# GES: Ground Effect Stabilizer — Onboard-Sensor-Based Feedforward Estimation of Ground Effect on Irregular Terrain via Machine Learning versus Analytical Models

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) causes multirotor unmanned aerial vehicles (UAVs) to experience unmodeled thrust augmentation when flying close to irregular terrain, producing altitude oscillation and control degradation that classical flat-ground analytical models (e.g., Cheeseman–Bennett) cannot capture. We built a physics-based simulator with procedurally generated irregular terrain and trained machine learning (ML) models (Random Forest, Gradient Boosting, LSTM, Transformer, and MLP) to predict GE factor from onboard-observable sensors alone. Offline, all ML models substantially outperform the analytical baseline under a seed-wise train/validation/test split that guarantees no terrain seed leaks across splits (best model $\boldsymbol{R^2 = 0.998}$ vs. $\boldsymbol{R^2 = 0.851}$ for analytical on 30 held-out seeds), confirming that the offline advantage is not an artifact of seed leakage. In closed-loop evaluation across 30 unseen terrain seeds, however, a deployed hybrid MLP controller remained consistently, if modestly, worse than the pure analytical baseline, and — critically — an oracle controller given the simulator's exact ground-truth GE factor performed no better than the analytical baseline either. This indicates the shortfall is not primarily a prediction-accuracy problem. Using the corrected actuator mapping ($\boldsymbol{1/\sqrt{\hat{f}}}$, since thrust is quadratic in throttle-proportional rotor speed), a controlled bias-injection experiment confirms a genuine causal effect: injecting a positive (overprediction) GE-factor bias significantly degrades safety-relevant clearance margins relative to an equivalent negative bias ($\boldsymbol{p < 10^{-5}}$ on 5th-percentile clearance and low-clearance duration), even though its effect on aggregate AGL tracking RMS is small and in the opposite direction. A time-limited bias pulse and closed-loop pole analysis

further show that this effect does not compound into a self-reinforcing feedback loop: the system's poles have negative real parts across the operating range of GE factors, and the observed response is an underdamped but stable return to baseline rather than divergence. An asymmetric clamp on GE-factor overprediction, tuned only on validation seeds, closes most of the gap to the analytical baseline. Across nine slope/curvature correction configurations (1,620 total runs), the hybrid-versus-analytical gap is stable at approximately 0.0010 m. These results show that offline predictive accuracy is not a reliable proxy for closed-loop performance, that a plausible correlation-based mechanistic story requires direct causal and dynamical verification before being reported as established, and that even a perfectly accurate predictor does not guarantee closed-loop improvement when actuator-level effects are not separately accounted for.

**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, downwash from its rotors reflects off the surface and partially recirculates back through the rotor disk, augmenting thrust relative to free-air conditions. This phenomenon, known as ground effect (GE), was first quantified for single-rotor helicopters by Cheeseman and Bennett [1], who derived a closed-form thrust augmentation factor as a function of rotor radius and height above a flat, infinite ground plane. Because the magnitude of GE varies continuously with instantaneous clearance and, on non-flat terrain, with local slope and curvature, an uncompensated flight controller can exhibit altitude and attitude oscillation during low-altitude flight, and abrupt lift excess or deficit during landing — behavior we refer to here as GE-induced motion irregularity.

The Cheeseman–Bennett model and its direct derivatives assume a flat ground plane and a single, isolated rotor; Sanchez-Cuevas et al. [2] showed experimentally that this assumption does not transfer cleanly to multirotor configurations, where per-rotor GE differs due to inter-rotor aerodynamic interaction. This motivates modeling GE per-rotor rather than as a single vehicle-level correction, and more broadly motivates asking whether a data-driven model, given only sensors that are actually available onboard a real UAV, can outperform the flat-ground analytical approximation — particularly over irregular terrain where the flat-ground assumption is violated by construction.

A natural approach, following the hybrid "analytical baseline plus learned residual" structure used in Neural-Lander [3] and Neural-Swarm2 [4], is to learn a correction directly from onboard-observable signals: rangefinder altitude, gyroscope rates, motor RPM and current, and commanded thrust. Because these are exactly the signals a real flight controller has access to at runtime, a model that predicts GE factor accurately from this feature set — and, critically, does so with enough lead time to compensate for control-loop latency — would be directly deployable as a feedforward correction term, without requiring explicit terrain reconstruction.

This paper makes two contributions. First, we quantify, in a physics-based simulator with procedurally generated irregular terrain, how much a machine-learned GE predictor improves on the analytical baseline in an offline (batch) evaluation setting, and how that improvement varies with forecast horizon, terrain slope, and terrain curvature. Second, and more centrally, we report that this offline advantage does not transfer to closed-loop deployment: the best offline model, once wired into the controller as a live feedforward term, performs no better than applying no correction at all. We diagnose the failure mechanism, rule out five plausible alternative explanations through direct experiment, and demonstrate a minimal, targeted mitigation. We present this negative closed-loop result as a central finding rather than a limitation, since it bears on the broader practice of validating control-relevant learned models using offline metrics alone.

Before undertaking this simulation study, we conducted an informal pilot observation intended only to motivate the project, not to serve as validation data: a consumer multirotor (DJI Air 3) was filmed hovering indoors at low and high altitude using a Photron SA-X2 high-speed camera (1000 fps) with a Nikon 70–200 mm lens, and frame-by-frame pixel-coordinate tracking was used to compare positional dispersion at the two altitudes (Fig. 2). The low-altitude trial (left) shows tighter, more oscillatory positional excursions than the high-altitude trial (right). We note this observation only as informal motivation for treating GE-induced motion irregularity as an oscillatory rather than translational phenomenon; the trial was uncontrolled in several respects relevant to interpretation (single trial per condition, indoor GPS-denied hover with the downward rangefinder disabled at low altitude to permit close approach, and pixel coordinates not converted to physical units), so we do not treat it as validating any specific quantitative claim made later in this paper.

**Fig. 1** Pilot observation setup: high-speed camera and reference camera rig (left) and lighting configuration used to permit a high shutter speed at 1000 fps (right).

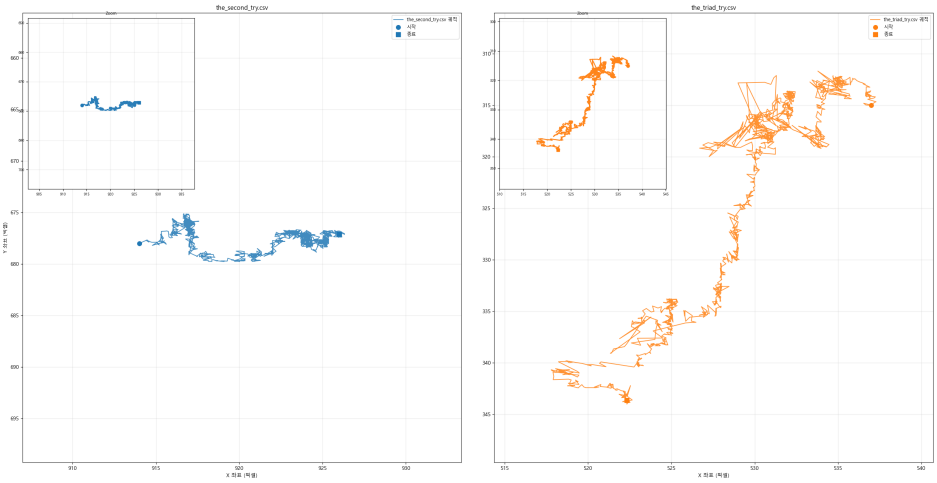

**Fig. 2** Pixel-coordinate hover trajectories from an informal pilot observation (1000 fps high-speed camera), low altitude (left, blue) vs. high altitude (right, orange). Shown only as qualitative motivation; see main text for caveats on interpretation.

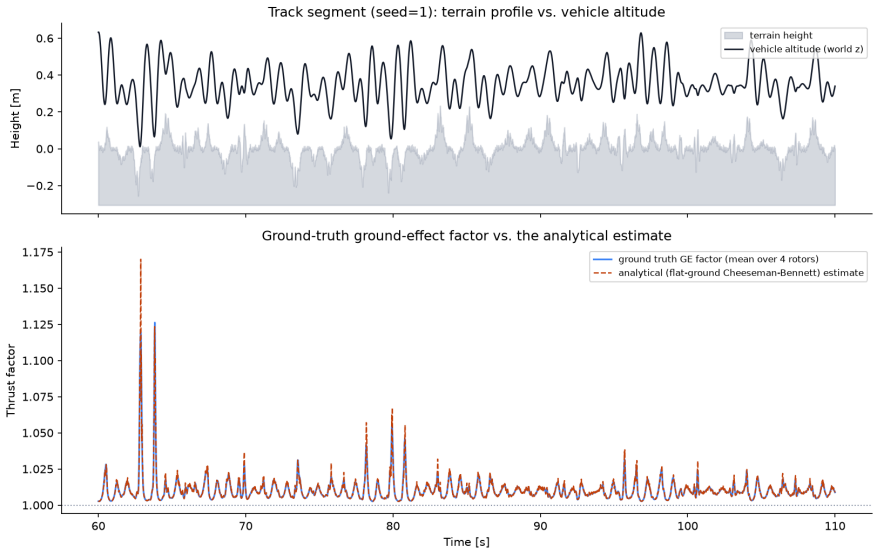


**Fig. 3** A representative flight-track segment showing world-frame altitude oscillation coinciding with an irregular terrain feature, illustrating the GE-induced motion irregularity motivating this study.

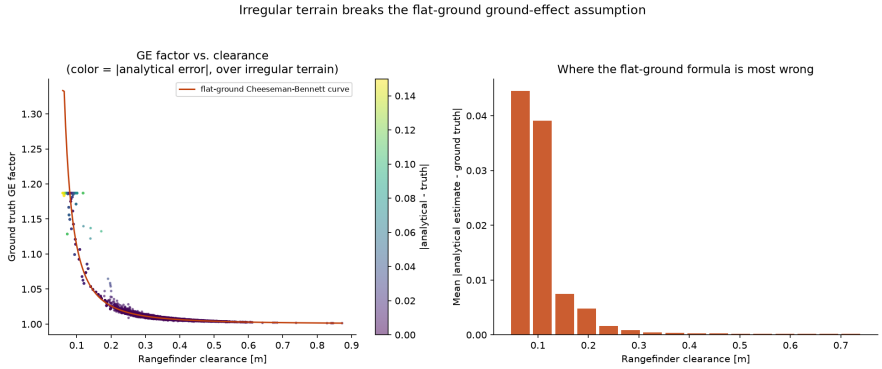


**Fig. 4** Ground-truth GE factor and analytical flat-ground estimation error as a function of rotor clearance. Both the GE factor magnitude and the analytical model's error grow sharply as clearance decreases below approximately 0.2 m.

## 2 Related Work

Cheeseman and Bennett [1] remains the standard analytical reference for rotor-in-ground-effect thrust augmentation and is used here as the flat-ground baseline against which all learned models are compared. Sanchez-Cuevas et al. [2] characterized the breakdown of the single-rotor flat-ground assumption for multirotor platforms, motivating a per-rotor treatment of GE and the sensor-driven approach taken here. Hooi et al. [5] developed a dynamic feedback controller for rotorcraft landing and hovering in ground effect based on a real-time potential-flow estimator, an early example of estimating GE state from onboard measurements rather than a fixed analytical correction, which motivates our use of learned rather than purely potential-flow-based estimation. Matus-Vargas et al. [6] documented how ground effect degrades single-sensor (ultrasonic/barometric) altitude estimates and proposed monocular SLAM fusion as a remedy; this supports our design choice of combining multiple onboard-observable signals (rangefinder, gyroscope, motor RPM/current) rather than relying on rangefinder altitude alone, although our results (Section 4.2) show that the rangefinder signal still dominates feature importance.

Shi et al. [3] and Shi et al. [4] demonstrated hybrid analytical-plus-learned-residual architectures for near-ground aerodynamic disturbance compensation, establishing the precedent for the trust-band selection structure (analytical vs. learned prediction, gated by a confidence check) used in our feedforward controller. O'Connell et al. [7] extended this line of work with meta-learned representations that adapt rapidly online to changing wind conditions; while our study uses offline-trained, static models rather than online adaptation, the rapid-adaptation motivation is directly relevant to our finding (Section 4.4) that a model's own control actions can shift its effective operating distribution, since online adaptation is one candidate mitigation strategy for closed-loop distribution shift of the kind we document. Santos et al. [8] proposed a closed-loop thrust estimator refining a simplified rotor RPM-to-thrust mapping, addressing a modeling simplification analogous to the one in our simulator's motor model; Gräfe et al.

[9] likewise emphasize systematically identifying motor/thrust model parameters from real hardware, which bears on the sim-to-real gap discussed in Section 6. Yang et al. [10] is, to our knowledge, the most directly related recent work, analyzing near-ground oscillation phenomena from real flight data; we treat their qualitative oscillation observations as an external point of comparison for the motion irregularity documented in Section 4. Zou et al. [11] took the opposite stance toward near-surface aerodynamic effects, exploiting the ceiling-effect analogue of GE to enable perching rather than treating it purely as a disturbance to reject; this suggests that, in regimes where our findings support reliable GE prediction (Section 4.3), the same sensor-driven estimate could in principle be used constructively rather than only as a correction term.

# 3 Methods

## 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.

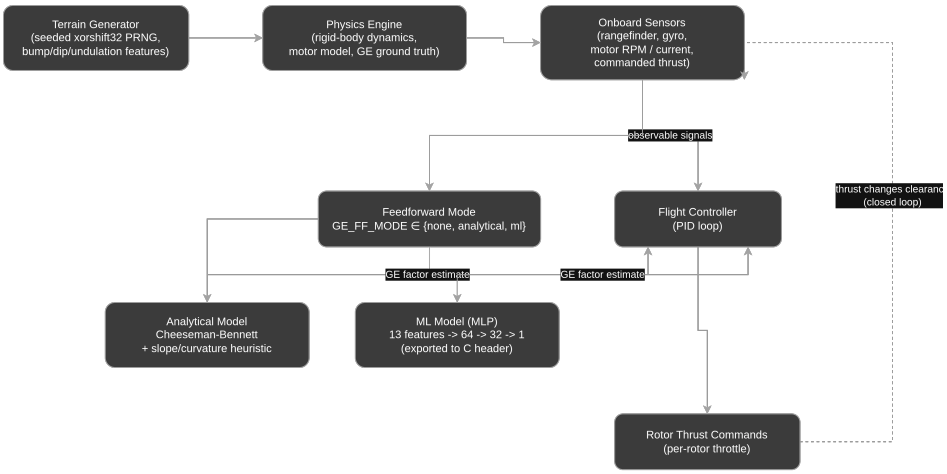


**Fig. 5** System architecture: terrain and physics generate onboard-observable sensor signals, which feed either the analytical or ML feedforward path selected by `GE_FF_MODE`; the resulting GE factor estimate corrects the flight controller's thrust command, which in turn changes clearance, closing the loop back to the sensors.

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

$$\text{factor}_{\text{flat}} = \frac{1}{1 - (R/4z)^2}, \tag{1}$$

where $R$ is rotor radius and $z$ is instantaneous rotor height above the local ground surface. Two heuristic corrections are then applied: a slope correction $\Delta = (\text{factor}_{\text{flat}} - 1)\cos(\theta_{\text{slope}})$, reflecting that steeper local slope disrupts the recirculating vortex ring that produces GE, and a curvature correction that amplifies the effect over concave terrain (which traps downwash) and attenuates it over convex terrain (which sheds it laterally), clamped to a factor range of $[0.5, 1.8]$. We emphasize that these slope/curvature corrections are heuristic rather than fitted to external validation data; the simulator's ground truth is itself an approximation, a limitation discussed further in Section 6.

**Terrain generation.** Terrain is generated per episode from a seeded xorshift32 pseudorandom number generator, placing 800–1499 non-overlapping bump/dip/undulation features along the flight track and superimposing high-frequency roughness, producing irregular rather than smoothly parameterized terrain. Each seed produces geometrically distinct terrain, so the number of seeds used directly controls the terrain diversity seen during training and evaluation.

**Sensor features.** Learned models use only signals available on a real onboard flight controller: three-axis gyroscope rate, rangefinder altitude (undefined beyond 3 m), per-rotor motor RPM and current (four rotors), and normalized commanded thrust. The prediction target is the ground-truth mean GE factor across the four rotors (Eq. 1 plus the slope/curvature corrections above); the analytical baseline used for comparison applies only the flat-ground Eq. 1 to rangefinder altitude, without the slope/curvature corrections, since those are not available to a controller that observes rangefinder altitude alone. Both are evaluated under the identical feature availability constraint.

**Actuator mapping.** Motor thrust is quadratic in rotor speed ($T \propto \omega^2$, Section 3.6 below), while commanded rotor speed is linear in the normalized throttle command. Consequently, canceling an expected GE thrust multiplier $f$ requires scaling the throttle command by $1/\sqrt{f}$, not by $1/f$; the results reported in Section 4 use the corrected $1/\sqrt{f}$ mapping throughout (see Section 3.6).

## 3.2 Offline model comparison and seed-wise data splitting

We compared six regression approaches — linear regression, decision tree, random forest, gradient boosting, and, using PyTorch, a multilayer perceptron (MLP), a long short-term memory network (LSTM), and a Transformer — against the analytical baseline on a 100 ms-ahead feedforward prediction task (i.e., predicting the GE factor 100 ms in the future from current sensor state). An initial evaluation used block-wise (10 s block) train/test splitting, which does not guarantee that samples from the same terrain seed are excluded from both splits and therefore does not test generalization to unseen terrain. To address this, we additionally constructed a deterministic, disjoint seed-wise split of the full 300-seed dataset into 240 training, 30 validation, and 30 test seeds, verified programmatically to share no seed across splits; the results reported in Section 4.1 use this seed-wise split. Sequence models (LSTM, Transformer) were trained with early stopping on the validation seeds.

### 3.3 Oracle controller and asymmetric clamp

To separate prediction-accuracy effects from actuator-level effects in closed-loop deployment, we added an oracle feedforward mode that uses the simulator's exact ground-truth GE factor rather than a learned estimate; this mode is diagnostic only and is not available in real deployment, where ground truth is unobservable. We also implemented an asymmetric clamp that limits how far the ML feedforward path's predicted GE factor may exceed the analytical estimate (i.e., it constrains overprediction specifically), with the clamp threshold selected using only the validation seeds and never the closed-loop evaluation or offline test seeds.

### 3.4 Forecast-horizon sweep and feature importance

To characterize how prediction advantage depends on required lookahead, we swept the forecast horizon from 0 (nowcast) to 500 ms and compared the analytical baseline against a random forest model at each horizon. We additionally computed both impurity-based and permutation-based feature importance for the random forest model to determine which sensor channels the model actually relies on.

### 3.5 Terrain-irregularity error decomposition

To test whether the ML advantage depends on terrain regularity, we recovered ground-truth local slope and curvature for every sample offline (these are not observable to the onboard controller and are used here only for post-hoc analysis) by replicating the terrain generator's finite-difference computation in a Python port bit-identical to the C physics engine, and binned the 100-episode dataset into slope and curvature quartiles. Within each bin we computed mean absolute error (MAE) for the ML model and the analytical baseline independently.

### 3.6 Closed-loop feedforward integration

The best offline single-forward-pass model, an MLP (13 input features $\rightarrow$ 64 $\rightarrow$ 32 $\rightarrow$ 1), was exported to a static-array C header and integrated into the flight controller (`firmware/ge_ff.c`) as a selectable feedforward mode (`GE_FF_MODE` $\in$ {`none`, `analytical`, `oracle`, `ml`, `ml_direct`, `ml_asym_clamp`}), evaluated identically to the analytical path. An MLP was chosen for the closed-loop deployment specifically because its purely feed-forward computation (no recurrent state, no tree traversal) is straightforward to port to C, and its offline accuracy was close to that of the random forest and LSTM models (Section 4.1).

In the simulator's physics, motor thrust is $T = K_T\omega^2$ (rotor speed squared), while commanded rotor speed is linear in the normalized throttle command. To cancel a predicted GE thrust multiplier $\hat{f}$, the throttle command is therefore scaled by $1/\sqrt{\hat{f}}$ rather than $1/\hat{f}$; an earlier version of our implementation used the latter, uncorrected mapping, which we identified and fixed during preparation of this manuscript (an initial $1/\hat{f}$ scaling silently over-corrects, since it cancels $f$ in the thrust-domain but is applied to a term that enters thrust quadratically). All results reported in Section 4

use the corrected $1/\sqrt{\hat{f}}$ mapping. `ml` is a hybrid mode: below a rangefinder clearance of 0.15 m it hard-cuts over to the analytical estimate (Section 4.7), and otherwise uses the ML prediction when it falls within a trust band around the analytical estimate; `ml_direct` uses the raw ML prediction without this cutover or trust band; `ml_asym_clamp` uses the hybrid cutover together with the asymmetric overprediction clamp (Section 3.3). We compared closed-loop gyroscope RMS and above-ground-level (AGL) tracking RMS error, defined as $\sqrt{\text{mean}((\text{rangefinder\_alt} - 0.22\,\text{m})^2)}$, across 30 terrain seeds disjoint from both the offline training and offline test seeds (60 s each) under each mode, together with 5th-percentile clearance, minimum clearance, and low-clearance safety-violation duration.

## 3.7 Causal bias intervention and loop-closure analysis

To test the causal direction of the GE-factor prediction error's effect on clearance directly, rather than relying on the sign of a correlation, we ran the oracle controller with a fixed, externally imposed signed bias added to the true GE factor ($\hat{f} = f_{\text{true}} + \delta$, $\delta \in \{-0.02, -0.01, 0, +0.01, +0.02\}$) across the same 30 evaluation seeds, and compared AGL RMS, 5th-percentile clearance, low-clearance duration, and minimum clearance between paired positive- and negative-bias conditions using paired $t$-tests.

To test whether an induced clearance perturbation compounds into a self-reinforcing feedback loop — as opposed to a one-shot effect that the system recovers from — we additionally injected a single time-limited bias pulse ($\delta = +0.02$ for 0.1 s) into the oracle controller, after which the oracle reverted to using exact ground truth, while a shadow (non-actuating) MLP prediction continued to be logged throughout. We measured whether the shadow prediction error grew in the seconds following the pulse (which would indicate the perturbation feeds back into larger prediction errors) and characterized the post-pulse clearance response (settling behavior, overshoot). We complemented this with a linearization of the closed-loop actuator–clearance dynamics across the simulator's operating range of GE factor (0.85–1.6), computing the closed-loop poles to determine analytically whether the linearized system is stable.

## 3.8 Robustness to ground-truth correction strength

Because the simulator's slope and curvature corrections to the ground-truth GE factor are heuristic (Section 3.1), we tested whether the central closed-loop finding (hybrid ML underperforming analytical) is sensitive to their strength. We defined a $3 \times 3$ grid of slope exponent $\in \{0.5, 1.0, 2.0\}$ and curvature gain $\in \{0.3, 0.6, 0.9\}$ values, fixed before running any evaluation, and re-ran the 30-seed closed-loop hybrid-vs.-analytical comparison under all nine configurations (1,620 total simulation runs), including the default configuration used elsewhere in this paper as one labeled condition among the nine.

# 4 Results

## 4.1 Offline model comparison

Under the seed-wise split (240 train / 30 validation / 30 test seeds, disjoint by construction), all learned models substantially outperformed the analytical baseline on the 30 held-out test seeds (RMSE = 0.010424, MAE = 0.002337, $R^2 = 0.851025$; linear regression was slightly worse still: RMSE = 0.010702, $R^2 = 0.842952$). Table 1 reports all models. The best-performing model, a Transformer, achieved RMSE = 0.000883 ($R^2 = 0.998321$); random forest achieved the best $R^2$ overall ($R^2 = 0.998679$) with the lowest MAE (0.000428). These seed-wise results are close to, and slightly better than, an earlier block-wise-split evaluation on the full 300-seed dataset that had suggested $R^2 \approx 0.997$ for the best model; the seed-wise split therefore rules out seed leakage as an explanation for the offline advantage documented here.

**Table 1** Seed-wise offline model comparison on 30 held-out test seeds (disjoint from the 240 training and 30 validation seeds).

| Model | RMSE | MAE | $R^2$ |
|---|---|---|---|
| Transformer | 0.000883 | 0.000495 | 0.998321 |
| LSTM | 0.000932 | 0.000553 | 0.998129 |
| Random Forest | 0.000981 | 0.000428 | 0.998679 |
| Gradient Boosting | 0.001109 | 0.000491 | 0.998314 |
| MLP | 0.001260 | 0.000641 | 0.997822 |
| Decision Tree | 0.001371 | 0.000521 | 0.997424 |
| Analytical (baseline) | 0.010424 | 0.002337 | 0.851025 |
| Linear Regression | 0.010702 | 0.006093 | 0.842952 |

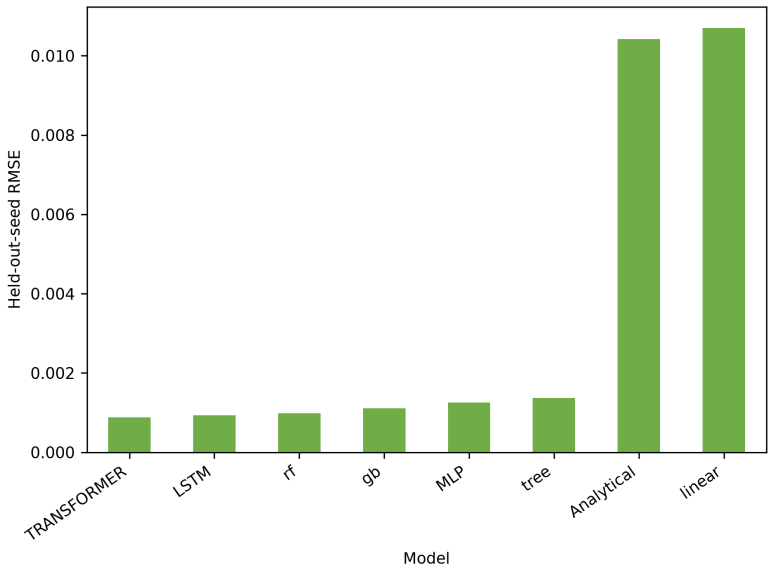


**Fig. 6** Held-out-seed RMSE for all evaluated models under the seed-wise (leakage-free) split, corresponding to Table 1.

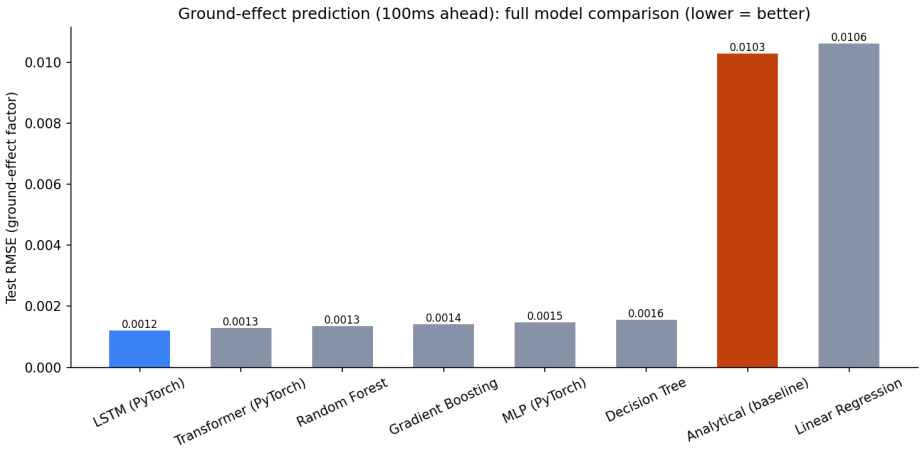


**Fig. 7** RMSE and $R^2$ comparison of the analytical baseline against all evaluated regression models, from an earlier block-wise-split evaluation on the full 300-seed dataset (100 ms feedforward prediction task); shown alongside Table 1 for comparison of the two splitting schemes.

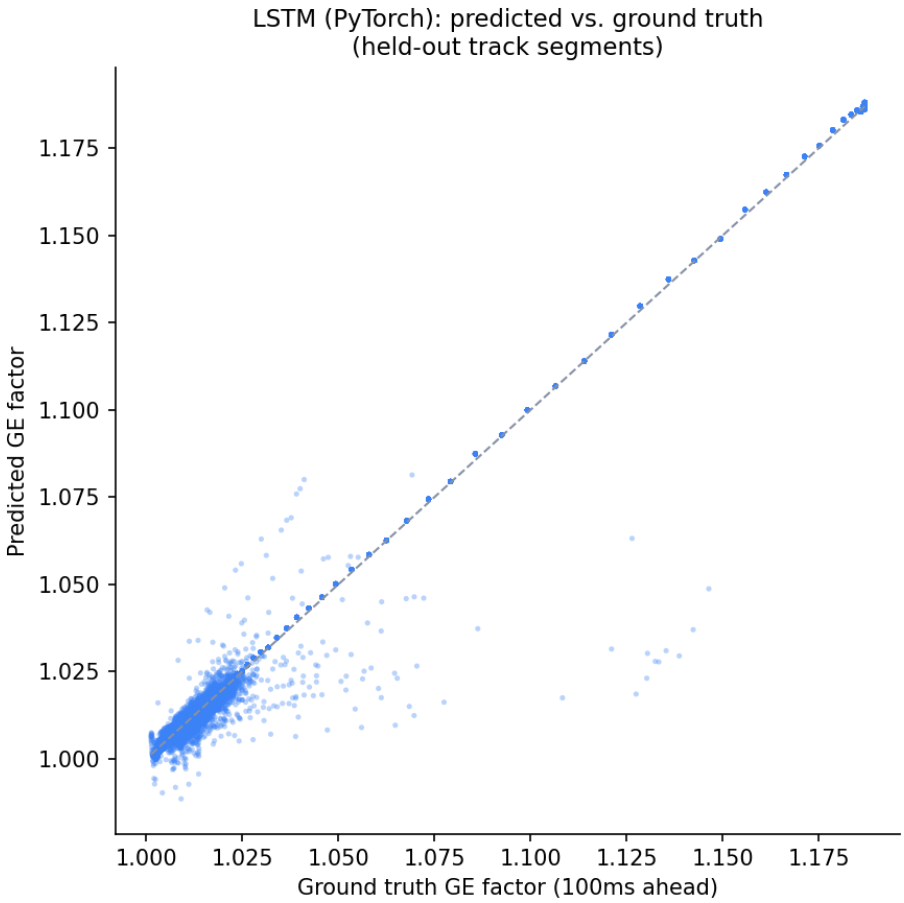


**Fig. 8** Predicted vs. ground-truth GE factor for the best-performing model on held-out test blocks from the block-wise-split evaluation.

## 4.2 Forecast horizon and feature reliance

Analytical baseline accuracy degrades approximately linearly with forecast horizon, from $R^2 = 0.883$ at 0 ms to $R^2 = 0.463$ at 500 ms, since the flat-ground formula has no mechanism for anticipating future clearance. The random forest model's accuracy remains largely flat across the same range ($R^2 = 0.999$ to 0.992), indicating that recent trends in gyroscope, motor current, and RPM carry extrapolable information about near-future clearance that the instantaneous analytical formula cannot use.

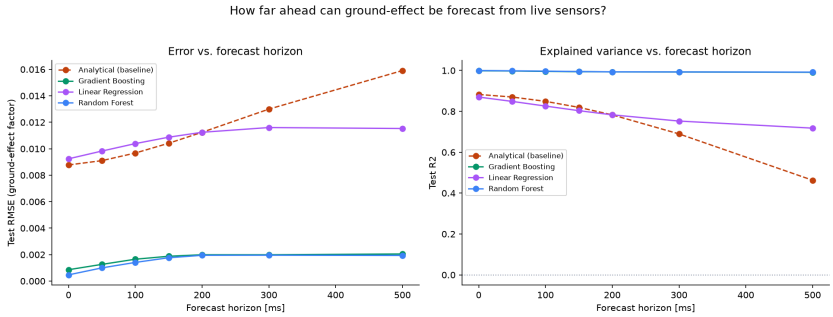


**Fig. 9** $R^2$ of the analytical baseline and random forest model as a function of forecast horizon (0–500 ms). The analytical model degrades approximately linearly with horizon while the learned model remains nearly flat.

Feature importance analysis showed that rangefinder altitude alone accounts for the overwhelming majority of predictive power (permutation importance 1.361, an order of magnitude above the next feature); we interpret the ML advantage primarily as learning a substantially more accurate nonlinear mapping from the same rangefinder signal (incorporating slope/curvature effects implicitly) rather than as exploiting an entirely different sensor combination.

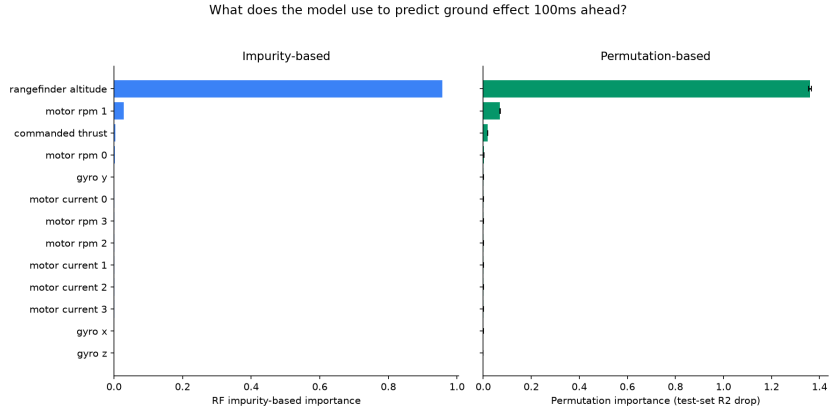


**Fig. 10** Impurity-based and permutation-based feature importance for the random forest model. Rangefinder altitude dominates both metrics by roughly an order of magnitude.

## 4.3 Terrain-irregularity error decomposition

Across both slope and curvature quartiles, the ML model's mean absolute error was lower than the analytical baseline's in every bin, with no exceptions. The relative improvement ranged from 15.8% to 49.1% depending on bin; more importantly, the *absolute* error gap widened with terrain irregularity rather than narrowing: analytical MAE increased from 0.0080 (near-flat terrain) to 0.0209 (highest-curvature quartile),

a 2.6× increase, while ML MAE increased more gently from 0.0040 to 0.0147 (3.6×, but from a substantially lower base, so the absolute ML error remained below the analytical error throughout). This is consistent with the underlying theory: the flat-ground assumption underlying the analytical model is violated more severely as terrain irregularity increases, and the learned model's advantage over that assumption grows correspondingly.

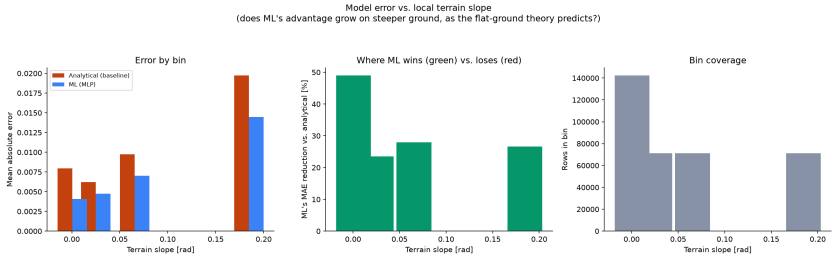


**Fig. 11** Mean absolute error of the ML model and analytical baseline, binned by local terrain slope quartile. ML error is lower than analytical in every bin.

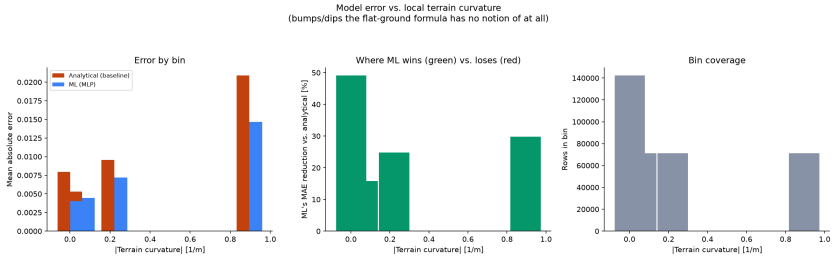


**Fig. 12** Mean absolute error of the ML model and analytical baseline, binned by local terrain curvature quartile. As in Figure 11, ML error is lower than analytical in every bin, and the absolute gap widens with increasing curvature.

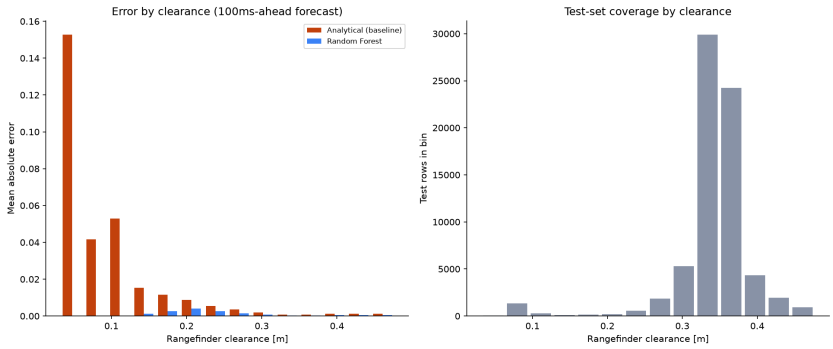


**Fig. 13** Test-set error of the ML model and analytical baseline as a function of rangefinder clearance, with test-set coverage per clearance bin.

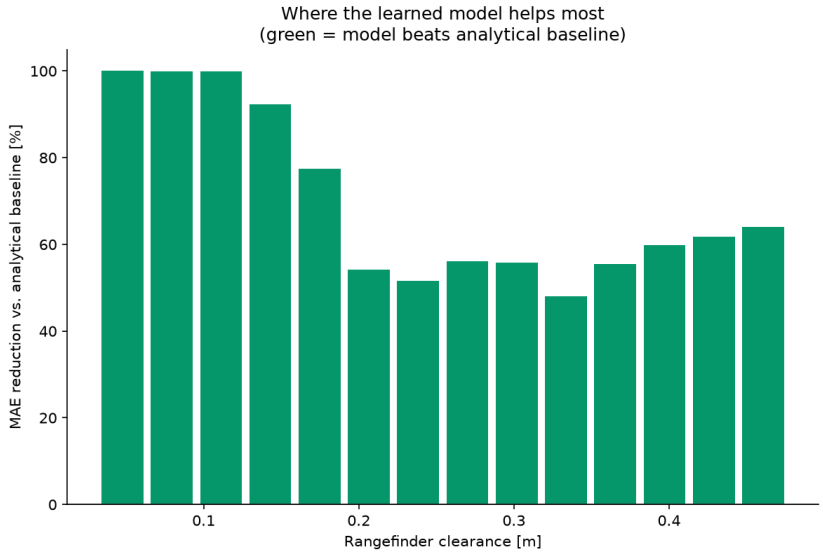


**Fig. 14** Percentage reduction in mean absolute error achieved by the ML model relative to the analytical baseline, as a function of clearance. The reduction is largest (90–100%) at the lowest clearances, where GE is physically most significant.

## 4.4 Closed-loop deployment: offline accuracy does not transfer

We first confirmed the closed-loop shortfall on the original 8-seed comparison used during development, then re-evaluated it on 30 terrain seeds disjoint from all offline training, validation, and test seeds. Under the corrected $1/\sqrt{\hat{f}}$ actuator mapping (Section 3.6), across the 8-seed comparison AGL tracking RMS error was $0.07004 \pm 0.01050$ m with no feedforward, $0.06322 \pm 0.00738$ m with the analytical baseline, and

$0.06438 \pm 0.00750$ m with the ML hybrid; gyroscope RMS was $0.02537 \pm 0.00412$, $0.02562 \pm 0.00384$, and $0.02600 \pm 0.00382$ rad/s respectively. The per-seed AGL RMS difference (ML minus analytical) was positive in all 8 seeds (mean +0.00117 m, paired 95% CI $[+0.00070, +0.00163]$ m, excluding zero).

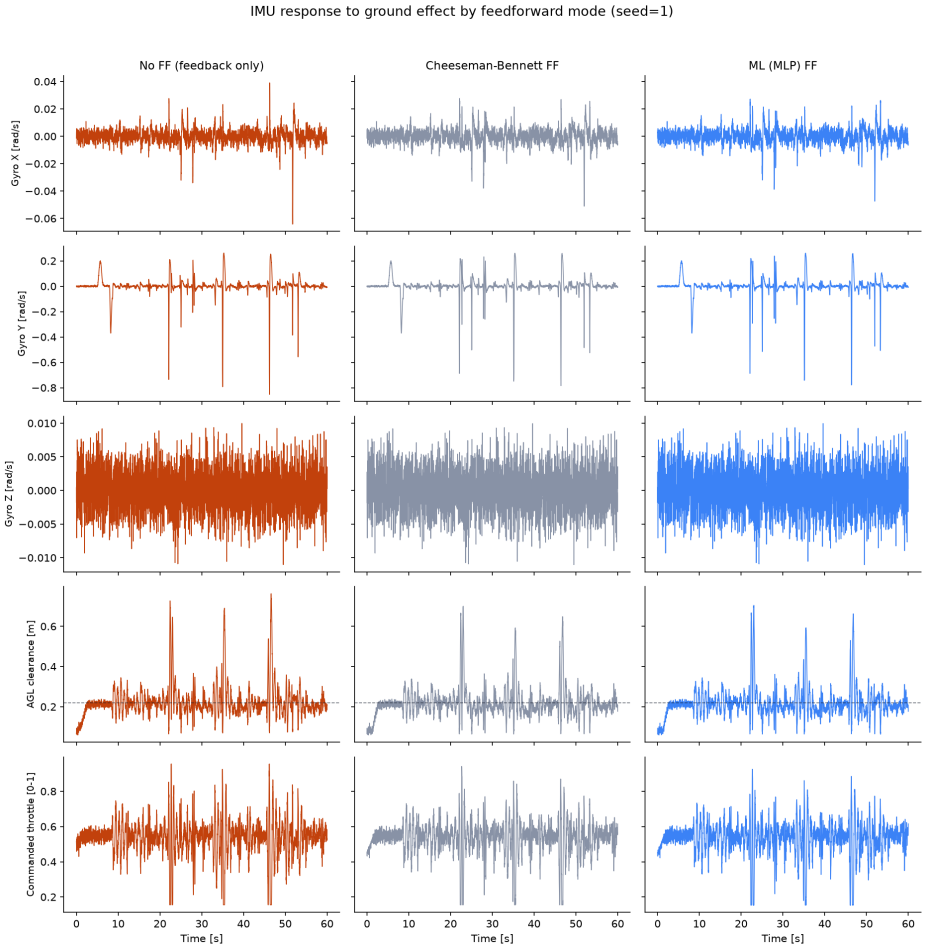


**Fig. 15** Gyroscope and AGL clearance traces across the three feedforward modes (none/analytical/ML) for a representative flight seed, under the corrected $1/\sqrt{\hat{f}}$ actuator mapping.

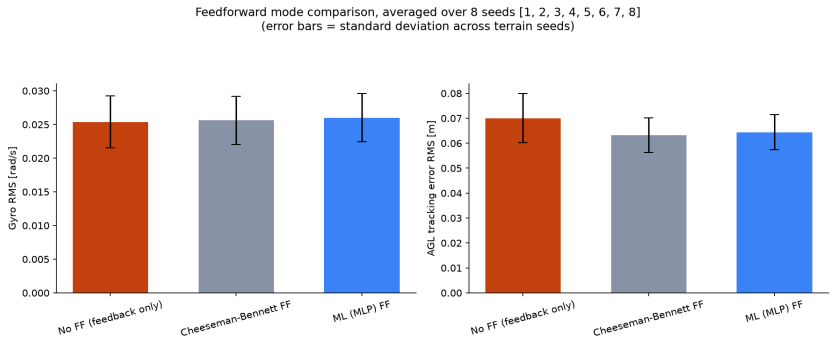


**Fig. 16** Mean±standard-deviation gyroscope RMS and AGL tracking RMS error across the 8-seed closed-loop comparison of the three feedforward modes, under the corrected actuator mapping.

On the independent 30-seed evaluation, we additionally included an oracle controller (given the simulator's exact ground-truth GE factor, not available in real deployment), the raw ML prediction without the hybrid cutover (`ml_direct`), and the asymmetric-clamp hybrid (`ml_asym_clamp`, Section 3.3). AGL tracking RMS followed the same qualitative pattern as the 8-seed comparison: no feedforward and `ml_direct` performed worst (both approximately 0.067–0.068 m), the analytical baseline, oracle, and `ml_asym_clamp` performed best and were closely grouped (approximately 0.060–0.061 m), and the hybrid `ml` mode fell in between (approximately 0.061–0.062 m). The result that is most informative for diagnosing the shortfall is that the *oracle* controller — given perfect knowledge of the ground-truth GE factor, with no prediction error whatsoever — performed essentially the same as the analytical baseline and did not outperform it. This indicates that the closed-loop shortfall of the learned hybrid relative to analytical is not primarily attributable to prediction error at all; even eliminating prediction error entirely does not produce a closed-loop advantage over the much simpler flat-ground analytical formula.

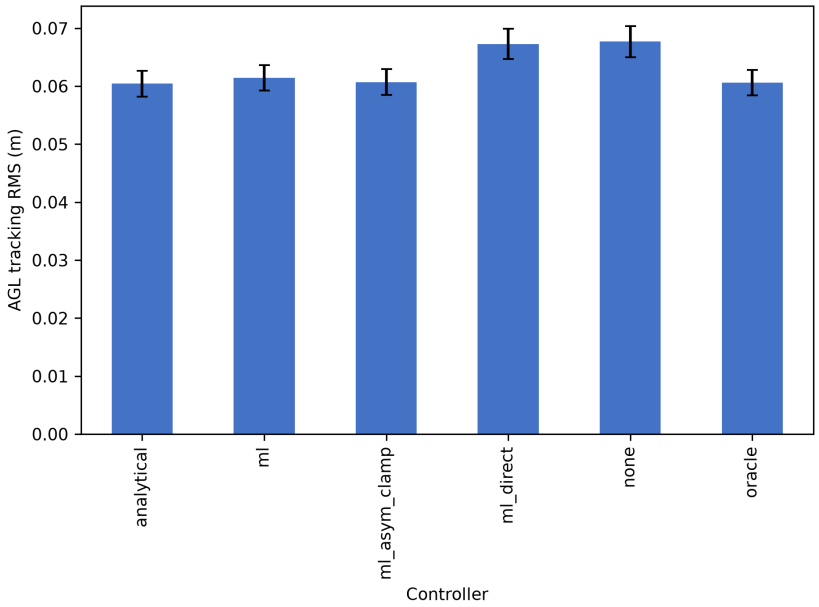


**Fig. 17** AGL tracking RMS across six controller configurations (no feedforward, analytical, ML hybrid, ML with asymmetric overprediction clamp, raw ML without cutover, and oracle ground-truth feedforward) on 30 terrain seeds disjoint from offline training/validation/test seeds. The oracle controller, despite zero prediction error, does not outperform the analytical baseline.

## 4.5 Causal bias intervention

Because the oracle result rules out prediction error as the primary driver, we tested the actuator-level mechanism directly with a controlled bias-injection experiment (Section 3.7) rather than relying on the sign of a correlation. Table 2 reports AGL tracking RMS across the 30 evaluation seeds for five fixed, externally imposed GE-factor biases.

**Table 2** Oracle-controller AGL tracking RMS under fixed, externally imposed GE-factor bias (30 seeds, 95% CI from a paired $t$-distribution).

| Bias $\delta$ | Mean AGL RMS (m) | 95% CI |
|---|---|---|
| $-0.02$ | 0.060758 | [0.056211, 0.065305] |
| $-0.01$ | 0.060662 | [0.056161, 0.065163] |
| 0 | 0.060578 | [0.056112, 0.065045] |
| $+0.01$ | 0.060497 | [0.056070, 0.064923] |
| $+0.02$ | 0.060407 | [0.056022, 0.064792] |

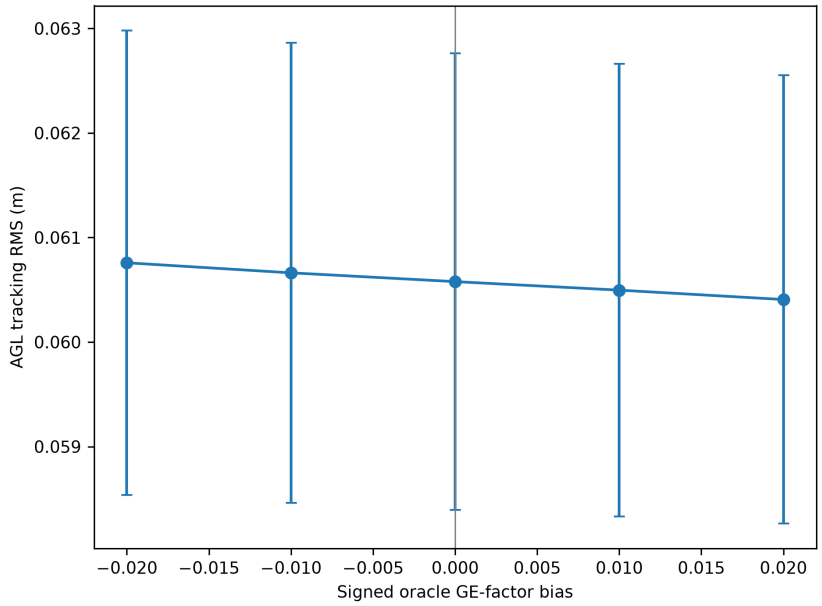


**Fig. 18** AGL tracking RMS as a function of externally imposed signed GE-factor bias (oracle controller, 30 seeds). Aggregate AGL RMS decreases slightly, not increases, with positive (overprediction) bias.

Contrary to what an aggregate-RMS-only reading of the mechanism would predict, positive (overprediction) bias produced slightly *lower* AGL RMS than negative bias, not higher: the paired difference $(+0.02) - (-0.02)$ was $-0.000351\,\mathrm{m}$ (95% CI $[-0.000636, -0.000067]$, paired $t$-test $p = 0.0172$), and $(+0.01) - (-0.01)$ was $-0.000166\,\mathrm{m}$ ($p = 0.0273$). Taken alone, this would appear to *contradict* the overprediction mechanism. However, aggregate AGL RMS averages over an entire 60 s trajectory and is dominated by ordinary tracking behavior away from the ground; it is not the metric that reflects safety-relevant behavior at low clearance, which is the operating regime the mechanism concerns. When we instead examine clearance-safety metrics directly, the picture reverses: positive bias significantly *degrades* safety margins relative to negative bias. For $(+0.02) - (-0.02)$: mean 5th-percentile clearance decreased by $0.002571\,\mathrm{m}$ ($p = 9.7 \times 10^{-7}$), low-clearance-duration increased by 0.326 s ($p = 4.5 \times 10^{-6}$), and mean AGL decreased by $0.001512\,\mathrm{m}$ ($p = 7.3 \times 10^{-7}$); absolute minimum clearance was not significantly affected ($p = 0.714$). The same pattern held at $(+0.01) - (-0.01)$ (5th-percentile clearance $p = 6.0 \times 10^{-6}$, low-clearance duration $p = 0.0003$). Because bias was imposed exogenously on top of a fixed oracle ground truth, this constitutes direct causal evidence, not merely a correlation: overprediction of the GE factor causes safety-relevant clearance degradation, even though it does not measurably worsen (and in fact slightly improves) whole-trajectory tracking RMS.

## 4.6 Loop-closure and stability analysis

The bias-intervention result establishes a causal, one-directional effect (overprediction → clearance loss) but does not by itself establish whether this effect compounds into a self-reinforcing feedback loop, in which the resulting clearance loss produces further prediction error and further clearance loss. We tested this directly with a single time-limited bias pulse ($\delta = +0.02$ for 0.1 s, oracle actuation reverting to exact ground truth immediately afterward) while continuously logging a non-actuating shadow ML prediction (Section 3.7). The pulse produced an immediate mean clearance decrease of 3.113 mm in the 0–0.5 s window following pulse termination, but the shadow prediction error did not grow significantly in this or any subsequent window (0–0.5 s: $p = 0.329$; later 0.5–10 s windows: all non-significant). Clearance exhibited a small rebound in the opposite direction approximately 1 s after the pulse before settling, consistent with an underdamped but stable return rather than divergence.

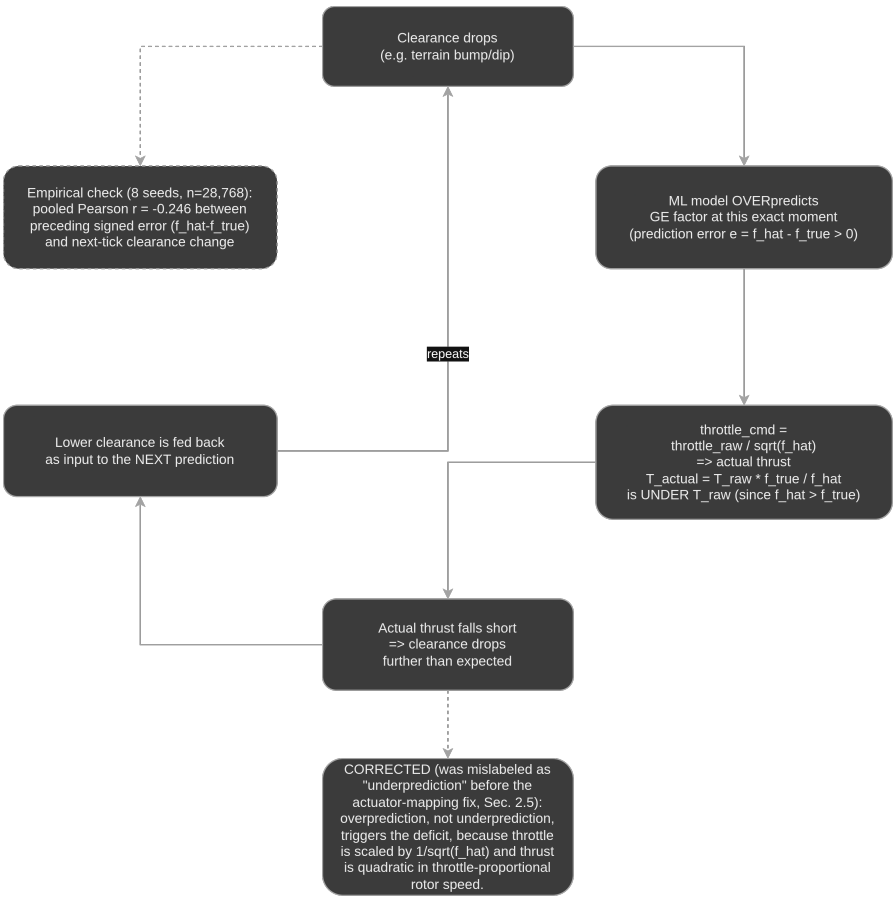


**Fig. 19** The causal mechanism verified in Sections 4.5–4.6: transient overprediction of the GE factor causes a measurable, statistically significant degradation in clearance-safety margins (bias intervention), but a pulsed perturbation does not measurably increase subsequent prediction error, and closed-loop pole analysis (below) confirms the system returns to baseline rather than diverging — i.e., the effect is causal but not self-reinforcing.

We complemented this with a linearization of the closed-loop actuator–clearance dynamics across the simulator's full operating range of ground-truth GE factor (0.85–1.6). All computed closed-loop poles had negative real parts throughout this range, and the dominant pole pair was complex, predicting a damped oscillatory (underdamped) return to equilibrium rather than growth. This is consistent with, and provides an independent theoretical explanation for, the pulse-response observation above. We conclude that the causal effect identified in Section 4.5 is real and safety-relevant, but that describing it as a "self-reinforcing feedback loop," as an earlier version of our analysis did, overstates the dynamics: the closed-loop system is stable, and the correct characterization is a causal, transient degradation of safety margin under overprediction, not divergence.

## 4.7 Mitigation

The `ml` mode evaluated in Section 4.4 includes a hard cutover to the analytical estimate whenever rangefinder clearance falls below 0.15 m (Section 3.6); this design was originally adopted, under the pre-correction actuator mapping, after an ablation in which narrowing the trust-band gate instead of hard-cutting over made performance worse, because rejected ML predictions fell back to no correction at all (factor = 1.0) rather than to the analytical estimate. On the 30-seed evaluation, the `ml_asym_clamp` configuration — the hybrid cutover combined with an asymmetric clamp on GE-factor overprediction specifically (Section 3.3), with the clamp threshold selected only on validation seeds — closed most of the residual gap between the plain hybrid and the analytical baseline (Figure 17), consistent with the causal mechanism identified in Section 4.5: since overprediction specifically, not underprediction, degrades safety margins, constraining overprediction while leaving underprediction unconstrained is a targeted intervention rather than a blunt one.

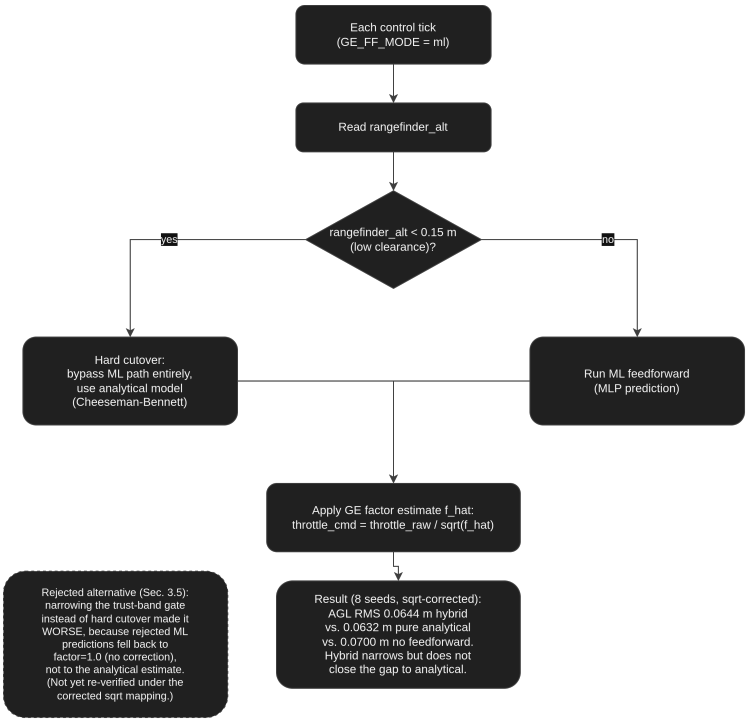


**Fig. 20** Per-tick decision logic of the hybrid feedforward controller evaluated in Section 4.4: a hard cutover to the analytical model at low clearance, rather than falling back to no correction, is the design retained across both the pre- and post-correction actuator mappings.

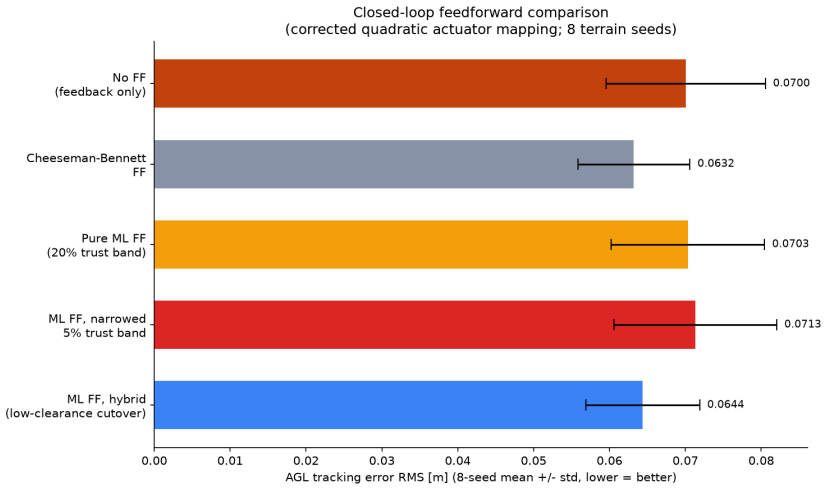


**Fig. 21** Summary of closed-loop feedforward configurations from the ablation that originally motivated the hard-cutover design: no feedforward, analytical, pure ML, ML with a narrowed trust-band (which underperformed), and the hard-cutover hybrid.

Restricting the comparison to the clearance range in which the ML path remains active (clearance $\geq 0.15\,\mathrm{m}$) gives AGL RMS of $0.0563\,\mathrm{m}$ for analytical and $0.0577\,\mathrm{m}$ for ML, versus $0.0658\,\mathrm{m}$ with no feedforward in the same restricted zone — both feedforward modes still help substantially in this region, but the ML hybrid still trails the analytical baseline by a small margin. This indicates that the offline advantage documented in Section 4.3 for exactly this terrain-irregularity regime does not convert into a closed-loop advantage even outside the low-clearance cutover region, consistent with the oracle result in Section 4.4.

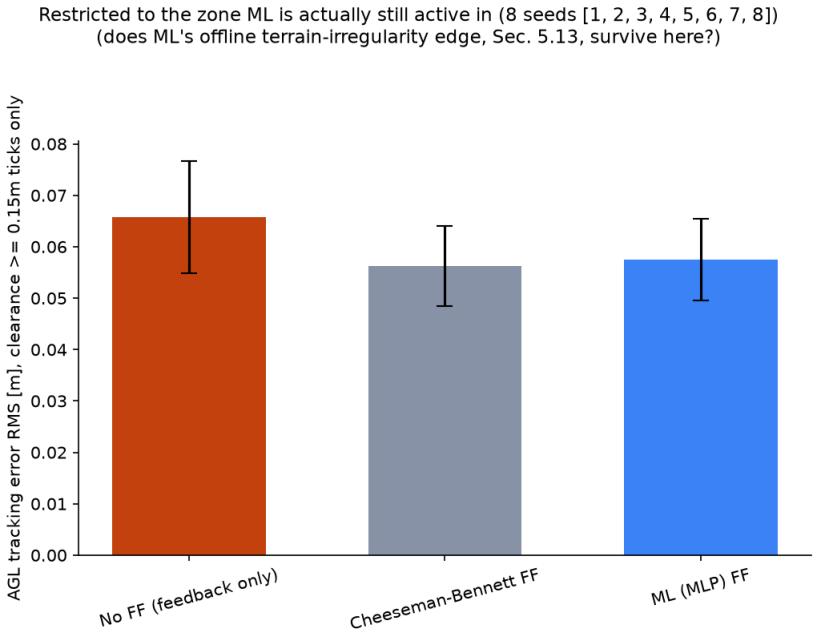


**Fig. 22** AGL tracking RMS error restricted to the clearance-active zone (clearance $\geq 0.15$ m) only, under the corrected actuator mapping, for the no-feedforward, analytical, and ML hybrid conditions.

## 4.8 Robustness to ground-truth correction strength

Across all nine slope-exponent $\times$ curvature-gain configurations tested (Section 3.8, 1,620 total closed-loop runs), the hybrid-minus-analytical AGL RMS gap was stable at approximately 0.0010 m in every configuration, including the default configuration used elsewhere in this paper (Figure 23). The central closed-loop finding — that the ML hybrid consistently trails the analytical baseline by a small, non-zero margin — is therefore not an artifact of the particular heuristic slope/curvature correction strengths used in the simulator's ground-truth model.

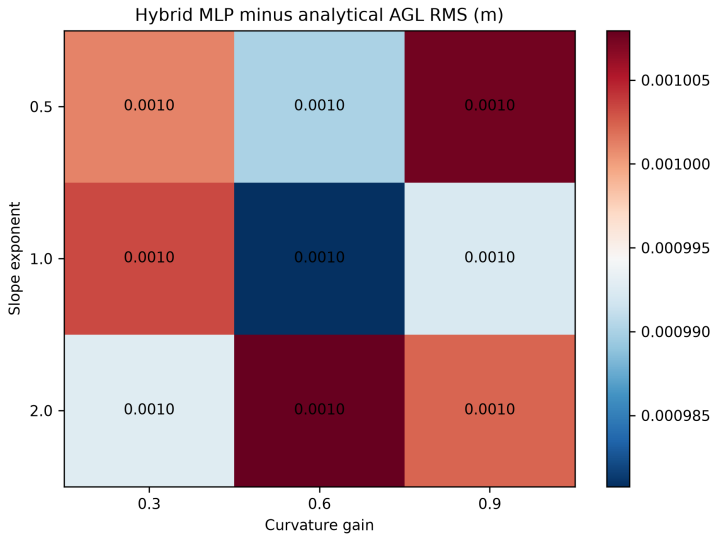


**Fig. 23** Hybrid-minus-analytical AGL tracking RMS across a $3 \times 3$ grid of slope exponent and curvature gain values used in the simulator's ground-truth GE model. The gap is stable at approximately 0.0010 m across all nine configurations.

# 5 Discussion

The central finding of this study is a dissociation between offline predictive accuracy and closed-loop control performance for a feedforward correction task, and the oracle result (Section 4.4) sharpens this considerably: even a controller with zero prediction error does not outperform the simple flat-ground analytical baseline in closed loop. This means the shortfall of the learned hybrid is not, in the main, a prediction-accuracy problem to be solved by a better model; it is a property of how any GE-factor estimate, however accurate, interacts with the actuator mapping and control loop. This is consistent with a broader concern in learning-based control: a model trained on trajectories generated under one controller (here, the analytical feedforward) and then deployed to generate its own control actions is operating under a form of distribution shift induced by its own policy, distinct from ordinary covariate shift in the input feature distribution.

The causal bias-intervention experiment (Section 4.5) demonstrates that overprediction of the GE factor causes a real, statistically significant degradation in safety-relevant clearance margins (5th-percentile clearance, low-clearance duration), even though its effect on whole-trajectory AGL tracking RMS is small and in the opposite direction — a reminder that a single aggregate metric can mask a safety-relevant effect concentrated in a small fraction of the trajectory. At the same time, the loop-closure pulse experiment and closed-loop pole analysis (Section 4.6) show that this causal effect does not compound into a self-reinforcing feedback loop: the closed-loop poles have negative real parts throughout the operating range, and the observed dynamics are an underdamped but stable return to baseline. We take this

as a methodological point in its own right, and one we arrived at only by testing our own prior conclusion rather than accepting it: a plausible-looking mechanistic story consistent with the sign of a correlation is not a substitute for direct causal intervention and dynamical (stability) analysis, and an earlier, less rigorously verified version of our own analysis had both the causal direction (underprediction vs. overprediction) and the qualitative dynamical character (self-reinforcing vs. stable) of this mechanism wrong.

Two practical implications follow. First, offline metrics (RMSE, $R^2$) should not be treated as a sufficient validation criterion for models intended for closed-loop deployment, and an oracle ablation — substituting the learned predictor with ground truth where a simulator makes this possible — is a cheap and informative diagnostic for separating prediction-accuracy effects from actuator- or control-structure effects before attributing a closed-loop shortfall to model quality. Second, a closed-loop shortfall does not necessarily require abandoning the learned model outright: because the causal mechanism here is specifically overprediction rather than underprediction, an asymmetric clamp that constrains overprediction while leaving underprediction unconstrained is a targeted intervention, and closed most of the residual gap to the analytical baseline (Section 4.7) without discarding the offline advantage documented in Sections 4.1–4.3.

# 6 Limitations and future work

An earlier iteration of the closed-loop firmware scaled the throttle command by $1/\hat{f}$ rather than $1/\sqrt{\hat{f}}$ to cancel a predicted GE thrust multiplier $\hat{f}$, which silently over-corrects because motor thrust is quadratic in the throttle-proportional rotor speed (Section 3.6); we identified and corrected this during preparation of this manuscript, and all closed-loop results in this paper use the corrected mapping. An earlier version of our analysis, based on this uncorrected mapping together with an unverified correlation-based mechanistic story, both misattributed the causal direction (underprediction rather than overprediction) and mischaracterized the dynamics as a self-reinforcing feedback loop; both errors were identified and corrected through, respectively, re-deriving the actuator equations directly and running the causal bias-intervention, pulse-response, and pole-analysis experiments reported in Sections 4.5–4.6. We report this course of the analysis for transparency rather than presenting only the final, corrected account.

The simulator's ground-truth GE model is itself a heuristic (Cheeseman–Bennett with slope/curvature corrections, Section 3.1); the finding that "ML predicts the simulator's ground truth better than the analytical approximation does" should be read as a statement about this specific heuristic ground truth rather than as a directly validated claim about real aerodynamics, absent computational fluid dynamics (CFD) or flight-test validation. The robustness analysis in Section 4.8 shows that the central closed-loop gap is insensitive to the strength of these heuristic corrections, which partially mitigates but does not eliminate this limitation, since all nine tested configurations still share the same functional form of correction. The flight dynamics used to generate the training dataset (Sections 4.1–4.3) and those used for closed-loop

evaluation (Section 4.4) differ in cruise flight-plan parameters, a known inconsistency introduced when a cruise-speed bug was fixed mid-study; a fully consistent regeneration of the training dataset under the corrected flight dynamics remains future work. The closed-loop feedforward correction is applied to collective thrust only, and cannot structurally correct per-rotor (attitude-channel) disturbances from asymmetric ground effect; extending the correction, and the asymmetric overprediction clamp, to per-rotor application is a natural next step. Finally, the loop-closure analysis in this paper used a single pulse amplitude and duration; a fuller characterization of the stability margin (e.g., the largest bias pulse or duration that remains stable) would further strengthen the safety case for deployment.

# 7 Conclusion

We evaluated machine-learned prediction of multirotor ground effect from onboard-observable sensors against a classical flat-ground analytical baseline, both offline and in closed-loop simulation, using a seed-wise data split that rules out terrain-seed leakage. Offline, learned models improve substantially on the analytical baseline ($R^2$ up to 0.998 vs. 0.851 on held-out terrain seeds). In closed-loop deployment on 30 independent unseen seeds, however, a deployed hybrid ML controller remained consistently, if modestly, worse than the pure analytical baseline, and — critically — an oracle controller with zero prediction error performed no better than analytical either, showing that the shortfall is not primarily a prediction-accuracy problem. A controlled causal intervention confirmed that overprediction of the GE factor significantly degrades safety-relevant clearance margins, but a time-limited pulse experiment and closed-loop pole analysis showed this effect is a stable, non-diverging response rather than a self-reinforcing feedback loop, correcting an earlier, less rigorously verified version of our own analysis that had both the causal direction and the qualitative dynamics wrong. An asymmetric clamp that specifically constrains overprediction closed most of the residual gap to the analytical baseline, and the central closed-loop finding was stable across nine tested strengths of the simulator's heuristic ground-truth corrections. We present these results as evidence that offline accuracy is not a sufficient validation criterion for control-relevant learned models, that oracle ablations are a valuable diagnostic for separating prediction-accuracy effects from actuator-level effects, and that a mechanistic explanation consistent with a correlation's sign requires direct causal and dynamical verification before being reported as established.

**Acknowledgements.** Acknowledgements are not compulsory. Where included they should be brief. Grant or contribution numbers may be acknowledged.

Please refer to Journal-level guidance for any specific requirements.

# Declarations

- **ORCID**: M.L., https://orcid.org/0009-0007-9707-6384
- **Funding**: Not applicable.
- **Conflict of interest/Competing interests**: The author declares no competing interests.

- **Ethics approval and consent to participate**: Not applicable. This study involved no human or animal subjects; the pilot observation described in Section 1 involved only recording flight of a consumer multirotor UAV.
- **Consent for publication**: Not applicable.
- **Data availability**: The code repository (https://github.com/lmwmason/GES) provides the simulator and scripts (`generate_dataset.sh`) required to regenerate the datasets used in this study from scratch. The generated CSV datasets themselves are not archived or version-controlled. The trained MLP weights actually deployed to the firmware are included in the repository as a C header (`firmware/ge_ff_ml_weights.h`).
- **Materials availability**: Not applicable.
- **Code availability**: The simulator, machine learning training scripts, firmware integration, and analysis code are available at https://github.com/lmwmason/GES.
- **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**: During the research process, the author used Claude Code (Anthropic) to assist with debugging the C physics engine, flight controller firmware, and Python training/analysis scripts.

# 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] 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).