Published-data validation of a consumer interior-ballistics engine: Calibration, unseen loads, and the limits of lookup

reloadingcalc.com

support@reloadingcalc.com

16 November 2026

Abstract

A calibrated lumped-parameter interior-ballistics engine gives muzzle-velocity mean absolute error (MAE) of 7.551% on 30,327 graded rows from published manuals. This fitted-corpus result includes powders used for calibration. A separate default-physics arm reports 13.66% MAE; unmatched reconstruction inputs limit causal attribution of the difference. Four graded populations now bound the engine's velocity accuracy side by side. Fitted-corpus rows give 5.64% MAE. Held-out loads within fitted powders, on an edition-split partition, give 5.05% (16,082 rows), the best estimate for the dominant user case, a new load on a powder with a fitted model. Clean never-seen loads (8,966 rows) give 11.74%, the applicable figure for a user whose powder has no fitted model; the default partition gives 11.93%. Published-data lookup wins on 4,656 identical rows with enough printed charge points: 6.29% MAE against the engine's 11.66%, a paired gap of 5.370 points with a 95% bootstrap interval of [4.161, 6.504]. Lookup scores only 52% of the never-seen partition. The engine extends calculation outside printed support and estimates pressure, whose accuracy remains insufficient for load-development safety claims. On a separate fitted-powder battery, leave-one-out ignition anchors give 5.28% max-charge MAE. Three no-exit runs are formally excluded and disclosed; none entered the graded corpus. Western-source error (10.37% MAE) remains genuine model error, including sensitivity to Accurate lot data; the stratum's 1,637 imputed-14-inch rows are its worst cell (MAE 14.58%, 27.2% beyond 20% error). The pressure-independent burn law and selection on calibration holdouts limit generalization.

1 Introduction

Published manuals give tested charge weights and velocities for specific combinations of cartridge, bullet and propellant. A calculator extends these tables to unprinted combinations, where its predictions need a separate error estimate. On the clean never-seen partition measured here, muzzle-velocity MAE is 11.74%, the regime a user whose powder has no fitted model should expect. The smaller 7.551% corpus figure includes calibration exposure and describes a different population. For the dominant case of a new load on a powder with a fitted model, a held-out edition-split cell inside the fitted powders measures 5.05% (Section 3.3).

The calibrated corpus gives 7.551% MAE and the separate default-physics arm gives 13.66%. This is a near-halving of reported error, but a matched-input rerun is required to attribute the entire difference to powder calibration. It does not show that the equations outperform

a printed-data predictor. A leave-one-out charge–velocity lookup wins where both predictors can be scored, with 6.29% against 11.66% MAE on identical never-seen rows; a paired cluster bootstrap puts the gap at 5.370 points with a 95% interval of $[4.161, 6.504]$. Its coverage is 52% of that partition; the engine can calculate the remaining combinations, with no evidence that their error is as small as the lookup-supported error.

QuickLOAD and GRT are consumer implementations of thermodynamic interior-ballistics models [1, 2]. The lumped-parameter tradition includes Corner's treatment and IBHVG2, an extension of the Baer–Frankle code [6, 9, 13]; the broader theory is described by Krier and Summerfield and by Carlucci and Jacobson [3, 4]. STANAG 4367 specifies a thermodynamic global-parameter model; no compliance claim is made here [14]. Spatially resolved two-phase codes such as XKTC include bed and flow physics absent from this zero-dimensional approximation [15]. This engine combines a first-law energy ledger with fitted burn-rate and loss terms. Its main burn path omits pressure feedback, so published-data calibration carries much of its velocity accuracy. No controlled QuickLOAD/GRT software comparison is available here, and no ranking against those programs follows from the results.

The reported quantities are errors against printed aggregates, not individual shots. Repeated editions can reprint a laboratory result; matching their text verifies transcription, without creating an independent experiment. Publishers also select which maximum loads to print, depleting the high-pressure tail. Pressure estimates outside printed support are a capability of the engine, but their availability does not establish their validity.

Sections 2 and 3 define the corpus and report scoped results. Section 5 retains the physical model and numerical details; Section 6 describes the calibration checks and their selection bias. Section 7 collects the limits, with the corrected leave-one-out anchor comparison in Section 7.3. The falsification ledger follows in Section 8.

2 Corpus construction and grading

2.1 Sources and row identities

The combined corpus contains Hodgdon/IMR Annual Manual records from 2017–2025, Winchester/Accurate/Ramshot records from 2022–2025, and Western Powders guide records with allied print captures [7, 8]. A printed start/max pair yields two charge-point observations, called half-rows below. Except for explicitly paired calibration-battery results, every count of rows below counts these single charge-point observations; rows and half-rows denote the same unit. Charge, cartridge and bullet identity determine the query; the printed barrel length determines geometry when available. The Western pipeline uses a documented 14-inch default for missing barrel length, which introduces uncertainty into those comparisons.

Positioned parsing follows each publisher's table grammar. Admission checks require start values not to exceed maximum values, velocity within 300–5,000 ft/s, and pressure no greater than 1.10 times the stated maximum average pressure (MAP). The pressure cap is a misparse filter, applied in the printed convention where resolved; it also removes observations most relevant to near-ceiling safety assessment. A complete count by upstream exclusion reason and unresolved convention is unavailable. Cross-edition-matched rows agree in at least two editions and are assigned to their earliest edition. The Western stratum has separate identity resolution and is not uniformly protected by that cross-edition rule.

The Western collection contains 10,883 observations, of which 9,898 resolve to a propellant and 985 do not. Charges are printed for 8,735, recovered for 775 and assumed for 388. The

latter categories remain in the corpus and are a source of input uncertainty; the corpus figure is not an as-printed-only estimate. Single-snapshot ADI data and the Vihtavuori guides are supplementary calibration campaigns, outside the combined-corpus denominator.

2.2 Grading rule and no-exit exclusion

For a printed velocity v_i and a simulated muzzle velocity $\widehat{v}_i$, the signed percentage error is

$$e_i = 100(\widehat{v}_i - v_i)/v_i, \qquad \text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |e_i|. \tag{1}$$

The within-10% rate is the fraction with $|e_i| \leq 10$. A no-exit run has no muzzle velocity: any velocity at the 5 ms cutoff is a mid-barrel state and is ineligible for grading. The hardened harness formally excludes these runs before computing summaries or worst-row tails. The 2026-11-16 rescan and the served corpus artifact disclose three: one Accurate 5744 .338 Lapua Magnum observation and two Ramshot True Blue .50 Browning observations. All three were already outside the graded set, so the exclusion leaves 30,327 graded rows and 7.551% MAE unchanged.

The combined record uses a velocity-only quarantine, $|e_i| > 60\%$, after excluding missing output. Pressure summaries and the never-seen clean partition additionally require printed-pressure error, where present, to remain within 60%. These are different grading rules. The 60% cutoff is a project data-screening convention; no public timestamp establishes that it preceded the first inspection of residuals. Conditioning a statistic on its own error truncates its tails, even if the removed fraction is small.

Table 1: Row reconciliation from the Phase-1 accounting JSON and corrected no-exit rescan. The Western no-exit run is one of the 11 Western velocity quarantines. The other two no-exit runs are the two no-output observations; no no-exit count is an additional subtraction.

Stage	Hodgdon/IMR	Win./Acc./Ram.	Western	Total
Admitted/resolved	16,445	4,366	9,898	30,709
Simulated	16,418	4,032	9,898	30,348
No output	0	2	0	2
Velocity quarantine	3	5	11	19
Graded velocity	16,415	4,025	9,887	30,327
Pressure-only quarantine	242	100	91	433
No exit (overlapping exclusions)	0	2	1	3

The Western pressure-clean subset has 9,796 rows after 102 either-channel exclusions. The artifact's quarantine list deduplicates those 102 half-rows to 83 unique cartridge/powder/charge keys. Thus 83 listed loads, 102 excluded half-rows and 21 combined-record exclusions describe different units or rules. Simulation parity counts refer to 30,348 attempts, while the velocity tables refer to 30,327 graded observations.

2.3 Quarantine sensitivity

Phase-1 JSON results give 7.382% MAE on 29,894 either-channel-clean rows, against 7.551% on the 30,327 velocity-clean rows. Tightening either-channel screening to 30% leaves 22,347 rows and 5.589% MAE. A stricter rule therefore improves the apparent accuracy while changing the population. The supplied sensitivity routine applies an internal velocity cutoff even to its output

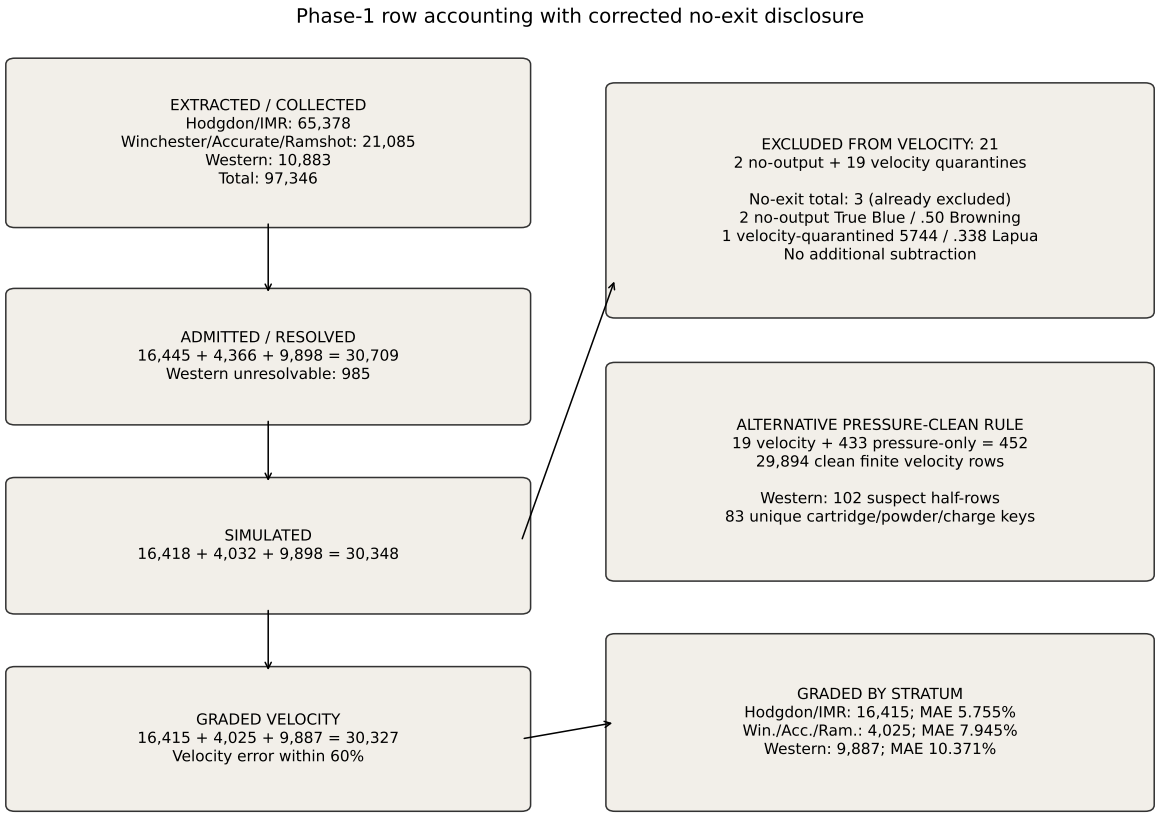


Figure 1: Phase-1 row-accounting flow for the corpus of record. The later no-exit rescan adds an explicit three-run disclosure within the already-excluded observations; Table 1 gives the overlapping counts. The diagram's "lenses" are alternative grading rules.

labeled "no quarantine"; that output repeats 7.551% and is not an untruncated estimate. No untruncated accuracy claim is based on that mislabeled block. A corrected calculation from the Phase-1 residual table, with the three no-exit outcomes removed, gives the following sensitivity estimates (revision JSON `quarantine_corrected.json`): The unscreened row count exceeds

Table 2: Corrected sensitivity to error screening. All rules first exclude no-exit predictions. "None" retains every finite exit-velocity residual.

Error screen	n	MAE (%)	Within 10% (%)
None	30,345	7.588	73.33
Velocity only, 60%	30,327	7.551	73.37
Either channel, 30%	22,347	5.589	84.63
Either channel, 60%	29,894	7.382	74.14
Either channel, 100%	30,301	7.571	73.39

the velocity-clean count by 18: these are the remaining velocity quarantines after removing the one overlapping Western no-exit run. The unscreened MAE exceeds the headline by 0.037 percentage points. The effect of the 60% velocity cutoff is small in aggregate, while the choice still conditions the target population on model output.

The tail rates of Table 4 are likewise near screen-neutral (revision analysis tail-rate JSON): The velocity-only screen removes only the beyond-60% tail, 0.06% of rows, and moves the

Table 3: Corpus-wide and never-seen tail rates under alternative error screens. All rows first exclude the three no-exit predictions.

Error screen	$P>10$ all (%)	$P>20$ all (%)	$P>10$ never-seen (%)	$P>20$ never-seen (%)
None	26.7	6.0	55.4	13.2
Velocity only, 60%	26.6	6.0	55.4	13.1
Either channel, 60%	25.9	5.4	54.9	12.1

aggregate $P>20$ rate by less than 0.1 points (13.2 to 13.1 on never-seen rows). The either-channel lens moves corpus-wide $P>10$ by 0.7 points and $P>20$ by 0.6 points. The quarantine is tail-neutral at the aggregate level and near-neutral at the safety-relevant 20% level. The never-seen signed median is -9.76% under the velocity-only lens and -9.64% on the clean partition; both figures indicate systematic under-prediction outside powder-calibration coverage, and the paper labels which lens each figure uses.

3 Velocity results and baselines

3.1 Calibrated-corpus result and its scope

The combined record measures 30,327 rows with 7.551% MAE, signed median -3.21%, and 73.4% within 10%. It uses the v1.1.14 engine and the 2026-11-11 corpus snapshot. The catalogue at the stated measurement context contains 45 graded powders. This is a fitted-corpus benchmark: rows from fitted powders account for most of its denominator, and some participated in calibration. It is not an unseen-load accuracy estimate.

The default-physics arm with fitted burn-rate rows removed reports 13.658% MAE against the calibrated record's 7.551% (Phase-1 baseline JSON). Calibration thus nearly halves the

Table 4: Corpus-of-record velocity errors, with the exact grading scope in Section 2.2. Tail rates are $P(|e_i|>10\%)$ and $P(|e_i|>20\%)$ under the velocity-only 60% screen; Section 2.3 gives their screen sensitivity. Fitted-corpus rows are observations from powders with fitted burn-rate rows; default-powder rows have no such row. Held-out rows within fitted powders are graded fitted-powder rows whose edition falls outside that powder's fitted-edition set (Section 3.3). Source: combined record, Phase-1 partition JSON and revision analysis JSONs.

Population	n	MAE (%)	Median (%)	Within 10% (%)	$P>10$ (%)	$P>20$ (%)
Combined	30,327	7.551	−3.21	73.4	26.6	6.0
Hodgdon/IMR	16,415	5.755	−2.09	84.1	15.9	3.1
Winchester/Accurate/Ramshot	4,025	7.945	−5.33	68.9	31.1	4.3
Western	9,887	10.371	−5.32	57.3	42.7	11.4
Fitted-corpus rows	21,110	5.640	−1.84	85.9	14.1	2.9
Held-out rows within fitted powders (edition split)	16,082	5.048	−1.81	89.3	10.7	1.7
Clean never-seen rows	8,966	11.736	−9.64	45.1	54.9	12.1
Default-powder rows	9,217	11.928	−9.76	44.6	55.4	13.1

reported corpus error, largely through fitted powders. This is a descriptive cross-arm comparison: it is not a paired causal estimate on identically screened rows. The reconstruction machinery used for the baseline did not reproduce all Western inputs, so a corrected matched-input default-physics rerun is required before attributing the entire difference to calibration.

3.2 Clean never-seen loads and dependence

The clean never-seen partition contains 8,966 rows whose powders have no fitted burn-rate row and which pass the 60% velocity screen and, where printed, the 60% pressure screen. "Clean" describes screening; global constants were selected using a different Hornady calibration battery. Its MAE is 11.736%, rounded to 11.74%, with signed median −9.64%, indicating systematic under-prediction outside powder-calibration coverage; only nine exact charge-point matches occur between that battery and the manual corpus. Fit-subsample membership is reconstructed at powder/edition level because the exact round-robin selection is not recoverable. Consequently a completely untouched prospective test set is still missing.

The corpus-analysis summaries (labeled Phase 1 in the accompanying JSONs) cover partitioning, row reconciliation and the lookup/default baselines. The anchor and ablation summaries are labeled Phase 2. Phase-1 uncertainty estimates use 10,000 seeded cluster-bootstrap resamples and percentile 95% intervals [17]. The corpus spans 65 powder clusters, including powders outside the 45-powder graded catalogue. Resampling these clusters gives combined-corpus MAE interval $[6.426, 8.725]$%; resampling 2,176 powder-by-cartridge clusters gives $[7.234, 7.870]$%. The wider powder interval reflects dependence shared across cartridges of a powder. These cluster counts are resampling units, not an exact effective sample size; publishers and reprinted physical loads can remain dependent across them.

The fitted-minus-default-powder MAE difference is −6.289 percentage points, with powder-by-cartridge interval $[-6.965, -5.601]$. The Hodgdon-minus-Western difference is −4.616 points, interval $[-5.346, -3.909]$. These are descriptive population differences, not effects of publisher or calibration assignment. A crossed publisher/physical-load bootstrap and prospective selection-adjusted intervals have not been measured.

3.3 Held-out loads within fitted powders

A fourth population grades the dominant user case: a new load on a powder that has a fitted model. Its cell holds the 16,082 graded rows whose powder carries a fitted burn-rate row and whose edition falls outside that powder's reconstructed fitted-edition set, so the fit's selection procedure never scored those editions. The partition is edition-level because fit-subsample row membership is not recoverable; the cell definition and its statistics were fixed in a written preregistration dated 16 November 2026, before the cell was computed [11]. MAE on the cell is 5.048%, rounded to 5.05%, with signed median −1.81% and 89.3% within 10%. That is below the fitted-corpus 5.64% and far below the never-seen 11.74%; the per-powder two-parameter fit generalizes across editions and cartridges, and the 11.7–11.9% error regime is a property of never-seen powders. Two qualifications bound the reading. Held-out editions of a powder still share its fitted parameters by design, and rows inside the fitted editions of the same powders score 7.53% MAE ($n = 5{,}028$), so edition membership is a noisy proxy for what the fit saw. The four populations now reported side by side in Table 4 are fitted-corpus rows at 5.64%, held-out rows within fitted powders at 5.05%, clean never-seen rows at 11.74%, and default-powder rows at 11.93%. A user adding a load to a powder with a fitted model sees about 5.05%; a user whose powder has no fitted model sees 11.74%.

3.4 Lookup wins within printed support

The lookup predictor omits each query's printed velocity and estimates it from other charges within powder × cartridge × bullet groups. It requires at least two other distinct charge points. It interpolates linearly within their hull and uses the nearest two points for linear extrapolation outside that hull. Thus "support" here means a group has enough printed charge points; the reported pooled result is not restricted to strict within-hull interpolation. Reference points are drawn only from the clean never-seen partition.

Table 5: Fair comparison on 4,656 identical never-seen rows (Phase-2 fair-fight JSON). Neither predictor has an additional predictor-specific residual cutoff.

Predictor	MAE (%)	Signed median (%)	Within 10% (%)
Printed-data lookup	6.293	0.00	83.0
Engine	11.663	−9.59	45.1

Published-data lookup wins this comparison, including in each publisher stratum. The engine-minus-lookup MAE difference is 5.370 percentage points. A paired cluster bootstrap over the identical rows (257 powder-by-cartridge clusters, mean cluster size 18.1, 10,000 resamples, percentile 95% intervals [17]) puts the gap at 5.370 points with interval $[4.161, 6.504]$, which excludes zero. Allowing reference groups from the unrestricted corpus gives lookup MAE 6.357% on the same rows and leaves the conclusion unchanged. The resampling unit is the half-row: one printed start/max pair yields two charge-point rows, so the corpus's 30,327 graded half-rows are about 15,086 distinct loads, and the 4,656 paired rows are about 2,331 loads. The interval is over half-rows; at load level the effective sample halves and the interval widens accordingly.

Lookup covers 4,656 of 8,966 rows, or 51.9% (about 52%). The engine's additional function is calculation where printed-data lookup has insufficient support and the generation of pressure estimates. Neither coverage nor the physical equations establish superior accuracy in that region. The applicable measured figure depends on the load's powder (Section 3.3). A new load on a

powder with no fitted model lands in the never-seen regime, 11.74%, with its screening and dependence qualifications; that population's signed median is −9.76% under the velocity-only screen and −9.64% on the clean partition. A new load on a powder with a fitted model lands near the fitted corpus, at about 5.05%.

4 Head-to-head comparison with Gordons Reloading Tool

This section answers one question: when both tools are handed the same handloads, how close does each come to the velocities the reloading manuals print? We measured it rather than argued it. 17,676 loads were taken from published manuals, entered once each, run through both engines unchanged, and each prediction was graded against the printed velocity. Everything below reports what that measurement says and what it means. The mechanics of data collection appear only where they affect how a number should be read; the capture records themselves are preserved with this paper's artifacts.

4.1 What was compared

The loads come from the Hodgdon and Winchester/Accurate manual families — 58.2% of our graded corpus, rifle and handgun smokeless loads. Both engines received the identical input file for every load, so any difference in output is the engine and its data, not the person entering the load.

The comparison is deliberately asymmetric, and the asymmetry must travel with every number in this section. GRT 2021.2040-nightly ran with its factory-default powder inputs; IBC ran its fitted calibration tables. That is how real users would run each tool, but it means the section measures the quality of each tool's *inputs* for these loads, not the superiority of either engine's physics. Two facts about our calibration sharpen the point: part of it was fitted to GRT's own pressure outputs, and part was fitted to the same manual pages this corpus comes from. This is a fit comparison between related tools, not an independent validation of either one. Where a result depends on that relationship, the text says so.

4.2 How to read the numbers

Three statistics carry the results. *Mean absolute error* (MAE) is the average size of a miss regardless of direction: an MAE of 5% means a tool's predictions typically land within 5% of the printed velocity. *Within-10%* (w10) is the share of loads predicted inside 10% of the book — the close-enough-to-plan-a-work-up measure. *Median bias* is the middle signed error: negative means the tool reads the books low on the typical load, positive means high. MAE says how far off; bias says which way.

Because identical loads went through both engines, the headline numbers are paired differences (IBC minus GRT) on the 15,421 loads both engines answered. Pairing is what makes the comparison fair: whatever error the manual pages themselves contribute, both engines see it, so the difference isolates the tools. The intervals around each difference come from resampling the data by powder — powder-to-powder variation, not load-to-load noise, is what limits confidence here — with a second interval family that also resamples by cartridge shown alongside. An interval that includes zero means the difference could be noise; we say so wherever it happens.

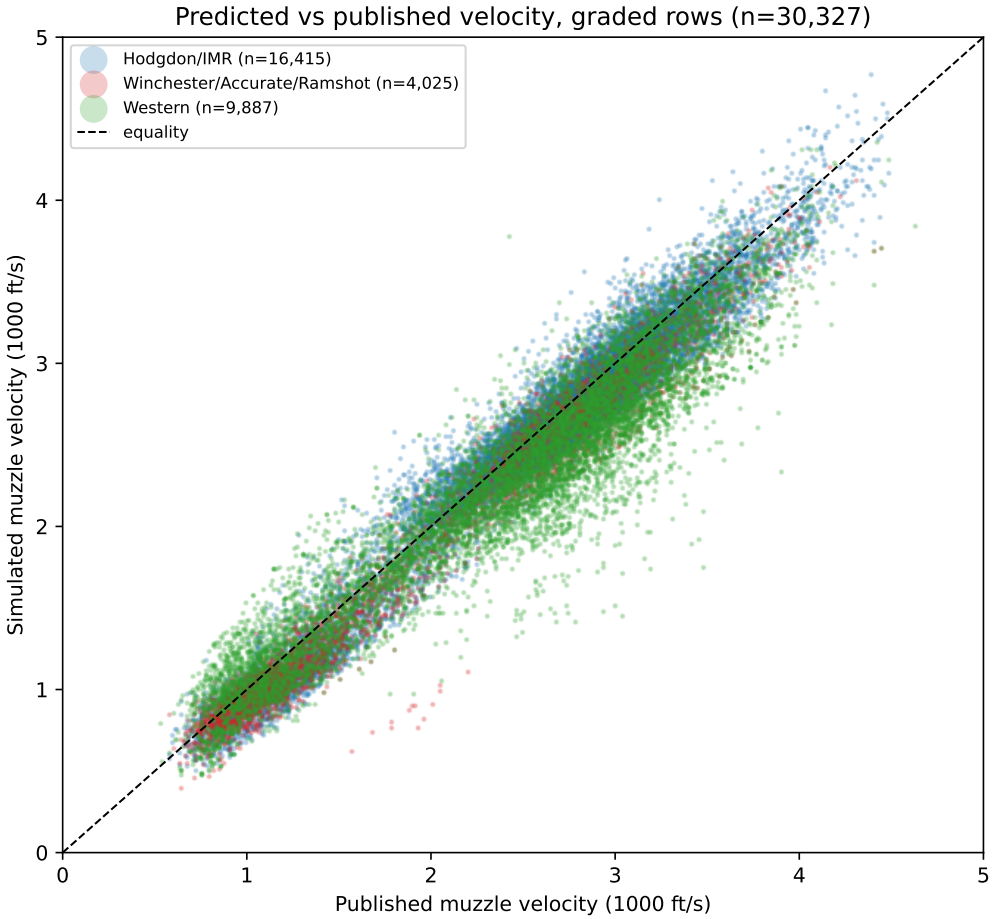


Figure 2: Predicted versus published velocity on the 30,327 graded rows, from the Phase-1 residual table. Colors identify publisher strata; the line is equality. Reprinted rows and start/max pairs are dependent.

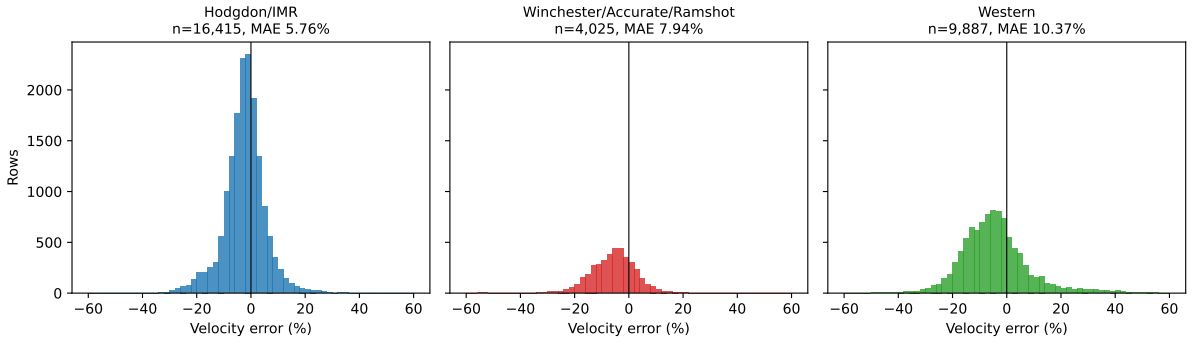


Figure 3: Velocity residual histograms by publisher stratum, with the velocity-only 60% rule. The Western distribution has larger error than the Hodgdon/IMR distribution.

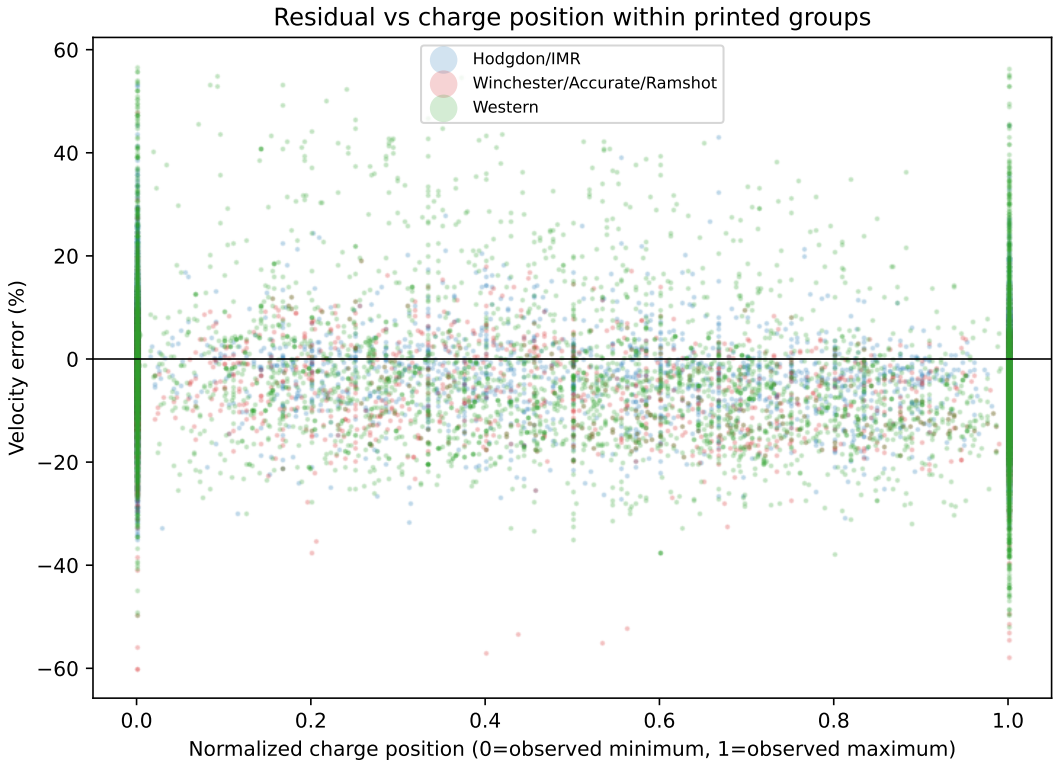


Figure 4: Residual versus normalized charge position within powder/cartridge/bullet groups. Zero and one denote observed group minimum and maximum, which need not be laboratory start/max endpoints. Single-charge groups are placed at one.

Table 6: Provenance pins for the head-to-head lane (full pins and per-powder provenance in the lane report, `f3_stats.json` and `f3d_provenance.json`).

Artifact	Pin
GRT arm	2021.2040-NIGHTLY, headless IPC, lean Pmax/v0 capture, 500 loads/session, batch driver restarts
IBC arm	`ballistics-calc` binary of record, sha256 `41ff483b19877bb515fe3381d84a1c62277bc1ea3ce7115964aef961202f25ea`, served ARMED env
IBC BA calibration	`powder_ba_calibration_grt.csv` sha256 `e8db7478...29d9` (63 data rows)
IBC peakloc windows	`peakloc_powder_grt.csv` sha256 `6c243c92...87c6` (123 rows); no window was fit to GRT outputs
Corpus snapshot	`state/revision/master_table.csv` sha256 `c8b36053...91f8`, 30,348 rows
Harness timeouts	REQ_TIMEOUT 30 s; socket read `settimeout(5)` governed (observed per-timeout wall 5.006 s median)
Vendored GRT caliber DB	`data/grt_database/caliber.csv` sha256 `bb37e5aa...e655`, 607 rows

4.3 Velocity results: near-parity on average, one clear exception

The average miss is nearly the same for both tools. Table 7 shows the paired differences: our mean absolute error runs 0.24 points higher than GRT's, with an interval that includes zero, so the data cannot distinguish the two average miss sizes. No equivalence margin was preregistered, so this is not a claim of equivalence — it is a statement that the measured gap is small and noisy. The one place the pair separates cleanly is w10: GRT was within 10% of the book on 4.9 points more loads. Both engines read the manuals low; our median bias sits closer to zero (-1.8% vs -3.8%), though the size of that advantage depends on how the data is grouped.

Table 7: Paired differences (IBC − GRT), percentage points, on 15,421 common graded rows. One-way CIs are percentile bootstraps clustered by powder (B = 10,000, primary of record); two-way CIs are Cameron–Gelbach–Miller normal approximations (B = 4,000 MAE / 2,000 signed mean and w10). Positive MAE gap = IBC worse.

Statistic (pp)	Point	CI95 powder (54 clusters)	CI95 powder×cartridge
mean signed diff (bias)	+1.594	[−0.164, +3.302]	[+1.141, +2.028]
MAE gap (accuracy)	+0.236	[−0.535, +1.063]	[−0.678, +1.150]
w10 gap	−4.948	[−10.699, +0.222]	[−7.056, −2.945]

Two features of the data matter when reading the table. The gap is tail-loaded: it shrinks under tighter lenses (to +0.08 points after discarding the 5% largest misses), meaning the average is set by a minority of bad loads rather than a uniform difference. And the engines' errors travel together — on 60.8% of loads both miss in the same direction, well above what chance would produce. That co-movement is expected, because the tools share case-capacity data and a burn-law family, and it means the shared lean toward low readings cannot be pinned on either engine alone.

4.4 Where the difference lives: calibrated vs never-seen powders

The gap explained in the last paragraph is mostly one thing: whether the powder is in our calibration table. A *fitted* powder is one whose burning inputs we have measured; a *never-seen* powder is one our engine must approximate from related powders, while GRT's database carries measured inputs for it. Table 8 splits the results accordingly.

Table 8: Cross-tab of MAE and w10 gaps (IBC − GRT; positive MAE gap = IBC worse), powder×caliber-class clustered percentile bootstraps, B = 4,000; the never-seen(all) row is powder-clustered (20 clusters, B = 4,000). Cell definitions are post hoc.

Cell	n	MAE gap [CI95]	w10 gap [CI95]
rifle | fitted	12,201	−0.27 [−1.17, +0.64]	−1.49 [−7.98, +4.38]
rifle | never-seen	1,162	+2.70 [+0.67, +4.49]	−22.03 [−38.0, −4.4]
pistol | fitted	465	+0.22 [−3.99, +2.39]	−0.65 [−16.7, +27.9]
pistol | never-seen	1,593	+2.32 [+0.78, +3.86]	−20.21 [−32.3, −8.0]
fitted (all)	12,666	−0.25 [−1.10, +0.67]	−1.46 [−7.90, +4.00]
never-seen (all)	2,755	+2.48 [+1.23, +3.74]	−20.98 [−31.5, −10.8]

On fitted powders the engines miss by the same amount. The small edge visible in our numbers is not trustworthy: 3,293 of those rows are exact calibration-fit loads from the same manual editions our calibration was fitted on, and when those rows are removed the edge reverses sign (Table 9) with every interval spanning zero. The correct reading is that the two tools are equivalent on powders we have calibrated, within what this design can detect.

Table 9: Exclusion sensitivity of the fitted cells (MAE gap, IBC − GRT, pp; powder×caliber-class clustering, B = 4,000, seeds 20261202/03).

Arm	rifle | fitted n	MAE gap [CI95]	fitted(all) n	MAE gap [CI95]
no drop	12,201	−0.27 [−1.19, +0.68]	12,666	−0.25 [−1.13, +0.67]
(a) drop 3,293 exact rows	8,952	+0.25 [−0.91, +1.24]	9,373	+0.24 [−0.84, +1.25]
(b) drop 3,478 fit-edition rows	8,769	+0.27 [−0.91, +1.28]	9,188	+0.27 [−0.85, +1.26]

On never-seen powders the separation is real and favors GRT: 2.3–2.7 points lower MAE and roughly 20 points higher w10, with intervals that exclude zero, robust to every test we applied (cluster-weighted bootstrap, leave-one-cluster-out, and dropping the dominant powder in each cell — Accurate 4350 in rifle, Hodgdon Longshot in pistol). The reading is the actionable part. The gap is a *coverage* gap, not a physics gap: where we have measured a powder's inputs we match GRT, and where we have not, guessed inputs produce guesses. Closing the powder table is therefore the direct response, and it is the expansion program this paper describes elsewhere.

4.5 Coverage: who answers, and who declines

Our engine produced a prediction for all 17,676 loads. GRT did not answer 2,254 of them, in three ways that mean different things and are kept separate for that reason. It explicitly refused 823 loads (4.66%), concentrated in charges above 100% case fill — declining there is defensible behavior under its own density model — and in eight cartridges whose entry in our authored files is unverified. It returned empty results for 29 (0.16%), reclassified as invalid. And 1,402

(7.93%) hit our harness's five-second read timeout; these are unresolved — they reflect a limit of our test rig, and a re-run on an unrestricted rig is registered as open work — not a verdict about GRT. We do not combine the three into one failure number because they do not share a meaning.

Table 10: Coverage outcomes on 17,676 paired loads (last-write-wins). The three GRT failure components are reported separately by design; no combined failure percentage is computed.

Outcome	GRT	IBC (ours)
answered	15,422/17,676 (87.25%)	17,676/17,676 (100.00%)
refusal	823 (4.66%)	0
timeout (UNRESOLVED, harness)	1,402 (7.93%)	0
empty capture (reclassified invalid)	29 (0.16%)	0
lens exclusions	0	1
scored	15,422 (87.25%)	17,675 (99.99%)

On the 2,254 loads GRT left unanswered, our engine produced a prediction for every one. Table context: 4.75% MAE on the fitted-powder rows of that subset, 8.55% on its never-seen rows, 3.54% on the overfill rows. Answering compressed charges that sit beyond printed support is reported here as a measurement; it is not scored as a virtue without accuracy and pressure evidence behind it.

4.6 Pressure: measurement only

Everything in this subsection is a measurement, not a safety statement: neither tool's pressure output is a measured chamber pressure, and neither supports a safety claim. Both engines under-read the published pressures by 13–17% on average, and only about 27–32% of loads land within 10% of the book figure. Table 11 reports each arm against the book and the two arms against each other, restricted to the 8,584 rows whose books print pressure in PSI as issued.

Table 11: Pressure measurements, 8,584 as-printed-PSI captured rows. Left: IBC vs GRT Pmax agreement. Right: each arm vs published pressure.

IBC − GRT Pmax agreement	
mean	+8.41% (CI95 powder [+1.05, +15.33])
median	+5.70%
|diff| MAE	25.25%
within 10%	25.77%
Arm vs published pressure	**mean signed / MAE / within 10%**
GRT	−16.48% / 20.70% / 26.96%
IBC	−13.02% / 20.21% / 32.40%

The paired gap between the arms (+3.47 points, interval spanning zero) is smaller than the shared distance from the books. On the subset GRT declined to answer, our engine's pressure sits −10.1% from the book on average. The shared under-reading mirrors the velocity result and is consistent with the low-bias discussion in Section 4.3; this data cannot settle whether the cause is book convention or model behavior.

4.7 Claims of record and open items

The lane report behind this section passed five rounds of adversarial review, each round remediating enumerated findings before the next; the seven sentences below are its claims of record, quoted verbatim, and no statement in or beyond this section extends them.

1. **Scope.** "On 15,421 paired loads (hodgdon + Winchester lanes; 58.2% of the graded corpus; velocity only; GRT 2021.2040-nightly run with default inputs; a protocol deviation from the feasibility gate is noted), . . . "

2. **Accuracy.** "IBC's mean absolute velocity error was 0.24 pp higher than GRT's (95% CI -0.54 to $+1.06$, clustered by powder; two-way powder $\times$ cartridge CI -0.68 to $+1.15$; cluster units chosen post hoc). The gap shrinks to $+0.08$ pp after trimming the 5% largest errors and does not change sign under any lens. No equivalence margin was prespecified, so equivalence is not claimed. GRT was within 10% of published velocity on 4.9 pp more loads."

3. **Bias.** "Both engines under-predicted published velocities on average. IBC's median bias was closer to zero (-1.8% vs -3.8%); the size of this difference is sensitive to the clustering choice."

4. **Rifle.** "On rifle loads, MAE point estimates were equal (5.10% vs 5.11%; gap CI spans zero; the w10 gap CI spans zero under powder clustering), while GRT was within 10% on 3.3 pp more loads. On fitted powders, 3,293 graded rows (3,288 in fit editions) are exact calibration-fit loads from the same Annual-Manual source (. . .), so the fitted-cell result is partly in-sample — and it is not sign-stable: dropping the exact rows flips the fitted-cell MAE gap from -0.27 pp to $+0.25$ pp, and dropping all 3,478 fit-edition rows gives $+0.27$ pp, with every CI spanning zero."

5. **Out-of-sample signal.** "IBC's velocity calibration draws on heterogeneous sources: a 216-load Hornady harness (pressure targets anchored to GRT's own Pmax), legacy and guide fits, and Annual-Manual fits that share source with this corpus (3,293 graded rows are exact fit-set loads). On powders absent from the calibration table, GRT's MAE was lower by 2.3–2.7 pp and its w10 higher by $\sim$20–22 pp (wild-cluster p = 0.008–0.041, 16–17 clusters, post-hoc cells), with the rifle effect concentrated in Accurate 4350 and the pistol effect in Hodgdon Longshot. The rifle MAE interval of record is the wild test-inverted $[+0.24, +4.62]$. With each cell's dominant powder removed the gaps persist at wild p = 0.017–0.036, test-inverted CIs still excluding zero (rifle MAE $[+0.22, +2.69]$; rifle w10 $[-23.66, -0.83]$; pistol MAE $[+0.24, +3.48]$; pistol w10 $[-29.86, -3.70]$)."

6. **Coverage.** "IBC returned an answer for all 17,676 authored loads. GRT, in our headless harness, returned empty captures on 0.16% (reclassified invalid), explicitly refused 4.66%, concentrated in charges above 100% fill and in eight cartridges whose authoring is unverified, and a further 7.93% hit a 5-second harness read timeout and are unresolved."

7. **Co-movement.** "The engines' errors are modestly correlated (Pearson r = 0.23, Spearman $\rho = 0.16$; same-sign excess $\approx$1.2 pp over a within-cell permutation null). Because they share case-capacity and component data and a burn-law family, and because IBC's

harness-refit pressure calibration was anchored to GRT outputs, and part of its velocity calibration was fit on this corpus's own Annual-Manual source, the common low bias cannot be attributed to book convention."

The timeout re-run is complete: all 1,402 were retried on a non-burstable rig with a 300-second cap, and 89.5% answered (5.8% refused, 4.6% still time out) — the attribution question is resolved. Two comparisons remain open, each preregistered and none preempted by anything above: a GUI-entry test of the eight refusal cartridges to separate our authoring from GRT's database; and leave-one-source-out recalibration refits, which would turn this fit comparison into the only out-of-sample test the design can support. A fourth item is closed as impossible for now: no current GRT release is available offline, so every GRT figure in this section describes version 2021.2040-NIGHTLY.

5 The engine

5.1 Model class, state, and component summary

The engine is a zero-dimensional, lumped-parameter, single-phase interior-ballistics solver written in C (version v1.1.14 for the corpus measurement). "Served" denotes the deployed configuration; experimental configurations are labeled separately. The default burn path uses a tabulated L-curve for powders with those records, while the fallback path uses a pressure-dependent Saint-Robert law for powders without them. L-values tabulate burn progression as the product of a burn-rate coefficient and a form function; the L-curve interpolates those values. It integrates the standard energy-and-motion system for a burning granular propellant charge pushing a projectile down a barrel [3, 4, 6]: the chamber plus swept bore is a single spatially uniform control volume, the projectile is a point mass, and the propellant bed is reduced to a burned-mass fraction $z(t)$ whose rate is set by a regression law times a burning-surface form function. The engine carries no two-phase granular-bed physics, no bed compaction, and no projectile rotational degree of freedom; Section 7 collects the consequences.

The state advanced per time step is (z, x, v) (burned fraction, projectile travel, projectile velocity) plus running accumulators (wall heat, escaped gas mass, primer energy and gas mass) and, in the shipped conservative mode, the internal energy U of the propellant gas. Pressure and temperature are not states: both are recomputed algebraically each step from the equation of state and the energy ledger. The stiffest coupling, the feedback between burn rate and pressure, is lagged one step, and the only true differential equations are the burn-fraction and projectile-dynamics equations; Section 5.9 reports the available time-step probes; these do not bound error across the full corpus.

The governing relations are:

- *Burn rate*: $\mathrm{d}z/\mathrm{d}t = B_{a,\mathrm{eff}}\,\phi(z)\,f_P(P)$, where $B_{a,\mathrm{eff}}$ is the effective burn-rate coefficient of the loaded propellant and f_P a pressure-coupling factor: the classical Saint-Robert factor $\max(P, 1\,\mathrm{bar})^n$ on the fallback path for powders without L-curve records, identically one on the served path (Sections 5.3, 7).

- *Gas energy*: in the shipped conservative mode the solver integrates $\mathrm{d}U = m_c Q_{ex}\,\mathrm{d}z + \mathrm{d}E_{\mathrm{primer}} - P\,\mathrm{d}V - \mathrm{d}Q_{\mathrm{wall}} - u\,\mathrm{d}m_{\mathrm{gap}}$ (Section 5.2).

Table 12: Canonical form versus served implementation, by component.

Component	Canonical form	Served implementation
Burn regression	Saint-Robert $r = aP^n$ with geometric form function	Pressure-independent at default ($f_P = 1$); P^n on the fallback path; empirical three-stage form function
Equation of state	Noble-Abel $P(V - bm) = mRT$	Shared covolume algebra with density-dependent covolume correction; dense corner latched and flagged
Energy equation	First law $\mathrm{d}U = Q_{ex}\,\mathrm{d}m - P\,\mathrm{d}V -$ losses	Served default; ledger instrumented and audit-closed (Section 5.9)
Wall heat	Convective $hA(T_{\mathrm{gas}} - T_{\mathrm{wall}})$	Canonical form; $h_{\mathrm{eff}} = 80{,}000$ W/(m^2 K), in-sample tuned (Sections 5.5, 7)
Gas inertia	Lagrange gradient or Sebert effective mass	Sebert only, $\lambda = 0.65$ default; no Lagrange correction, no explicit gas kinetic-energy sink
Bore friction	Quasi-static or constant Coulomb	Empirical velocity-dependent model with shot-start-calibrated breakaway; stall pathology
Engraving	Travel-bounded, shot-start-anchored	Sine profile over the engraving interval, peak 0.25× shot-start pressure (fitted)
Primer/ignition	Initial condition or ignition delay	Timed energy and gas-mass injection with shape functions; per-load ignition anchor (quantified leak, Section 7)
Blowby (cylinder gap)	Choked orifice with stagnation-enthalpy carry-off	Mass matches the choked-orifice result; carry-off uses internal energy, not enthalpy

- *Equation of state*: chamber pressure follows the Noble-Abel relation with covolume applied to the free volume, consistent with the energy ledger [4], with a density-dependent covolume correction and a latched guard on the dense corner of the EOS (Section 5.2).
- *Projectile motion*: $m_{\mathrm{eff}}\,\mathrm{d}v/\mathrm{d}t = P\,A - F_{\mathrm{friction}}(v) - F_{\mathrm{engraving}}(x)$, with a velocity-dependent friction model, a travel-bounded engraving force, a Sebert effective mass, and per-cartridge SAAMI test-barrel geometry from bundled data files (Section 5.7).

Table 12 compares the classical form with the implementation. The pressure-independent burn law and fitted loss constants restrict extrapolation (Section 7).

5.2 Conserved-gas energy bookkeeping

In the shipped conservative mode, U is a true state variable, seeded at ignition to $z_{\mathrm{init}}\,m_c Q_{ex}$ and integrated each step as the full ledger:

$$\mathrm{d}U \;=\; m_c Q_{ex}\,\mathrm{d}z \;+\; \mathrm{d}E_{\mathrm{primer}} \;-\; P\,\mathrm{d}V \;-\; \mathrm{d}Q_{\mathrm{wall}} \;-\; u\,\mathrm{d}m_{\mathrm{gap}}, \tag{2}$$

the five terms being chemical release, primer injection, expansion work delivered at the projectile base, convective wall loss, and internal energy carried off by gap gas, with a non-negativity floor on U.

Pressure comes from the Noble-Abel relation applied to this state:

$$P \;=\; (k-1)\,\frac{U}{V_{\text{avail}}}, \qquad V_{\text{avail}} = V - b_{\text{eff}}\, m_{\text{gas}}, \tag{3}$$

where V is chamber plus swept-bore volume and V_{avail} is the chamber free volume less the covolume term, with base covolume $b = 0.001$ m^3/kg, a per-propellant catalog covolume ratio, k the specific heat ratio, and m_{gas} the burned gas mass plus injected primer gas minus escaped mass. The served surface carries a density-dependent covolume correction,

$$b_{\text{eff}} \;=\; b\,(1 + \beta\,(\rho_g/\rho_{g,0})^p), \tag{4}$$

with compiled defaults $\beta = -0.05$, $p = 2$, and $\rho_{g,0} = 300$ kg/m^3, where $\rho_g = m_{\text{gas}}/V_{\text{avail}}$ is gas density in SI units. The correction gives a covolume that shrinks as the gas densifies, a first-order stand-in for the attractive term the two-parameter Noble-Abel form omits, fitted on the observation that a constant-covolume gas is too stiff exactly where dense loads over-predict pressure. Gas temperature is algebraic, $T = U/(m_{\text{gas}}\, C_v)$ capped at 5,000 K, with $C_v = R/(k-1)$ and a single global specific gas constant $R = 330$ J/(kg K); per-powder molecular weights are not tracked (Section 7).

The free-volume expression carries a numerical floor at 1% of chamber volume; reaching it means the load sits beyond the Noble-Abel validity envelope, at loading densities past the dense-gas limit, and the computed peak pressure is a clamp artifact. The engine latches the condition, reports it as an explicit dense-extrapolation flag on its output contract, and fails the pressure-safety verdict closed to an "unverified, dense" classification (Section 5.8).

The October energy-ledger audit (Section 5.9) is a bookkeeping check: it measured the unaccounted residual of this ledger over 206 live loads at a mean of 0.014% and a maximum of 0.048% of input energy, against a 3% adoption gate for any physics change. The conservative mode was not the original default: an earlier variant recomputed gas energy from the burn fraction each step without the expansion-work debit, violating the first law at −31.8% to −39.3% non-closure on a 12-load audit; that variant was retired to a regression-detection role and never serves.

5.3 Burn model and form function

The served burn law is pressure-independent at default, the engine's largest structural divergence from classical theory. The pressure exponent on the served path is zero, so the pressure-coupling factor f_P of Section 5.1 is identically one; a Saint-Robert law with an interpolated per-powder pressure exponent runs only on the fallback path for powders without L-curve records. The served rate is

$$\frac{\mathrm{d}z}{\mathrm{d}t} \;=\; 1000\, B_a\, s_{ba}\, M_{ba}\, \phi(z), \tag{5}$$

where B_a is the catalog burn-rate constant in ms^{-1} (the factor 1000 converts to s^{-1}), s_{ba} is a global scalar equal to 1.10 that the engine source describes as compensating for under-modeled energy losses, and M_{ba} is the per-powder calibration multiplier of Section 6.1, carrying a weak load-condition selectivity through a fill-density term

$$M_{ba} \;=\; s_{\text{mult}} \cdot \min\!\left(1,\; \rho_f/\rho_{f,0}\right)^{0.20}. \tag{6}$$

Here ρ_f is charge mass in grains divided by case volume in cm^3, and $\rho_{f,0}$ uses the same gr/cm^3 units. It is distinct from the gas-density reference $\rho_{g,0}$. The fill-density term is active below

saturation on 7,689 of 21,114 fitted rows in the Phase-2 activation audit (36.4%); its units are internally consistent. An environment hook exists to restore a pressure exponent; it is unset on every served surface. Served accuracy is carried by per-powder tables fitted to published data, so inside calibration coverage the engine behaves as a physics-shaped lookup machine, and outside that coverage the pressure dependence of burn rate is not modeled.

The form function $\phi(z)$ is not derived from grain geometry. It is a three-stage empirical construction: a progressive ignition stage $\phi = 1+a_0 z$ for z below z_1 (defaults $z_1 = 0.2$, $z_2 = 0.85$); a main stage interpolating the nine catalog L-values sampled at $z = 0.1$ through 0.9, normalized before use because catalog L-values encode the product $B_a\phi$ rather than ϕ alone; and a degressive tail beyond z_2. A smoothstep blend transitions between the polynomial and L-curve branches, and ϕ is hard-clamped to $[0.1, 10]$ as a numerical guard. This is the empirical analogue of the textbook progressivity concept, with burn progression carried by tabulated data rather than a grain model. The z-independent multiplier M_{ba} cannot express errors in progression shape; the one z-shaped calibration lever is the per-powder peak-location table of Section 5.9.

5.4 Ignition and primer

Shot start is initialized by solving the initial burned fraction from the equation of state at the shot-start pressure, ignoring covolume at small z and clamping the result to a physical band between 0.005 and 0.05.

Primers are configured from a per-cartridge table carrying energy, gas mass, injection duration, and a rate-profile shape. Within the injection window, a normalized rate of rectangular, truncated-Gaussian, or triangular shape drives simultaneous energy and gas-mass injection, and under conservation the same rate injects directly into the state U; this exceeds the canonical minimum treatment, in which a zero-dimensional model seeds ignition as an initial condition or a fixed delay. A per-powder ignition-transient table can further gate both burn paths with an exponential delayed-onset factor under fill-density validity gates.

On the serving path, an ignition anchor computes the ignition delay from the load's own maximum-charge fill density where one is available (otherwise from the powder's catalog reference density), through

$$\tau \;=\; \tau_0 \left((\rho_{\text{anchor}}/\rho_f)^Q - 1 \right), \tag{7}$$

capped at 0.30 ms with $\tau_0 = 0.30$ ms; both densities use the fill-density convention. The construction sharpens maximum-charge prediction and ties each start charge to a quantity the calibration pipeline has already seen; the resulting self-recall is quantified in Section 7. The ignition transient itself remains a named unmodeled channel: the engine project's preregistered flame-spread studies tested a segmented spread law, a constant-velocity spread law, and a convective-ignition law, and measurement rejected all three, bracketing the gap between a constant fast spread that over-delays magnum populations and convective spread two to three orders of magnitude too slow at published grain properties.

5.5 Wall heat transfer

The default wall-loss law matches the canonical convective form:

$$\frac{\mathrm{d}Q}{\mathrm{d}t} \;=\; h_{\text{eff}}\, A(x)\, \max\!\left(0,\; T_{\text{gas}} - T_{\text{wall}}\right), \qquad A(x) = A_{\text{chamber}} + \pi D x, \tag{8}$$

where the wetted area grows as the projectile exposes bore, $h_{\text{eff}} = 80{,}000$ W/(m^2 K), and the wall temperature is 300 K. The coefficient was selected against the engine's regression anchors,

with burn-rate band multipliers adjusted under the same constraint. These anchors reuse tuned loads, so the coefficient has no independent heat-transfer validation (Section 7).

Two alternative correlations are implemented behind environment gates, default-off after rejection on measurement: a pressure-dependent coefficient form, swept across fifteen configurations, traded peak pressure against velocity in roughly a 1:3 ratio with no jointly feasible setting; and a radiative-temperature hybrid with decay-weighted wetted area. The engine project's diagnostic reading: near-peak over-pressure on the worst loads is burn-progression curve shape at matched expansion work, not a wall-heat magnitude defect.

5.6 Cylinder gap and bore friction

For revolver cartridges, gas escapes through the cylinder gap. The mass side matches the canonical choked-orifice result,

$$\frac{\mathrm{d}m}{\mathrm{d}t} \;=\; C_d\, A\, P \sqrt{\frac{k}{R\,T}} \left(\frac{2}{k+1}\right)^{\frac{k+1}{2(k-1)}}, \tag{9}$$

with a discharge coefficient of 0.65, a default gap width of 0.15 mm, revolver auto-detection with per-load override, gas temperature floored at 300 K and capped at 5,000 K, and a per-step cap of 10% of remaining gas mass. The energy side deviates from the canonical stagnation-enthalpy carry-off: the ledger removes the escaped gas's specific internal energy $u\,\mathrm{d}m$, omitting the nozzle-flow work term, a stated approximation with revolver loads as the sensitive population (Section 7).

Bore friction replaces the quasi-static textbook picture with an empirical velocity-dependent form in which friction decreases with velocity, parameterized in four jacket-material sets (copper, gilding metal, lead, monolithic). The classical shot-start concept re-enters through a calibrated breakaway variant that anchors the v-to-zero friction maximum to the shot-start pressure, together with a static/kinetic split that scales friction by 0.6 while the projectile slides; this calibration is load-bearing on the served surface, because without it low-fill-density loads stall with systematic safety-direction errors. The model carries a known pathology: because its friction rises as v approaches zero and the integrator floors v at zero, a projectile decelerating after pressure collapse is trapped at full breakaway-level friction, an absorbing stall state. A Coulomb-consistency cap prevents single-step sign flips; it does not remove the absorbing state (Section 7).

5.7 Projectile dynamics, engraving, gas inertia

Dynamics follow the net force through the effective mass,

$$m_{\mathrm{eff}}\,\frac{\mathrm{d}v}{\mathrm{d}t} \;=\; P\,A - F_{\mathrm{friction}}(v) - F_{\mathrm{engraving}}(x), \tag{10}$$

with a semi-implicit Euler update (velocity from the current-step force, position from the new velocity), a $v \geq 0$ floor, and the Coulomb-consistency impulse cap of Section 5.6.

The effective mass is the Sebert form

$$m_{\mathrm{eff}} \;=\; m_p + \lambda\, m_g, \tag{11}$$

with λ taken from cartridge and caliber tables where present and defaulting to 0.65, above the classical 1/2 and best read as a fitted absorber for omitted zero-dimensional effects (Section 7).

An optional dynamic-mass mode scales the gas term with the burned fraction plus injected primer gas, so full gas inertia is not charged at $t = 0$. Pressure acts on the projectile base at the chamber-average value: there is no Lagrange base-pressure correction and no explicit gas kinetic-energy sink in the energy budget. The audit instrumentation treats gas kinetic energy as the implied half-λ term; the dynamics sub-ledger (gas work against bullet and gas kinetic energy plus friction and engraving work) closed within 0.5% on all twelve loads of the October 2026 audit, with the bullet's share of gas work between 75.5% and 86.3%. There is no rotational degree of freedom.

Engraving is a sine-profile force over the engraving interval, from the freebore jump through a 2 mm engagement length:

$$F_{\text{eng}} \;=\; F_{\text{peak}}\,\sin\!\big(\pi(x - x_0)/L_{\text{eng}}\big), \qquad F_{\text{peak}} = 0.25\,p_s A, \tag{12}$$

with the 0.25 multiplier fitted on the engine project's harness sweep. This matches the classical travel-bounded, shot-start-anchored picture; the sinusoidal shape and the fitted multiplier are refinements without canonical status. Engraving work is debited in the ledger, though it is small (0.04 to 0.36 J in the audit loads). Shot-start, rifling-contact, and engraving-complete milestones are recorded on the output trace.

Barrel exit terminates the simulation: the loop breaks when travel reaches barrel length, capturing muzzle velocity, exit burn fraction, and barrel time; an exit burn fraction below 0.98 raises an incomplete-burn warning. There is no post-exit expansion phase. A run whose projectile has not exited by the end of the fixed 5 ms window reports an explicit truncation flag rather than presenting a mid-bore velocity as a muzzle velocity.

5.8 Pressure standards and fail-closed safety verdicts

The pressure-ceiling lookup uses exact matching before fuzzy matching: a source-backed override table of verified transducer-era ceilings, then exact matches against the legacy table (raw and normalized), then a prefix-fuzzy pass with ambiguity checking. Anything unresolved returns an unknown ceiling, and the safety verdict fails closed: the verdict reads unknown, its validity flag reads false, and the percent-of-ceiling figure is omitted, never reported as safe. The ordering exists because a longest-substring matcher once resolved a +P cartridge onto the non-+P ceiling row.

Ceiling entries carry a unit-standard classification (crusher-era CUP versus transducer psi). Conversion between the two systems runs only against verified pairs transcribed from the rifle and pistol/revolver SAAMI standards [5, 16], by normalized-exact lookup only; where no verified pair exists, no conversion is applied. On the current table the machinery applies one pair whose ratio is 1.0, so no served ceiling value moves.

5.9 Numerical algorithms: integrator, guards, convergence, energy closure

Integration uses a fixed-step semi-implicit Euler scheme: velocity from the current-step force, position from the new velocity (Section 5.7); burn fraction by forward Euler on the current-step rate; gas energy by explicit accumulation of the five ledger terms of Section 5.2. The default step is 0.5 microseconds over a fixed 5 ms window, not adaptive; an environment probe can override the step size with a window-preserving step budget. Algebraic pressure and temperature reduce the differential state dimension. Energy closure tests accounting consistency; it cannot bound discretization error or validate the burn law [12].

Two historical convergence probes cover limited populations. An internal study on the engine project's 216-load harness found headline metrics unchanged from 0.5 down to 0.05 microseconds (median per-load velocity drift 0.000%, mean at most 0.02%). A later probe reran five representative loads at 0.25, 0.5, 1, and 2 microseconds and moved peak pressure by at most 0.23% and velocity by at most 0.08% across the fourfold range. These probes support small numerical error on the tested loads. A stratified convergence study covering dense, guard-active and stalled cases over the full corpus is still missing.

Stability guards, all in code: the covolume floor of Section 5.2 (latched and flagged when hit); non-negativity floors on the gas energy state; the 5,000 K temperature cap and a gas-mass floor; a pressure-coupling clamp band on the fallback burn path; the Coulomb-consistency friction impulse cap; the 10% per-step blowby cap; the $v \geq 0$ floor; the form-function clamp $[0.1, 10]$; and non-finite guards on the optional heat-law candidates. The energy floors can mask unphysical states, a correctness-review item the engine project records against itself; the friction cap and velocity floor jointly create the stall state of Section 5.6.

Energy closure is instrumented permanently. Write-only audit accumulators integrate gas work, friction work, engraving work, and gap carry-off alongside the state each step and never feed physics; the end-of-run block computes energy in (chemical release plus primer energy) against energy out (bullet and gas kinetic energy, wall heat, friction and engraving work, gap carry-off, residual gas energy) and reports the unaccounted residual. The adoption gate for any physics change is an absolute residual under 3%; the historical regression threshold is 0.07% on the 206-load ledger audit, whose observed maximum was 0.048% (Section 5.2). The separate 12-load audit covers the dynamics sub-ledger and the retired energy variant; the 216-load battery is a velocity-calibration population. A sensitivity pre-pass measures powder temperature sensitivity by two guarded solves; per-powder primer-impulse, ignition-transient, and peak-location tables load from data files alongside the main databases; the loaders fail loud, so a missing or malformed file disables its layer rather than silently substituting defaults.

5.10 Base data and its provenance

Cartridge dimensions, projectile records, and the propellant thermodynamic records (Q_{ex}, covolume, burn-rate coefficient B_a, pressure exponent, form-function L-tables, and per-lot metadata) derive from the GRT database [2]. The engine ingests the GRT propellant file and its lot metadata directly; each powder may carry multiple lots with date stamps and a quality flag, and only lots in the servable mode are usable for serving. the records carry lot dates and quality flags. Redistribution permission and an archival source version have not been established for this submission. The engine's own additions above that base are the calibration layer of Section 6, the pressure-convention bridge, and the safety verdict machinery that consumes the results.

5.11 What the model is

The engine is a consumer-class simulator in the QuickLOAD/GRT tradition: a thermodynamic bulk-charge model, not a multi-dimensional reactive-flow computation. It resolves no intra-chamber flow, no grain-level flame spreading, and no case-head expansion. Its outputs at the muzzle are velocity and peak pressure. This places it in the same model class as the tools handloaders already use; the reported errors apply to the published-load populations in Section 3.

6 Calibration protocol and selection

6.1 Parameters, data access and checks

The burn-rate table carries a scalar s_{mult} and a fill-density reference $\rho_{f,0}$ for each powder. A powder without a native usable propellant record can be represented by a donor's thermodynamic record, called a carrier lot. This record must reproduce the donor's outputs before fitting. Four carrier checks in the evidence reproduce 105/105, 115/115, 103/103 and 119/119 sampled outputs for H414, IMR 4955, H380 and MagPro respectively. Powder-keyed ignition and burn-shape tables must accompany the carrier, because those lookups change predictions even when thermodynamic fields agree.

The coarse scalar grid is $\{0.70, 1.00, 1.30, 1.60, 1.90, 2.20, 2.50\}$ with a local refinement at ± 0.15. The density grid for the cited campaign is $\{4.50, 7.76\}$ gr/cm^3. The coarse grid reaches 1.60 exactly; the grid is not arithmetically defective. Donor values may lie outside the grid, so donor-versus-grid comparison is a search-policy comparison over unequal domains. The objective minimizes the absolute signed median velocity error on a bounded fit subsample; tails enter later checks. This objective can hide opposing errors across cartridges.

6.2 Adoption checks and deviations

The project protocol requests at least 150 cross-edition-matched half-rows in the candidate's total corpus and caps the cartridge-stratified fit subsample at 150. The original wording incorrectly required 150 fit rows; observed fit sets of 41 (MagPro), 24 (IMR 8133) and 10 (IMR 4955) do not meet such a minimum. Their thin fit sets are recorded deviations from that wording, without evidence of prospectively registered exemptions. An external timestamped protocol registry is unavailable. The checks are therefore described as project-prespecified, not independently verified preregistration [11].

Fit editions supply the grid objective; other editions supply the set historically called the holdout. Control, donor and grid arms are all scored on it. Adoption requires absolute pooled velocity median at most 10%, non-increasing quarantine counts and worst velocity error, and no deterioration of either printed-pressure series. A fit/control sign flip with both magnitudes exceeding 1% fails. Cartridge medians at or below -5% coexisting with medians at or above $+5\%$ also fail. Missing pressure series make the prescribed test unevaluable; a resulting hold is not a measured pressure failure.

A full-corpus rerun checks determinism and regression, including exact reproduction of selection-set statistics. Because that corpus contains the selection set, the rerun supplies no independent confirmation. Selecting the best arm on the holdout makes that set a selection set, with winner's optimism [10]. No three-way untouched test or nested selection correction was run. All following candidate statistics are therefore retrospective selected-arm descriptions.

6.3 Selected candidates and holds

MagPro has 355 cross-edition-matched half-rows. Its selection set has 314 rows across 17 cartridge clusters. The unfitted control gives 11.97% MAE, the grid row $(1.60, 7.76)$ gives 3.73%, and the Magnum donor row $(1.5152, 13.4070)$ gives 3.11% with signed median $+0.21\%$. The control and donor are different arms: the Phase-1 check finds zero exact matches across their 314 predictions. The donor's MAE advantage over the grid is 0.625 percentage points; its paired cartridge-bootstrap interval is $[-1.33, 0.16]$ for donor minus grid, which includes zero. The

Table 13: Protocol deviations and their interpretation.

Issue	Resolution and remaining limit
MagPro fit editions	AM21 has no Ramshot rows. The effective fit source is AM23 alone (41 half-rows); 314 other half-rows supply selection.
Thin fit slices	Below the originally stated minimum; no retrospective claim of a prospectively authorized exception.
Carrier lookups	Ignition/shape tables must be cloned with the lot; earlier partial clones cannot establish donor equivalence.
ADI single snapshot	A row split within one capture replaces an edition split. Its 12- and 6-row selection subsets are weak tests.
Earlier catalogue additions	The 45-powder catalogue does not imply that every historical addition passed the later protocol unchanged.

grid and H1000 donor coincide in parameter values and outputs, so they are not independent evidence. All 18 full-corpus regression checks pass for the selected Magnum row.

IMR 8133 is a second documented accepted candidate; the two accepted cases do not account for every historical addition to the catalogue. IMR 4955 remains held despite 99.3% within 10% on 448 selection rows because opposing cartridge medians trigger the sign rule. A worst error of 13.01% is consistent with that within-10% rate, which allows some rows outside the interval. H414 and H380 have no usable fit-edition rows under the specified design and remain unevaluable without an amendment.

The ADI APS 450 and APS 650 candidates fail full-corpus regression checks after selection on small subsets. Vihtavuori candidates lack printed pressure and cannot complete the prescribed adoption test. No velocity-only success from these campaigns establishes pressure accuracy. Their supplementary data are not added to the combined-corpus denominator.

7 Limitations

7.1 Western model error and inherited lot data

The Western stratum's 10.371% MAE is genuine prediction error against its stated target data. The Accurate-lot data drift is a model/data sensitivity finding: a thermodynamic record chosen from a different lot or fit state changes predictions even when a powder name stays the same. The uncertainty includes inherited burn-shape records and assumed geometry, as well as calibration coverage. Default-powder Western rows give 13.356% MAE against 8.092% for fitted-powder Western rows. Neither the lot sensitivity nor the absence of a native fitted row excuses the resulting error.

The Western pipeline's documented 14-inch barrel default leaves 1,637 of the stratum's 9,887 graded rows, 16.6%, simulated at an imputed length, and that cell is the stratum's worst: its MAE is 14.580% with 27.2% of rows beyond 20% error, against 9.536% and 8.3% for the 8,250 rows with a printed barrel. A preregistered decision rule, fixed before the audit ran, asked whether the resolved-only MAE fell below 8.871% (the stratum figure minus 1.5 points); at 9.536% it does not, so the genuine-model-error reading of this section stands, with the imputed subset disclosed as the stratum's largest measured weakness. The audit scores the committed freeze-era residuals, and its verdict is scoped to that battery of record.

The corrected full-pipeline no-exit rescan reproduces the Western record byte-exactly and

finds one no-exit observation in 9,898 attempts, already excluded. Together with the two no-output observations in the other stratum, the three no-exit cases leave the graded corpus clean of mid-barrel velocities. The Western error cannot be attributed to failed-exit predictions entering the reported metric. Earlier reconstruction diagnostics with non-finite missing-barrel inputs are superseded by the corrected rescan; their apparent shift is not a revision of the Western accuracy estimate.

Printed-source maxima under-represent pressures near and beyond the standard ceiling. The upstream 1.10-MAP filter further narrows that region. Missing per-shot dispersion and the 775 recovered and 388 assumed Western charges prevent a decomposition into model-form error, input error and laboratory variation. A printed velocity can also be rounded or revised across editions.

7.2 Pressure-independent burning and component ablations

The L-curve burn law uses pressure exponent zero. At a fixed burn fraction and fitted coefficient, a pressure change does not change its burning rate. This removes the classical Saint-Robert feedback needed to model how burning responds when charge or geometry changes outside calibration coverage. Replacing it by the existing pressure-dependent fallback while retaining shipped calibration gives 49.116% MAE in the Phase-2 screened arm; only 13,171 of 30,348 attempts survive its grading rule. This failed law-swap experiment is a dead end under those coefficients. It does not refute a separately refitted Saint-Robert model, because the two paths use differently scaled burn coefficients and the swap reuses L-curve-era calibration. The reconstruction-input caveat for corpus arms also applies here.

Table 14: Individual component removals on the separate 207-success fitted-powder battery. Positive change means larger velocity MAE. Source: Phase-2 harness-ablation JSON.

Removal	Max MAE change (points)	Start MAE change (points)
Convective heat loss	+0.89	+1.33
Density-dependent covolume	+0.10	+0.02
Engraving model	0.00	0.00
Primer impulse table	−0.05	+0.04

Heat loss is the only component in this ablation set with an approximately one-point aggregate effect. Primer, engraving and density-dependent covolume move battery MAE by at most 0.10 points each, but each powder's fit allocates a two-parameter grid against a median-error objective, so a component's contribution can be absorbed into the fitted multiplier. Under that fit these ablations are uninformative about the components' physical significance; they are not evidence that engraving, primer impulse, or density-dependent covolume are negligible. Component-level conclusions require a fit whose parameters cannot absorb a component's contribution, or independent measurement. These component removals were not completed across the full manual corpus.

Changing the global burn scalar from 1.10 to 0.99 or 1.21 moves max-charge battery MAE from 4.37% to 4.53% or 4.99%. Varying the heat coefficient from 80,000 to 60,000 or 100,000 $\mathrm{W/(m^2\,K)}$ gives 4.50% or 4.29%. Appendix A tabulates the full grid and its caveat. The constants were chosen using regression anchors; their sensitivity does not validate them on untouched data. The model also omits a Lagrange base-pressure correction and projectile rotation. Gap

loss removes internal energy with the escaping mass, without a stagnation-enthalpy flow-work term; revolver predictions are sensitive to that approximation.

7.3 Leave-one-out ignition anchors

The per-load ignition anchor uses the load's own maximum-charge fill density when known (Section 5.4). That couples a start-charge prediction to its printed maximum and introduces self-recall. The fitted-powder battery consists of 216 Hornady reference load pairs, with 207 successful pairs reported in the served artifact. It was used for calibration and is distinct from the manual corpus, apart from the nine exact charge-point matches noted in Section 3.2. The artifact reports nine unsuccessful attempts without a per-case failure classification. The current Phase-2 experiment first reproduces all 207 max-charge errors in this battery exactly, then excludes each load's own anchor pair before sibling selection. It uses the current 0.30/0.30 ms ignition time constant/cap, rather than the older 0.45/0.45 ms state.

Table 15: Current anchor-policy comparison. All errors are velocity MAE in percent on the same 207 successful pairs from 216 attempted loads. Source: Phase-2 anchor-variants JSON.

Anchor policy	Max-charge MAE	Start-charge MAE
Own printed pair (served)	4.37	5.65
Leave-one-out sibling	5.28	6.52
Interpolated sibling	5.28	6.52
Catalogue only	5.42	6.17

The corrected leave-one-out max-charge result is 5.28%, with 86.0% within 10%. The 6.52% figure is the current start-charge MAE; its earlier use as a max-charge headline mixed calibration states. Excluding the own pair costs 0.91 percentage points at maximum charge and 0.87 at start charge. The current fallback count is 192 of 216 attempts, superseding the historical 201 count. Linear interpolation through siblings gives identical outputs because each available group has at most one distinct-bullet sibling, reducing interpolation to the sibling choice.

For a new load without its own book pair, the exclusion policy searches a same-powder raw-cartridge sibling, then the CIP-normalized cartridge group, then catalogue density. The 5.28% result measures this anchor-exclusion policy within the fitted-powder battery. Other fitted parameters remain exposed to these loads, so it is not the generalization estimate for never-seen powders. The manual-corpus 11.74% and battery 5.28% figures must not be pooled.

7.4 Pressure estimates and ceiling classifications

Raw pressure comparisons in the served corpus artifact are primary: CUP-printed rows have MAE 23.46% and median -16.00% ($n = 1,679$); psi-printed rows have MAE 24.32% and median -22.48% ($n = 3,615$). These are either-channel-screened comparisons, which omit the worst pressure residuals. CUP and piezoelectric psi are instrument conventions, not interchangeable pressure units [5]. The data cannot establish charge–pressure slope fidelity or the error rate of near-ceiling classifications.

The display bridge in the population artifact applies a fixed factor $c = 1.2783$, derived from a curated reference-set median shortfall of 21.77% as $c \simeq 1/(1 - 0.2177)$. In percentage notation it maps

$$e_{\text{display}} = 100\{c(1 + e_{\text{raw}}/100) - 1\}. \tag{13}$$

Its near-zero corrected reference median is a calibration construction, not evidence of unbiased pressure prediction. Different screening and CUP treatment also distinguish the population artifact from the raw corpus summaries; they are not two independent estimates on an identical population. No universal CUP-to-psi conversion follows from the bridge.

External evidence bounds the pure convention component of the bridge. Lagrange gradient relations place a breech-side transducer above the space-mean pressure by roughly 4–8% at typical fill ratios [4, 6], and the SAAMI transducer locations sit near the chamber in both the rifle and pistol standards [5, 16], so a larger dialect correction is not supported. The curated reference set behind c consists of transducer measurements, and its 21.77% median shortfall against them matches the raw psi-printed deficit; public IBHVG2 validation studies report lumped-parameter codes within roughly 8–10% of measured gauges [9]. The bridge is therefore retained as conservative safety machinery rather than claimed as unbiased prediction: it compensates a measured model under-drive, applied so that an over-max load cannot display as compliant.

The exact-match-first SAAMI/CIP ceiling lookup (Section 5.8) prevents some identification errors and fails closed for unknown ceilings. It does not calibrate a pressure prediction. No validated per-load prediction interval or conditional wrong-side-of-ceiling rate is available, and a displayed percentage cannot be read as a probability of safety. Pressure remains indicative only; published load data and established load-development practice remain necessary.

7.5 Selection, numerics and reproducibility limits

The cluster intervals describe sampling variation under chosen clusters. They do not remove selection bias, account for all publisher dependence, or provide an individual-load prediction band. The calibration holdout was reused for arm choice; the full-corpus rerun is a determinism/regression check. The clean never-seen partition is retrospective and screened on observed error. A prospective untouched test is needed to measure the performance of the complete selection procedure.

The historical time-step studies cover the calibration battery and a five-load diagnostic set. No corpus-wide stratified convergence proof follows from them. The corrected no-exit rescan establishes exit status for all corpus attempts, but it does not establish the guard-activation or exit-burn-fraction distribution. The earlier Phase-2 instrumentation reconstruction had non-finite barrel inputs and cannot supply corpus-wide activation counts. Energy accounting can close even when the physical burn model is wrong.

Source-identifying manuals are cited, but the copyrighted row corpus has not been released with redistribution permission. Private source data and mutable per-deployment URLs prevent complete independent reproduction. The revision includes machine-readable summary snapshots, full file digests and figure inputs at summary level; no DOI deposit or public engine repository/licence release is claimed. GRT base-data redistribution rights also remain unresolved. The present result is an auditable error report with restricted inputs, not a fully reproducible open benchmark.

8 Falsification ledger

The engine-project record describes ten attempts to obtain held-out pressure accuracy from published book data. The available records supply a retrospective sequence without externally timestamped preregistration. All ten failed their recorded adoption criteria. The repeated failure

was a pooled level improvement alongside dense maximum-charge tails outside their caps; this prevents a median improvement from establishing pressure validity.

Table 16: Pressure-falsification record. Early attempts have no individually recoverable hypothesis/gate specification in the supplied summary; those details are not reconstructed from outcomes.

Attempt	Hypothesis or tested axis	Recorded gate/outcome
1–4	Four early book-data statistical corrections; individual definitions unavailable	All failed recorded holdout criteria; exact per-attempt gates unavailable
5	Powder calibration from book data	Level/tail criteria failed; improved median did not control dense tails
6	Revised book-data correction	Rejected on holdout; individual gate values unavailable
7	Load-density gradient correction	Dense-tail criteria failed despite level improvement
8	Web-shape and lot-choice axes	No held-out transfer from tested book-measurable axes
9	Two-phase bed-lag form	Best lag at zero grid floor; proposed mechanism inactive
10	Two-level hierarchical pooling	Dense-tail criteria failed; recorded stopping rule invoked

This sequence supports rejection of the tested statistical corrections under their recorded criteria. It cannot prove that no physical model or future pressure measurement can improve accuracy. The pressure-dependent fallback experiment adds a separate failed law swap, with incompatible inherited calibration; the numerical failure is not evidence against pressure-dependent burning itself.

Velocity remains the measured output used for calibration. The pressure record motivates an indicative-only label and the dense-extrapolation guard, whose flag prevents a clamp-limited pressure from being presented as validated. Independent instrumented pressure traces are still required to test progression shape and pressure response to charge.

9 Conclusion

The calibrated engine reports 7.551% velocity MAE on 30,327 graded published observations, compared with 13.66% in the reported default-physics arm. The comparison is calibration-dependent and requires a matched-input rerun for a fully paired attribution. Four populations now bound the velocity result: fitted-corpus rows at 5.64%, held-out loads within fitted powders at 5.05%, clean never-seen loads at 11.74%, and the default partition at 11.93%. A new load on a powder with a fitted model is the dominant user case, and 5.05% is the best estimate for it; 11.74% applies to a user whose powder has no fitted model. Lookup wins where printed charge support permits both predictors, with 6.29% against 11.66% MAE on 4,656 identical rows, a paired gap of 5.370 points (95% interval $[4.161, 6.504]$), but scores only 52% of the never-seen partition.

The engine extends calculation beyond printed support and supplies indicative pressure estimates. The corrected leave-one-out anchor result is 5.28% max-charge MAE on its separate fitted-powder battery. Western error remains a model limitation, including inherited Accurate-

lot sensitivity and the disclosed 1,637-row imputed-14-inch cell; all three no-exit runs are excluded and disclosed without changing the graded total. Pressure-independent burning and selection on the calibration holdout prevent stronger predictive or safety claims.

A Constants-sensitivity grid

The grid re-scores the 216-load served battery (207 successful pairs) under single-constant changes around the served point; the served arm reproduces the served artifact byte-for-byte. Only the measured arms are reported: the cross terms between the burn scalar and the heat coefficient were never run, so this is a one-at-a-time local response, never a full factorial. Table 17 gives the arms; values in parentheses are signed medians.

Table 17: Constants-sensitivity arms on the 207-success served battery. MAE in % with signed median in parentheses. Source: revision constants-grid JSON over the Phase-2 harness data.

Arm	Max MAE (median)	Start MAE (median)
Served (s_{ba} 1.10, h_{eff} 80,000)	4.37 (+0.92)	5.65 (+0.69)
s_{ba} 0.99	4.53 (−1.52)	5.94 (−1.59)
s_{ba} 1.21	4.99 (+2.99)	6.05 (+2.49)
h_{eff} 60,000	4.50 (+1.39)	5.86 (+1.41)
h_{eff} 100,000	4.29 (+0.38)	5.54 (−0.04)

A 10% change in the burn scalar costs 0.16 or 0.62 points of max-charge MAE and flips the signed median across zero. A 25% change in the heat coefficient moves max-charge MAE by at most 0.13 points, and the 100,000 setting scores marginally better than the served 80,000 pin, which is a regression-anchor choice. All sensitivity values are conditional on the remaining global constants (and every per-powder grid fit) having been tuned on the same 207-load battery that supplies the headline; they measure local response around a selection-optimal point, not sensitivity of an untuned model.

Data availability

The measurement endpoints are https://reloadingcalc.com/accuracy.json (fitted-powder battery) and https://reloadingcalc.com/corpus-accuracy.json (raw corpus summaries, quarantine keys and no-exit disclosure). The combined record is https://reloadingcalc.com/freeze-record.json; the pressure bridge is documented in https://reloadingcalc.com/population-accuracy.json. Values were retrieved on 16 November 2026; the combined record retains its stated 11 November measurement date. A per-deployment endpoint can change after citation.

Phase-1 partition, bootstrap and reconciliation summaries and Phase-2 anchor, fair-comparison and ablation JSONs accompany this revision, together with a digest manifest. Raw copyrighted rows and engine source are not publicly deposited; hashes identify files without granting access to them. No public DOI archive or reproducibility claim beyond the released summaries is made.

References

[1] Broemel, H. G., *QuickLOAD: Interior Ballistics Program for Ammunition Development and for Experienced Handloaders*, Dipl.-Ing. H. G. Broemel, Babenhausen, Germany (first Windows release 1995).

[2] *Gordons Reloading Tool (GRT)*: interior-ballistics simulation software for handloaders and ammunition designers; user manual and handbook, https://grtools.de/doku.php?id=en:doku:start (accessed 2026-11).

[3] Krier, H., and Summerfield, M. (eds.), *Interior Ballistics of Guns*, Progress in Astronautics and Aeronautics, Vol. 66, AIAA, New York, 1979. ISBN 0-915928-32-9.

[4] Carlucci, D. E., and Jacobson, S. S., *Ballistics: Theory and Design of Guns and Ammunition*, 4th ed., CRC Press, Boca Raton, 2025. DOI 10.1201/9781003560661.

[5] Sporting Arms and Ammunition Manufacturers' Institute, *SAAMI Z299.4-2025: Voluntary Industry Performance Standards for Pressure and Velocity of Centerfire Rifle Ammunition for the Use of Commercial Manufacturers*, SAAMI, 2025.

[6] Corner, J., *Theory of the Interior Ballistics of Guns*, John Wiley & Sons, New York, 1950.

[7] Hodgdon Powder Company, *Hodgdon Annual Manual*, annual editions 2017–2025. Hodgdon/IMR tables and Winchester/Accurate/Ramshot tables as specified in Section 2.1.

[8] Western Powders, *Western Powders Handloading Guide*, source print captures and allied records identified in the restricted corpus. Exact edition/page identities are not released; this limits reconstruction.

[9] Anderson, R. D., and Fickie, K. D., *IBHVG2: A User's Guide*, BRL-TR-2829, U.S. Army Ballistic Research Laboratory, July 1987. NTIS ADB117104.

[10] Cawley, G. C., and Talbot, N. L. C., "On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation," *Journal of Machine Learning Research* 11 (2010), 2079–2107.

[11] Nosek, B. A., Ebersole, C. R., DeHaven, A. C., and Mellor, D. T., "The preregistration revolution," *PNAS* 115 (2018), 2600–2606. DOI 10.1073/pnas.1708274114.

[12] Oberkampf, W. L., and Roy, C. J., *Verification and Validation in Scientific Computing*, Cambridge University Press, 2010. DOI 10.1017/CBO9780511760396.

[13] Baer, P. G., and Frankle, J. M., *The Simulation of Interior Ballistic Performance of Guns by Digital Computer Program*, BRL Report 1183, December 1962. DTIC AD0299980.

[14] NATO, *Thermodynamic Interior Ballistic Model with Global Parameters*, STANAG 4367, revision 3, 14 May 2012.

[15] Gough, P. S., *The XNOVAKTC Code*, BRL-CR-627, 1990. NTIS ADA220153.

[16] Sporting Arms and Ammunition Manufacturers' Institute, *SAAMI Z299.3-2022: Voluntary Industry Performance Standards for Pressure and Velocity of Centerfire Pistol and Revolver Ammunition for the Use of Commercial Manufacturers*, 2022.

[17] Efron, B., and Tibshirani, R. J., *An Introduction to the Bootstrap*, Chapman & Hall, 1993.