



— FIELDHAND AI · FIELD TRIAL SOIL REPORT · PUBLIC SAMPLE

# Repeat grid soil sampling, temporal change and baseline backtesting on two fields in **Northern Iowa**

Sample Farm, Northern Iowa · Field A (about 70 ac) and Field B · Trial  
sampling March 2024 – April 2025; pre-trial baseline June 2022 · Names  
removed, map mirrored

PREPARED BY

Aiterra · Fieldhand AI

REPORT DATE

24 October 2026

EDITION

Public sample of report v2.2

FIELDHAND AI · FIELD TRIAL SOIL REPORT · PUBLIC SAMPLE

# Repeat grid soil sampling, temporal change and baseline backtesting on two fields in Northern Iowa

Sample Farm, Northern Iowa · Field A (about 70 ac) and Field B  
Trial sampling March 2024 – April 2025; pre-trial baseline June 2022



Field A boundary, traced from the farm's 2019 prescription export. Shape mirrored; no coordinates or imagery.

|              |                                                                                                                          |
|--------------|--------------------------------------------------------------------------------------------------------------------------|
| Prepared by  | Aiterra · Fieldhand AI                                                                                                   |
| Grower       | Sample Farm, Northern Iowa (identity withheld)                                                                           |
| Laboratories | Laboratory A (Field A, 2024–2025); Laboratory B (Field B, 2022 baseline)                                                 |
| Report date  | 24 October 2026                                                                                                          |
| Edition      | Public sample of report v2.2. Real data and analysis; names, places, IDs, photos and lab numbers removed; maps mirrored. |

## Summary

This report documents grid soil sampling at Sample Farm, Northern Iowa, under a regional on-farm trial program [1]. The trial-period data are three grid sampling events on Field A (about 70 ac): 23 Mar 2024, 13 Oct 2024 and 20 Apr 2025. Each event used the same 35 planned grid points, and all samples were analysed by Laboratory A (commercial) at 0–6 in. A 71-sample grid on Field B (about 155 ac) from June 2022 serves as a pre-trial baseline. The report covers the trial design, data quality control, the observed variability and change in pH, phosphorus (P), potassium (K), organic matter (OM) and CEC, and a hold-out backtest on the repeat grid: an independent reimplementations of the platform's temporal hold-out protocol, scored against two simple reference estimators that set the benchmarks platform estimates must beat.

## Principal findings

- **Within-field variability.** In April 2025, Field A ranged from pH 5.6–7.3, K 98–278 mg/kg, P1 17–69 mg/kg and OM 1.0–5.5% across 35 samples. On Field B in 2022, pH ranged 4.8–6.8 (median 5.6).
- **Change at the same points.** Between spring 2024 and spring 2025, pH rose at 28 of 35 Field A points and fell at 3 (median +0.2). K rose at 25 points (median +21 mg/kg) and CEC fell at 32. No lime, fertilizer or yield records were available, so the changes cannot be attributed.
- **Data quality.** Coordinate headers were swapped in every laboratory export, and not-run analyses were stored as zeros. In the platform, a Haney soil-health phosphorus value had been taken as the field's phosphorus value. Coordinates and not-run values are corrected in the data used here. In the platform, missing-value handling and the phosphorus record were corrected in October 2026, and weak Bray P1 is now Field A's phosphorus of record [8].
- **Platform backtest.** Under an independent reimplementations of the platform's temporal hold-out protocol, 197 of 210 held-out Field A laboratory values (0.938) fell inside their nominal 90% intervals: pH 63/70, P 66/70, K 68/70. All 13 misses lay above the interval. The intervals were therefore close to nominal and, where they failed, failed on the side of under-estimating later values. The values come from one field and two time steps and cluster by grid point, so the effective sample size is smaller than 210 [13].
- **Reference estimators.** For each later date, a nominal 90% interval taken from the earlier date's field range covered 155 of 175 and 158 of 175 later values on the two forward steps. Coverage fell to 28/35 for pH as the upper end of the field's pH range rose. Predicting each point from its own earlier value cut mean absolute error by 13%–72% per forward step, 7%–66% over the full span, for every property. Point-level information therefore carries over between tests on this field, and the platform's field-level estimates do not yet use it.
- **Zones versus a single rate (new).** On held-out samples, a map beat the field average in all 18 Field A tests (median skill +0.29), and 2-zone maps were confidently better in 5: OM (Mar 2024), K (Apr 2025), OM (Apr 2025), buffer pH (Apr 2025), pH (Apr 2025). The Field A patterns held between dates for Mg, OM, K and pH (rank correlation 0.95–0.96, 0.94, 0.83 and 0.75–0.93) but not for P (0.43–0.46). On Field B (one 2022 date) no zone map was confidently better than a single rate. A single rate is recommended on Field B until it is re-sampled [14].

**Table 1.** Key figures.

| Item                    | Value                                                                                                                                                                                                              |
|-------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Fields                  | 2 (Field A about 70 ac; Field B about 155 ac; about 225 ac in platform)                                                                                                                                            |
| Trial sampling          | Field A: 3 grid events on the same 35 planned points, 23 Mar 2024 – 20 Apr 2025; 105 samples                                                                                                                       |
| Supplementary tests     | Haney soil-health test on 4 Field A samples, April 2025 [3]                                                                                                                                                        |
| Pre-trial baseline      | Field B: 71 grid samples, 15 Jun 2022; no trial-period sampling                                                                                                                                                    |
| Paired change, Field A  | Spring 2024 to spring 2025: median pH +0.2, K +21 mg/kg, P1 +4 mg/kg (computed for this report)                                                                                                                    |
| Platform backtest       | 197 of 210 held-out laboratory values (0.938) inside nominal 90% intervals; pH 0.90, P 0.94, K 0.97 (Section 4.2; reproducibility kit [13])                                                                        |
| Reference estimators    | Nominal 90% field-range intervals: coverage 155/175 and 158/175 on the two forward steps; same-point value beat the field median for 5 of 5 properties (Section 4.3)                                               |
| Zone test (new in v2.2) | 24 held-out field-date × nutrient tests: a map beat a single field-average rate in 18 of 18 Field A tests (median skill +0.29); zones confidently better in 5 tests, all on Field A; Field B (one date): none [14] |

## 1 Trial design and site

### 1.1 Program and objectives

The trial was run with a regional on-farm trial program and the grower, for the trial period May 2024 to December 2025 [1]. The first Field A sampling (March 2024) came before the trial period began, and it is included here as the first event. The scope of work called for soil-fertility zone maps built from 2.5-acre grid samples pre-plant and post-harvest, with model validation by post-harvest grid sampling of the same sites [1].

### 1.2 Site and fields

Both fields are in northern Iowa. Field A was in corn from 2019 to 2023 and in soybeans in 2024, with corn planned for 2025. Corn is strip-tilled and soybeans no-tilled. The grower reported rye and a four-way mix as cover crops, with limited growth in drought, and described the removal of fence lines and the installation of tile drainage, which may explain some anomalies in the zone maps. A recurring rectangular pattern was flagged for further investigation. The 35 Field A grid points span about 990 m by 330 m (computed for this report).

### 1.3 Sampling and laboratory methods

Field A samples were taken at 0–6 in and analysed by Laboratory A (commercial) . Methods, as printed on the reports, were: OM by loss on ignition; P1 by weak Bray and P2 by strong Bray (both 1:7); K, Mg, Ca and Na by neutral ammonium acetate; pH in 1:1 water; buffer index; S by ICAP; Zn, Mn, Fe and Cu by DTPA; and B by sorbitol-DTPA [3][4]. In April 2025 four samples (6, 14, 25 and 27) also received the Haney soil-health test (H3A extraction). Coordinates are the planned grid points, identical to six decimal places on all three dates, so results can be paired by point. Where cores were actually pulled was not recorded. Field B was sampled on 15 June 2022 (Laboratory B, via an a mapping service Map Book). Its depth and methods are not stated in the source records [3].

**Table 2.** Sampling events, given as range (median); K and P1 in mg/kg, OM in %.

| Field   | Sampled     | Laboratory                | Report    | Completed              | n  | Depth (in) | pH            | K             | P1         | OM            |
|---------|-------------|---------------------------|-----------|------------------------|----|------------|---------------|---------------|------------|---------------|
| Field B | 15 Jun 2022 | Laboratory B              | [removed] | 29 Jun 2022 (CSV date) | 71 | n/s        | 4.8–6.8 (5.6) | 101–454 (152) | not used   | 2.1–7.6 (4.5) |
| Field A | 23 Mar 2024 | Laboratory A (commercial) | [removed] | 2 Apr 2024             | 35 | 0–6        | 5.3–7.0 (5.9) | 80–375 (155)  | 11–85 (26) | 1.1–5.8 (3.4) |
| Field A | 13 Oct 2024 | Laboratory A (commercial) | [removed] | 23 Oct 2024            | 35 | 0–6        | 5.5–6.9 (6.1) | 85–338 (160)  | 12–87 (35) | 1.1–6.1 (2.9) |
| Field A | 20 Apr 2025 | Laboratory A (commercial) | [removed] | 29 Apr 2025            | 35 | 0–6        | 5.6–7.3 (6.1) | 98–278 (169)  | 17–69 (35) | 1.0–5.5 (3.1) |

n/s = not stated. Statistics were computed for this report from the raw workbook [3]. Field B P is not reported because its methods are not stated (Section 2.3).

## 1.4 Scope of work delivered

Work delivered: three grid sampling events on Field A (105 samples) with a Haney test on 4 samples; a zone report for Field A reviewed online with the grower and partners; two online consultations; the platform account with corrected Field A records; the zone-map evaluation of October 2026; and this report. Dates and details are in Table 11 (Appendix B).

# 2 Data and quality control

## 2.1 Provenance

The data come from five laboratory files: one Field B CSV, three Field A standard-panel exports and one Field A Haney export. Each file is kept as an Aiterra internal record, available under NDA [3]. For this report the five files were retrieved again and all five matched that record. Apart from the missing-value rule below, no values were imputed or edited.

## 2.2 Quality-control findings

**Table 3.** Quality-control findings and their treatment in this report.

| Finding                      | Scope                                                            | Treatment                                                                                                                                                                         |
|------------------------------|------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Coordinate headers swapped   | All three Laboratory A (commercial) exports                      | "Latitude" holds longitude and vice versa. Corrected coordinates were used throughout [3].                                                                                        |
| Not-run analyses stored as 0 | Apr 2025: Na, NO <sub>3</sub> -N, Mn, Fe, Cu, B in 31 of 35 rows | These were reported only for the four Haney samples (6, 14, 25 and 27). Zeros were treated as missing [3].                                                                        |
| Buffer pH stored as 0        | Mar 2024: 4 samples                                              | All four have pH 6.7–7.0 (computed for this report). Treated as missing. The trial package lists only one such value, so this report corrects it.                                 |
| Zinc outlier                 | Mar 2024, sample 1                                               | 42.9 mg/kg against a maximum of 7.1 in the other 34 samples (computed for this report). Retained and flagged; Zn is not used in Section 3.                                        |
| Phosphorus method mix        | Field A, Apr 2025                                                | Haney (H3A) P had been loaded as the phosphorus value, and H3A cations had replaced NH <sub>4</sub> OAc values at the Haney points [3]. Corrected October 2026 (Section 2.4) [8]. |
| Field B P methods not stated | Field B, 2022                                                    | Columns P_B, P and P_O have no stated method [3]. Field B P is not reported here.                                                                                                 |
| Misleading file names        | Research folder                                                  | Two files named as field boundaries hold the 35 Field A and 71 Field B sample points, not boundaries (checked for this report).                                                   |

## 2.3 Phosphorus method of record

Field A carries three phosphorus measurements: weak Bray P1 and strong Bray P2 on every date, and Haney H3A P on four samples in 2025. They are not interchangeable. In October 2026 the April 2025 weak Bray P1 (report [lab report no. removed]) was set as the Field A phosphorus of record, with Haney results kept separately as soil-health measurements. Blank laboratory cells are now stored as missing rather than zero. Field A Mn, B and Fe, measured on only four samples, are withheld from zone maps and recommendations [8]. This report uses P1 throughout.

## 2.4 Superseded figures

An earlier Field A zone report (March 2025) gave the field area as about 97 ac and listed percentage “accuracy” scores for several properties [6]. Neither the area nor the scores can be traced to the platform record or to a documented analysis, so both are superseded by this report, as are performance figures in an earlier progress update.

# 3 Findings

All statistics in this section are descriptive and were computed for this report from the raw workbook. The Field A events are paired by grid point, while Field B has a single event. With 35 samples per event (about 2 ac per sample), field-level statements are indicative.

## 3.1 Within-field variability

Both fields vary widely. On Field A in April 2025, K ranged 98–278 mg/kg and P1 17–69 mg/kg. OM ranged 1.0–5.5% and CEC 5.5–24.7 meq/100 g. The pattern repeats from date to date (Figure 1): one end of Field A holds the five lowest April 2025 pH values (5.6–5.7). Field B was more acidic in 2022 (median pH 5.6, range 4.8–6.8; Figure 3), with higher OM (median 4.5%) and CEC (median 20.4 meq/100 g) than Field A (Table 2).

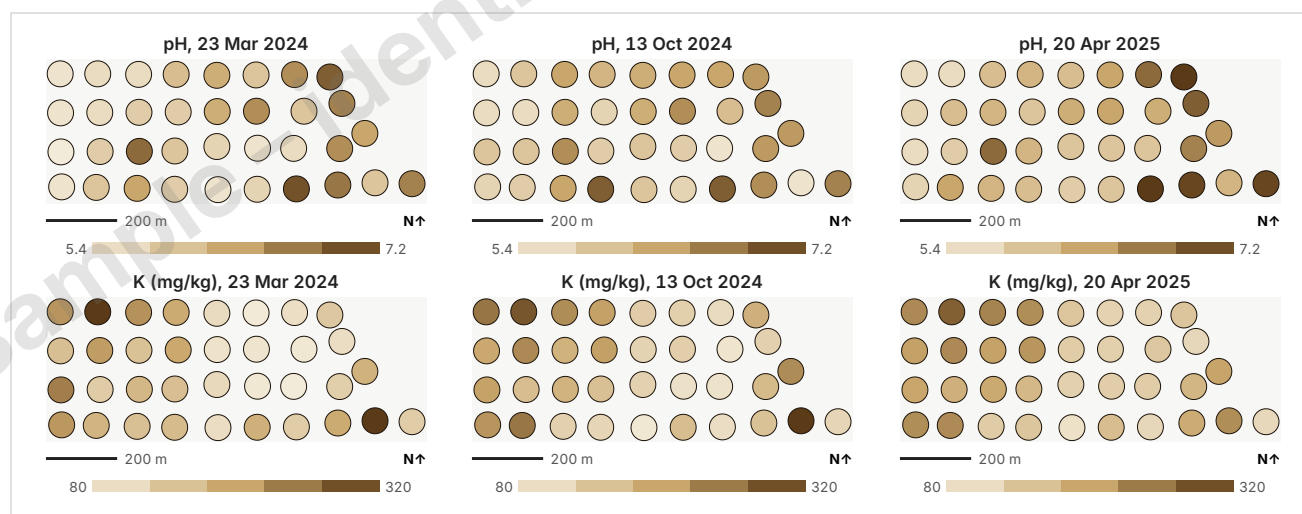


**Figure 1.** Field A: pH (top) and K in mg/kg (bottom) at the 35 grid points on each sampling date. Each point is shaded by its laboratory value (light = low, dark = high; values beyond the legend range take the end colour); nothing is interpolated. Plotted from the corrected coordinates in local metres.

## 3.2 Change between sampling dates

Because the three Field A events share the same planned points, change can be measured point by point (Table 4). Between spring 2024 and spring 2025, pH rose at 28 points and fell at 3. K rose at 25 points (median +21 mg/kg) and P1 at 22 (median +4 mg/kg). OM showed no net change (median 0), and CEC fell at 32 of 35 points (median –0.9 meq/100 g). No lime, fertilizer, manure or yield records were available for 2024–2025, so these

changes cannot be separated from management, sampling-position or laboratory variation. They are still large enough to move a field from one fertility class to another, and they show that a single test goes out of date between seasons.

**Table 4.** Field A, paired change from 23 Mar 2024 to 20 Apr 2025 at the 35 grid points (computed for this report).

| Property                       | Median: Mar 2024 / Oct 2024 / Apr 2025 | Up | Down | No change | Median change |
|--------------------------------|----------------------------------------|----|------|-----------|---------------|
| pH (1:1 water)                 | 5.9 / 6.1 / 6.1                        | 28 | 3    | 4         | +0.2          |
| Organic matter (%)             | 3.4 / 2.9 / 3.1                        | 14 | 17   | 4         | 0             |
| P, weak Bray P1 (mg/kg)        | 26 / 35 / 35                           | 22 | 13   | 0         | +4            |
| K, NH <sub>4</sub> OAc (mg/kg) | 155 / 160 / 169                        | 25 | 10   | 0         | +21           |
| CEC (meq/100 g)                | 12.9 / 12.2 / 12.1                     | 3  | 32   | 0         | -0.9          |

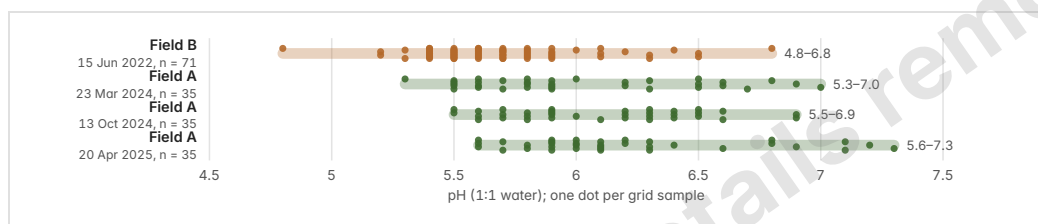


**Figure 2.** pH (1:1 water) in every sample by event; the bar spans the minimum to the maximum.

### 3.3 Field B baseline (2022)

The June 2022 Field B grid (71 samples, about 2.2 ac per sample) predates the trial and has not been repeated. pH ranged 4.8–6.8 (median 5.6), K 101–454 mg/kg (median 152), OM 2.1–7.6% and CEC 10.0–31.3 meq/100 g (Figure 3). Because its laboratory methods and depth are not stated, Field B values are shown for context only. They are not compared directly with Field A, and are not used in Section 4. A trial-period sample on the same grid would be needed before Field B can serve as a second validation field.

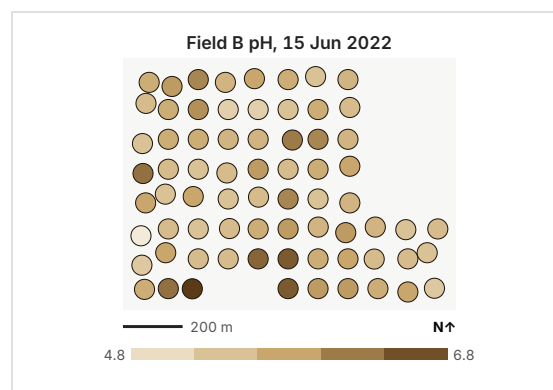


**Figure 3.** Field B pH at the 71 grid points, 15 June 2022, plotted from the source coordinates. The field boundary is not available.

## 4 Model validation and backtesting

### 4.1 Objective and data partition

Every soil estimate in the Aiterra platform carries a stated 90% uncertainty range. A backtest checks whether those ranges can be trusted. The platform's validation protocol includes a temporal field hold-out arm, which is the test used here [12]. For each field and property, each pair of consecutive sampling dates is scored. The estimate for every later sample is the field-level estimate from the earlier date: the earlier date's field mean, with an interval that combines the earlier date's within-field variability, an allowance for the time elapsed and laboratory measurement error. Every laboratory value at the later date is held out and scored. Nothing from the later date is used to form the estimate it is scored against. If the intervals are calibrated, about 90% of held-out values fall inside them [10].

**Partition for this farm.** Training is 23 Mar 2024 and testing 13 Oct 2024; then training 13 Oct 2024 and testing 20 Apr 2025, with 35 later values per property and step. Properties are pH, phosphorus (weak Bray P1, the method of record) and K, the three for which the protocol is defined. Field B is not scored, because it has a single sampling date. The protocol's spatial arm was not run (Section 5.2). The run used an independent reimplementations of the platform protocol, made because the platform harness was not available for this report. Before use, it reproduced a documented reference result on another grower's data exactly (three coverages and their point counts) [12]. Frozen laboratory data, the code and every individual check are an Aiterra internal record, available under NDA, so every figure can be reproduced from them [13].

## 4.2 Platform backtest results

**Table 5.** Platform backtest, Field A (independent reimplementations of the platform's temporal hold-out protocol [12]): estimate  $\pm$  90% half-width, values covered at each step, and pooled results.

| Property                       | Step 1<br>estimate | Covered | Step 2<br>estimate | Covered | Pooled<br>coverage        | 95% CI<br>(naive) | MAE  | Mean CRPS |
|--------------------------------|--------------------|---------|--------------------|---------|---------------------------|-------------------|------|-----------|
| pH (1:1 water)                 | 6.01 $\pm$ 0.86    | 33/35   | 6.10 $\pm$ 0.72    | 30/35   | 63/70 (0.90)              | 0.81–0.95         | 0.35 | 0.25      |
| P, weak Bray P1<br>(mg/kg)     | 32.5 $\pm$ 32.5    | 32/35   | 36.1 $\pm$ 30.8    | 34/35   | 66/70 (0.94)              | 0.86–0.98         | 11.1 | 8.2       |
| K, NH <sub>4</sub> OAc (mg/kg) | 160.2 $\pm$ 125.7  | 33/35   | 168.5 $\pm$ 116.3  | 35/35   | 68/70 (0.97)              | 0.90–0.99         | 44.9 | 31.2      |
| <b>All three</b>               |                    |         |                    |         | <b>197/210</b><br>(0.938) | 0.90–0.96         |      |           |

Step 1: 23 Mar  $\rightarrow$  13 Oct 2024 (0.56 yr). Step 2: 13 Oct 2024  $\rightarrow$  20 Apr 2025 (0.52 yr). The CI is a Wilson interval that ignores clustering by point and step, so it is optimistic. MAE = mean absolute error of the estimate; CRPS = continuous ranked probability score (lower is better) [10].

Pooled coverage was 0.938, against a nominal 0.90. pH reached exactly nominal coverage (63/70), P 0.94 and K 0.97. All 13 misses were above the upper bound of the interval. pH misses rose from 2 at step 1 to 5 at step 2, consistent with the upward shift at the high end of the pH range (Section 4.4). The intervals are wide: the mean pH interval spans 1.57 pH units and the mean K interval 242 mg/kg. Their coverage comes from that width, not from point-level precision.

## 4.3 Reference estimators

Two simple reference estimators, computed for this report, put the platform result in context. Each uses only the earlier date and is scored on the same later values.

- **Field range (no location).** Every point at the later date is predicted by the earlier date's field median. The nominal 90% interval is the earlier date's 5th–95th percentile range.
- **Same point (persistence).** Each point is predicted by its own value at the earlier date. For the October 2024  $\rightarrow$  April 2025 step, the interval half-width is the 90th percentile of absolute change at the 35 points between March and October 2024, so it is set before the scored data are seen.
- **Metrics.** Skill = 1 – MAE(same point)  $\div$  MAE(field median); positive skill means location-specific information helps.

**Table 6.** Reference estimators, step 2: train 13 Oct 2024, score 20 Apr 2025 (n = 35 per property; computed for this report).

| Property                       | Field median [90% range] | Coverage | MAE, field | MAE, same point | Skill | Same-point interval | Coverage |
|--------------------------------|--------------------------|----------|------------|-----------------|-------|---------------------|----------|
| pH (1:1 water)                 | 6.1 [5.57, 6.69]         | 28/35    | 0.35       | 0.27            | 25%   | ±0.46               | 28/35    |
| Organic matter (%)             | 2.9 [1.50, 4.56]         | 30/35    | 0.85       | 0.30            | 65%   | ±0.50               | 30/35    |
| P, weak Bray P1 (mg/kg)        | 35 [16.8, 68.0]          | 34/35    | 9.46       | 8.20            | 13%   | ±22.60              | 35/35    |
| K, NH <sub>4</sub> OAc (mg/kg) | 160 [94.8, 264.5]        | 34/35    | 41.46      | 18.91           | 54%   | ±43.60              | 33/35    |
| CEC (meq/100 g)                | 12.2 [7.20, 19.13]       | 32/35    | 3.05       | 0.89            | 71%   | ±1.90               | 31/35    |

Step 1 (23 Mar → 13 Oct 2024), same estimators: field-range coverage 155/175; same-point skill pH 39%, OM 72%, P1 26%, K 46%, CEC 69%. Over the full span (March 2024 → April 2025) the same-point estimate again beat the field median for all five properties (skill 7%–66%).

#### 4.4 What the results show and do not show

**They show** that on Field A the platform's stated intervals were honest: 197 of 210 held-out values (0.938) fell inside nominal 90% intervals, close to the 0.89–0.90 of the field-range reference. They also show where those intervals can be improved. Every miss was on the high side, and pH coverage fell when the upper end of the field's pH range rose: 7 reference-interval misses in April 2025 against 2 the previous autumn. Most importantly, knowing a point's earlier value cut error relative to the field median for every property and pair of dates (skill 13%–72% on each forward step and 7%–66% over the full span). The same-point estimate also beat the platform's field-level estimate: pH MAE 0.21–0.27 against 0.35, and K 19–26 against 45 mg/kg. Spatial, point-level estimates that keep this coverage while narrowing the intervals are the target for this field.

**They do not show** how accurate platform soil maps are between sample points; the spatial arm was not run. The 210 scored values cluster by point, property and step, so the effective sample size is smaller, and no cluster-robust interval can be formed from one field. Differences between dates combine true change, the unrecorded positions of the cores and laboratory variation. The analysis covers one field over 13 months, used a verified reimplementations rather than the platform harness itself, and has not been independently replicated.

## 5 Zone maps: do zones beat a single rate?

**Method.** Aiterra ran an independent held-out test on the farm's own laboratory files [14]: each sampling event with at least 12 samples was analysed on its own (4 events, 176 samples, 24 event × nutrient tests; Field B P by an unstated method). One sample at a time was held out and predicted from the others by the field mean (flat rate) and by standard reference methods: a smooth map (inverse-distance weighting), ordinary kriging, 2- and 3-zone maps, and a regression on public covariates (USGS 3DEP elevation; SSURGO). **Skill** = 1 – squared error of the method ÷ squared error of the flat rate; 0 is no better than a single rate, +0.5 halves the error. 95% confidence intervals (CI) are from 2,000 paired bootstrap resamples with a fixed seed. Full tables are in Appendix A.

**Results.** On Field A, a smooth map beat the flat rate in 18 of 18 tests (median skill +0.29), and 2-zone maps had a CI above zero in 5: OM (Mar 2024), K (Apr 2025), OM (Apr 2025), buffer pH (Apr 2025), pH (Apr 2025). By nutrient, median 2-zone skill on Field A was OM +0.32, Mg +0.23, P +0.24, K +0.14, buffer pH +0.22 and pH +0.16 (Table 7, Figure 4). The March 2024 pattern held in October 2024 and April 2025 for Mg (rank correlation 0.95–0.96), OM (0.94), K (0.83) and pH (0.75–0.93), but only weakly for P (0.43–0.46) (Table 8). On Field B (71 samples, June 2022) smooth maps beat the flat rate with a CI above zero for Mg, OM, P and pH, but no 2-zone map was confidently better and there is no second date to test stability.

**Uniform, or too few samples to tell?** A field-nutrient is called “Uniform: flat rate is right” only when four tests hold: agronomic tolerance, no held-out skill, weak spatial structure and, where repeat data exist, no stable pattern over time (Table 9) [15]. 2 of 24 tests met the tolerance rule (Field A buffer pH, March and October 2024) and 0 met all of tests 1–3, so no field-nutrient on this farm is shown to be uniform. Where zones are not proven, the flat rate means “not proven either way”; Table 10 gives the state of every test.

## 5.1 Zone recommendations

- **Field A: use zones for pH (and lime), K and OM.** Each was confidently better than a single rate on at least one date, and each pattern held across all three dates. Mg zones are supported by a stable pattern (rank correlation  $\geq 0.95$ ) although no single date proved them on its own.
- **Field A phosphorus: keep a single rate** until more samples prove a pattern; no 2-zone P map was confidently better on any date, and the P pattern was the least stable.
- **Field B: re-sample before zoning.** The only data are from 2022; a new grid at the current density would allow the same test and a stability check. Until then, a single rate is recommended.
- **Label every rate:** "Zoned (skill  $\pm$  CI)", "Single rate: field is uniform" or "Single rate: not enough samples".

**Economic value.** Next proof point: we are working with the grower to bring in yield, fertilizer and lime records. When they arrive, value will be reported in \$/acre/year with a five-year view, every input sourced.

## 6 Limitations and planned validation

### 6.1 Limitations

- **One field with trial-period data.** Field B has only the June 2022 baseline, with no trial-period sampling. Its depth and methods are not stated, so it is not comparable with Field A for P.
- **Sample size.** 35 samples per Field A event (about 2 ac per sample), and 4 Haney samples. Field-level statements are indicative.
- **Management records.** No 2024–2025 lime, fertilizer, manure or yield records were available, so the changes in Section 3.2 cannot be attributed.
- **Boundaries and area.** Recorded areas disagree (about 70 vs 97 ac for Field A; about 155 vs 170 ac for Field B) [3], and no boundary polygon was available, so densities are approximate.
- **Sampling positions.** Coordinates are planned points; where the cores were actually pulled was not recorded. Haney units and depth are not stated in the export [3].
- **Zone test.** 35–71 samples per event; at least 100 points are recommended for a reliable variogram, so nugget:sill and range are indicative only.

### 6.2 Planned validation

- Run the protocol's spatial arm on Field A with the platform harness itself: score point-level estimates for April 2025 trained on March and October 2024, reporting coverage, interval width, CRPS and MAE against the reference estimators, whatever they show.
- Confirm both field boundaries and areas with the grower, regenerate the zone maps on them and repeat the zone test; obtain 2024–2025 lime, fertilizer and yield records so that changes on Field A can be interpreted.

## 7 Platform implications

The trial's findings shaped the Aiterra platform and its Fieldhand application: blank laboratory cells are stored as missing, never as zero; each value is held with its extraction method and sampling date, with a stated phosphorus value of record; properties with too little support are withheld from zones; and Field A is a calibration benchmark that platform estimates must meet. Following the zone-map evaluation, zones are tested per sampling date and nutrient and a field without zones states why. Table 12 (Appendix B) gives the outcome for each finding; it does not describe how the platform works.

## 8 Getting started and contact

The farm's Aiterra account holds the Field B and Field A fields. Fieldhand runs in a web browser on a computer, phone or tablet at [fieldhand.aiterra.ag](https://fieldhand.aiterra.ag) and can be installed from the browser [11]. Access is by an email invitation to the grower's address. The link is valid for 30 days and works only for the invited address. The main sections are Fields (map, field boundaries, zones, weather and soil layers), Plans (a soil test turned into a multi-year plan with tracked goals), Tasks (field work and recommendations), Library (uploaded soil tests, boundaries and reports) and the Fieldhand AI assistant [11]. Fieldhand AI is an AI assistant and can make mistakes; its answers should be checked before buying or applying inputs [11].

Support: [support@aiterra.ag](mailto:support@aiterra.ag) · Aiterra, Austin, TX.

## References and data sources

1. Regional on-farm trial program agreement (private).
2. Earlier trial agreement, 2024 (private).
3. Aiterra trial results package with raw data workbook, October 2026. Aiterra internal record, available under NDA.
4. Laboratory A (commercial), laboratory reports with CSV/XLSX exports (report numbers removed); Laboratory B via a mapping service, Field B soil test export, 2022.
5. Results review with the grower, March 2025 (private).
6. Fieldhand EarnedZones report for Sample Farm, Field A, from the March 2024 samples (presentation, March and April 2025).
7. Grower intake record, 2025 (private).
8. Aiterra engineering record: Field A phosphorus method of record, missing-value handling and analyte holds, October 2026. Aiterra internal record, available under NDA.
9. Online meeting with the grower and a drainage-data partner, April 2025 (private).
10. Gneiting, T., Balabdaoui, F. and Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness. J. R. Stat. Soc. B 69, 243–268.
11. Aiterra Fieldhand AI web application, <https://fieldhand.aiterra.ag> (accessed 18 October 2026).
12. Aiterra platform validation protocol: temporal field hold-out arm of the calibration backtest, October 2026. Aiterra internal record, available under NDA.
13. Aiterra. Temporal hold-out backtest reproducibility kit v1.0, October 2026: frozen laboratory data, code and every check. Aiterra internal record, available under NDA.
14. Aiterra. Zone-map held-out evaluation, Sample Farm, October 2026: analysis scripts with fixed seeds and result files; public covariates from USGS 3DEP (EPQS) and USDA-NRCS SSURGO (Soil Data Access). Aiterra internal record, available under NDA.
15. Cambardella, C. A. et al. (1994). Field-scale variability of soil properties in central Iowa soils. Soil Sci. Soc. Am. J. 58, 1501–1511.

## Appendix A. Zone-map evaluation tables

Supporting tables for Section 5. Every figure is computed from the frozen zone-evaluation result files, an Aiterra internal record, available under NDA [14].

**Table 7.** Held-out skill against a single field-average rate, by nutrient, over 24 tests (both fields) [14].

| Nutrient  | Tests | Smooth map: median skill | Smooth map better than flat | 2-zone: median skill | 2-zone CI above 0 | Median nugget:sill |
|-----------|-------|--------------------------|-----------------------------|----------------------|-------------------|--------------------|
| Mg        | 4     | +0.23                    | 4/4                         | +0.22                | 0/4               | 0.20               |
| K         | 4     | +0.26                    | 4/4                         | +0.14                | 1/4               | 0.04               |
| OM        | 4     | +0.33                    | 4/4                         | +0.26                | 2/4               | 0.40               |
| P         | 4     | +0.23                    | 4/4                         | +0.21                | 0/4               | 0.00               |
| Buffer pH | 4     | +0.27                    | 4/4                         | +0.12                | 1/4               | 0.12               |
| pH        | 4     | +0.23                    | 4/4                         | +0.09                | 1/4               | 0.19               |

Across all 24 tests, median skill was +0.25 for the smooth map (CI above 0 in 71%), +0.17 for kriging (25%), +0.19 for 2 zones (21%), +0.18 for 3 zones (17%) and +0.10 for public covariates alone (29%). Field A P is weak Bray P1; Field B P method not stated.

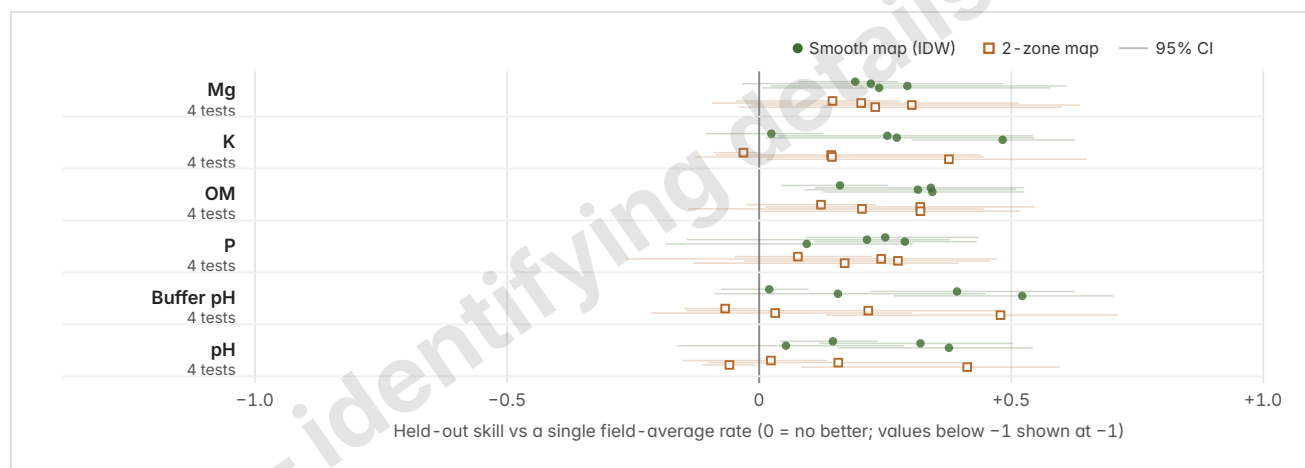


**Figure 4.** Held-out skill of each test, by nutrient, with its 95% CI. Points right of zero beat a single field-average rate; an interval entirely right of zero is a confident gain. Computed for this report [14].

**Table 8.** Temporal transfer on Field A: March 2024 map against later samples, pattern only (each event's mean removed) [14].

| Nutrient  | Rank correlation, Oct 2024 | Rank correlation, Apr 2025 | 3-zone pattern skill, Oct 2024 | 3-zone pattern skill, Apr 2025 |
|-----------|----------------------------|----------------------------|--------------------------------|--------------------------------|
| Mg        | 0.95                       | 0.96                       | +0.41                          | +0.43                          |
| K         | 0.83                       | 0.83                       | +0.58                          | +0.62                          |
| OM        | 0.94                       | 0.94                       | +0.59                          | +0.57                          |
| P         | 0.46                       | 0.43                       | -0.06                          | -0.34                          |
| Buffer pH | 0.76                       | 0.85                       | +0.48                          | +0.45                          |
| pH        | 0.75                       | 0.93                       | +0.43                          | +0.59                          |

Spearman rank correlation between the March 2024 smooth map and the later samples at their locations.

**Table 9.** Objective uniformity test and how many of the 24 tests pass each part [14][15].

| Test                           | Rule                                                                                                                                     | Tests passing                                                |
|--------------------------------|------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------|
| 1. Agronomic tolerance         | Standard deviation $\leq$ tolerance $\div$ 1.96; tolerance $\pm 0.3$ for pH and buffer pH; otherwise coefficient of variation $\leq$ 10% | 2 of 24                                                      |
| 2. No held-out skill           | Upper 95% confidence bound of 2-zone skill below 0.1; separates "uniform" from "cannot tell"                                             | 3 of 24                                                      |
| 3. Weak spatial structure      | Nugget:sill above 0.75, or variogram range shorter than the sample spacing                                                               | 1 of 24                                                      |
| 4. No stable pattern over time | Where repeat samples exist, cross-date rank correlation below 0.3                                                                        | Field A March 2024 map vs two later dates: 0 of 6 tests pass |
| All of 1–3                     |                                                                                                                                          | <b>0 of 24</b>                                               |

Sample spacing for test 3 =  $\sqrt{(\text{field area} \div \text{samples})}$ , computed for this report.

**Table 10.** Zone state by sampling event and nutrient: 2-zone held-out skill and label [14].

| Event                | n  | Mg      | K       | OM      | P       | Buffer pH | pH      |
|----------------------|----|---------|---------|---------|---------|-----------|---------|
| Field B, 15 Jun 2022 | 71 | +0.15 ? | −0.03 ? | +0.12 ? | +0.08 ? | −0.07 ?   | +0.02 ? |
| Field A, 23 Mar 2024 | 35 | +0.20 ? | +0.14 ? | +0.32 Z | +0.24 ? | +0.22 ?   | +0.16 ? |
| Field A, 13 Oct 2024 | 35 | +0.30 ? | +0.14 ? | +0.20 ? | +0.28 ? | +0.03 ?   | −0.06 ? |
| Field A, 20 Apr 2025 | 35 | +0.23 ? | +0.38 Z | +0.32 Z | +0.17 ? | +0.48 Z   | +0.41 Z |

Z = zones beat a flat rate (2-zone 95% CI above 0). U = uniform (tests 1–3 met). ? = not proven either way; flat rate by default.

## Appendix B. Supporting material

**Table 11.** Work delivered under the trial.

| Item                        | Date                         | Description                                                                                                                                                                     |
|-----------------------------|------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Grid soil sampling, Field A | Mar 2024; Oct 2024; Apr 2025 | Three events on the same 35 planned grid points, 105 samples analysed by Laboratory A (commercial); Haney soil-health test on 4 samples in April 2025 [3][4].                   |
| Zone report, Field A        | 28 March 2025                | Soil-fertility zone report built from the March 2024 samples, presented and reviewed online with the grower and program partners [5][6].                                        |
| Consultations               | 28 March 2025; 4 April 2025  | Online results review; online meeting with the grower and a drainage-data partner on tile and water data [5][9].                                                                |
| Platform account            | Oct 2026                     | Field B and Field A fields set up; Field A April 2025 laboratory results re-recorded with their extraction methods, and the phosphorus value of record fixed (Section 2.4) [8]. |
| Zone-map evaluation         | 24 Oct 2026                  | Held-out test of zone maps against a single field-average rate (24 tests) (Section 5) [14].                                                                                     |
| Trial report                | 24 Oct 2026 (v2.2)           | This document, with the backtest in Section 4 computed from the raw laboratory files.                                                                                           |

**Table 12.** Trial findings and resulting platform outcomes.

| Finding                                         | Platform outcome                                                                                                                                                                                  |
|-------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Not-run analyses exported as 0                  | Blank laboratory cells are stored as missing, never as zero [8].                                                                                                                                  |
| Several phosphorus methods on one field         | Each laboratory value is held with its extraction method. Haney soil-health results are kept separate from standard fertility values, and each field has a stated phosphorus value of record [8]. |
| Micronutrients measured on 4 of 35 samples      | Properties with too little support are withheld from zone maps and recommendations [8].                                                                                                           |
| Soil changed between seasons at the same points | Every value is held with its sampling date. Coverage is checked on later tests, because intervals from old tests can be overconfident (Section 4).                                                |
| Field-level intervals honest but wide           | Field A is a calibration benchmark: 0.938 coverage of nominal 90% intervals is the bar to keep while narrowing them (Section 4.2).                                                                |
| Point-level pattern persists between tests      | Platform estimates on Field A must beat the same-point reference estimator, which the current field-level estimate does not (Section 4.4).                                                        |