative to the projected midpoint of MDF motion, median signed differences were 19.5 ms (range 2–41 ms) for rising and 66.5 ms (38–91 ms) for falling edges. At the 25% and 75% motion references, the rising-edge medians were 47.0 and −39.0 ms, and the falling-edge medians were 129.0 and −6.5 ms (Fig. 19). Using the first sustained 5-pixel movement of the projected MDF edge as an earlier visual reference gave medians of 80.0 and 204.5 ms, respectively. The sign changes across the motion references show how strongly the timing estimate depends on the reference. A marked crossing of the sensor's optical field would be needed to estimate response latency for a known physical step.

Table 9 reports the added-delay simulation outcomes. The zero-delay condition reproduced all six original per-seed metrics for every controller. At 20 ms, analytical tracking RMS increased from 60.436 to 82.843 mm, and the hybrid-minus-analytical difference increased from 0.847 to 1.936 mm (paired 95% CI 1.383–2.488 mm). At 70 ms, the analytical and hybrid means were 710.406 and 723.712 mm, respectively, giving a difference of 13.306 mm (7.129–19.483 mm). Fifth-percentile clearance for analytical feedforward declined from 133.845 to 98.343 mm across these endpoint conditions. The delay changes the range feedback available to all modes, including the oracle controller's feedback path. These large tracking errors quantify degradation

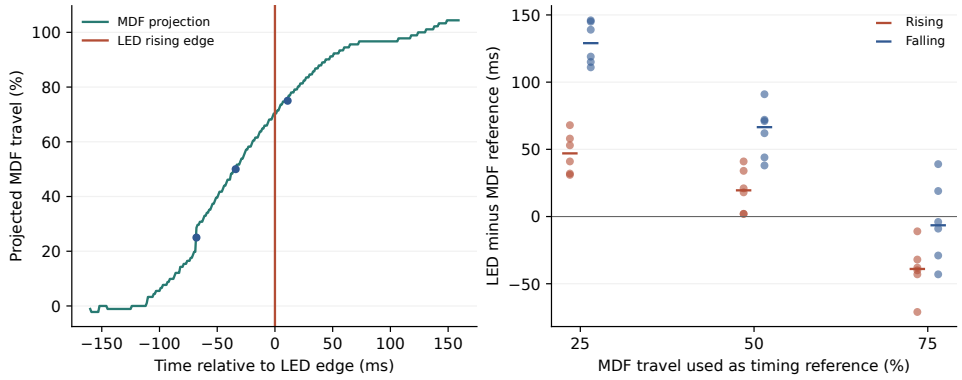


Fig. 19 LED transition time relative to image-based MDF motion. The left panel shows one rising transition, with markers at 25%, 50% and 75% of the apparent motion. The right panel shows events and direction-specific medians for each reference. Negative values indicate that the LED edge precedes the selected milestone.

on the tested trajectories. They do not establish a nonlinear stability boundary or a measured hardware delay.

Table 9 Mean AGL tracking RMS (mm) across the same 30 terrain seeds per controller and added rangefinder delay. The sampling period was 1 ms at all settings.

Controller	0 ms	20 ms	70 ms
None	67.688	103.635	841.488
Analytical	60.436	82.843	710.406
Current oracle	60.578	84.665	739.183
Pure MLP	67.082	100.194	810.706
Hybrid MLP	61.283	84.778	723.712
Asymmetrically clamped MLP	60.676	83.293	721.299

5 Discussion

Identification for control distinguishes model fit from downstream control performance [13, 14]. Sequential imitation learning likewise identifies a shift in the visited states when policy outputs influence subsequent observations [15]. The present study provides a controlled multirotor ground-effect case with matched-seed oracle and bias diagnostics. Offline RMSE and R^2 quantify forecast errors without weighting their closed-loop consequences. The offline-leading LSTM was evaluated as a predictor, and the deployed MLP improved tracking relative to no feedforward but trailed analytical feedforward. The default analytical and oracle means were close, while the Hayden sensitivity showed an oracle advantage. Prediction-error structure, timing, motor lag,

and collective aggregation can all affect the usefulness of an estimate. The bias intervention shows that an error can worsen clearance even when aggregate tracking RMS decreases slightly. The active pulse grid provides evidence on finite perturbations, and the pole and delay calculations characterize the specified local model.

The results distinguish three evaluation questions. The held-out prediction experiment asks whether onboard observations forecast the simulated GE target under the data-collection policy. The matched controller comparison asks whether acting on those forecasts improves a specified outcome. The bias interventions ask which outcomes change when the factor supplied to the controller is perturbed. A model can rank well on the first question and still produce errors with substantial consequences during the short periods that dominate the third. Conversely, the closed-loop experiment compares whole controllers, including their fallback and cutover rules. It cannot separate the effects of individual prediction errors from those rules without further interventions.

The apparent closeness of the default analytical and oracle means is specific to the tested outcome and dynamics. Reducing the altitude PID gains did not reveal an oracle advantage at the tested scales, whereas changing the base aerodynamic model did. Thus a one-factor explanation based on feedback rejection alone has limited support here. The weighting experiment also changes only the collective GE factor. Even a thrust-weighted ratio that exactly describes total instantaneous thrust cannot convey the independent rotor adjustments needed to correct a GE-induced torque. The analytical–oracle comparison is best read as a diagnostic of the available improvement in this controller and simulator, with the achievable margin sensitive to the assumed plant.

6 Limitations and future work

The heuristic aerodynamic model and sensor settings limit extrapolation to physical vehicles. The 0.847 mm tracking difference and 2.571 mm clearance contrast are about 0.67% and 2.0% of the 127 mm rotor radius. The two contrasts are smaller than the simulator's per-sample range noise standard deviation (10 mm). The evaluated outcomes use simulator geometry and aggregate many samples, so this comparison provides a scale reference. The moving-target measurement did not quantify range resolution or deployment uncertainty. The bench force estimates depend on extrapolation from a 30 g load, and the RPM mismatches limit aerodynamic interpretation. The rangefinder-delay perturbations quantify sensitivity to specified simulator timing, with the MDF measurement providing only an approximate time scale. Both training and evaluation use the corrected simulator, though deploying a different controller can still shift the input distribution. Collective compensation does not resolve differential rotor disturbances. A nonlinear input-to-state stability analysis across switching regimes remains open. Independent validation would require range steps with a defined optical reference, calibrated thrust at matched rotor speeds, and multirotor tests over irregular terrain.

7 Conclusion

We evaluated machine-learned prediction of multirotor ground effect from onboard-observable sensors against classical flat-ground analytical estimation, both offline and in closed-loop simulation, using corrected flight dynamics and disjoint terrain-seed partitions. Offline, learned models substantially outperformed the analytical baseline, with the LSTM achieving an RMSE of 0.007417 ($R^2 = 0.919927$), compared with 0.024537 ($R^2 = 0.123749$) for the analytical baseline on 106,920 common held-out samples at a 100 ms prediction horizon. In closed-loop evaluation across 30 independent unseen terrain seeds, however, the deployed hybrid MLP controller achieved an AGL tracking RMS error of 0.061283 m, compared with 0.060436 m for analytical feedforward, demonstrating that improved offline prediction accuracy did not translate into superior tracking under the investigated collective-thrust compensation architecture.

The oracle controller, supplied with the simulator's current ground-truth GE factor, achieved an AGL tracking RMS error of 0.060578 m under the default Cheeseman-based configuration, with no statistically detectable difference from analytical feedforward after Holm correction. However, under the Hayden-based configuration, oracle compensation improved tracking RMS by 0.001466 m relative to analytical feedforward, demonstrating that the apparent proximity of analytical and oracle performance depends on the assumed aerodynamic model. Controlled bias interventions further showed that GE-factor overprediction reduced fifth-percentile clearance and increased low-clearance exposure while yielding a slightly lower mean tracking RMS. The tracking contrast was inconclusive after Holm adjustment. Active-controller pulse experiments demonstrated transient coupling between prediction perturbations and vehicle trajectories but did not establish sustained self-reinforcing prediction-error growth. Moreover, although the delay-free linearized model exhibited poles with negative real parts at the investigated operating points, the nominal 100 ms-delay analysis yielded a negative phase margin of $-30.37°$, precluding general claims of closed-loop stability.

An asymmetric clamp constraining GE-factor overprediction reduced the hybrid controller's tracking disadvantage relative to analytical feedforward, while the hybrid-minus-analytical difference remained positive across all nine tested slope/curvature correction configurations. These findings indicate that offline prediction metrics alone are insufficient to establish the effectiveness of learned aerodynamic compensation in closed-loop control. They also demonstrate the value of oracle comparisons, controlled interventions, and delay-sensitive analysis for distinguishing predictive performance from controller-level behavior. Because the aerodynamic corrections are heuristic and the observed tracking differences are small relative to potential modeling and sensing uncertainties, the conclusions remain restricted to the investigated synthetic simulation configurations. The bench measurements complement the simulation by illustrating challenges in matching rotor speed and defining a physical timing reference. Independent aerodynamic and flight-level controller validation remain necessary before extending the closed-loop findings to physical multirotor operation.

A Numerical details for the offline assessment

Table 10 records the common target count and all three pooled metrics at each lead. The target count is shared across methods at each lead and changes with lead as samples near the trajectory boundary become unavailable. The 100 ms sweep estimates below were obtained from separate fits and are used for within-sweep comparisons. The primary 100 ms comparison in Table 2 uses the broader set of candidate models.

Table 10 Held-out performance of the four horizon-sweep estimators.

Lead (ms)	Model	Targets	RMSE	MAE	R^2
0	LSTM	107100	0.001859	0.000926	0.995255
0	MLP	107100	0.003679	0.001696	0.981412
0	Random forest	107100	0.003716	0.001419	0.981030
0	Analytical	107100	0.010714	0.003623	0.842318
50	LSTM	107010	0.005251	0.002367	0.961030
50	MLP	107010	0.009231	0.004083	0.879555
50	Random forest	107010	0.009278	0.003798	0.878347
50	Analytical	107010	0.017909	0.006834	0.546690
100	LSTM	106920	0.007613	0.003590	0.915652
100	Random forest	106920	0.013927	0.006079	0.717709
100	MLP	106920	0.014022	0.006733	0.713825
100	Analytical	106920	0.024537	0.010253	0.123749
200	LSTM	106740	0.010013	0.004446	0.844770
200	MLP	106740	0.016549	0.007825	0.575974
200	Random forest	106740	0.016563	0.007767	0.575279
200	Analytical	106740	0.030208	0.014138	-0.412774

A.1 Terrain subgroup support

The terrain comparison conditions on the source-row clearance and terrain class. For each class and model, an error metric is first calculated within each seed and then averaged across contributing seeds. Table 11 reports the contributing row count for the analytical estimate and LSTM in the same clearance band used by the main terrain figure. A row belongs to a single terrain class. Within this subset, estimates with more variation in the number of per-seed rows do not automatically receive more weight in the seed-mean metric.

B Full contrasts against analytical feedforward

Tables 12–17 show all five nonreference controllers against analytical feedforward. Each difference is formed within a seed, and the table shows its mean, ordinary 95% paired t interval, and Holm-adjusted p value. A positive clearance difference indicates a larger clearance, while positive duration and event differences indicate more exposure

Table 11 Offline subgroup results at 0.15–0.25 m clearance.

Terrain	Model	Rows	Seeds	MAE	RMSE
Flat	Analytical	29687	30	0.003552	0.006222
Flat	LSTM	29687	30	0.001353	0.001997
Bump	Analytical	23979	30	0.006415	0.009530
Bump	LSTM	23979	30	0.003247	0.005431
Dip	Analytical	19478	30	0.007769	0.016337
Dip	LSTM	19478	30	0.003011	0.004914
Ripple	Analytical	6550	30	0.020598	0.039275
Ripple	LSTM	6550	30	0.008291	0.016219

or events. Holm families are defined across the full set of pairwise controller comparisons within each metric. The reported confidence intervals are not simultaneous intervals.

Table 12 AGL tracking RMS (mm). Controller minus analytical, paired by seed.

Controller	Difference	95% CI	Adjusted p
No feedforward	7.252	[6.038, 8.465]	6.98e-12
Oracle	0.142	[0.000, 0.284]	0.0986
Direct MLP	6.646	[5.621, 7.671]	1.15e-12
Hybrid MLP	0.847	[0.681, 1.012]	1.43e-10
Asymmetric clamp	0.240	[0.194, 0.286]	1.1e-10

Table 13 Gyroscope RMS (rad/s). Controller minus analytical, paired by seed.

Controller	Difference	95% CI	Adjusted p
No feedforward	-0.000050	[-0.001091, 0.000991]	1
Oracle	-0.000274	[-0.000386, -0.000162]	0.000283
Direct MLP	0.000731	[-0.000248, 0.001710]	0.826
Hybrid MLP	0.000890	[0.000634, 0.001147]	1.16e-06
Asymmetric clamp	0.000083	[0.000038, 0.000128]	0.00779

Table 14 Minimum clearance (mm). Controller minus analytical, paired by seed.

Controller	Difference	95% CI	Adjusted p
No feedforward	-0.126	[-0.384, 0.132]	1
Oracle	0.023	[-0.024, 0.070]	1
Direct MLP	-0.064	[-0.196, 0.067]	1
Hybrid MLP	-0.055	[-0.169, 0.058]	1
Asymmetric clamp	-0.008	[-0.024, 0.008]	1

Table 15 Fifth-percentile clearance (mm). Controller minus analytical, paired by seed.

Controller	Difference	95% CI	Adjusted p
No feedforward	3.423	[2.303, 4.543]	6.4e-06
Oracle	0.025	[-0.213, 0.263]	0.829
Direct MLP	-1.223	[-1.937, -0.509]	0.00858
Hybrid MLP	-1.888	[-2.251, -1.525]	2.1e-10
Asymmetric clamp	-0.140	[-0.329, 0.049]	0.419

Table 16 Low-clearance duration (s). Controller minus analytical, paired by seed.

Controller	Difference	95% CI	Adjusted p
No feedforward	-0.377	[-0.518, -0.236]	6.17e-05
Oracle	-0.011	[-0.035, 0.012]	0.991
Direct MLP	0.170	[0.084, 0.256]	0.00182
Hybrid MLP	0.173	[0.142, 0.204]	3.67e-11
Asymmetric clamp	0.005	[-0.019, 0.029]	1

Table 17 Low-clearance events (count). Controller minus analytical, paired by seed.

Controller	Difference	95% CI	Adjusted p
No feedforward	-1.133	[-2.170, -0.097]	0.299
Oracle	-0.133	[-0.368, 0.101]	1
Direct MLP	1.367	[0.520, 2.213]	0.0306
Hybrid MLP	0.233	[-0.254, 0.721]	1
Asymmetric clamp	-0.067	[-0.343, 0.210]	1

C Correction-strength and delay sensitivity

Table 18 reports the mean AGL tracking RMS for every controller and every correction cell. The default is the middle cell. The table separates the trend within the Cheeseman-based correction family from switching to another flat-ground term. The same 30 terrain seeds are used in each cell.

Table 18 All 54 correction-cell/controller means of AGL tracking RMS (mm).

p	g	None	Analytical	Oracle	Direct	Hybrid	Clamp
0.5	0.3	67.793	60.530	60.536	67.211	61.388	60.761
0.5	0.6	67.771	60.513	60.540	67.179	61.358	60.755
0.5	0.9	67.749	60.487	60.546	67.133	61.335	60.726
1	0.3	67.719	60.446	60.568	67.104	61.308	60.693
1	0.6	67.688	60.436	60.578	67.082	61.283	60.676
1	0.9	67.678	60.410	60.574	67.045	61.256	60.652
2	0.3	67.574	60.298	60.611	66.898	61.154	60.540
2	0.6	67.542	60.270	60.622	66.871	61.138	60.515
2	0.9	67.518	60.258	60.622	66.836	61.113	60.485

C.1 Companion matrices and delay margins

The first four rows of the companion form are shared across operating points. Table 19 gives the variable final row. Substituting it into Eq. 12 yields the numerical matrix for that GE factor. The full-precision values are included with the results files. Table 20 shows the nominal-gain margin results and omits the upper ratio when the nominal linear system fails the stability check. Each entry refers to a fixed GE operating point and an added pure delay.

Table 19 Final row of the companion matrix at each GE factor.

GE	a_0	a_1	a_2	a_3	a_4
0.85	-1085.50914874	-4504.86296727	-2460.48740381	-222.22222222	-40.00000000
1.0	-1177.39998301	-4886.20992951	-2668.77329483	-222.22222222	-40.00000000
1.2	-1289.77705981	-5352.57479819	-2923.49466889	-222.22222222	-40.00000000
1.6	-1489.30626535	-6180.62100119	-3375.76086812	-222.22222222	-40.00000000

Table 20 Nominal altitude gains. Phase margin in degrees and upper gain ratio.

GE	Delay (ms)	Phase margin	Upper ratio
0.85	0	15.50	2.426
0.85	30	2.79	1.120
0.85	100	-26.88	–
1.0	0	14.15	2.236
1.0	30	0.80	1.033
1.0	100	-30.37	–
1.2	0	12.61	2.042
1.2	30	-1.49	–
1.2	100	-34.39	–
1.6	0	10.13	1.768
1.6	30	-5.20	–
1.6	100	-40.98	–

D Active prediction pulse follow-up

Table 21 reports the positive 0.02, 0.25 s pulse used as the illustrative contrast in the Results. Each row is a post-pulse paired difference from the zero-bias trajectory, averaged within the time window and then across the 30 terrains. Clearance is expressed in millimetres. Absolute prediction error is in GE-factor units. The table shows point estimates with ordinary 95% intervals for each window. The remaining durations, signs and amplitudes are supplied in the same machine-readable format. The error diagnostic compares the naturally predicted factor with same-tick simulator truth in each trajectory; it is distinct from the 100 ms-ahead offline forecast score.

Table 21 Paired post-pulse response to a 0.25 s, +0.02 prediction pulse.

Window (s)	Metric	Mean difference	95% CI
0.0–0.5	Clearance (mm)	-2.391044	[-2.694159, -2.087930]
0.0–0.5	Absolute error	0.000655	[0.000310, 0.000999]
0.5–1.0	Clearance (mm)	-0.081756	[-0.235717, 0.072206]
0.5–1.0	Absolute error	0.000011	[-0.000058, 0.000080]
1.0–2.0	Clearance (mm)	0.189194	[0.095931, 0.282458]
1.0–2.0	Absolute error	-0.000008	[-0.000036, 0.000020]
2.0–5.0	Clearance (mm)	0.194759	[0.149588, 0.239930]
2.0–5.0	Absolute error	0.000001	[-0.000010, 0.000013]
5.0–10.0	Clearance (mm)	0.070059	[0.040283, 0.099834]
5.0–10.0	Absolute error	0.000002	[-0.000005, 0.000009]

E Exact seed assignments

The terrain sets below are disjoint. Offline partitions refer to the corrected 300-episode dataset. The controller-validation set selects the upper prediction cap and the controller-evaluation set is the independent set used for the comparative experiments.

Offline training (240 seeds).

1, 4, 5, 7, 8, 9, 10, 11, 13, 14, 15, 16, 17, 18, 19, 21, 22, 23, 24, 25, 26, 28, 30, 31, 33, 34, 35, 37, 38, 40, 41, 42, 43, 44, 45, 46, 48, 50, 51, 52, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 73, 75, 76, 77, 78, 79, 80, 82, 83, 84, 85, 86, 87, 88, 90, 91, 92, 95, 96, 97, 98, 101, 102, 103, 104, 105, 107, 108, 109, 110, 111, 112, 113, 114, 115, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 132, 133, 134, 135, 136, 137, 139, 141, 142, 144, 145, 147, 149, 150, 152, 153, 154, 155, 156, 157, 158, 159, 161, 162, 163, 165, 166, 167, 168, 169, 170, 171, 172, 174, 175, 176, 177, 180, 181, 182, 183, 186, 187, 190, 191, 192, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 222, 223, 224, 225, 226, 229, 230, 232, 235, 236, 238, 239, 240, 241, 243, 244, 245, 246, 247, 248, 249, 251, 252, 254, 256, 257, 258, 259, 260, 261, 262, 263, 264, 266, 267, 268, 270, 271, 272, 273, 274, 275, 276, 277, 279, 280, 281, 282, 283, 284, 285, 286, 287, 289, 291, 292, 293, 294, 295, 296, 298, 299, 300.

Offline validation (30 seeds).

2, 6, 29, 32, 74, 93, 100, 106, 138, 146, 148, 151, 179, 184, 185, 188, 189, 193, 221, 228, 231, 233, 234, 250, 253, 265, 269, 278, 288, 290.

Offline test (30 seeds).

3, 12, 20, 27, 36, 39, 47, 49, 53, 71, 72, 81, 89, 94, 99, 116, 131, 140, 143, 160, 164, 173, 178, 219, 220, 227, 237, 242, 255, 297.

Controller validation (30 seeds).

301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330.

Final controller evaluation (30 seeds).

331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360.

Acknowledgements. The author acknowledges the use of computational resources for the simulation and machine-learning experiments.

Declarations

Ethical approval

Not applicable. The study measured simulation trajectories, a multirotor pilot observation, and hardware bench responses. It did not collect human or animal subject data.

Consent to participate

Not applicable.

Consent to publish

Not applicable.

- **ORCID**: M.L., https://orcid.org/0009-0007-9707-6384
- **Funding**: The author reports no specific funding for this study.
- **Conflict of interest/Competing interests**: The author declares no competing interests.
- **Data availability**: The version 1.1 dataset is publicly available on Zenodo under CC BY 4.0 (https://doi.org/10.5281/zenodo.24383190) [16]. The deposit includes the corrected 300-seed offline dataset, raw telemetry for the 540 rangefinder-delay sensitivity runs, per-seed simulation outcomes, statistical summaries, run manifests, exact seed lists, and the bench force telemetry, camera timing metadata, derived optical traces, and analysis scripts. The original high-speed recordings are held by the corresponding author and are available upon reasonable request. The public repository (https://github.com/lmwmason/GES) contains the simulator, deployed MLP weights, and corrected data-generation and evaluation commands in `experiments/README.md`.
- **Materials availability**: Not applicable.
- **Code availability**: The simulator, machine learning training scripts, and firmware integration are available at https://github.com/lmwmason/GES. The rangefinder-delay sensitivity experiment is reproducible from commit `73adaf5`.
- **Author contribution**: M.L. conceived the study, developed the simulator and firmware, trained and evaluated all models, conducted all analyses, and wrote the manuscript.
- **Generative AI use**: The study used Claude Code (Anthropic) for assistance with simulator, firmware, and analysis code. OpenAI Codex assisted with bench data processing, figure generation, and manuscript revision.

References

[1] Cheeseman IC, Bennett WE. The Effect of the Ground on a Helicopter Rotor in Forward Flight. Aeronautical Research Council; 1955. 3021.

[2] Sanchez-Cuevas P, Heredia G, Ollero A. Characterization of the Aerodynamic Ground Effect and Its Influence in Multirotor Control. Journal of Sensors. 2017;2017:1–17.

[3] Shi G, Shi X, O'Connell M, Yu R, Azizzadenesheli K, Anandkumar A, et al. Neural Lander: Stable Drone Landing Control Using Learned Dynamics. In: 2019 International Conference on Robotics and Automation (ICRA). IEEE; 2019. p. 9784–9790.

[4] Shi G, Hönig W, Shi X, Yue Y, Chung SJ. Neural-Swarm2: Planning and Control of Heterogeneous Multirotor Swarms Using Learned Interactions. IEEE Transactions on Robotics. 2021;38(2):1063–1079.

[5] Hooi CG, Lagor FD, Paley DA. Flow Sensing, Estimation and Control for Rotorcraft in Ground Effect. In: 2015 IEEE Aerospace Conference. IEEE; 2015. p. 1–8.

[6] Matus-Vargas A, Rodriguez-Gomez G, Martinez-Carranza J. A Monocular SLAM-based Controller for Multirotors with Sensor Faults under Ground Effect. Sensors. 2019;19(22):4917.

[7] O'Connell M, Shi G, Shi X, Azizzadenesheli K, Anandkumar A, Yue Y, et al. Neural-Fly Enables Rapid Learning for Agile Flight in Strong Winds. Science Robotics. 2022;7(66):eabm6597.

[8] Santos D, Saska M, Nascimento T. A Generalized Thrust Estimation and Control Approach for Multirotors Micro Aerial Vehicles. IEEE Robotics and Automation Letters. 2024;9(10):8889–8896. https://doi.org/10.1109/LRA.2024.3433749.

[9] Gräfe A, Scherer C, Hönig W, Trimpe S. How to Model Your Crazyflie Brushless. arXiv preprint arXiv:260305944. 2026;Accepted for publication at the 2026 IEEE International Conference on Robotics and Automation (ICRA).

[10] Yang T, Chai K, Ji J, Wu Y, Xu C, Gao F. Ground-Effect-Aware Modeling and Control for Multicopters. arXiv preprint arXiv:250619424. 2025;.

[11] Zou Y, Li H, Ren Y, Xu W, Li Y, Cai Y, et al. Perch a Quadrotor on Planes by the Ceiling Effect. arXiv preprint arXiv:230700861. 2023;Presented at the 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).

[12] Johnson W. NDARC: NASA Design and Analysis of Rotorcraft. NASA; 2022. NASA/TP–20220000355. Available from: https://ntrs.nasa.gov/citations/20220000355.

[13] Gevers M. Identification for Control: From the Early Achievements to the Revival of Experiment Design. European Journal of Control. 2005;11(4–5):335–352. https://doi.org/10.3166/ejc.11.335-352.

[14] Hjalmarsson H. From Experiment Design to Closed-Loop Control. Automatica. 2005;41(3):393–438. https://doi.org/10.1016/j.automatica.2004.11.021.

[15] Ross S, Gordon G, Bagnell D. A Reduction of Imitation Learning and Structured Prediction to No-Regret Online Learning. In: Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics. vol. 15 of Proceedings of Machine Learning Research; 2011. p. 627–635. Available from: https://proceedings.mlr.press/v15/ross11a.html.

[16] Lee M.: Ground Effect Simulation and Bench Measurements for Multirotor Feedforward Evaluation. Zenodo dataset, version 1.1 (2026). Available from: https://doi.org/10.5281/zenodo.24383190.