SCALE FROM PRIORS, LOCATION FROM PIXELS: THE LIMITS OF MONOCULAR PITCH RECONSTRUCTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Professional baseball measures pitch trajectories with multi-camera tracking systems. Most games have only one broadcast camera behind the pitcher. We ask how much of a pitch's 3D trajectory this single view can recover, and where the accuracy of a working system comes from. We build a leakage-audited benchmark of 12,140 MLB pitches from OpenCommand, with solved camera poses and Statcast-derived reference trajectories. For the constant-acceleration trajectory models used in practice, metric depth has an exact scale ambiguity from a fixed camera. Gravity does not resolve it, because pitched baseballs are not gravity-only projectiles. For these models, absolute scale must come from prior information. Anchoring depth at the release and plate planes brings median 3D error from tens of meters to about one meter and plate-crossing error to 7 cm. We then measure what the observed ball track adds once calibration, temporal support, and priors are held fixed. A learned population prior with no pixel coordinates already beats the anchored physics fit in depth. In matched ablations tested once in each of the 31 MLB venues, adding the track cuts plate error from 30 to 5.4 cm and speed error from 3.6 to 1.9 mph. Depth error falls from 71 to 47 cm. Every venue improves on all three. These figures use the camera calibration supplied with the dataset. For the anchored physics fit, a reference-free audit with calibration and cohort built from inputs alone gives 8.7 cm median plate error. Priors set the scale, and the pixels set the location.

1 INTRODUCTION

Professional baseball measures pitch trajectories with multi-camera optical tracking systems such as Statcast and Hawk-Eye (Owens et al., 2003). Outside the top leagues, only a single broadcast camera behind the pitcher is available. This work asks how much 3D trajectory information that one view contains. Monocular trajectory reconstruction is a classical vision problem, and absolute scale is its known weak point (Ribnick et al., 2008; Chen et al., 2009; Hartley et al., 2003). The usual remedy is to fit projectile dynamics and let gravity supply the metric scale (Ribnick et al., 2008; Van Zandycke & De Vleeschouwer, 2022). Recent work applies physics-based and learned variants of this idea to sports balls (Liu & Wang, 2022; Ponglertnapakorn & Suwajanakorn, 2025; Grad et al., 2026). For real MLB pitches, we show that this remedy fails.

The difficulty is geometric. Scaling a trajectory about the camera center leaves its pinhole projection unchanged. Constant-acceleration trajectory families with zero or free acceleration are closed under this scaling. Their absolute depth is therefore not identifiable from pixels, at any frame rate and with any detector. Known gravity seems to fix this, but a pitched baseball is not a gravity-only projectile. Spin and drag change the acceleration in all three axes (Nathan, 2008; Adair, 2002). A gravity-only fit turns this mismatch into depth errors of tens of meters (Figure 1).

Metric scale must therefore come from prior information. We show that anchoring depth at the release and plate planes is enough. It brings median 3D error from tens of meters to about one meter with a median plate error of 7 cm. Once priors supply the scale, we ask what the pixels still add. We answer with matched ablations. All variants share the architecture, target, loss, and training schedule, and only the input set changes. Each of the 31 MLB venues is tested once by a model that never saw it. A learned population prior with no pixel coordinates already places depth to 71 cm, better than the anchored physics fit. The observed track then cuts plate error from 30 to 5.4 cm and

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

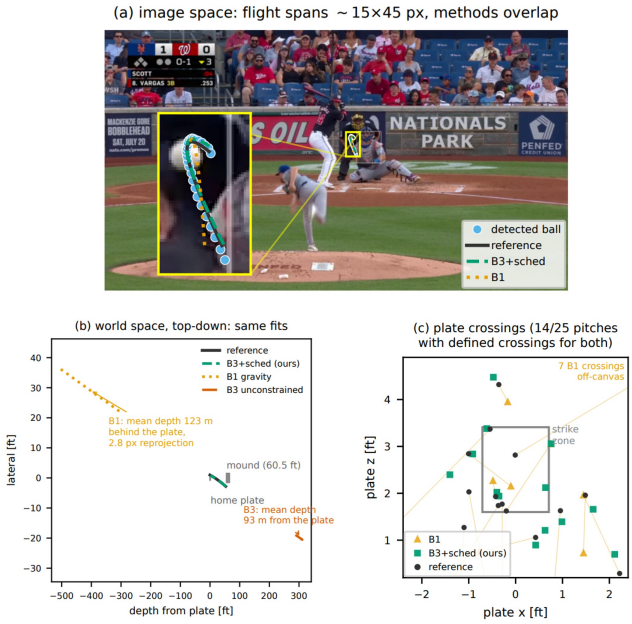


Figure 1: Monocular reconstruction from an MLB center-field broadcast. **(a)** Several reconstructions match the observed ball track in image space. **(b)** The same reconstructions land tens of meters from the plate, on opposite sides of it. **(c)** Plate crossings for a random sample of 25 benchmark pitches, shown for the 14 pitches where both estimators define a crossing. Trajectory priors bring the crossing to within centimeters. Frame source: MLB Baseball Savant. Broadcast image © the respective rights holder, reproduced for scholarly analysis.

speed error from 3.6 to 1.9 mph, but depth error only from 71 to 47 cm. The pixels determine where the ball crosses the plate. They refine depth only within the scale that the priors have set. A Fisher analysis shows why pixels alone cannot supply depth. Even with both endpoint depths known, the pixel likelihood leaves a mid-flight depth uncertainty of meters (§5).

Contributions.

1. A leakage-audited benchmark for monocular pitch reconstruction built on OpenCommand (Kim, 2026), with 12,140 pitches and audits of every coupling with the reference (§3).
2. An observability analysis that states when absolute scale is unidentifiable and why gravity does not recover it for pitched baseballs. A Fisher analysis bounds the mid-flight depth information that remains once scale is anchored (§5).
3. A controlled measurement of what the observed track adds once calibration, temporal support, and priors are fixed. The track improves plate location, speed, and depth in every one of 31 held-out venues, and the resulting estimator reaches about 5 cm plate error, 1.9 mph speed error, and 47 cm 3D error (§7).

All accuracy figures use the camera calibration supplied with the dataset and Statcast-derived reference trajectories (§3.1). A reference-free audit checks the plate accuracy without either (§6.2).

2 Related Work

Monocular 3D ball reconstruction. Trajectory triangulation recovers a moving point from one camera by restricting the motion to a parametric family (Avidan & Shashua, 2000; Kaminski & Teicher, 2004), and later work studies when trajectory priors make the problem well-posed (Park et al., 2010; Valmadre & Lucey, 2012). A fixed camera is the degenerate case of that analysis. Ribnick et al. (2008) analyze when projectile positions and velocities can be recovered from one

view. Gravity-only reconstruction has been applied to basketball and soccer (Chen et al., 2009; Kim et al., 1998). MonoTrack (Liu & Wang, 2022) reconstructs badminton shuttles with a drag-dominated model and court geometry. Van Zandycke & De Vleeschouwer (2022) use the known ball diameter, which is not usable here because a baseball spans about 4 px. Closest to our work, Grad et al. (2026) benchmark physics models on monocular soccer video and find that observation noise matters more than model expressiveness. Ponglertnapakorn & Suwajanakorn (2025) learn 3D trajectories from 2D tracks. Our learned baseline N0 follows the same direct regression route. The uniqueness conditions of Ribnick et al. (2008) assume dynamics of known magnitude that supply the scale. For trajectory families closed under camera-centered scaling no such conditions exist, and the known-gravity route is unreliable for pitched baseballs. Our contribution is to audit where the accuracy of a working system comes from. We separate calibration, temporal support, anchors, and learned population regularities with matched ablations, and we tie the result to a Fisher analysis. Ball detection and tracking (Kamble et al., 2019; Huang et al., 2019; Sun et al., 2020; Tarashima et al., 2023; Cui et al., 2023) are upstream of our problem, and the same missing-scale problem appears in monocular depth and SLAM (Eigen et al., 2014; Godard et al., 2019; Mur-Artal et al., 2015).

Physics-informed estimation and datasets. Embedding physics improves data efficiency when the physics is right (Raissi et al., 2019; Karniadakis et al., 2021; Greydanus et al., 2019). We contribute a quantified counterexample. A biased dynamics assumption ($a_z = -g$) is transferred into geometry, and a weaker model with correct geometric priors performs far better. Beyond broadcast understanding (Giancola et al., 2018) and detection assets (Drummey & Marcano, 2024), OpenCommand (Kim, 2026) released per-pitch detections, poses, and Statcast metadata for 2024–2026. We turn it into a monocular 3D benchmark with an explicit leakage protocol.

3 BENCHMARK

3.1 DATA AND TASK

We use OpenCommand (Kim, 2026), which provides broadcast-video ball tracks, solved camera poses, and Statcast metadata. World coordinates are plate-centered in feet, with x lateral, y toward the pitcher, and z upward.

- **Ball observations.** Each pitch contains per-frame ball-center detections from a YOLO detector on 1280×720 broadcast video at approximately 59.6 fps.
- **Camera pose.** Each clip includes a solved pinhole camera with position C, pan, tilt, roll, and focal length. The pose is estimated upstream from strike-zone geometry and detected ball runs. We treat calibration as a separate problem, so all reconstruction results are conditional on the supplied poses.
- **Reference trajectory.** Statcast metadata provides a 9-parameter constant-acceleration trajectory $(\mathbf{p}_0, \mathbf{v}_0, \mathbf{a})$ and the plate crossing. This is the evaluation reference. It is a Statcast-derived trajectory summary rather than raw Hawk-Eye tracking. B3 shares this trajectory form.
- **Reconstruction task.** For each pitch, the estimator receives only the detected pixels $\{\mathbf{u}_k\}$, relative timestamps $\{s_k - s_0\}$, and the supplied camera pose. It must recover the 3D trajectory over the observed window. No pitch-specific Statcast quantity reaches the estimator. Population-level constants used as priors are listed in §4.
- **Evaluation alignment.** The upstream data do not preserve the offset between video time and Statcast trajectory time. We estimate one continuous, sub-frame temporal offset per pitch by aligning the projected reference to the detections. This alignment is used only for evaluation. After alignment, the reference reprojects with a median error of 1.44 px and a p90 of 2.41 px. One pixel at the plate is about 1 cm. Appendix E bounds the depth error that the alignment itself induces.
- **Benchmark sample.** We sample 60 games across venues and retain 12,140 pitches after quality filtering. Each retained pitch has a valid pose, reference alignment below 3 px, and at least eight consecutive detections. These gates use the reference and remove 30% of

clips (Appendix J). The median pitch contains 23 observations over 0.44 s. The benchmark evaluation is a fixed random subset of 5,000 of these pitches, scored with the four metrics of §3.3, and Table 1 gives the numbers to beat. Learned entrants train on the 234 games of a larger pool that are not benchmark games, as the N0 and N1 rows of Table 1 do (Appendix L.4). The matched ablations of §7 are an analysis on the full 294-game pool, which contains the benchmark games.

The benchmark uses the 2024 season and covers 31 venues, the 30 home parks and London Stadium. The reference dependence of the supplied poses, alignment, and sample selection is audited in §3.2.

3.2 Leakage Protocol and Known Couplings

Estimators receive exactly three inputs: pixels, relative timestamps, and the camera pose. Every other quantity is classified in a written audit (cf. Kapoor & Narayanan, 2023) as camera information, field geometry, population-level prior, or evaluation-only. Three couplings with the reference remain. We disclose and quantify each.

1. **Camera poses.** OpenCommand solved each game's mount position using ball runs and Statcast trajectories. It solved per-clip pan, tilt, roll, and focal length from strike-zone boxes whose world size uses the Statcast strike-zone heights. Poses therefore inherit Statcast information at the game and clip level. We treat the pose as given calibration, identical across methods. We measure its influence by pose perturbation (Fig. 6) and by recalibration without Statcast input (§6.2).
2. **Run selection.** The benchmark picks the detection run that best matches the reference flight window. A reference-free heuristic picks an overlapping run on 98.2% of 1,500 audited pitches, where overlapping means that at least 80% of the detections coincide (Appendix N).
3. **Arrival prior.** The arrival window is a 2024 population statistic with a median gap of 0.00 s. It is never set per pitch.

3.3 Metrics and Statistics

All metrics compare against the reference at the observed frame times. When a fit uses a degraded subset of frames, we still evaluate on all frames. We report 3D and per-axis RMSE in cm, plate-crossing error at y=17/12 ft in cm, speed MAE at the first observation in mph, and reprojection RMSE in px. Plate error is conditional on a defined crossing. Anchored methods produce a crossing for 100% of pitches. Prior-free B0–B3 produce one for about 57%, and each table notes this. A fit with fewer than three observations or an optimizer error counts as a failure and is reported, not dropped.

Medians come with 95% confidence intervals from a game-clustered bootstrap, since pitches within a game share camera geometry. Method comparisons use paired per-pitch differences and win rates.

4 Estimators and Priors

R0: rays + depth schedule. R0 back-projects each detection ray onto a depth schedule linearly interpolated from the release plane ($y = 54.5$ ft) to the plate plane ($y = 17/12$ ft). A quadratic readout gives velocity and plate crossing. R0 uses no physics and no optimization, so its metric depth comes entirely from the schedule prior.

Ballistic models. B0–B3 fit $\mathbf{p}(\tau) = \mathbf{p}_0 + \mathbf{v}_0\tau + \frac{1}{2}\mathbf{a}\tau^2$ to the observed pixels using robust least squares, with

$$\text{B0} : \mathbf{a} = 0, \qquad \text{B1} : \mathbf{a} = (0, 0, -g), \qquad \text{B2} : \mathbf{a} = (0, 0, a_z), \qquad \text{B3} : \mathbf{a} \text{ free}.$$

B1-pop is B1 with the acceleration fixed at the population mean of the reference trajectories instead of gravity. B3 matches the reference trajectory family, so its failure reflects observability rather than model mismatch. B4 is a speed-dependent aerodynamic model, $\mathbf{a} = (0, 0, -g) - k_d|\mathbf{v}|\mathbf{v} + \mathbf{m}$, with the drag constant k_d fixed from known ball constants and C_d=0.35, and a free constant lift vector $\mathbf{m}$ (Appendix L).

Priors. We consider six independently switchable population-level priors: *release passage*, *plate geometry*, *mound direction*, *arrival window*, *aerodynamic bounds*, and a soft *depth schedule*. All constants are fixed from league geometry or population statistics and never use test-pitch information. Exact bounds are given in Appendix K. We denote B3 with the release, plate, direction, and arrival priors as B3+geo, and B3 with the depth schedule, direction, and aerodynamic priors as B3+sched.

5 Observability Analysis

Setting. The argument applies to a fixed camera center, ball-center observations without size or blur cues, and a trajectory model closed under scaling about that center. All three hold for the center-field broadcast and for B0, B2, and B3. Models whose acceleration depends on speed, such as B4 with its known drag term, are not closed under scaling. B4 tests that route below.

Exact scale ambiguity. Let C be the camera center and consider the scaling

$$\mathcal{H}_\lambda(\mathbf{p}(t)) = C + \lambda(\mathbf{p}(t) - C), \qquad \lambda > 0.$$

Every transformed point stays on the same camera ray, so its image projection is unchanged. If a trajectory family is closed under this transformation, its metric depth is unobservable from monocular pixels.

For constant-acceleration motion, the transformation becomes

$$(\mathbf{p}_0, \mathbf{v}_0, \mathbf{a}) \mapsto \big(C + \lambda(\mathbf{p}_0 - C), \lambda\mathbf{v}_0, \lambda\mathbf{a}\big).$$

B0, B2, and B3 remain in their model classes for any λ. Each therefore has an exact one-dimensional scale ambiguity, independent of noise, frame count, or frame rate (cf. Triggs et al., 1999).

Why gravity does not recover scale. B1 fixes $\mathbf{a} = (0, 0, -g)$, which removes the scaling freedom. This would recover metric scale if the ball moved under gravity alone. A pitched baseball does not. Drag slows it along its path, and spin produces a Magnus force with lateral and vertical components. For the common backspin fastball, the vertical acceleration is well below g in magnitude (Nathan, 2008; Adair, 2002). For other pitch types all three components differ from the gravity-only values in ways that depend on the spin axis. B1 fits the wrong acceleration in every axis, and the fit converts that mismatch into a systematic depth bias of tens of meters. The size and sign of the bias follow from the scaling map. To first order, the fit picks the scale $\lambda \approx g/|a_z|$ at which the true vertical acceleration matches g. A fastball with $|a_z|$ near $0.6g$ is placed about 1.6 times farther from a camera roughly 130 m away. That puts it roughly 80 m behind the plate, the same order as the 123 m in Figure 1b. A pitch with $|a_z|$ above g is placed in front of the plate, which is why B1 scatters to both sides of the field in Appendix Figure 4. Our experiments establish that gravity-only dynamics are misspecified for this purpose. They do not separate the contributions of lift, drag, and the lateral component.

Replacing gravity with a better constant does not help. B1-pop fixes the acceleration at the population mean of the reference trajectories, $(-2.5, 27.2, -23.9)$ ft/s^2, and still misses depth by 36 m (Table 1). The reason is a lever arm. A fixed acceleration sets the scale through the ratio of the true to the assumed acceleration. Across pitches that ratio varies by about a third, and the camera sits about 130 m from the ball. A third of 130 m is about 40 m. Acceleration priors act on depth through this lever arm. Position anchors at the release and plate planes act on depth directly, and succeed where acceleration priors fail.

Known drag does not help either. Under the scaling map the drag term of B4 scales as λ^2 while everything else scales as λ, so B4 is not closed under scaling. In practice the difference is too small to use. Over a 0.4 s flight the drag deceleration changes by only a few ft/s^2, so the drag mismatch at a wrong scale is nearly constant and the free lift vector absorbs it. Imposing depth offsets of ± 10 m at the first observation changes the best attainable pixel RMSE of B4 by less than 0.1 px for every one of 300 pitches, and B4 misses depth by 81 m (Table 1). Known drag with free lift is scale-ambiguous in practice.

Geometric and aerodynamic priors also break the scaling symmetry, by adding information that the pixel likelihood does not contain. Bounds on the acceleration or on the release and plate boxes only

Table 1: Benchmark results at full observation density (5,000 pitches, medians, supplied calibration). The learned models N0 (direct regression) and N1 (residual on top of B3+sched) are trained on the 234 pool games that are not benchmark games (§7, Appendix L.4). All estimators with game-clustered 95% CIs are in Appendix Table 3. B3 has a free scale, so its depth error is where the optimizer stopped along a continuum of equal image fits, and B4 behaves the same way (§5). B1 and B1-pop fix the acceleration at gravity and at the population mean, and their errors reflect model mismatch. Plate error is conditional on a defined crossing (100% for R0 and B3+sched, 57–58% for B1 and B3, 75% for B1-pop, 28% for B4).

Method	3D [cm]	depth [cm]	lateral [cm]	vertical [cm]	plate [cm]	speed [mph]	reproj [px]
R0 ray+schedule	96.2	95.7	4.2	7.4	6.6	8.3	0.5
B1 gravity	3846.6	3831.5	178.9	280.7	60.5	136.4	0.7
B1-pop mean accel.	3585.3	3567.1	175.1	250.4	215.7	93.4	0.8
B3 const. accel.	5189.9	5166.9	210.9	353.8	114.4	61.6	0.4
B4 drag + free lift	8168.7	8124.8	350.9	555.8	72.8	71.3	0.4
B3+geo	199.1	198.2	9.7	14.3	16.9	27.5	0.5
B3+sched	90.3	89.8	4.0	6.9	6.6	7.8	0.5
N0 learned direct	47.2	47.0	2.8	3.6	4.4	1.9	1.8
N1 learned residual	**45.6**	**45.3**	**2.7**	**3.4**	**4.1**	**1.9**	1.7

restrict the feasible scale to an interval. Tight anchors such as the depth schedule select a single scale. Figure 2b shows the ambiguity directly. Depth changes of ± 10 m leave the best attainable image fit nearly unchanged.

How much depth information remains. We quantify the residual information with a Fisher analysis (Kanatani, 2005). Each trajectory is represented by its 3D positions at three time knots, $\mathbf{P}_0$, $\mathbf{P}_{T/2}$, and $\mathbf{P}_T$, a reparameterization of the nine constant-acceleration parameters. We linearize the pixel observation model at the reference trajectory with central differences, whiten by $\sigma{=}1$ px, and express all nine coordinates in feet. Across 3,000 pitches, the Jacobian is numerically rank 8. Its single null direction matches the analytic scaling gauge and lies almost entirely in depth.

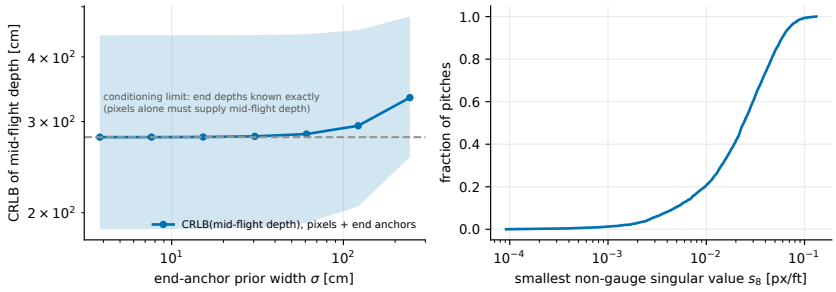
Even when both endpoint depths are known exactly, the mid-flight depth remains poorly determined. We condition on the endpoint depths by removing their two columns from the Jacobian and read the mid-knot depth entry of the inverse Fisher matrix. Writing $\mathrm{CRLB}(y_{T/2})$ for this Cramér–Rao variance bound, $\sqrt{\mathrm{CRLB}(y_{T/2})}$ has a median of 2.8 m across pitches at 1 px observation noise, with an IQR of 1.9–4.4 m (Fig. 2a). After the global scale is anchored, the pixels alone still carry only weak information about mid-flight depth.

The bound has an empirical counterpart. B3 with the depth schedule alone, the two anchors and nothing else, leaves the mid-flight depth to the pixels, and its depth RMSE is 1.9 m (Appendix Table 6). A mid-knot uncertainty of 2.8 m spread over the flight as the null-mode profile gives an RMSE of about 2 m, consistent with the fit. Every working estimator adds information the bound does not include. R0 interpolates depth between two anchor planes. B3+sched adds bounds on the acceleration, and the aerodynamic bound on a_y alone limits the mid-flight depth deviation from a straight schedule to about 30 cm. The learned models of §7 add population regularities. The bound therefore does not cap their depth error. It says what the pixel likelihood contributes on its own. It determines the image-plane quantities, lateral position, height, and the plate crossing, and it refines depth only within a scale that a prior has set. §7 measures this pattern empirically.

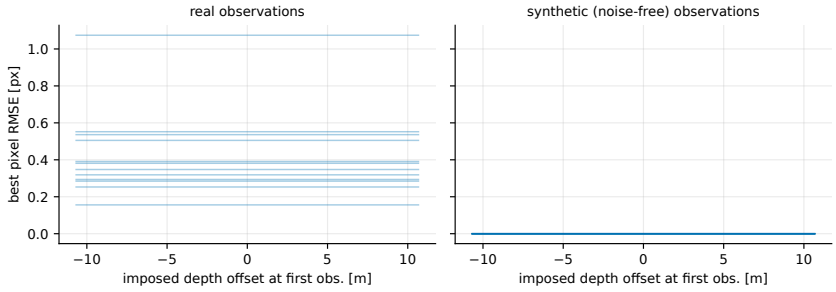
6 Experiments

6.1 Main Comparison

Table 1 compares representative estimators. The full set, prior ablations, and confidence intervals are in Appendix H.



(a) Mid-knot depth CRLB and s_8 CDF.



(b) Depth profile likelihood.

Figure 2: Observability of monocular pitch reconstruction. **(a)** Left, the mid-flight depth bound stays at meters even when both endpoint depths are known. Right, the smallest non-gauge singular value is nonzero for every pitch, so the only null direction is the analytic scaling gauge. **(b)** Best attainable pixel RMSE against an imposed depth offset at the first observation, one curve per pitch, for real (left) and synthetic (right) observations. Pixel error stays nearly unchanged across ± 10 m, so image fit alone does not determine metric depth.

- **Prior-free fits fail.** B3 reaches sub-pixel reprojection error while missing depth by 52 m. Reprojection error is therefore a poor indicator of 3D accuracy. The 52 m figure is where the optimizer stopped along the continuum of equally good image fits in Figure 2b.
- **Fixed accelerations do not help.** B1 has no scale freedom, yet it misses depth by 38 m. B1-pop, with the population-mean acceleration in place of gravity, misses by 36 m. Both are the lever-arm failure of §5.
- **Geometric bounds narrow the error to 2 m.** B3+geo restricts the feasible scale to an interval but does not select a point in it.
- **Depth anchors bring the error to about 1 m.** R0 interpolates depth between the release and plate planes with no physics and reaches 96 cm. B3+sched adds the same two anchors as soft residuals, together with the direction and aerodynamic bounds, and reaches 90 cm. Both have lateral and vertical errors below 8 cm and plate-crossing errors of 6.6 cm.
- **Learned priors halve the remaining error.** Trained on games disjoint from the benchmark, N0 reaches 47 cm and N1 reaches 46 cm, with 4 cm plate error and 1.9 mph speed error (§7).

The soft anchors need the bounds. With the two anchors alone, the pixels must supply the mid-flight depth, and the error stays at 1.9 m, the value the Fisher bound predicts (§5). The anchors also assume that the detection run starts near release and ends near the plate. The failure cases in Appendix I are runs that violate it. The anchored fits still have speed errors near 8 mph. Speed is mostly depth velocity, and the schedule fixes depth only to within its anchor widths. The learned estimators of §7 bring speed error to 1.9 mph. The ballistic model earns its place under missing observations. With 75% of detections removed, R0 degrades from 96 to 299 cm while B3+sched goes from 90 to 160 cm (Appendix B).

Table 2: What the observed track adds once calibration, temporal support, and priors are fixed. All variants share architecture, target, loss, and training schedule. Each of the 31 venues is tested once by a model that never saw it (5-fold venue cross-validation, two seeds averaged per pitch, supplied calibration). Top: medians with game-clustered 95% CIs over all 58,188 test pitches, with the anchored physics fit B3+sched on the same pitches. Bottom: paired per-pitch difference of the full input against each reduced input, summarized per venue, with median and IQR over the 31 venues and the number of venues where full is better (negative means full is better). meta+cam: track duration, observation count, and camera pose, no pixel coordinates. +pix: pixel track with its release-plane and plate-plane back-projections. +P: B3+sched estimate.

Input	3D RMSE [cm]	depth RMSE [cm]	plate error [cm]	speed MAE [mph]
B3+sched (physics fit)	93.3 [89.4, 97.2]	92.8 [88.9, 96.7]	6.6 [6.4, 6.9]	8.1 [7.8, 8.4]
meta+cam	80.5 [78.1, 83.1]	70.6 [67.9, 73.6]	30.2 [30.0, 30.5]	3.63 [3.55, 3.70]
meta+cam+pix	50.7 [49.3, 52.4]	50.3 [48.9, 51.9]	5.0 [4.9, 5.2]	2.03 [1.97, 2.08]
meta+cam+P	46.6 [45.5, 48.0]	46.2 [45.0, 47.5]	4.8 [4.6, 4.9]	1.88 [1.84, 1.92]
full (meta+cam+pix+P)	47.1 [45.6, 48.4]	46.6 [45.2, 48.0]	5.4 [5.2, 5.5]	1.91 [1.87, 1.95]

full minus	metric	venue median	IQR over venues	venues improved
meta+cam	depth RMSE [cm]	-17.1	$[-24.3, -11.7]$	31/31
meta+cam	plate error [cm]	-23.5	$[-24.8, -22.0]$	31/31
meta+cam	speed MAE [mph]	-1.57	$[-1.80, -1.33]$	31/31
meta+cam+pix	depth RMSE [cm]	-2.1	$[-3.1, -1.0]$	27/31
meta+cam+pix	plate error [cm]	$+0.2$	$[-0.1, +0.5]$	9/31
meta+cam+pix	speed MAE [mph]	-0.10	$[-0.14, -0.02]$	25/31
meta+cam+P	depth RMSE [cm]	$+0.6$	$[-0.2, +1.7]$	10/31
meta+cam+P	plate error [cm]	$+0.5$	$[+0.3, +0.7]$	1/31
meta+cam+P	speed MAE [mph]	$+0.04$	$[+0.01, +0.07]$	6/31

6.2 Calibration Audits

Perturbing the supplied poses by small random amounts changes B3+sched by less than 2% in 3D error (Appendix C). This does not test a bias shared by all Statcast-assisted poses, so we audit that directly.

Removing per-clip Statcast scale. We replace the pitch-specific Statcast zone heights with fixed league-average heights when solving each clip. On pitches that survive both calibration pipelines, the change is below 1 cm in 3D and plate error.

Reference-free cohort audit. We further remove the game-level reference-assisted calibration and build the evaluation cohort from input information alone, using four fixed population constants (Appendix L.5). This retains 91.6% of eligible clips. On this ungated cohort, B3+sched has a median plate-crossing error of 8.7 cm, against 6.6 cm on the benchmark cohort. The p90 is 35 cm. The tail comes from clips with poor reference-free calibration that reference-based filtering would have removed. Typical plate accuracy therefore survives without reference-assisted calibration or selection. Speed MAE is 8.2 mph, against 7.8 mph on the benchmark. Re-estimating the four constants from the 2025 season changes the median plate error by 0.1 cm (Appendix M). The audit scores plate crossing and speed under this fixed calibration. The 3D errors elsewhere in the paper use the supplied calibration and the benchmark cohort.

7 What the Track Adds: Matched Ablations

Design. We train small MLPs on 58,188 pitches from 294 games. At test time they use only monocular inputs: the resampled image track, its back-projections onto the release and plate planes, the track duration and observation count, and the camera pose (Appendix L.3). Every variant in Table 2 shares the architecture, the target (the nine reference trajectory parameters at the first observation), the loss, and the training schedule. Only the input set changes. The reduced input meta+cam

contains the track duration, the observation count, and the camera pose, and no pixel coordinates. It encodes calibration, temporal support, and, through the learned prior, scale. The two track routes are the pixel track with its back-projections onto the release and plate planes (+pix) and the B3+sched estimate computed from the same pixels (+P). In Table 1 and the appendix, N0 denotes the direct regressor from meta+cam+pix and N1 the residual model on top of B3+sched. Evaluation is a 5-fold venue cross-validation, so every venue is tested once by a model that never saw a pitch from it. All numbers use the supplied calibration (§3.2).

Result. Table 2 reads as a ladder. A training-set mean with no input has 150 cm depth error (Appendix Table 4). With no pixel coordinates, meta+cam places depth to 71 cm, better than B3+sched, the anchored physics fit that reads the ball track, at 93 cm on the same pitches. This is the strongest evidence that scale comes from priors. Duration is informative because the detection run starts at a variable distance from release (Appendix D). Adding the track improves every metric in all 31 venues, and the improvements are uneven.

- **Plate location.** Error falls from 30.2 to 5.4 cm. The venue-median paired improvement is 23.5 cm.
- **Speed at the first observation.** Error falls from 3.63 to 1.91 mph. The venue-median paired improvement is 1.6 mph.
- **Depth.** Error falls from 70.6 to 46.6 cm. The venue-median paired improvement is 17.1 cm.

The pixels remove most of the plate error, about half of the speed error, and a third of the depth error. The full model reaches 47 cm median 3D error, half the 93 cm of B3+sched on the same pitches. P alone matches or slightly beats the full input on every metric, since P is computed from the same pixels, and we quote the full input as the conservative choice. The small plate degradation from adding raw pixels on top of P is systematic, and we do not tune the shared schedule per variant. The priors supply the absolute scale, and the track then adds consistent information about location, speed, and depth within that scale.

Interpretation. Three points follow from the design. First, the 30 cm plate error of meta+cam is the error of a population guess, the typical crossing for that venue and duration. The track removes most of that error, so the plate crossing is a quantity the pixels measure. Second, the depth reduction is smaller but present in every venue, so the track refines depth within a scale the prior has set. Third, the comparison is matched and tested per venue, so the improvement comes from the pixel track and not from capacity or a venue-specific fit. Shuffling the pixel features within venue and duration bins removes the entire gain (Appendix Table 4).

Scope of the learned prior. Held-out venue performance is generalization within the MLB population, not extrapolation. When N1 is trained only on pitches at or below 92 mph, it underestimates held-out pitches at or above 96 mph (Appendix F). Speed is therefore a prior refined by the track, like depth, and unlike the plate crossing, which the pixels measure directly. The learned prior also needs reference-labeled training pitches from the same population, which the physics fits do not.

8 Conclusion and Limitations

For constant-acceleration models observed from a fixed camera, absolute depth is not observable from the pixel likelihood, and fixed accelerations do not resolve the ambiguity for pitched baseballs. Absolute scale must come from prior information. Anchoring depth at the release and plate planes brings 3D error to about one meter, and a learned population prior with no pixel coordinates does better still. Once that scale is supplied, the track determines the plate crossing, halves the speed error, and refines depth by a third in all 31 held-out venues. Priors set the scale, and the pixels set the location.

Limitations. (i) The reference shares the constant-acceleration form of B3, so absolute errors are relative to that reference. (ii) The supplied calibration has Statcast provenance. The reference-free audit covers the physics fit's plate crossing and speed, so all 3D errors and learned-model results are conditional on that calibration. (iii) The arrival and schedule priors encode this detector and broadcast pairing, and we study one camera family.

REPRODUCIBILITY STATEMENT

All experiments run from fixed-seed configurations and write manifests, splits, per-pitch results, and the leakage audit to disk. Appendix L lists optimizer settings, prior weights, network architecture, training schedule, splits, and the reference-free calibration as implemented. Code and per-pitch results will be released with the camera-ready version. The dataset is public (CC BY-NC-SA 4.0).

AI USE STATEMENT

Large language models assisted with experiment code, manuscript editing, and internal review. The authors designed the experiments and verified all results against the code and per-pitch results.

REFERENCES

Robert K Adair. *The physics of baseball*. HarperCollins, New York, 2002.

Shai Avidan and Amnon Shashua. Trajectory triangulation: 3d reconstruction of moving points from a monocular image sequence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(4):348–357, 2000.

Hua-Tsung Chen, Ming-Chun Tien, Yi-Wen Chen, Wen-Jiin Tsai, and Suh-Yin Lee. Physics-based ball tracking and 3d trajectory reconstruction with applications to shooting location estimation in basketball video. *Journal of Visual Communication and Image Representation*, 20(3):204–216, 2009.

Yutao Cui, Chenkai Zeng, Xiaoyu Zhao, Yichun Yang, Gangshan Wu, and Limin Wang. Sportsmot: A large multi-object tracking dataset in multiple sports scenes. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9887–9897. IEEE, 2023.

Dylan Drummey and Carlos Marcano. Baseballcv: A collection of tools and models designed to aid in the use of computer vision in baseball. `https://github.com/dylandru/BaseballCV`, 2024.

David Eigen, Christian Puhrsch, and Rob Fergus. Depth map prediction from a single image using a multi-scale deep network. *Advances in neural information processing systems*, 27, 2014.

Silvio Giancola, Mohieddine Amine, Tarek Dghaily, and Bernard Ghanem. Soccernet: A scalable dataset for action spotting in soccer videos. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1711–1721. IEEE, 2018.

Clément Godard, Oisin Mac Aodha, Michael Firman, and Gabriel J Brostow. Digging into self-supervised monocular depth estimation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 3828–3838, 2019.

Lukasz Grad, Krzysztof M Czajkowski, and Aliaksandr Varashylau. Physics-based 3d ball trajectory reconstruction from monocular soccer video: A multi-model benchmark. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 9940–9948, 2026.

Samuel Greydanus, Misko Dzamba, and Jason Yosinski. Hamiltonian neural networks. *Advances in neural information processing systems*, 32, 2019.

Richard Hartley, Andrew Zisserman, et al. *Multiple view geometry in computer vision*, volume 2. Cambridge university press Cambridge, 2003.

Yu-Chuan Huang, I-No Liao, Ching-Hsuan Chen, Tsì-Uí İk, and Wen-Chih Peng. Tracknet: A deep learning network for tracking high-speed and tiny objects in sports applications. In *2019 16th IEEE international conference on advanced video and signal based surveillance (AVSS)*, pp. 1–8. IEEE, 2019.

Paresh R Kamble, Avinash G Keskar, and Kishor M Bhurchandi. Ball tracking in sports: a survey. *Artificial Intelligence Review*, 52(3):1655–1705, 2019.

Jeremy Yirmeyahu Kaminski and Mina Teicher. A general framework for trajectory triangulation. *Journal of Mathematical Imaging and Vision*, 21(1):27–41, 2004.

Kenichi Kanatani. *Statistical optimization for geometric computation: theory and practice*. Courier Corporation, 2005.

Sayash Kapoor and Arvind Narayanan. Leakage and the reproducibility crisis in machine-learning-based science. *Patterns*, 4(9), 2023.

George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.

Taeone Kim, Yongduek Seo, and Ki-Sang Hong. Physics-based 3d position analysis of a soccer ball from monocular image sequences. In *Sixth International Conference on Computer Vision (IEEE Cat. No. 98CH36271)*, pp. 721–726. IEEE, 1998.

Tom Kim. Opencommand: Open catcher-target and pitcher-command data for mlb, August 2026. URL `https://open-command.com`. Dataset. Code repository: https://github.com/tomdoyo/open-command.

Paul Liu and Jui-Hsien Wang. Monotrack: Shuttle trajectory reconstruction from monocular badminton video. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 3512–3521. IEEE, 2022.

Raul Mur-Artal, Jose Maria Martinez Montiel, and Juan D Tardos. Orb-slam: A versatile and accurate monocular slam system. *IEEE transactions on robotics*, 31(5):1147–1163, 2015.

Alan M Nathan. The effect of spin on the flight of a baseball. *American Journal of Physics*, 76(2): 119–124, 2008.

Neil Owens, C Harris, and C Stennett. Hawk-eye tennis system. In *2003 international conference on visual information engineering VIE 2003*, pp. 182–185. IET, 2003.

Hyun Soo Park, Takaaki Shiratori, Iain Matthews, and Yaser Sheikh. 3d reconstruction of a moving point from a series of 2d projections. In *European Conference on Computer Vision (ECCV)*, pp. 158–171, 2010.

Puntawat Ponglertnapakorn and Supasorn Suwajanakorn. Where is the ball: 3d ball trajectory estimation from 2d monocular tracking. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 6112–6121. IEEE, 2025.

Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.

Evan Ribnick, Stefan Atev, and Nikolaos P Papanikolopoulos. Estimating 3d positions and velocities of projectiles from monocular views. *IEEE transactions on pattern analysis and machine intelligence*, 31(5):938–944, 2008.

Nien-En Sun, Yu-Ching Lin, Shao-Ping Chuang, Tzu-Han Hsu, Dung-Ru Yu, Ho-Yi Chung, and Tsì-Uí Ík. Tracknetv2: Efficient shuttlecock tracking network. In *2020 International Conference on Pervasive Artificial Intelligence (ICPAI)*, pp. 86–91. IEEE, 2020.

Shuhei Tarashima, Muhammad Abdul Haq, Yushan Wang, and Norio Tagawa. Widely applicable strong baseline for sports ball detection and tracking. *arXiv preprint arXiv:2311.05237*, 2023.

Bill Triggs, Philip F McLauchlan, Richard I Hartley, and Andrew W Fitzgibbon. Bundle adjustment—a modern synthesis. In *International workshop on vision algorithms*, pp. 298–372. Springer, 1999.

Jack Valmadre and Simon Lucey. General trajectory prior for non-rigid reconstruction. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1394–1401, 2012.

Gabriel Van Zandycke and Christophe De Vleeschouwer. 3d ball localization from a single calibrated image. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 3471–3479. IEEE, 2022.

A Qualitative Results

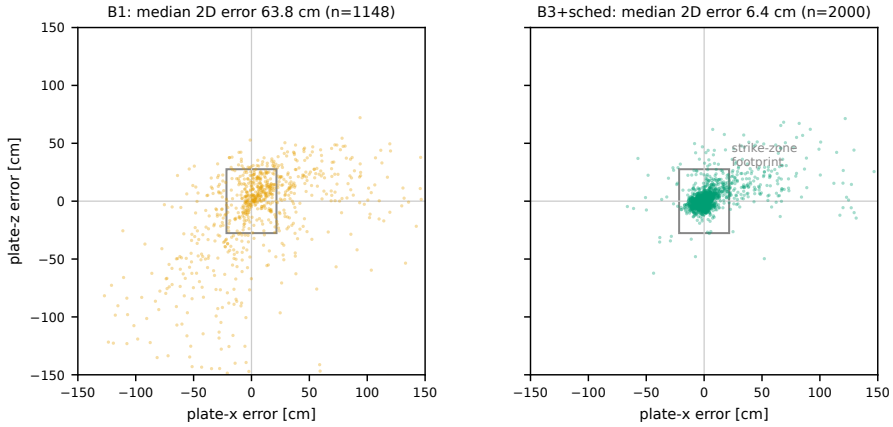


Figure 3: Signed plate-crossing errors (prediction minus reference) for a fixed sample of 2,000 benchmark pitches, with the strike-zone footprint drawn at the same scale. Fits with undefined crossings are omitted, which leaves n=1,148 for B1. Gravity anchoring (B1) scatters across the zone, while B3+sched concentrates within centimeters.

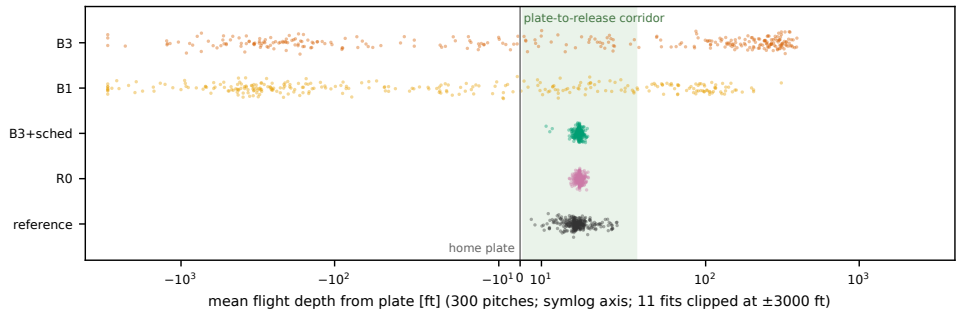


Figure 4: Where each method places the pitch: distribution of fitted mean flight depth over 300 pitches on a field-scale (symlog) axis. The reference, R0, and B3+sched occupy the plate-to-release corridor. B1 and unconstrained B3 scatter tens of meters away on both sides of the field.

B Robustness Sweeps

Sampling and missing observations. Reducing temporal density has little effect on anchored depth until sampling becomes severe (Table 5). At 12.5% density, B3+sched reaches 167 cm error and its success rate falls to 80.7%. Random dropout shows a clearer benefit from the ballistic model. With 75% of observations removed, R0 degrades from 96 to 299 cm, while B3+sched increases from 90 to 160 cm (Fig. 6c). The ballistic model helps mainly with missing observations and trajectory smoothing rather than with metric scale. The prior-free fits show smaller errors at the lowest density. With few observations the optimizer moves less from its initialization, which then acts as an implicit anchor. These entries describe the optimizer, not the ambiguity.

Pixel noise. Figure 5 evaluates synthetic observations with controlled Gaussian noise. Unconstrained B3 remains scale-ambiguous even with zero noise, and small perturbations move the recovered trajectory tens of meters in depth. R0 and B3+sched remain nearly constant across the noise sweep. Better detections improve image localization but do not resolve the monocular depth ambiguity.

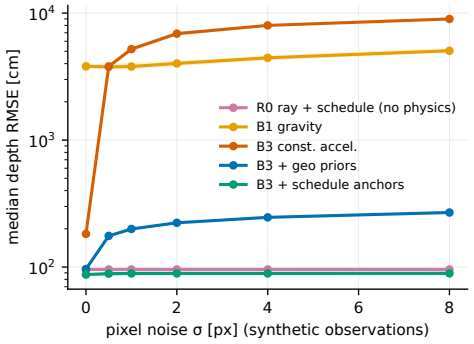


Figure 5: Depth error under synthetic pixel noise (5,000 pitches, medians, game-clustered 95% bands). Anchored methods remain nearly constant across the sweep.

C Calibration Audit Details

Perturbing the supplied camera poses at 0.5–4 times the solver's nominal uncertainty changes B3+sched by less than 2% in 3D RMSE (90.3 to 91.7 cm) and by 3 cm in plate error at the extreme (Fig. 6e). This bounds sensitivity to small random pose errors. A systematic bias shared by all Statcast-assisted poses would leave it unchanged, which is why the fixed-zone and reference-free audits of §6.2 and Appendix L.5 are needed.

D Release Anchor Placement

The depth schedule anchors the first observation at the release plane, $y{=}54.5$ ft. The reference depth at the first observation has a population median of 53.1 ft with an IQR of 49.3 to 54.4 ft, because the detector picks up the ball at a variable distance from release. We re-fit the Table 1 subset with the anchor moved to this population median, a reference-derived population constant. The 3D error of R0 falls from 96 to 87 cm and that of B3+sched from 90 to 83 cm. Plate error is unchanged at 6.6 cm, and speed error rises by 0.4 to 1.3 mph. The remaining depth error reflects the pitch-to-pitch spread of the first-observation depth, which no fixed anchor can capture.

A hand-crafted conditional anchor does not capture it either. The first-observation depth correlates with track duration at $r{=}0.63$, and a population regression gives $y_{\text{first}} = 29.0{+}56.1\,\tau$ ft for duration τ in seconds, with a residual standard deviation of 6.3 ft. Setting the release anchor per pitch from this regression leaves the 3D error of R0 at 95 cm and that of B3+sched at 91 cm, and improves the speed error of B3+sched from 7.8 to 6.6 mph. The learned no-pixel prior of §7 reaches 71 cm depth

error from duration, observation count, and camera pose together, so it captures more than a linear duration effect.

E Alignment Floor

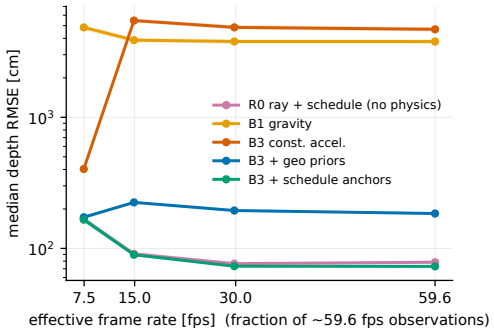
The evaluation reference is aligned in time to the detections once per pitch (§3.1). An offset error δt shifts the reference depth by about $|v_y|\,\delta t$, with $|v_y|$ near 131 ft/s, so one frame (16.8 ms) would be 67 cm. We measure how precisely the offset is recovered with a synthetic test on the Table 1 subset. For each pitch we project the reference at its fitted offset, add independent Gaussian pixel noise, and re-run the same grid search and bounded minimization. At 1.44 px noise, the median of the reference residual, the recovered offset is off by 0.6 ms at the median and 1.9 ms at the p90, which induces a depth RMSE of 2.3 cm at the median and 7.1 cm at the p90. At 2.41 px noise, the p90 of the residual, the induced depth RMSE is 3.8 cm at the median and 12 cm at the p90. The offset is well determined because the ball moves several pixels per frame and every frame constrains it. Correlated detection errors could raise this floor, but it sits an order of magnitude below the 47 to 93 cm depth errors that the paper compares.

F Population Shift

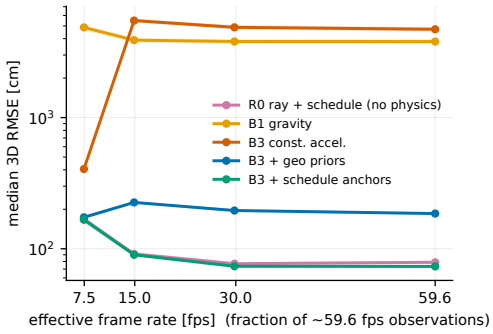
We train N1 only on pitches at or below 92 mph using the venue split and test on held-out venues. In population, the speed MAE is 1.76 mph with a median signed bias of $+0.02$ mph. On held-out pitches at or above 96 mph, the MAE is 6.58 mph with a median signed bias of -6.58 mph (IQR $[-7.9, -5.4]$), and all 997 such pitches are predicted slow. Plate location on the same pitches remains at 5.5 cm. The learned speed prior therefore does not extrapolate beyond the training range, while plate location is unaffected. Table 10 reports N1 error by pitch type on the venue split. Common types lie at 45–49 cm depth error and 1.8–2.4 mph. The knuckleball subset (n=16) shows 109 cm and 10.4 mph. With 16 pitches we report the number without attributing a cause.

G Camera Geometry and Additional Panels

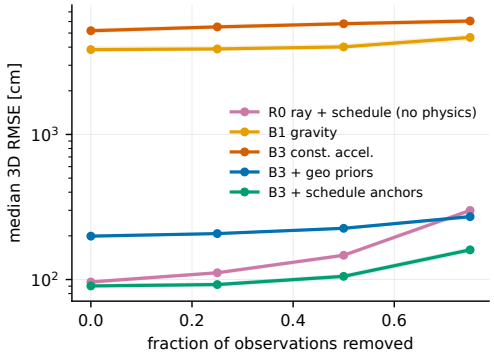
Venue-median depth error for B3+sched correlates with the camera's lateral offset $|C_x|$ at $r = -0.49$ (Fig. 6d). An off-axis mount rotates part of the depth direction into the image plane, which improves conditioning once scale is anchored, and the per-pitch CRLB reproduces this effect. The effect spans only tens of centimeters. Every MLB center-field viewpoint remains far inside the regime where absolute depth is not observable, so viewpoint matters only in this limited sense.



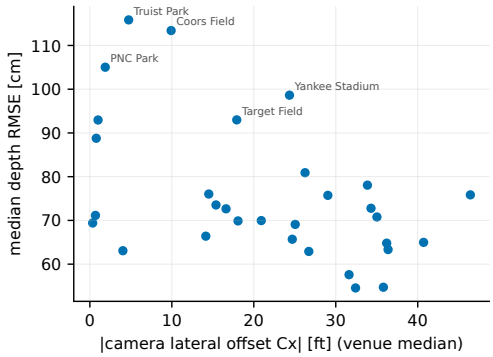
(a) Observation density vs. depth RMSE.



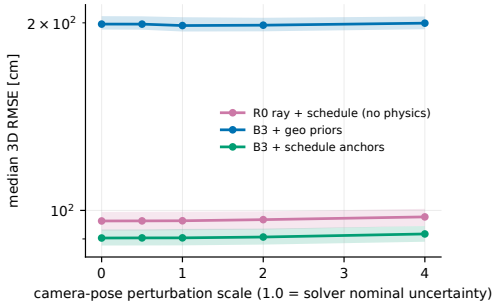
(b) Observation density vs. 3D RMSE.



(c) Random detection dropout.



(d) Venue camera offset vs. depth error ($r = -0.49$).



(e) Pose-perturbation sensitivity.

Figure 6: Additional robustness and camera-geometry panels (B3+sched unless noted).

H ADDITIONAL TABLES

Table 3: Table 1 with game-clustered bootstrap 95% CIs (5,000 pitches, full density). N0 and N1 are trained on the 234 non-benchmark pool games.

Method	3D RMSE [cm]	depth [cm]	lateral [cm]	vertical [cm]	plate 2D [cm]	speed [mph]	reproj [px]
R0 ray+schedule	96.2 [88.2, 105.1]	95.7 [87.8, 104.6]	4.2 [3.1, 5.2]	7.4 [6.5, 8.4]	6.6 [6.1, 7.1]	8.3 [7.9, 8.7]	0.5 [0.5, 0.5]
B0 const. velocity	5327.3 [4993.6, 5730.3]	5307.0 [4973.5, 5710.1]	228.1 [177.9, 270.8]	374.2 [332.5, 416.0]	56.2 [47.8, 64.6]	149.7 [116.8, 190.6]	3.9 [3.6, 4.1]
B1 gravity	3846.6 [3576.6, 4156.0]	3831.5 [3557.1, 4136.7]	178.9 [140.7, 226.2]	280.7 [255.2, 307.0]	60.5 [51.5, 72.4]	136.4 [119.7, 161.4]	0.7 [0.7, 0.8]
B1-pop mean accel.	3585.3 [3394.6, 3802.7]	3567.1 [3378.9, 3786.3]	175.1 [139.6, 206.4]	250.4 [230.7, 276.3]	215.7 [167.2, 315.9]	93.4 [83.4, 111.6]	0.8 [0.7, 0.8]
B2 fitted a_z	4614.2 [4336.5, 4902.9]	4592.5 [4316.1, 4876.9]	186.8 [136.2, 233.2]	318.7 [292.1, 356.2]	82.2 [72.5, 90.5]	79.2 [75.6, 85.0]	0.7 [0.7, 0.8]
B3 const. accel.	5189.9 [4990.8, 5433.7]	5166.9 [4963.8, 5402.0]	210.9 [137.4, 276.2]	353.8 [327.7, 384.7]	114.4 [97.2, 138.1]	61.6 [59.3, 63.6]	0.4 [0.4, 0.5]
B4 drag + free lift	8168.7 [7923.1, 8416.7]	8124.8 [7874.2, 8371.3]	350.9 [248.4, 449.0]	555.8 [510.3, 609.1]	72.8 [66.0, 79.8]	71.3 [69.2, 73.2]	0.4 [0.4, 0.5]
B3+geo	199.1 [191.7, 209.1]	198.2 [190.6, 208.0]	9.7 [7.1, 12.0]	14.3 [13.3, 15.2]	16.9 [15.0, 19.2]	27.5 [26.6, 28.2]	0.5 [0.4, 0.5]
B3+geo+aero	196.2 [187.0, 205.0]	195.2 [185.9, 203.8]	9.5 [7.0, 11.6]	13.9 [13.0, 14.9]	17.2 [14.8, 19.8]	22.5 [21.5, 23.4]	0.5 [0.4, 0.5]
B3+sched	90.3 [83.5, 98.3]	89.8 [83.0, 97.9]	4.0 [3.0, 4.9]	6.9 [6.2, 7.8]	6.6 [6.2, 7.2]	7.8 [7.2, 8.5]	0.5 [0.5, 0.5]
N0 learned, direct	47.2 [44.2, 50.4]	47.0 [43.9, 49.9]	2.8 [2.5, 3.2]	3.6 [3.3, 3.9]	4.4 [4.1, 4.7]	1.9 [1.8, 2.1]	1.8 [1.7, 1.9]
N1 learned, residual	45.6 [42.9, 49.0]	45.3 [42.7, 48.7]	2.7 [2.3, 3.1]	3.4 [3.1, 3.8]	4.1 [3.8, 4.4]	1.9 [1.8, 2.0]	1.7 [1.6, 1.8]

Table 4: Matched input ladder and shuffle controls on one venue fold (split seed 42, training seed 0, medians on the held-out venues). All rows share architecture, target, and training. 'mean' predicts the training-set mean. 'P' is the B3+sched estimate. Shuffles permute the named input block across pitches within (venue, duration-quartile) bins, so the block keeps its marginal distribution but loses its pitch-specific content.

Input	depth RMSE [cm]	plate error [cm]	speed MAE [mph]
mean	150	30.6	4.6
meta	69	30.0	3.6
meta+cam	68	30.2	3.6
meta+cam+pix	48	5.6	2.0
meta+cam+P	46	5.4	2.0
full	47	5.3	1.9
full, pixels shuffled	46	5.5	2.0
full, pixels and P shuffled	65	29.7	3.6

Table 5: Density sweep (medians, game-clustered 95% CIs). Success rate counts fits with ≥ 3 observations that returned a result; errors are conditional on success.

Method	density	success	3D RMSE [cm]	depth [cm]	plate [cm]	speed [mph]
R0 ray+schedule	1	100.0%	96.2 [88.2, 105.1]	95.7 [87.8, 104.6]	6.6 [6.1, 7.1]	8.3 [7.9, 8.7]
R0 ray+schedule	0.5	100.0%	94.4 [87.3, 104.3]	93.9 [86.9, 103.8]	6.1 [5.5, 6.8]	7.9 [7.3, 8.6]
R0 ray+schedule	0.25	98.0%	114.6 [103.4, 127.4]	114.0 [103.0, 126.8]	7.0 [6.0, 8.2]	9.7 [8.4, 11.5]
R0 ray+schedule	0.125	80.7%	168.0 [147.6, 185.5]	167.2 [146.8, 184.7]	10.1 [8.6, 11.9]	14.8 [11.2, 18.3]
B0 const. velocity	1	100.0%	5327.3 [4993.6, 5730.3]	5307.0 [4973.5, 5710.1]	56.2 [47.8, 64.6]	149.7 [116.8, 190.6]
B0 const. velocity	0.5	100.0%	5093.2 [4738.8, 5419.8]	5067.2 [4717.3, 5394.1]	50.1 [42.8, 59.9]	153.9 [119.6, 197.5]
B0 const. velocity	0.25	98.0%	5000.3 [4626.5, 5388.3]	4979.7 [4608.4, 5363.0]	49.3 [42.2, 57.4]	155.5 [118.3, 201.9]
B0 const. velocity	0.125	80.7%	5821.8 [5381.0, 6308.7]	5796.6 [5355.4, 6274.7]	42.7 [36.9, 48.8]	203.5 [159.2, 264.5]
B1 gravity	1	100.0%	3846.6 [3576.6, 4156.0]	3831.5 [3557.1, 4136.7]	60.5 [51.5, 72.4]	136.4 [119.7, 161.4]
B1 gravity	0.5	100.0%	3855.1 [3599.8, 4225.8]	3834.2 [3580.2, 4204.7]	62.3 [52.7, 74.1]	135.4 [119.8, 162.4]
B1 gravity	0.25	98.0%	3969.2 [3680.6, 4360.7]	3947.7 [3662.5, 4333.5]	60.0 [51.1, 72.2]	142.3 [125.7, 172.6]
B1 gravity	0.125	80.7%	4869.3 [4480.0, 5319.1]	4846.3 [4458.1, 5291.6]	52.9 [44.9, 62.7]	180.8 [149.3, 219.1]
B2 fitted a_z	1	100.0%	4614.2 [4336.5, 4902.9]	4592.5 [4316.1, 4876.9]	82.2 [72.5, 90.5]	79.2 [75.6, 85.0]
B2 fitted a_z	0.5	100.0%	4556.1 [4300.9, 4865.7]	4533.2 [4281.8, 4845.9]	86.7 [73.9, 100.8]	79.1 [75.5, 83.4]
B2 fitted a_z	0.25	98.0%	4108.1 [3791.1, 4430.0]	4092.0 [3776.2, 4413.6]	79.7 [69.8, 88.9]	81.5 [77.7, 89.9]
B2 fitted a_z	0.125	80.7%	2856.3 [2490.8, 3320.8]	2845.0 [2476.1, 3301.3]	51.9 [46.7, 58.1]	113.1 [95.7, 137.4]
B3 const. accel.	1	100.0%	5189.9 [4990.8, 5433.7]	5166.9 [4963.8, 5402.0]	114.4 [97.2, 138.1]	61.6 [59.3, 63.6]
B3 const. accel.	0.5	100.0%	5425.4 [5214.0, 5625.0]	5390.6 [5188.4, 5602.3]	110.8 [97.6, 126.7]	62.5 [60.1, 64.6]
B3 const. accel.	0.25	98.0%	4881.5 [4620.8, 5193.6]	4865.7 [4595.1, 5167.2]	85.5 [76.9, 94.2]	67.3 [64.5, 69.5]
B3 const. accel.	0.125	80.7%	405.6 [385.2, 431.1]	403.3 [383.1, 428.8]	24.6 [21.2, 28.0]	37.3 [36.5, 38.1]
B3+geo	1	100.0%	199.1 [191.7, 209.1]	198.2 [190.6, 208.0]	16.9 [15.0, 19.2]	27.5 [26.6, 28.2]
B3+geo	0.5	100.0%	211.7 [203.7, 219.9]	210.7 [202.6, 218.6]	18.8 [16.7, 21.0]	27.8 [26.8, 28.5]
B3+geo	0.25	98.0%	235.7 [229.2, 242.5]	234.5 [228.0, 241.3]	20.2 [17.9, 22.0]	26.8 [25.9, 27.6]
B3+geo	0.125	80.7%	173.7 [165.9, 181.9]	172.9 [165.3, 181.0]	12.9 [11.9, 14.3]	17.2 [16.0, 18.2]
B3+geo+aero	1	100.0%	196.2 [187.0, 205.0]	195.2 [185.9, 203.8]	17.2 [14.8, 19.8]	22.5 [21.5, 23.4]
B3+geo+aero	0.5	100.0%	202.6 [193.2, 210.2]	201.5 [192.4, 209.2]	18.3 [16.2, 20.8]	22.7 [22.0, 23.7]
B3+geo+aero	0.25	98.0%	214.0 [207.1, 219.8]	213.0 [206.1, 218.6]	19.6 [17.4, 21.8]	23.6 [22.9, 24.1]
B3+geo+aero	0.125	80.7%	181.4 [173.2, 188.9]	180.8 [172.4, 188.1]	13.4 [12.1, 14.9]	16.3 [15.3, 17.2]
B3+sched	1	100.0%	90.3 [83.5, 98.3]	89.8 [83.0, 97.9]	6.6 [6.2, 7.2]	7.8 [7.2, 8.5]
B3+sched	0.5	100.0%	90.6 [83.4, 99.7]	90.2 [83.0, 99.4]	6.1 [5.5, 6.7]	8.5 [7.7, 9.5]
B3+sched	0.25	98.0%	112.8 [101.3, 124.6]	112.3 [100.8, 123.9]	7.0 [6.0, 8.2]	12.3 [10.6, 13.8]
B3+sched	0.125	80.7%	166.6 [146.9, 183.5]	165.9 [146.1, 182.5]	10.1 [8.6, 11.8]	17.3 [14.8, 19.9]

Table 6: Prior ablation on B3 (full density; medians, game-clustered 95% CIs). The aero-only fit keeps a free scale, and its plate entry is not meaningful because most of its crossings are undefined or far outside the field.

Priors	3D RMSE [cm]	depth [cm]	plate [cm]	speed [mph]
aero	6070.6 [5841.5, 6335.4]	6039.2 [5812.7, 6309.0]	67198.3 [8846.0, 341081.7]	59.8 [57.9, 61.6]
direction	5425.0 [5076.5, 5852.6]	5402.6 [5048.9, 5823.1]	99.9 [82.6, 117.5]	29.1 [28.3, 29.8]
none	5189.9 [4990.8, 5433.7]	5166.9 [4963.8, 5402.0]	114.4 [97.2, 138.1]	61.6 [59.3, 63.6]
plate	860.1 [816.9, 904.7]	855.8 [813.2, 900.5]	58.3 [55.5, 60.9]	48.7 [46.8, 50.8]
release	771.8 [738.0, 814.0]	768.0 [733.8, 809.3]	37.6 [33.6, 41.8]	71.7 [69.4, 73.7]
arrival	666.6 [635.5, 691.2]	662.7 [632.7, 688.5]	28.4 [25.5, 31.1]	55.2 [52.2, 57.3]
release, plate	472.4 [456.9, 494.1]	470.1 [454.2, 491.5]	47.9 [42.5, 52.5]	50.4 [48.5, 52.6]
release, arrival	266.3 [252.6, 280.9]	265.0 [251.4, 279.6]	20.2 [17.2, 23.6]	39.6 [37.8, 41.4]
release, plate, direction, arrival	199.1 [191.7, 209.1]	198.2 [190.6, 208.0]	16.9 [15.0, 19.2]	27.5 [26.6, 28.2]
release, plate, direction, arrival, aero	196.2 [187.0, 205.0]	195.2 [185.9, 203.8]	17.2 [14.8, 19.8]	22.5 [21.5, 23.4]
depth schedule (two anchors)	191.6 [177.6, 205.8]	190.6 [176.5, 204.9]	7.1 [6.6, 7.6]	36.6 [34.1, 39.7]
depth schedule, direction, aero (B3+sched)	90.3 [83.5, 98.3]	89.8 [83.0, 97.9]	6.6 [6.2, 7.2]	7.8 [7.2, 8.5]

I Example Reconstructions

Fig. 7 shows a typical pitch and the worst B3+sched failure in the evaluation set, in both world coordinates and image space. Failure cases are dominated by late-starting or early-ending detection runs, for which the schedule anchors are placed at the wrong depths.

J Sample Accounting

From 17,396 clips in the 60 sampled games: 3,563 excluded for pose quality (upstream ball-fit RMSE ≥ 3 px), 1,150 for reference-alignment RMSE ≥ 3 px, 543 for missing joins, leaving 12,140 pitches (69.8%). The large training set applies the same filter to 294 games (58,188 kept). The preprocessing script logs every exclusion with its reason. The 294-game pool includes the 60 benchmark games (Appendix L.4).

Table 7: Combined stress tests (medians, game-clustered 95% CIs).

Method	condition	3D RMSE [cm]	plate [cm]	speed [mph]
B3 const. accel.	density=0.25+dropout=0.25	3315.8 [2939.2, 3790.4]	50.5 [44.6, 56.4]	69.9 [67.4, 73.8]
B3 const. accel.	density=0.25+noise=2px	7541.3 [7242.1, 7841.5]	60.1 [50.5, 68.8]	85.0 [81.4, 89.7]
B3 const. accel.	noise=2px+dropout=0.5	8000.0 [7789.1, 8251.7]	64.8 [56.2, 72.0]	82.0 [78.4, 86.1]
B3+sched	density=0.25+dropout=0.25	159.6 [145.4, 173.4]	9.8 [8.2, 11.4]	17.6 [15.9, 19.7]
B3+sched	density=0.25+noise=2px	113.0 [101.7, 125.3]	7.4 [6.5, 8.5]	12.5 [10.7, 14.0]
B3+sched	noise=2px+dropout=0.5	104.5 [94.1, 115.1]	7.2 [6.3, 8.1]	9.8 [8.8, 11.4]
B3+geo	density=0.25+dropout=0.25	243.2 [235.5, 252.1]	19.5 [17.3, 21.2]	25.4 [24.6, 26.1]
B3+geo	density=0.25+noise=2px	278.1 [269.6, 287.7]	24.9 [22.7, 26.8]	31.2 [30.3, 32.4]
B3+geo	noise=2px+dropout=0.5	260.5 [253.1, 269.4]	24.2 [21.9, 26.3]	32.2 [31.2, 33.3]
R0 ray+schedule	density=0.25+dropout=0.25	219.6 [206.1, 232.8]	9.8 [8.2, 11.5]	22.0 [18.6, 24.9]
R0 ray+schedule	density=0.25+noise=2px	114.6 [103.4, 127.4]	7.4 [6.5, 8.6]	9.7 [8.4, 11.5]
R0 ray+schedule	noise=2px+dropout=0.5	147.0 [135.9, 159.4]	7.2 [6.4, 8.1]	11.8 [10.3, 13.9]

Table 8: Paired per-pitch comparisons at full density (median of A−B with game-clustered 95% CI; win rate = fraction of pitches where A beats B).

A	B	metric	median diff	win rate A
B3+sched	R0 ray+schedule	rmse 3d cm	-2.2 [-2.3, -2.1]	70.2%
B3+sched	R0 ray+schedule	rmse depth cm	-2.2 [-2.3, -2.1]	70.2%
B3+sched	R0 ray+schedule	plate err 2d cm	0.0 [-0.0, 0.0]	49.3%
B3+sched	R0 ray+schedule	speed err mph	-0.7 [-0.7, -0.7]	61.7%
B3+sched	B3+geo	rmse 3d cm	-98.1 [-104.6, -92.7]	82.6%
B3+sched	B3+geo	rmse depth cm	-97.6 [-104.2, -92.2]	82.6%
B3+sched	B3+geo	plate err 2d cm	-7.6 [-8.6, -6.7]	73.0%
B3+sched	B3+geo	speed err mph	-13.6 [-15.6, -12.4]	71.7%
B3+geo	B3 const. accel.	rmse 3d cm	-4923.2 [-5188.8, -4733.8]	98.9%
B3+geo	B3 const. accel.	rmse depth cm	-4900.8 [-5159.5, -4711.9]	98.9%
B3+geo	B3 const. accel.	plate err 2d cm	-93.2 [-114.3, -78.5]	88.5%
B3+geo	B3 const. accel.	speed err mph	-28.3 [-30.4, -26.7]	79.3%
R0 ray+schedule	B3 const. accel.	rmse 3d cm	-5039.6 [-5262.0, -4834.9]	99.6%
R0 ray+schedule	B3 const. accel.	rmse depth cm	-5015.7 [-5238.2, -4813.5]	99.6%
R0 ray+schedule	B3 const. accel.	plate err 2d cm	-97.0 [-124.5, -84.6]	92.9%
R0 ray+schedule	B3 const. accel.	speed err mph	-47.3 [-49.8, -44.5]	85.7%

K Prior Constants

All constants are global. None is derived from any test pitch.

- **Release passage.** Release plane y=54.5 ft with box $x \in [-4.5, 4.5]$ ft, $z \in [2, 8.5]$ ft. The trajectory is extended backward on a 60-point grid over $[-0.35, +0.05]$ s and the release-plane crossing is interpolated linearly. Hinge penalties on the crossing x and z with weight 12 per ft, and on the crossing time outside $[-0.30, +0.02]$ s with weight 60 per s.
- **Plate geometry.** Plate plane y=17/12 ft with box $x \in [-3, 3]$ ft, $z \in [-0.5, 6]$ ft. Hinge penalties on the crossing x and z with weight 12 per ft.
- **Arrival window.** Hinge penalty on $t_{\text{cross}} - t_{\text{last}}$ outside $[-0.06, +0.20]$ s with weight 60 per s. The window is a 2024 population statistic (p1 -0.05, median 0.00, p95 $+0.18$ s).
- **Mound direction.** Hinge penalties on the initial velocity. Depth velocity toward the plate (weight 120), slopes $|v_x/v_y|, |v_z/v_y| \leq 0.25$ (weight 3), and speed within 60–165 ft/s, about 41–112 mph (weight 0.3).
- **Aerodynamic bounds.** Hinge penalties on $|a_x| \leq 35$, $a_y \in [5, 45]$, $a_z \in [-45, -5]$ ft/s^2, each with weight 1.5.
- **Depth schedule.** Two Gaussian anchors written as linear residuals, $3\,(y(0) - 54.5)/1.5$ and $3\,(y(t_{\text{last}}) - 17/12)/1.0$ in ft. Their effective standard deviations are 0.5 ft at release and 0.33 ft at the plate.

Hinge means $\max(\ell - v, 0) + \max(v - h, 0)$ for a bound $[\ell, h]$. All prior residuals are appended to the pixel residual vector and pass through the same robust loss.

Table 9: Paired per-pitch differences between matched variants (split seed 42, six held-out venues, so the venue-clustered 95% CIs are indicative only. Median A−B, win rate = fraction of pitches where A beats B).

A	B	metric	median diff	win A
full	meta+cam+P	rmse depth cm	0.70 [-0.44, 1.83]	0.48
full	meta+cam+P	plate err 2d cm	-0.00 [-0.34, 0.21]	0.50
full	meta+cam+P	speed err mph	-0.02 [-0.08, 0.03]	0.51
full	meta+cam+pix	rmse depth cm	-0.02 [-1.22, 0.88]	0.50
full	meta+cam+pix	plate err 2d cm	-0.18 [-0.51, 0.07]	0.53
full	meta+cam+pix	speed err mph	-0.05 [-0.14, 0.05]	0.52
full	full/shufpix	rmse depth cm	0.41 [-0.84, 2.06]	0.49
full	full/shufpix	plate err 2d cm	-0.09 [-0.25, 0.09]	0.51
full	full/shufpix	speed err mph	-0.01 [-0.05, 0.04]	0.51
meta+cam+P	meta+cam	rmse depth cm	-17.74 [-20.99, -14.48]	0.69
meta+cam+P	meta+cam	plate err 2d cm	-23.05 [-24.11, -22.07]	0.95
meta+cam+P	meta+cam	speed err mph	-1.57 [-1.79, -1.30]	0.71
full/shufboth	meta+cam	rmse depth cm	-1.13 [-5.44, 1.97]	0.52
full/shufboth	meta+cam	plate err 2d cm	-0.16 [-0.45, 0.08]	0.52
full/shufboth	meta+cam	speed err mph	0.02 [-0.01, 0.05]	0.49

Table 10: N1 test error by pitch type (venue split, medians). Rare types have small samples, noted in the n column.

Type	n	depth [cm]	speed [mph]	plate [cm]
FF	3336	45.8	1.77	3.8
SI	1925	48.0	2.36	4.4
SL	1466	46.1	1.93	4.9
CH	1195	47.6	2.20	4.6
FC	1025	47.2	2.24	4.4
ST	795	48.8	2.35	5.8
CU	653	45.1	1.91	6.2
FS	365	49.5	1.83	4.4
KC	146	44.8	1.87	6.0
SV	59	49.2	2.07	10.3
KN	16	108.9	10.35	9.6
FO	5	39.1	2.11	5.5
CS	2	144.0	9.29	32.1

L Implementation Details

L.1 Ballistic Fits

B0–B3 and their prior-augmented variants use `scipy.optimize.least_squares` with the trust-region-reflective method, the `soft_l1` loss with `f_scale` of 2.0 px, and at most 400 function evaluations. Tolerances are the library defaults. The parameters are unbounded. Parameter scaling uses 5 ft for position, 20 ft/s for velocity, and 10 ft/s^2 for acceleration. The residual vector contains the u and v pixel residuals of every observation followed by the prior residuals of Appendix K.

Initialization is the same for all classes. The first detection is back-projected to the y=53 ft plane and the last to the plate plane. The initial velocity is the displacement between them divided by the observed span, corrected by half of a nominal acceleration $(0, 25, -20)$ ft/s^2 times the span. B2 initializes a_z at $-g$ with g=32.174 ft/s^2. B3 initializes $\mathbf{a}$ at the same nominal value. B1-pop fixes $\mathbf{a}$ at the mean reference acceleration of the 294-game pool, $(-2.46, 27.24, -23.93)$ ft/s^2, with per-axis standard deviations $(11.0, 4.0, 8.7)$ ft/s^2. Time is measured from the first observation.

B4 integrates $\dot{\mathbf{v}} = (0, 0, -g) - k_d|\mathbf{v}|\mathbf{v} + \mathbf{m}$ with fourth-order Runge–Kutta at two substeps per frame. The drag constant is $k_d = \rho C_d A/(2m_{\text{ball}})$ with sea-level air density ρ=0.002377 slug/ft^3, C_d=0.35, ball mass 5.125 oz, and diameter 2.9 in, which gives k_d=0.00192 per ft and a drag deceleration of 32 ft/s^2 at 130 ft/s. The nine parameters are $\mathbf{p}_0$, $\mathbf{v}_0$, and the constant lift vector $\mathbf{m}$. Initialization, loss, and scaling follow B3, with $\mathbf{m}$ set so that the initial total acceleration equals the B3 nominal value. Changing C_d to 0.30 or 0.40 moves the median depth error by less than 3 m.

L.2 R0

R0 assigns each observation a depth on the line from y=54.5 ft at the first detection to y=17/12 ft at the last, linear in time, and back-projects the pixel ray onto that depth plane. Velocity, acceleration,

Table 11: Learned models, game-level split (medians with game-clustered 95% CIs on the held-out test set).

Method	3D RMSE [cm]	depth [cm]	plate [cm]	speed [mph]
B3+sched(testset)	97.3 [88.0, 109.1]	96.8 [87.7, 108.7]	6.7 [6.0, 7.6]	8.3 [7.5, 9.5]
N0	49.8 [46.2, 53.5]	49.3 [46.0, 53.1]	4.8 [4.3, 5.2]	2.0 [1.8, 2.1]
N1	46.9 [44.0, 50.1]	46.6 [43.8, 49.8]	4.2 [3.8, 4.6]	1.9 [1.8, 2.0]

Table 12: Learned models, venue-level split (medians with game-clustered 95% CIs on the held-out test set).

Method	3D RMSE [cm]	depth [cm]	plate [cm]	speed [mph]
B3+sched(testset)	95.2 [85.8, 108.1]	94.7 [85.3, 107.6]	5.9 [5.5, 6.3]	7.9 [7.3, 8.8]
N0	50.3 [47.0, 54.2]	49.9 [46.7, 53.8]	4.9 [4.6, 5.2]	2.0 [1.9, 2.2]
N1	47.1 [43.8, 50.8]	46.8 [43.5, 50.4]	4.5 [4.2, 4.8]	2.0 [1.9, 2.2]

and plate crossing come from a per-axis least-squares fit of a quadratic in time to the resulting 3D points. The plate crossing is the smallest positive root of the fitted depth polynomial.

L.3 Learned Models

Inputs. Each track is resampled to 16 points by linear interpolation in time. The 105-dimensional input consists of the 16 pixel coordinates scaled to $[(u - 640)/1280,\ (v - 360)/720]$, the lateral and vertical coordinates of the 16 rays back-projected onto the release plane and the plate plane divided by 10 ft, three track features (duration, observation count divided by 30, and the first relative timestamp, which is zero by construction), and six camera features (mount x and z divided by 100 ft, focal length divided by 20,000 px, and pan, tilt, and roll multiplied by 50). The matched ablations select subsets of these blocks.

Targets. The target is the reference trajectory re-parameterized at the first observation, nine values. Targets are standardized with fixed constants (mean $(0, 52, 5.5, 0, -130, -5, 0, 25, -18)$ and scale $(5, 20, 3, 15, 40, 10, 20, 10, 10)$ in ft, ft/s, and ft/s^2). N0 predicts the standardized target. N1 receives the standardized B3+sched estimate as nine extra inputs and predicts the standardized residual to it. Pitches where B3+sched returns a non-finite estimate are dropped from N1 training and evaluation.

Architecture and training. The network is an MLP with three hidden layers of width 256 and GELU activations, without dropout or normalization layers. Training uses the Huber loss with δ=1, AdamW with learning rate 10^{-3} and weight decay 10^{-4}, a cosine learning-rate schedule, and batch size 512. N0 and N1 train for 200 epochs. There is no early stopping. We keep the checkpoint with the lowest validation loss. The matched ablations train for 150 epochs, the venue cross-validation and the speed-shift experiment for 120 epochs. Training seeds are 42 for N0, N1, and the speed-shift experiment, 0 for the matched ablations, and 42 and 43 for the venue cross-validation. In the venue cross-validation and the matched ablations every variant predicts the standardized target directly, and the B3+sched estimate enters only as an input where the variant name includes P.

L.4 Splits and Cohorts

The benchmark manifest samples two games per home park (60 games, 12,140 pitches). A second manifest samples eight games per home park (240 games). Their union is the 294-game pool, which yields 58,188 pitches under the same quality filter. The pool includes the 60 benchmark games.

Learned models use splits of the 294-game pool with fractions 70/15/15, drawn with a fixed seed. The game split partitions games (205 train, 44 validation, 45 test games, or 40,109 / 9,365 / 8,714 pitches). The venue split partitions venues (21 / 4 / 6 venues, or 39,477 / 7,723 / 10,988 pitches). The game split guarantees that no game appears in more than one part, while venues can recur across parts. The venue split additionally holds out whole venues, so its test pitches come from camera mounts never seen in training. All learned-model results are reported on the split test sets, and the B3+sched numbers quoted alongside them are computed on the same test sets. The 5-fold venue cross-validation permutes the 31 venues with a fixed seed and assigns every fifth venue to one fold,

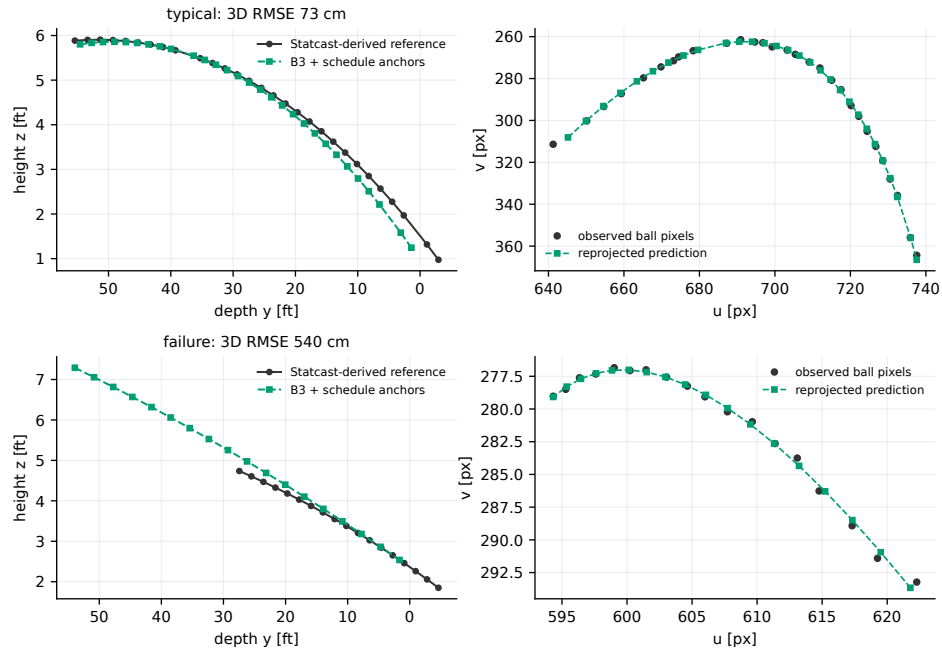


Figure 7: Typical (top) and failure-case (bottom) reconstructions: side view (left, depth and height plane) and image space (right).

so each venue is tested exactly once. Within a fold, four of the remaining venues serve as validation and the rest as training. The speed-shift experiment uses the venue split and trains only on pitches at or below 92 mph.

The benchmark protocol for learned methods (Table 1) trains on the 234 pool games that are not benchmark games (46,048 pitches), holds out 10% of those games (4,395 pitches) for checkpoint selection, and tests on the 5,000-pitch benchmark subset. N0 and N1 there use the architecture, loss, and 200-epoch schedule above with seed 42 and a single model per row.

L.5 Reference-Free Calibration and Cohort

The reference-free audit solves each clip's pan, tilt, roll, and focal length from the detected strike-zone box alone, given four fixed population constants. The constants are the mean reference strike-zone top and bottom heights over the season and the median mount x and z over all parks. The mount depth is fixed at 400 ft, as in the supplied poses. The world box is placed at the plate plane with the league plate width, and a bounded least-squares fit from a zero-angle initialization matches its four corners to the detected ones. A clip with residual above 3 px is marked as a calibration failure and stays in the cohort accounting.

Run selection is reference-free. Detection runs are split at gaps of more than four frames. A run is a candidate if it has at least eight detections and starts within 0.6 s of the clip's release hint. The longest candidate is kept. The release hint is a per-clip timing field in the OpenCommand per-frame track table. It is video-derived clip metadata and is not taken from the Statcast trajectory. It is used only for run selection and never passed to an estimator. Cohort status (no box, no track, no flight run, calibration failure, or frozen) is decided for every clip before any reference is read. References are then used only to score the frozen cohort. The season-holdout variant (below) re-estimates the four constants from the 2025 season.

M SEASON-HOLDOUT CALIBRATION CONSTANTS

We re-estimate the four population calibration constants from the 2025 season (zone heights 3.432 and 1.604 ft, mount $\bar{C}_x = -23.6$ ft and $\bar{C}_z = 30.1$ ft) and re-run the audit on 2024. The median plate error is 8.8 cm with game-clustered 95% CI $[8.2, 9.4]$, the p90 is 35.2 cm, 56.4% of pitches fall within 10 cm, and the cohort is 15,904 of 17,355 eligible clips. This matches the run with 2024 constants (8.7 cm), so the constants carry no material information about the evaluated season.

N RUN-SELECTION AUDIT

The benchmark selects the candidate detection run best aligned with the reference. The reference-free heuristic of Appendix L.5 selects an overlapping run on 98.2% of 1,500 audited pitches drawn with a fixed seed from the benchmark. Two selections count as overlapping when at least 80% of their detections coincide. The audit log will be released with the code.