

Guidance-track orientation under geometric uncertainty: modest coverage gains without demonstrated benefit from confirmation

Xin Li, Bojian Guo, Asel Kartanova

Changing straight guidance-track orientation can reduce geometric coverage deviation, but whether a separate confirmation stage improves orientation decisions has remained unclear. Under an externally timestamped protocol, we evaluated 360 synthetic fields and 350 real field boundaries with simulated execution, reserved headlands, prespecified operational budgets and separate calibration, confirmation, reference-search and holdout scenarios. Relative to the nominal longest-edge direction, the full rule reduced mean summed geometric deviation by 0.102–0.120 percentage points (10.2–12.0 m²/ha) across the four primary cohort/error settings. However, confirmation did not improve on calibration-only selection: under the low-position-error model, both rules chose identical angles on all 710 fields; under moderate error, they differed on only 3/360 synthetic fields and 5/350 real boundaries, with slightly unfavorable full-rule point estimates and paired intervals reaching or crossing zero. No point-adverse primary switches were observed for the full rule, calibration-only rule or budgeted deterministic comparator, so prevention of harmful switches was not demonstrated. Weighted-loss, angle-grid, swath-phase and budget sensitivities reveal that results depend on geometric and operational assumptions. A retrospective polygon-overlay correction invalidated an earlier harmful-comparator example. The findings support modest, model-conditional benefits of reorientation, but no incremental value from confirmation or measured machinery performance.

Keywords

agricultural machinery; coverage planning; guidance orientation; geometric uncertainty; conditional value-at-risk

Guidance-track direction affects the coverage of agricultural fields, but practical optimization must balance geometric deviation against non-working travel and transitions. Previous approaches optimized orientation together with path length and headland turns (JIN and TANG 2010, 2011), or combined orientation with region segmentation (YANG et al. 2025). Here we ask a narrower question: once admissible directions have been fixed, does adding an independent confirmation stage improve the decision to retain or replace the nominal longest-edge direction?

Coverage quality also depends on application control. LUCK et al. (2010) examined map-based boom-section control, while LUCK et al. (2011) related overlap to field-shape factors under different section-control configurations. Missed, repeatedly covered and outside-field areas are therefore useful geometric descriptors, but their agronomic and environmental costs are not interchangeable. Their unweighted sum provides a transparent geometric metric; it does not directly estimate chemical use, operating cost or environmental harm.

Uncertainty-aware coverage planning is also established. SCHIRMER et al. (2019), for example, selected parallel coverage lanes using localizability information and predicted localization uncertainty. The question considered here is narrower: given a nominal orientation and a set of operationally admissible candidates, when should an alternative replace the nominal plan if its apparent coverage benefit is exposed to uncertain execution? We study a rule that retains the nominal orientation unless prespecified evidence criteria are met in separate calibration and confirmation samples. Independent holdout scenarios then evaluate the frozen decisions.

Polygon overlay supplies the coverage measurements used by this rule; it is a standard computational operation rather than the methodological novelty. The orientation decision is also distinct from complete path planning. The open-source library Fields2Cover separates headland generation, swath generation, routing and path planning (MIER et al. 2023), and dedicated headland planners address boundary and continuous-curvature constraints (MIER et al. 2025b). A reserved headland and Euclidean connections between straight work segments do not, by themselves, establish a feasible machine trajectory.

We compare a confirmation-based rule with calibration-only, winner-then-confirm and operationally matched deterministic selection, rather than treating improvements over nominal as evidence of value from confirmation. The externally timestamped benchmark covers 360 synthetic geometries and 350 observed boundaries from the Fields2Benchmark dataset (MIER et al. 2025a), with simulated execution, reserved headlands and independent evaluation samples. Its main finding is negative for the additional stage: it rarely changes decisions and has no demonstrated holdout advantage. Pre-specified grid, phase, width, weighting and operational-budget sensitivities identify which geometric assumptions matter; a separate historical numerical correction is reported transparently.

Material and methods

Design and evidence separation

The revision separates a retrospective correction of the submitted 65-field analysis from a new, fixed-protocol simulation. The historical results and development-only pilots were known before the new protocol. The new protocol received a verified RFC 3161 timestamp from DigiCert at 23:47:49 UTC on 26 September 2026, before the recorded benchmark start. Its SHA-256 digest is retained with the timestamp response, verification record and code manifest. This establishes the existence of that protocol version by the signed time; it does not retrospectively preregister the original study.

For each field, four separately seeded pseudorandom streams supply 96 calibration, 96 confirmation, 96 reference-search and 256 holdout scenarios. Candidate angles use common latent scalar draws in their own orientation frames; configurations reuse those draws and are not independent replications. Policy and reference-witness choices are frozen before holdout generation; neither uses holdout values. Seed derivation, candidate ranking, operational eligibility and tie rules are specified here; additional implementation details and frozen scenario arrays belong to the archived research record.

Field material and coordinates

The prepared inventory contains 360 synthetic fields and all 350 Fields2Benchmark source boundaries (MIER et al. 2025a). Synthetic fields comprise six families—regular, tapered, convex-irregular, L-shaped, notched and multi-bay—with 60 fields per family, equally allocated to 0.5–3, 3–10 and

10–30 ha. Seeds were fixed without outcome screening. Real boundaries provide observed geometry with simulated execution, not farm-operation trials. Table 1 summarizes the field inventory; Figure 1 shows the separation of decision and evaluation evidence.

Table 1: Field material and role in the revised study

Material	Fields	Role
Historical archive	65	retrospective numerical correction; 16 development, 24 original-test, 24 replication and 1 public demonstration
New synthetic geometries	360	six families and three area bands; separate primary cohort
Fields2Benchmark geometries	350	all source boundaries; simulated execution on observed geometry
Sensitivity subset	120	prespecified 60 synthetic and 60 real fields from the new inventory
Fine-angle subset	24	prespecified hash-ranked subset of the sensitivity inventory

Note: Repeated uncertainty and geometry configurations reuse fields and are not independent field replicates. Ineligible configurations are reported separately. Source: This study; Fields2Benchmark (MIER et al. 2025a).

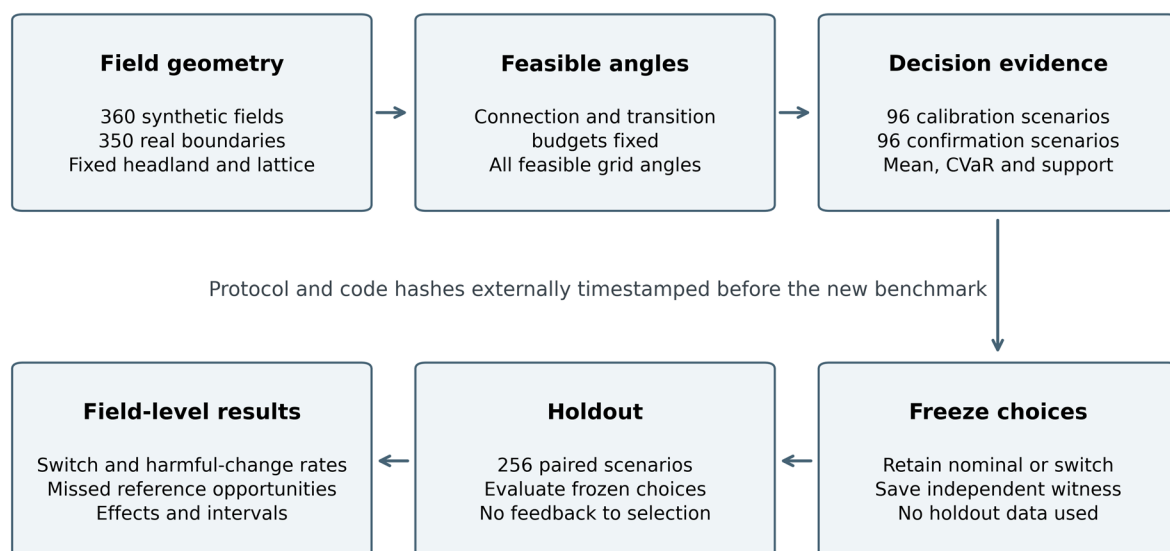


Figure 1: Evidence separation in the expanded geometric simulation; the independent reference witness detects opportunities for a fixed reference policy; it is not an exhaustive oracle

Source WKT coordinates lack an explicit datum. We state a WGS84 working assumption, project each field into its local UTM zone and translate to a local metric frame. An ETRS89 sensitivity and per-field projection records are retained. Two invalid source boundaries were repaired while preserving polygonal components and holes. Empty eroded targets or absence of nominal swaths are recorded as ineligible with explicit denominators; source boundaries remain in the inventory.

Headlands, swaths and operational admissibility

Let F be the physical field and Q its mainland target after a mitre-join inward buffer of width h . The reserved headland is $H = F \setminus Q$. Primary configurations use $w = 6$ m and $h = 2w$. Disconnected parts of Q are retained; an empty target or absent nominal swath is recorded as ineligible. Mainland loss

excludes reserved H; outside area is measured against F. Reservation does not imply headland coverage. Figure 2 illustrates the reserved target and fixed swath lattice.

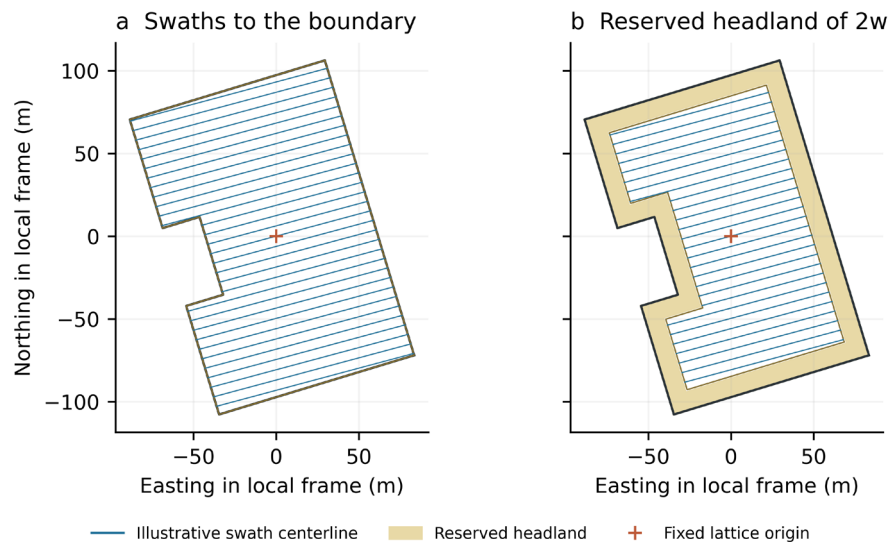


Figure 2: Centroid-anchored swaths on a historical development field at 15°; the 12 m reserved headland defines the main-field target for a 6 m implement; the shaded area is reserved but not assumed to be covered; full geometric headland passes are evaluated separately

Angles span 0°–175° at 5° increments. The nominal angle follows the longest exterior edge, rounded to the nearest circular grid angle. In a frame centered on the field centroid and rotated by the candidate angle, centerlines lie at $y_k = \phi + kw$ for integer k ; primary phase ϕ is zero. Centerline spacing remains w when footprint width changes. Lines are clipped to Q ; fragments shorter than $0.25w$ are discarded. Fragments are sorted by row and left endpoint, with direction alternating by fragment ordinal.

Productive distance sums fragment lengths; connection distance C sums Euclidean distances between successive endpoints; transition count T counts those connections. An admissible candidate has nonempty mainland and at least one swath, $C \leq 1.05C_0$ and $T \leq T_0 + 1$, with a 10^{-8} m allowance on C . These shared budget proxies omit vehicle-feasible turns and headland-pass connection costs, including in the full-headland sensitivity.

The full-headland sensitivity generates closed centerline contours at inward distances $(j + 1/2)w$ for each reserved headland pass. Their buffered footprints are combined with mainland footprints, including deposition multiplicity at the interface, and the policy is reselected for that full-field objective. These are explicit geometric coverage loops. They do not satisfy a specified vehicle curvature bound, plan obstacle-aware connections or simulate application on actual turns; consequently, this sensitivity cannot be read as a complete machinery-operation model.

$$Q = \text{buffer}(F, -h), H = F \setminus Q; y_k = \phi + kw \quad (\text{Eq. 1})$$

Uncertainty model

Each scenario draws independent normal marginals for width, lateral displacement and heading, shared across all swaths. Effective width is w times $\text{clip}(1 + 0.025Z, 0.92, 1.08)$. The whole pattern rotates about the field centroid by the heading draw and then translates along its nominal normal. This is a systematic field-wide pattern error, not independently accumulating pass-by-pass drift.

Primary low- and moderate-position scenarios have lateral standard deviations of 0.02 and 0.20 m. Angular standard deviation in degrees is $(180/\pi) \text{atan}[\sigma_{\text{position}}/\max(R, 1 \text{ m})]$, where R is the maximum centroid-to-boundary radius. Position and angle are clipped at ± 4 standard deviations. This scales the angular endpoint displacement to the stipulated positional magnitude. It is a modeling convention, not a measured steering-heading distribution.

RTK and DGNSS provide context for the scale contrast, not fitted error classes. Dynamic RTK tests provide centimetre-scale context (FRISCHHUT and NOACK 2025). NOVATEL (2026) lists typical horizontal root-mean-square (RMS) accuracies of 1 cm + 1 ppm for RTK and 0.40 m for DGNSS. These receiver metrics are not one-dimensional lateral standard deviations or complete implement-tracking errors. The 0.02/0.20 m standard deviations are therefore stipulated centimetre/decimetre sensitivities, not validated RTK/DGNSS distributions. The 2.5% width-factor standard deviation remains an assumption and can dominate low-position scenarios. A secondary historical stress class retains 0.16 m lateral standard deviation clipped at ± 0.65 m and 0.8° heading standard deviation clipped at $\pm 3.2^\circ$. A separate row-residual sensitivity adds local perturbations. All clipped normals differ from conditional truncated normals.

Coverage metric and numerical model

$A(\cdot)$ denotes planar area in square metres. Each clipped centerline forms a flat-ended rectangular footprint B_i of scenario width. With union $U = \bigcup_i B_i$, missed mainland area is $M = A(Q \setminus U)$, overlap multiplicity is $O = \sum_i A(Q \cap B_i) - A(Q \cap U)$, and outside-field area is $X = A(U \setminus F)$. Triple coverage contributes twice to O . Mainland deposition inside reserved H is not classified as outside F .

The primary endpoint is $D = 100(M + O + X)/A(F)$, expressed relative to original physical field area F . Missed area M is measured in the mainland target Q ; O counts overlap multiplicity and X is footprint area outside F . Reserved headland H is outside Q and is not counted as missed mainland coverage. The equally weighted sum is a geometric index, not unique area saved or an estimate of chemical use. Sensitivities use $A(Q)$ normalization or weight the three components differently; explicit headland-footprint coverage is evaluated as a distinct endpoint. The primary coefficients in Equation 3 are $\lambda_M = \lambda_O = \lambda_X = 1$.

Overlay uses a common 10^{-6} m precision grid in a local frame. An inverse rigid transformation keeps primary rectangles axis-aligned. A decrease in field or target polygon-component count during initial plan precision reduction causes failure; this is not a topology guarantee for every later overlay. Numerical convergence and geometry identities are audited rather than claiming mathematical exactness.

$$\begin{aligned} U &= \bigcup_i B_i; M = A(Q \setminus U); X = A(U \setminus F) \\ O &= \sum_i A(Q \cap B_i) - A(Q \cap U) \end{aligned} \quad (\text{Eq. 2})$$

$$D = 100 \frac{\lambda_M M + \lambda_O O + \lambda_X X}{A(F)} \quad (\text{Eq. 3})$$

Decision rules and upper-tail evidence

CVaR90 is estimated by the mean of the largest $\text{ceil}(0.10N)$ scenario values: 10 values in each 96-scenario decision sample and 26 in the 256-scenario holdout. The budgeted deterministic rule minimizes zero-error D over admissible angles. The calibration-only and full gate rules evaluate every admissible grid angle, removing reliance on the former top-six shortlist. In Equation 4, $D(j)$ denotes the j -th value in ascending order and N is the scenario count.

An alternative must reduce mean D by at least 0.3% relative and CVaR90 by at least 0.2%, win in at least 55% of paired scenarios, and have positive nominal one-sided 90% percentile-bootstrap lower bounds for both nominal-minus-candidate endpoints. Bounds use 1,000 paired resamples shared across candidates and endpoints. The calibration-only rule applies these tests to calibration and maximizes the smaller relative endpoint gain.

The full gate requires the tests in both samples and maximizes the smallest of four gains across endpoints and samples. If none qualifies, it retains nominal. Confirmation thus affects ranking as well as eligibility; pointwise bounds provide no family-wise 90% guarantee over searched angles. A winner-then-confirm ablation freezes the calibration winner, tests that angle on confirmation and retains nominal if it fails, without substitution. Unconstrained deterministic and minimum-transition rules are separate comparison anchors that do not expand gate feasibility. Relative to calibration-only, the full rule changes sample count, evidence requirements and ranking; it does not isolate the effect of sample independence alone.

Deterministic ties select fewer transitions, then shorter connections, then smaller angle. Gate-score ties select the larger angle. A zero nominal endpoint yields zero relative gain, avoiding division by zero. Within eligibility and bootstrap tests, the numerical tolerance is zero, distinct from the 10^{-4} percentage-point holdout audit tolerance. Both calibration and confirmation must satisfy the mean, CVaR90, paired-win and one-sided bootstrap criteria; the complete candidate pool is all budget-admissible 5° angles. In Equation 5, E is the set passing both samples; a_0 is nominal and $g_{\mu,s}$ and $g_{\tau,s}$ are relative reductions in sample s .

$$\text{CVaR}_{0.90}(D) = \frac{1}{k} \sum_{j=N-k+1}^N D_{(j)}, k = \text{ceil}(0.10N) \quad (\text{Eq. 4})$$

$$a^* = \underset{a \in E}{\text{argmax}} \min\{g_{\mu,\text{cal}}(a), g_{\tau,\text{cal}}(a), g_{\mu,\text{conf}}(a), g_{\tau,\text{conf}}(a)\} \quad \text{if } E \neq \emptyset$$

$$a^* = a_0 \quad \text{if } E = \emptyset \quad (\text{Eq. 5})$$

Holdout effects and adverse-switch audit

Summaries weight fields equally within each configuration and report synthetic and real cohorts separately; the overall mixture is descriptive. Direct paired contrasts compare policies on the same fields. Two-sided 95% percentile intervals use 10,000 paired field resamples; repeated configurations and scenarios are not extra replicate fields. Relative reduction divides the difference of field means by the reference mean. Absolute reduction is reference minus candidate, in percentage points; multiplying by 100 gives m²/ha of summed deviation. Area-weighted summaries are secondary and descriptive.

A point-estimated adverse switch increases either holdout mean D or CVaR90 by more than 10⁻⁴ percentage points, a numerical rather than agronomic tolerance. A secondary audit uses 2,000 paired resamples and Bonferroni-adjusted percentile intervals across both endpoints at nominal simultaneous 95% confidence. Labels are harmful if either lower limit exceeds tolerance, non-worsening if both upper limits are within tolerance, and uncertain otherwise. These bootstrap approximations do not establish true harm or safety.

For each conditional harm rate we also report the range from supported harmful switches to supported harmful plus uncertain switches. This is a classification range, not a confidence interval. The Wilson intervals describe observed field labels under a binomial approximation and do not remove simulation-model uncertainty or dependence between configurations.

Switching and missed reference opportunities

Switch rate uses successfully evaluated eligible fields. Failures, pending jobs and geometric ineligibility are reported separately rather than classified as retained plans. Each policy sensitivity uses its own prespecified subset denominator. The requested false-change rate is operationalized as the point-adverse-switch rate defined above. It is reported among switches and, separately, among evaluated fields; point estimates and secondary evidence labels remain distinct. Wilson intervals accompany explicit counts; zero-switch conditional rates are undefined. Retained fields are ties by construction.

An independent 96-scenario reference search freezes one admissible alternative maximizing its smaller relative mean/CVaR gain, even if that score is negative. Holdout assesses the fixed primary thresholds, 0.3% and 0.2%, also for threshold sensitivities. An opportunity is supported when both threshold-adjusted bootstrap lower limits exceed zero. If either upper limit is below zero, that witness is classified no_opportunity; this does not establish absence of other opportunities. No-admissible-alternative and uncertain cases are distinguished.

The primary missed-opportunity rate counts retained fields with supported witnesses divided by all supported reference opportunities. Point-estimated opportunities and rates per retained field are secondary. This measures abstention cost relative to the independent reference policy, not an exhaustive-angle oracle or regret after switching. Curated boundaries are not a random sample of all farms; no population-safety claim follows.

$$R_{\text{miss}} = \frac{\#(\text{retain nominal} \cap \text{reference opportunity})}{\#(\text{reference opportunity})} \quad (\text{Eq. 6})$$

Sensitivity and reproducibility

A prespecified subset contains 60 synthetic and 60 real fields; 24 receive the 1° grid. Sensitivities cover headlands, width, phase, residual errors, weights, evidence and budgets. The finer grid also re-

rounds nominal and changes the budget reference. Changing w to 3 or 12 m also changes $h = 2w$ and the absolute width-error SD; these are whole-configuration comparisons. Top-six/top-twelve pools and nested 300/1,000 bootstrap counts are compared with the primary rule. Choices precede hold-out; outcomes do not retune settings.

Full-headland D and its harm audit use the full-field endpoint; stored M/O/X columns still describe mainland swaths. Weighted objectives have separate weighted holdout audits. The unweighted reference witness is not used to claim weighted missed-opportunity rates.

The frozen protocol, timestamp evidence, code manifest, candidate diagnostics, decisions, scenario arrays and numerical outputs are deposited in the versioned reproducibility archive (Zenodo v5.1; <https://doi.org/10.5281/zenodo.22995389>). A downloaded copy of this archive passed its internal SHA-256 manifest verification. Third-party field boundaries are obtained from Fields2Benchmark (MIER et al. 2025a) under the original source terms. The numerical evaluation concerns simulated geometric decisions, not measured machinery performance.

Results

Primary benchmark

The completed registry contains 3,354 field-configuration attempts across the 710-field inventory. For both low- and moderate-position models, all 360 synthetic and 350 real-boundary fields were evaluated without primary ineligibility. Repeated configurations reuse these same fields. The full field-status inventory and prespecified denominators are retained in the versioned research record; geometric ineligibility is not counted as retaining nominal.

With the primary 12 m headland reservation, mean reserved-area fractions $A(H)/A(F)$ were 26.3% (range 8.6–69.0%) for synthetic fields and 39.5% (9.5–86.5%) for real boundaries. The primary endpoint D uses $A(F)$ as its denominator; the separately archived $A(Q)$ -normalized diagnostic uses the mainland area. Neither endpoint treats reserved headland H as missed mainland coverage.

The operational budget sometimes left no alternative to nominal (low/synthetic: 209/360; low/real boundaries: 166/350; moderate/synthetic: 209/360; moderate/real boundaries: 166/350). Thus an overall retention count combines forced retention with evidence-based abstention. Table 2 and Figure 3 include all evaluated fields, including unchanged decisions.

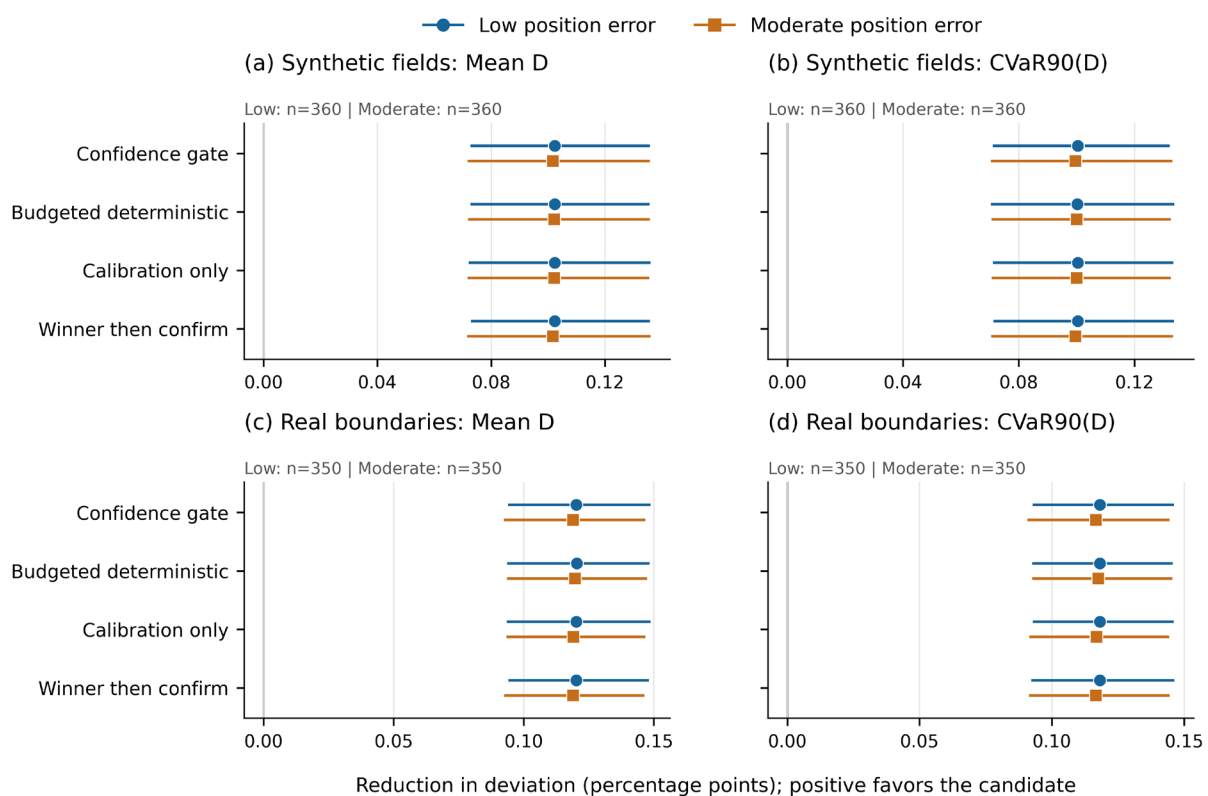
Table 2: Equal-field mean and upper-tail reductions relative to nominal, in percentage points

Position class	Cohort	Policy	Mean reduction (95% CI)	CVaR90 reduction (95% CI)
Low	Synthetic	Gate	0.1024 [0.0728, 0.1358]	0.1005 [0.0710, 0.1322]
Low	Synthetic	Budgeted deterministic	0.1024 [0.0728, 0.1357]	0.1002 [0.0703, 0.1338]
Low	Synthetic	Calibration only	0.1024 [0.0721, 0.1360]	0.1005 [0.0711, 0.1335]
Low	Synthetic	Winner then confirm	0.1024 [0.0729, 0.1358]	0.1005 [0.0712, 0.1336]
Low	Real boundaries	Gate	0.1201 [0.0939, 0.1488]	0.1182 [0.0927, 0.1462]
Low	Real boundaries	Budgeted deterministic	0.1204 [0.0937, 0.1484]	0.1182 [0.0925, 0.1458]
Low	Real boundaries	Calibration only	0.1201 [0.0935, 0.1487]	0.1182 [0.0929, 0.1460]
Low	Real boundaries	Winner then confirm	0.1201 [0.0940, 0.1482]	0.1182 [0.0923, 0.1463]

Position class	Cohort	Policy	Mean reduction (95% CI)	CVaR90 reduction (95% CI)
Moderate	Synthetic	Gate	0.1017 [0.0718, 0.1358]	0.0996 [0.0704, 0.1331]
Moderate	Synthetic	Budgeted deterministic	0.1022 [0.0719, 0.1358]	0.1000 [0.0705, 0.1325]
Moderate	Synthetic	Calibration only	0.1022 [0.0718, 0.1357]	0.1001 [0.0707, 0.1326]
Moderate	Synthetic	Winner then confirm	0.1017 [0.0716, 0.1361]	0.0996 [0.0705, 0.1334]
Moderate	Real boundaries	Gate	0.1189 [0.0923, 0.1467]	0.1167 [0.0907, 0.1445]
Moderate	Real boundaries	Budgeted deterministic	0.1197 [0.0935, 0.1474]	0.1176 [0.0925, 0.1455]
Moderate	Real boundaries	Calibration only	0.1191 [0.0933, 0.1467]	0.1169 [0.0915, 0.1444]
Moderate	Real boundaries	Winner then confirm	0.1189 [0.0925, 0.1464]	0.1167 [0.0914, 0.1445]

Note: Positive values favor the policy. CVaR90 reductions average field-specific scenario tail means, rather than pooling scenarios across fields. Intervals use 10,000 paired field bootstrap resamples; policy-specific random seeds can cause small Monte Carlo differences between marginal intervals for identical outcomes. Direct contrasts of identical policies are exactly zero. Multiplying every value by 100 gives m²/ha of summed geometric deviation; this is not unique area saved. Low and moderate reuse fields and latent scenario draws.

Absolute coverage improvement relative to nominal



Equal weight per field; paired field-bootstrap 95% intervals. Cohorts and configurations are not pooled.

Figure 3: Mean and CVaR90 reductions from nominal by cohort and primary position class; points and paired field-bootstrap intervals include retained fields; positive reductions indicate lower geometric deviation

Decision selectivity and ablations

Table 3 reports switching, point-estimated adverse outcomes and retention despite supported reference opportunities, with explicit denominators.

Table 3: Switching, point-estimated adverse changes and missed supported reference opportunities

Class/cohort	Policy	Switches/evaluated, % (95% CI)	False changes/ switches, % (95% CI)	Missed/supported opportunities, % (95% CI)
Low/Synthetic	Gate	88/360; 24.4% [20.3, 29.1]	0/88; 0.0% [0.0, 4.2]	1/87; 1.1% [0.2, 6.2]
Low/Synthetic	Budgeted deterministic	89/360; 24.7% [20.5, 29.4]	0/89; 0.0% [0.0, 4.1]	2/87; 2.3% [0.6, 8.0]
Low/Synthetic	Calibration only	88/360; 24.4% [20.3, 29.1]	0/88; 0.0% [0.0, 4.2]	1/87; 1.1% [0.2, 6.2]
Low/Real boundaries	Gate	106/350; 30.3% [25.7, 35.3]	0/106; 0.0% [0.0, 3.5]	0/103; 0.0% [0.0, 3.6]
Low/Real boundaries	Budgeted deterministic	111/350; 31.7% [27.1, 36.8]	0/111; 0.0% [0.0, 3.3]	0/103; 0.0% [0.0, 3.6]
Low/Real boundaries	Calibration only	106/350; 30.3% [25.7, 35.3]	0/106; 0.0% [0.0, 3.5]	0/103; 0.0% [0.0, 3.6]
Moderate/Syn- thetic	Gate	84/360; 23.3% [19.3, 28.0]	0/84; 0.0% [0.0, 4.4]	2/84; 2.4% [0.7, 8.3]
Moderate/Syn- thetic	Budgeted deterministic	89/360; 24.7% [20.5, 29.4]	0/89; 0.0% [0.0, 4.1]	1/84; 1.2% [0.2, 6.4]
Moderate/Syn- thetic	Calibration only	87/360; 24.2% [20.0, 28.8]	0/87; 0.0% [0.0, 4.2]	1/84; 1.2% [0.2, 6.4]
Moderate/Real boundaries	Gate	102/350; 29.1% [24.6, 34.1]	0/102; 0.0% [0.0, 3.6]	4/103; 3.9% [1.5, 9.6]
Moderate/Real boundaries	Budgeted deterministic	111/350; 31.7% [27.1, 36.8]	0/111; 0.0% [0.0, 3.3]	0/103; 0.0% [0.0, 3.6]
Moderate/Real boundaries	Calibration only	105/350; 30.0% [25.4, 35.0]	0/105; 0.0% [0.0, 3.5]	2/103; 1.9% [0.5, 6.8]

Note: Wilson intervals describe observed labels. False change means a point-adverse switch: either endpoint increases by $>10^{-4}$ pp. This operational label is not proof of true harm. The independent reference witness uses fixed 0.3% mean and 0.2% tail thresholds. Missed opportunity measures retaining nominal despite that witness; it is not an exhaustive oracle or regret after switching. Secondary evidence labels and rates per retained field are supplied separately.

The secondary paired-scenario classification gave low/synthetic: 0 supported harmful, 0 uncertain, among 88 switches; low/real boundaries: 0 supported harmful, 1 uncertain, among 106 switches; moderate/synthetic: 0 supported harmful, 0 uncertain, among 84 switches; moderate/real boundaries: 0 supported harmful, 1 uncertain, among 102 switches. Uncertain outcomes are not counted as safe. These classifications are conditional on the supplied model and finite holdout; the point-estimated and evidence-supported adverse rates answer different questions. Figure 4 shows every Gate switch under the moderate class, with retained fields explicitly accounted for.

Field-level effects of the confidence gate: moderate position error

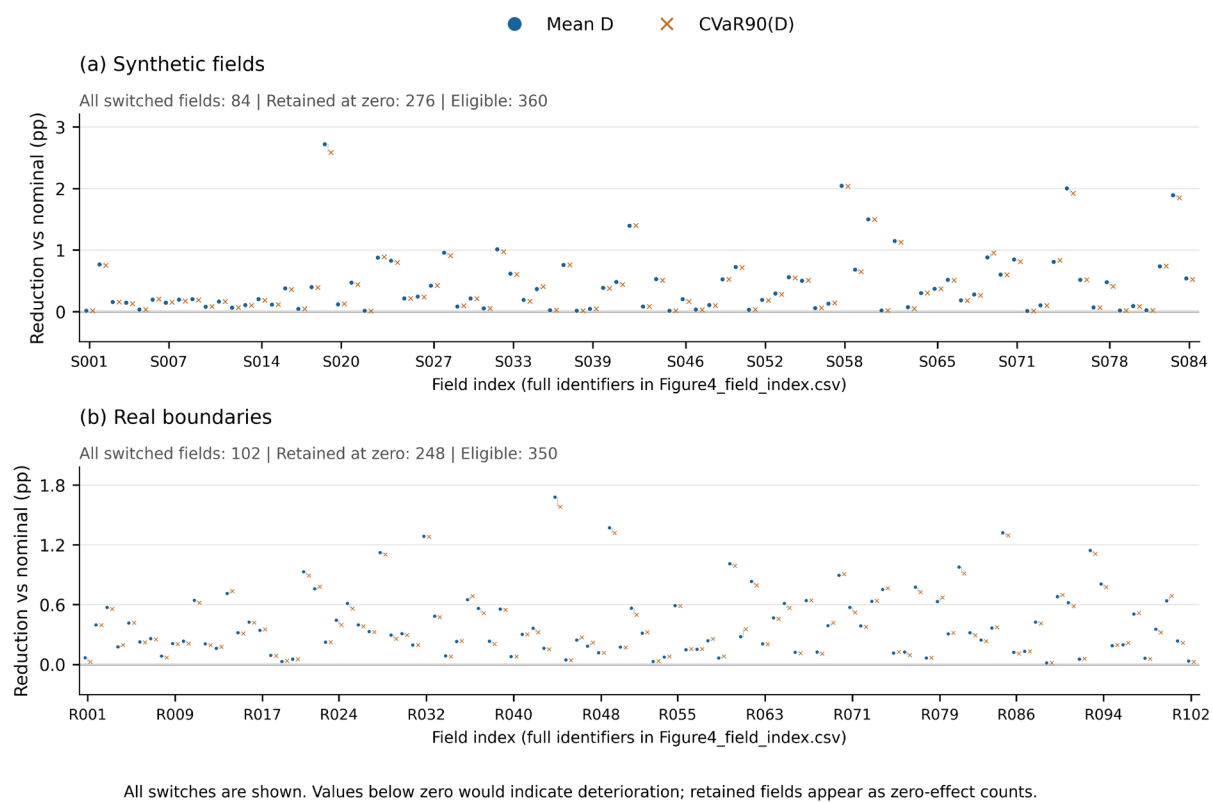


Figure 4: Per-field mean and CVaR90 reductions for every Gate switch in the moderate-position class; negative values indicate deterioration; retained fields have identical nominal and policy outcomes and are counted in panel annotations; full field identifiers and unchanged-decision rows are available in the versioned research record

Table 4 compares Gate directly with budgeted deterministic and calibration-only selection on the same fields, separating improvement from nominal from added value over these alternatives.

Table 4: Direct paired reduction achieved by Gate relative to the stated comparator, in percentage points (pp) of summed geometric deviation

Class	Cohort	Comparator	Mean reduction [95% CI]	CVaR90 reduction [95% CI]
Low	Synthetic	Budgeted deterministic	0.000064 [-0.000069, 0.000281]	0.000237 [-0.000062, 0.000786]
Low	Synthetic	Calibration only	0.000000 [0.000000, 0.000000]	0.000000 [0.000000, 0.000000]
Low	Real boundaries	Budgeted deterministic	-0.000227 [-0.000429, -0.000076]	-0.000048 [-0.000188, 0.000087]
Low	Real boundaries	Calibration only	0.000000 [0.000000, 0.000000]	0.000000 [0.000000, 0.000000]

Class	Cohort	Comparator	Mean reduction [95% CI]	CVaR90 reduction [95% CI]
Moderate	Synthetic	Budgeted deterministic	-0.000459 [-0.001261, -0.000009]	-0.000466 [-0.001201, 0.000028]
Moderate	Synthetic	Calibration only	-0.000461 [-0.001301, 0.000000]	-0.000524 [-0.001343, 0.000000]
Moderate	Real boundaries	Budgeted deterministic	-0.000742 [-0.001341, -0.000261]	-0.000924 [-0.001964, -0.000156]
Moderate	Real boundaries	Calibration only	-0.000163 [-0.000407, 0.000006]	-0.000245 [-0.000610, 0.000041]

Note: All values are in percentage points (pp) of the field-normalized summed geometric-deviation metric. Positive values favor Gate and negative values favor the comparator. Every contrast uses the same fields and paired scenario draws. One pp corresponds to 100 m²/ha of summed deviation: the largest negative Table 4 point estimate (-0.000924 pp) equals only -0.0924 m²/ha, not a change of 0.0924 percentage points. This table compares Gate with operationally matched alternatives, rather than reporting its larger reductions from nominal.

Sensitivity to geometry and uncertainty

Headland reservation changes both the mainland coverage target and selected orientation. Relative to the primary two-width reservation, zero-width headlands changed 12/60 synthetic and 25/60 real-boundary selections; one-width and three-width reservations each changed 4/60 synthetic and 16/60 real-boundary choices. These are whole-configuration effects, not a headland-width-only causal effect: admissibility and swath fragments also change. A smaller untreated mainland target can reduce measured mainland deviation without improving full-field coverage. The paired comparisons are restricted to common eligible fields; Table 5 separately reports the full geometric headland objective.

Table 5: Gate effects within the full geometric headland objective.

Cohort	n	Mean reduction from nominal	CVaR90 reduction from nominal	Switches; point-adverse; uncertain
Synthetic	60	0.1236 [0.0486, 0.2182] 12.36 m ² /ha	0.1217 [0.0461, 0.2135] 12.17 m ² /ha	12; 0; 0
Real boundaries	60	0.1047 [0.0557, 0.1620] 10.47 m ² /ha	0.1085 [0.0574, 0.1670] 10.85 m ² /ha	18; 0; 0

Note: Effects are in pp (95% paired field-bootstrap CI) and m²/ha of summed deviation. Full-field and mainland objectives have different targets and are not subtracted from one another. C and T budgets still describe mainland connections and transitions; headland loops and their connections are not included. Geometric contour coverage does not establish feasible vehicle turns.

Within the full geometric objective, every direct mean and CVaR90 comparison of Gate with the budgeted deterministic, calibration-only and winner-then-confirm rules had an interval containing zero. Thus the positive effects from nominal did not demonstrate an advantage over those alternatives.

All 710 primary-field decisions matched the all-feasible rule when candidate pools were restricted to the six best budget-admissible deterministic angles; restricting to twelve also changed no selections. These are empirical agreements within the evaluated scenarios, not proofs that a deterministic shortlist is sufficient in general.

Changing the swath lattice phase by 0.25w, 0.50w or 0.75w altered 19–32 of 60 selected directions across the synthetic and real-boundary subsets, although all paired mean and CVaR90 confidence intervals for their performance differences included zero. Refining the angle grid from 5° to 1° changed 7/11 synthetic and 10/13 real-boundary selections. The matched synthetic mean-deviation change was -0.163 pp [95% CI -0.284, -0.060], whereas the real-boundary change was -0.203 pp [-0.535,

0.124]; negative differences favor the finer-grid configuration. The finer grid changes nominal-angle rounding and budget references, and does not establish continuous-angle optimality.

Adding the prespecified row-residual errors did not change any selected direction among the 60 matched fields in either cohort. Nevertheless, mean geometric deviation increased by 0.111 pp [95% CI 0.087, 0.136] in the synthetic subset and by 0.046 pp [0.037, 0.056] in the real-boundary subset. Thus stable choices do not imply stable realized coverage outcomes.

Implement width had a larger influence on CVaR90(D) than most gate thresholds. Relative to the 6 m configuration on matched fields, 3 m increased CVaR90(D) by 0.335 pp (95% CI 0.244, 0.446; $n = 60$) in synthetic fields and by 0.451 pp (0.344, 0.558; $n = 60$) in real-boundary fields. Conversely, 12 m reduced CVaR90(D) by 0.702 pp (0.484, 0.940; $n = 60$) and 1.101 pp (0.782, 1.434; $n = 57$), respectively. For 12 m, mean-D reductions were 0.035 pp (−0.148, 0.224) in synthetic fields and 0.224 pp (−0.076, 0.530) in real-boundary fields; neither mean-D interval excluded zero. These whole-configuration contrasts jointly modify lattice pitch, effective-width variation, cutoff and reserved headland width. Three real boundaries were geometrically ineligible in the 12 m configuration; its comparison used 57 matched fields.

Without headland reservation, Gate mean outside-field coverage was 0.9682/1.5726 pp for synthetic/real boundaries (96.82/157.26 m²/ha). The corresponding component was zero in the primary and reserved-headland width, phase, headland and row-residual configurations. Their benefits therefore concern the other modeled components; they do not establish reduced physical off-field deposition or zero deployment risk. Across the eight headland, width and phase comparisons, every direct Gate-versus-budgeted-deterministic, calibration-only and winner-then-confirm interval touched or crossed zero.

Additional sensitivity evidence: weighting and operating budgets

Prespecified weighted-loss comparisons used $(\lambda_M, \lambda_O, \lambda_X) = (2, 1, 1)$ and $(1, 0.5, 1.5)$, rather than assuming that missed, overlapping and off-field coverage have equal costs. In each of the four 60-field cohort/error subsets, each weighting changed at most one selected angle from the equal-weight rule (zero in low-position synthetic fields). Each weighted policy reduced its own weighted mean endpoint from its weighted nominal by 8.83–24.54 m²/ha across the settings; this is a coefficient-weighted area equivalent, not an agronomic benefit. When the weighted choices were scored again using the unweighted endpoint, their mean deviation differed from the equal-weight rule by 0 to −0.041 m²/ha. Because objectives differ, weighted reductions must not be compared numerically with the equal-weight reductions as though they were the same loss.

Across the four 60-field parameter subsets, varying nominal confidence from 0.80 to 0.95 or paired-win support from 0.50 to 0.60 altered no selected direction. Adjusting the connection-distance allowance from 0 to 10% altered at most 3/60 directions, whereas permitting 0, 1 or 2 additional transitions changed up to 8/60 directions. Varying the bootstrap resample count from 1,000 to 300 altered no selections in the four matched subsets. These are configuration-specific checks, not general threshold invariance. The larger sensitivity to operational budgets and geometric discretization, together with the lack of confirmation gain in the primary analyses, supports the use of the simplest empirically adequate rule as a practical starting point. Table 6 summarizes the selected sensitivity results.

Table 6: Selected sensitivity results incorporated from the deposited analyses.

Sensitivity	Matched sample	Changed choices	Paired effect or main interpretation
1° vs 5° angle grid	Synthetic n=11	7/11	Mean $\Delta D = -0.163$ pp [-0.284, -0.060]
1° vs 5° angle grid	Real boundaries n=13	10/13	Mean $\Delta D = -0.203$ pp [-0.535, 0.124]
Phase 0.25w–0.75w	Each cohort n=60	19–32/60	Mean and CVaR90 intervals included 0
Weighted (2,1,1) or (1,0.5,1.5)	Each cohort n=60	0–1/60	Unweighted mean difference 0 to -0.041 m ² /ha
Confidence 0.80–0.95; win support 0.50–0.60	Each cohort n=60	0/60	No selection changes in these settings
Connection budget 0–10%; transition allowance 0–2	Each cohort n=60	up to 3/60; up to 8/60	Operational constraints altered more choices
300 vs 1,000 bootstrap resamples	Each cohort n=60	0/60	Identical selected directions

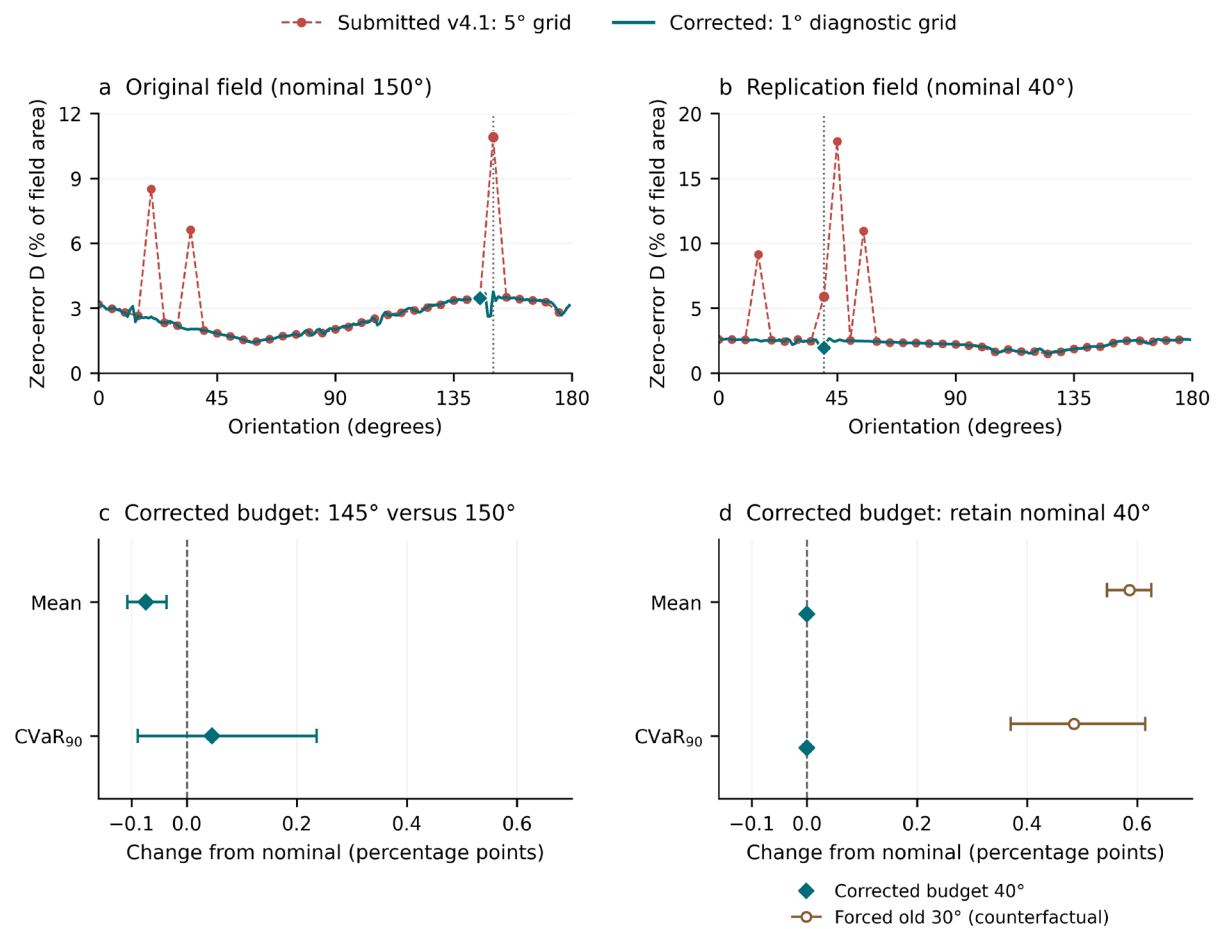
ΔD denotes candidate minus baseline mean geometric deviation for the stated matched fields; negative values indicate a reduction. Confidence intervals are paired field-bootstrap 95% intervals. Other rows report choices under separately defined endpoints and must not be combined into one effect estimate. Subset samples and repeated configurations are not independent field replications. Underlying sensitivity outputs retained in the versioned research archive.

Under the secondary, more severe historical-stress model, the full rule reduced mean deviation from nominal by 0.099 pp [95% CI 0.069, 0.133] in 360 synthetic fields and by 0.118 pp [0.092, 0.146] in 350 real-boundary fields. It switched direction on 72 and 99 fields, respectively; neither cohort showed a point-adverse full-rule switch, although each had one uncertain switch. These results remain conditional on the stipulated, rather than field-calibrated, error model.

In the same stress scenario, the budgeted deterministic rule produced 6/89 point-adverse synthetic switches and 3/111 real-boundary switches, compared with 1/78 and 0/104 for calibration-only selection. The full rule showed 0/72 and 0/99 point-adverse switches, but also had slightly worse mean and upper-tail outcomes than calibration only and retained nominal more frequently despite a supported reference opportunity. These descriptive contrasts do not establish lower true harm rates or isolate confirmation as a causal factor.

Correction of historical cases

The historical polygon-union audit is reported separately from the frozen expanded benchmark. Correcting zero-error coverage areas changed some deterministic comparator rankings and nine selections across five of the 65 historical fields. In the formerly highlighted replication case, the corrected budgeted rule retains the nominal 40° direction rather than the previously selected 30° angle; therefore the claimed harmful policy switch and its associated superiority claim are withdrawn. The other highlighted case remains uncertain in its upper-tail effect. Figure 5 documents both numerical corrections; the archived audit retains the full before-and-after calculations.



Positive change = larger deviation; bars: paired bootstrap intervals (256 scenarios).

Figure 5: Retrospective correction of historical geometric coverage estimates; panels (a,b) compare archived and corrected zero-error direction curves in two selected fields; nominal and corrected budgeted choices are marked; panels (c,d) show mean and CVaR₉₀ differences from nominal using the same historical holdout scenarios with corrected overlay; the replication policy now retains the nominal 40° direction; its former 30° selection is shown only as a counterfactual; intervals are Bonferroni-adjusted paired-scenario bootstrap intervals for these preselected examples and do not provide selection-adjusted inference

All four historical-test and three replication Gate switches remained unchanged. In the corrected 24-field replication, mean and CVaR₉₀ reductions from nominal were 0.103 pp [95% CI 0.000, 0.255] and 0.123 pp [0.000, 0.322], respectively. Both intervals include zero; they cannot support a historical superiority claim.

Discussion

What the confirmation stage added

The direct comparisons separate two questions: whether an admissible alternative improves on longest-edge guidance, and whether the full gate improves on a simpler selection rule. Reductions from nominal alone cannot answer the second question. Table 4 gives direct paired effects against the budgeted deterministic and calibration-only comparators, including negative advantages. These comparisons use the same fields and distinguish orientation benefits from the additional value of the full rule.

Under the low-position model, confirmation changed no selected angle in either cohort. The additional confirmation sample therefore provided no observed decision or holdout benefit in that setting. The full rule also changes ranking and uses more scenarios than calibration only, so differences elsewhere should not be attributed solely to sample independence. The winner-then-confirm comparison makes the selection order explicit, but none of these pointwise bootstrap rules supplies a simultaneous safety guarantee over all searched directions.

Under the moderate-position model, the full gate differed from calibration only on 3/360 synthetic fields and 5/350 fields with real boundaries. Its mean and CVaR90 point estimates were slightly worse in both cohorts, with direct paired-field intervals reaching or crossing zero. Thus the primary comparisons did not demonstrate an added benefit of the full confirmation-based rule. No point-adverse switches were observed for Gate, calibration only or the budgeted deterministic rule in either primary class. Differences in uncertain classifications should therefore not be described as demonstrated prevention of harmful switches.

The direct mean-D contrasts with the budgeted deterministic rule favored that comparator for low-position real boundaries and both moderate-position cohorts; their intervals exclude zero in that direction. The moderate-position real-boundary CVaR90 contrast also favored the comparator. These direct differences are small relative to the improvements from nominal; Table 4 now reports them in the same percentage-point units as Table 2. They nevertheless preclude a general claim of superiority over the operationally matched deterministic rule.

Practical interpretation and limits

The reported m^2/ha values express a summed geometric loss. Overlap is counted with multiplicity, and the metric combines missed, repeated and outside deposition; its reduction is not an equal reduction in chemical use or unique land area. Reserved headlands change the mainland target, whereas contour passes add a separate full-field coverage calculation. Real machine evaluation would additionally require measured tracking and implement errors, feasible headland turns, application-control logic and task-specific costs. The present comparison can inform which candidate directions merit such evaluation.

The evidence is conditional on six synthetic families, a curated boundary collection and prescribed error models. Separate random scenario streams prevent the held-out simulation outcomes from selecting the reported policies; they do not validate the model against farm measurements. The 2.5% width variability was assumed and may dominate the low-position model. Its variance was not independently calibrated or swept. Correlations between swaths, orientation and positioning error, and systematic implement behavior could change both relative performance and adverse-switch rates.

Width and fine-grid sensitivities alter several aspects of the configuration and should not be read as isolated causal effects.

Retaining nominal can be required by the operational budget or can result from insufficient gate evidence. The feasibility inventory distinguishes these situations. The opportunity audit tests a separately selected reference witness, which avoids selecting an angle with holdout results but cannot detect every useful alternative. Wilson and bootstrap intervals quantify uncertainty in the observed labels and effects under their resampling assumptions; they do not remove these geometric and statistical limits.

Consequences of the historical correction

The historical overlay correction removes a previously claimed harmful deterministic switch and an accompanying argument for the gate. Although the gate's historical angle choices were unchanged, its comparative advantage is no longer established. The retrospective correction and the prospective benchmark answer different questions and are retained as distinct, auditable evidence.

Conclusions

Within the tested simulations, changing an admissible guidance direction modestly reduced geometric deviation relative to the nominal longest-edge rule, but adding confirmation did not improve on simpler selection. All 710 low-position selections matched calibration only; under moderate error, only 3/360 synthetic and 5/350 real-boundary choices differed, with slightly worse full-rule point estimates and paired intervals reaching or crossing zero. The budgeted deterministic comparator was also competitive, and the primary models showed no point-adverse switches for any of these three policies. The prespecified sensitivity analyses summarized in the text show that angle lattice, phase, implement width, geometric loss weights and especially operating budgets influence results more visibly than some confidence thresholds. A historical numerical correction invalidated an earlier favorable gate comparison. These outcomes are conditional on geometries and stipulated uncertainty laws; no machine trials, deployment safety or economic benefits have been demonstrated.

References

- Frischhut, M.; Noack, P. O. (2025): Validation of the dynamic accuracy of different GNSS receivers. *agricultural engineering.eu* 80(3), pp. 189–201, <https://doi.org/10.1515/ae.2025.3338>
- Jin, J.; Tang, L. (2010): Optimal coverage path planning for arable farming on 2D surfaces. *Transactions of the ASABE* 53(1), pp. 283–295, <https://doi.org/10.13031/2013.29488>
- Jin, J.; Tang, L. (2011): Coverage path planning on three-dimensional terrain for arable farming. *Journal of Field Robotics* 28(3), pp. 424–440, <https://doi.org/10.1002/rob.20388>
- Luck, J. D.; Pitla, S. K.; Shearer, S. A.; Mueller, T. G.; Dillon, C. R.; Fulton, J. P.; Higgins, S. F. (2010): Potential for pesticide and nutrient savings via map-based automatic boom section control of spray nozzles. *Computers and Electronics in Agriculture* 70(1), pp. 19–26, <https://doi.org/10.1016/j.compag.2009.08.003>
- Luck, J. D.; Zandonadi, R. S.; Shearer, S. A. (2011): A case study to evaluate field shape factors for estimating overlap errors with manual and automatic section control. *Transactions of the ASABE* 54(4), pp. 1237–1243, <https://doi.org/10.13031/2013.39022>
- Mier, G.; Casado Faulí, A. M.; Valente, J.; de Bruin, S. (2025a): Fields2Benchmark: An open-source benchmark for coverage path planning methods in agriculture. *Smart Agricultural Technology* 12, Article 101156, <https://doi.org/10.1016/j.atech.2025.101156>

- Mier, G.; Fennema, R.; Valente, J.; de Bruin, S. (2025b): Continuous curvature path planning for headland coverage with agricultural robots. *Journal of Field Robotics* 42(3), pp. 641–656, <https://doi.org/10.1002/rob.22489>
- Mier, G.; Valente, J.; de Bruin, S. (2023): Fields2Cover: An open-source coverage path planning library for unmanned agricultural vehicles. *IEEE Robotics and Automation Letters* 8(4), pp. 2166–2172, <https://doi.org/10.1109/LRA.2023.3248439>
- NovAtel (2026): Positioning. OEM7 User Documentation, D100402 v6A, September 2026. <https://docs.novatel.com/OEM7/Content/ReceiverOperation/PositioningOverview.htm>, accessed on 27 Sept 2026
- Schirmer, R.; Biber, P.; Stachniss, C. (2019): Coverage path planning in belief space. 2019 International Conference on Robotics and Automation (ICRA), pp. 7604–7610, <https://doi.org/10.1109/ICRA.2019.8793969>
- Yang, T.; Du, X.; Zhang, B.; Wang, X.; Zhang, Z.; Wu, C. (2025): Coverage path planning based on region segmentation and path orientation optimization. *Agriculture* 15(14), Article 1479, <https://doi.org/10.3390/agriculture15141479>

Authors

Xin Li (ORCID 0009-0005-0670-5006) is a researcher and corresponding author, **Bojian Guo** (ORCID 0009-0005-8353-2436) is a researcher, and **Asel Kartanova** (ORCID 0000-0003-1479-0747) holds a Candidate of Sciences degree in Physics and Mathematics and is an Associate Professor at the Department of Information Systems in Economics, Higher School of Economics and Business, I. Razzakov Kyrgyz State Technical University, 66 Ch. Aitmatov Ave., Bishkek 720044, Kyrgyz Republic, e-mail: lixin26@kstu.kg

Declarations

Author contributions: Xin Li - conceptualization, methodology, software, formal analysis, visualization and original draft; Bojian Guo - methodological review, data curation and result verification; Asel Kartanova - supervision, study-design review and manuscript revision.

Funding: This research received no specific external funding. Conflict of interest: The authors declare no conflict of interest. Ethics: The study involved no human participants, animal experiments or personal data.

Data and code availability: Reproducibility materials are published at <https://doi.org/10.5281/zenodo.22995389> (Zenodo version 5.1). The downloadable archive contains the fixed protocol, timestamp evidence, author code, synthetic geometries, source provenance, field manifests, frozen decisions and numerical results. An obtained archive copy passed verification of all 11,200 listed SHA-256 checksums; seven key published field-bootstrap intervals were independently regenerated from archived field-level outputs using the deposited function and frozen seeds. This checks data integrity and statistical aggregation, not a complete rerun of the field geometry simulations. New author-written code is under the MIT License and author-generated materials under CC BY 4.0. The 350 third-party Fields2Benchmark boundary geometries are not redistributed; readers must obtain those from <https://doi.org/10.5281/zenodo.14524735> under their original terms.

Declaration on generative AI: OpenAI ChatGPT assisted with code revision, computation, analysis auditing, consistency checks, figure and document preparation, and language revision. The revision includes automated checks and documented visual inspection. The authors remain responsible for reviewing the revised calculations, interpretations and references. The tool is not an author; the scientific figures were generated from code and data, not a generative-image model.