Evaluation of the ZoomAgri varietal identification technology

For Grains Technology Australia

TECHNICAL REPORT SERIES

Version 1.1

AUTHORS

Fonti Kar
Australian National University
✉ fonti.kar@anu.edu.au
iD 0000-0002-2760-3974

Emi Tanaka
Australian National University
✉ emi.tanaka@anu.edu.au
iD 0000-0002-1455-259X

Patrick Li
Australian National University
✉ patrick.li@anu.edu.au
iD 0000-0003-4959-106X

DATE
22nd July 2025 *(Last modified: 7th August 2025)*

KEYWORDS
barley, goodness-of-fit test

Table of contents

List of Figures

List of Tables

1. Overview

Grain Trade Australia (GTA) is interested in the performance of the ZoomAgri's "Zoom One" variety determination device. This device is designed to identify the varietal(s) of barley (*Hordeum vulgare*) kernels using computer vision and machine learning methods from the scan of the seeds laid out on a tray. An experiment was conducted to evaluate the efficacy and consistency of the devices using 1) samples containing a single variety and 2) samples containing two admixed varieties.

In this report, the overall structure of the supplied data and its visualisation is presented in Section 2, a detailed description of the statistical method used is in Section 3 with key results shown in Section 4 and we conclude with Section 5.

2. Data Description

The data analysed in this report was collected at the GrainCorp laboratory facility on February 26–27, 2025 and summarised in Table 2.1. For each sample, multiple independent evaluations were conducted by different devices. Each evaluation involved classifying seeds that were considered valid, that is, seeds not discarded during pre-processing and deemed both healthy and legitimate barley seeds. For each evaluation, the device predicted the percentage of seeds belonging to the most prevalent variety, as well as, when applicable, the second most prevalent variety. Occasionally the second or third prevalent variety remain unclassified and is labelled as "Others".

Table 2.1. Overview of the barley sample dataset used in this report.

	Total number
Devices	10
Samples	24
Pure samples	12
Admix samples	12
Varieties	7
Evaluations	243

Note that the original data contained a total of 31 samples, however 7 were irrelevant (with no known expected variety composition) and thus removed for analysis in this report.

2.1. Sample structure

For this report, the dataset consisted of 24 samples, including both single variety ("Pure") samples and admixtures of two varieties ("Admix"). Table 2.2 presents the sample codes along with the expected percentages for each of the 7 varieties.

Table 2.2. Expected percentage breakdown of varieties for each sample code. Dots (".") indicate 0% presence of the corresponding variety in that sample.

Sample code	Bass	La Trobe	Maximus CL	Commodus CL	Spartacus CL	Westminster	Planet
GTAPure1	100	.	.	.	.	.	.
GTAPure2	.	100	.	.	.	.	.
GTAPure3	.	.	100	.	.	.	.
GTAPure4	.	.	.	100	.	.	.
GTAPure5	.	.	.	.	100	.	.
GTAPure6	.	.	.	.	.	100	.
GTAPure7	.	.	.	.	.	.	100
GTAPure8	100	.	.	.	.	.	.
GTAPure9	.	100	.	.	.	.	.
GTAPure10	.	.	100	.	.	.	.
GTAPure11	.	.	.	100	.	.	.
GTAPure12	.	.	.	.	100	.	.
GTAMix1	85	.	15	.	.	.	.
GTAMix2	15	.	.	.	.	.	85
GTAMix3	20	.	.	.	80	.	.
GTAMix4	.	80	20	.	.	.	.
GTAMix5	.	75	.	.	.	.	25
GTAMix6	.	.	75	.	.	.	25
GTAMix7	.	.	.	.	70	.	30
GTAMix8	.	.	30	.	.	.	70
GTAMix9	65	35	.	.	.	.	.
GTAMix10	35	.	65	.	.	.	.
GTAMix11	.	40	60	.	.	.	.
GTAMix12	.	.	.	.	50	.	50

2.2. Data structure

All 24 samples were evaluated once on each of the 10 Zoom One devices, except for the sample GTAPure10, which was evaluated four times for the device Quambatook (device ID 354). This yielded a total of 243 evaluations, as shown in Figure 2.1. Visual inspection of the data shown in Figure 2.1 suggests the devices generally appear to perform effectively and consistently, except for some samples. However, a formal statistical test is described in Section 3, with the corresponding results presented in Section 4.

A notable variation in the number of classified seeds is observed across both samples and devices as shown in Figure 2.2. Some samples, such as GTAPure2 and GTAMix2, exhibit
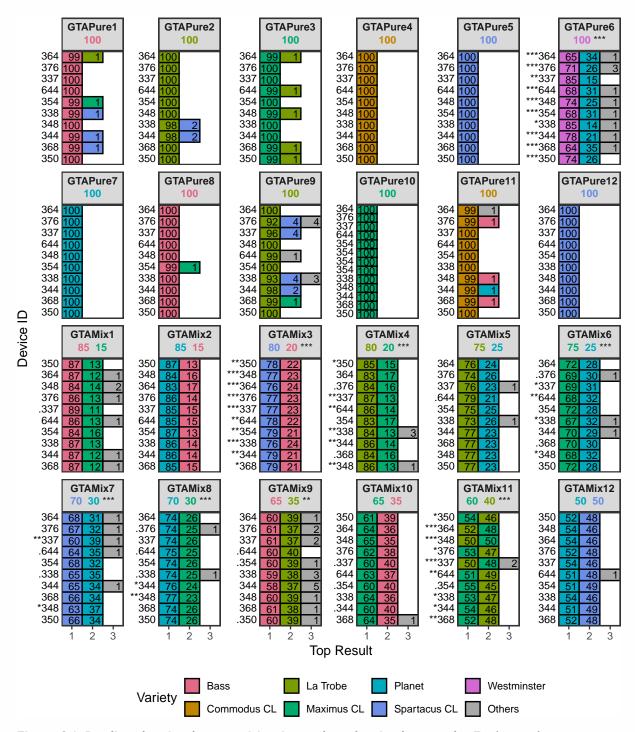
Figure 2.1. Predicted varietal composition for each evaluation by sample. Each panel represents a sample with rows corresponding to individual evaluations (row labelled with device ID) and columns indicating the top predicted varieties in order of abundance. Colored cells show the predicted variety and its classification percentage. The panel title shows the sample code with the expected percentage of variety as indicated by the text color. The color corresponds to the variety as shown in the legend. The symbols next to device ID indicate observation-level performance, while sample-level performance is shown near the sample code label (see Section 4 for performance code explanation).

high counts, while others like GTAPure12 have much lower counts, suggesting uneven seed classification activity across samples. Additionally, the distribution of the classified seed across devices varies (e.g. device Donald Bunker generally has higher classified seeds than Warracknabeal). The latter is formally tested as described in Section 3.5 with results in Section 4.3.
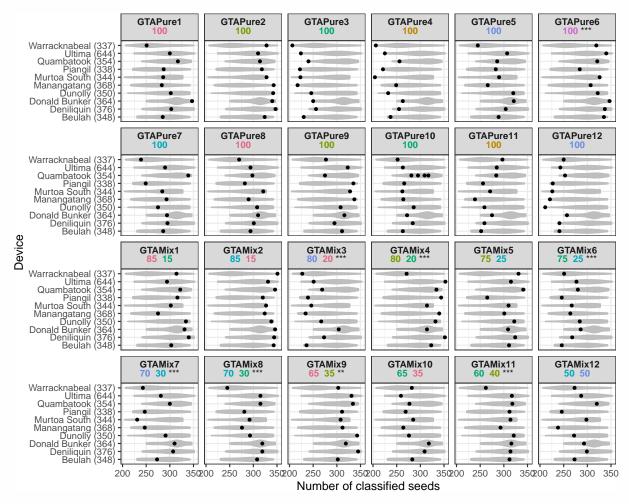

Figure 2.2. Distribution of the number of classified seeds across devices by sample. Within each panel, violin plots show the distribution of classified seed counts across all samples by device name (with ID in bracket) with the black dots marking the observed values for the corresponding sample. See Figure 2.1 for the interpretation of the panel titles.

2.3. Predicted vs expected seed numbers

The predicted and expected seed numbers for each variety at every evaluation was estimated as per Section 3.2. Figure 2.3 visually presents the performance of varietal predictions. We can see, for example, in Figure 2.3 that GTAPure6 is particularly poorly predicted. The efficacy and consistency of these predictions are formally tested as described in Section 3.3 and Section 3.4, respectively, with results in Section 4.1 and Section 4.2.
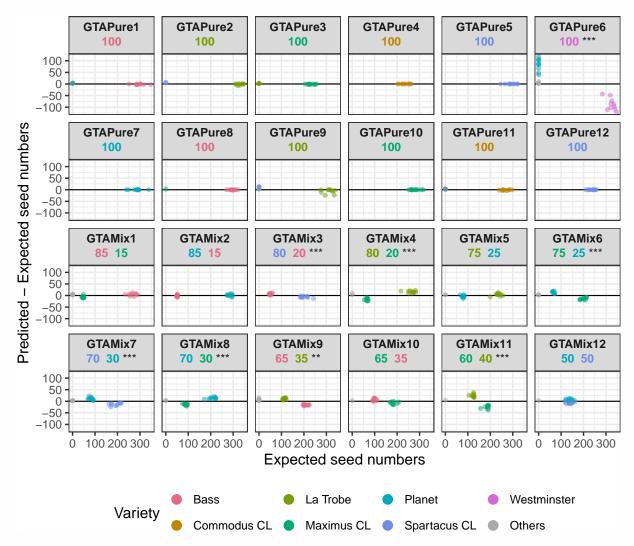
Figure 2.3. The difference in the expected and predicted seed number against the expected seed numbers by sample. The points corresponding to observed predictions colored by the corresponding variety. Like Figure 2.1, the title of each panel shows the sample code and the corresponding percentage of variety mix.

3. Statistical Methods

There are three main statistical tests conducted as follows:

- To test the **efficacy** of the devices, we used a χ^2 goodness-of-fit test to compare the predicted percentages of variety to its expected distribution. For those who are mathematically or statistically inclined, the setting and notations are described in Section 3.1 with further details of the statistical tests to determine prediction efficacy described in Section 3.3.

- To test for **consistency** of the results across devices, we used a Poisson regression to test for the statistical significance of the device effect. Further details of test is given in Section 3.4.

- To test if the average number of **classified seeds** differ across devices, we used a Poisson regression in a similar manner to test for consistency, except the response corresponds to the number of classified seeds and we do not consider differences between samples. Specific details are presented in Section 3.5.

3.1. Settings

Let $n_v = 7, n_s = 24, n_d = 10$, and $n = 243$ be the numbers of varieties, samples, devices and evaluations, respectively. We denote the number of replicates for sample s tested with a device d as r_{sd}. Note that only one sample at one device was replicated 4 times and remaining are unreplicated so $r_{sd} = 1$ for all but one sample.

For convenience, we let $V_s = \{1, \cdots, n_v | \pi_{vs} > 0\}, S = \{1, \cdots, n_s\}, D = \{1, \cdots, n_d\}$ and $R_{sd} = \{1, \dots, r_{sd}\}$ denote the set of index for non-zero varieties for sample s, set of index for samples, set of index for devices and set of replicate number for sample s and device d.

Suppose that for variety v of replicate r of sample s at device d:

- p_{vsdr} and π_{vs} are the predicted and expected proportion of seed count, respectively,
- n_{vsdr} and μ_{vsdr} are the predicted and expected seed count, respectively, and
- $n_{\cdot sdr} = \sum_{v=1}^{n_v} n_{vsdr}$ is the total number of classified seed counts.

A single evaluation, given by the 3-tuple (s, d, r), corresponds to a unique combination of replicate r of sample s at device d with an observation corresponding to 1-, 2- or 3- tuple of the proportion of classified seeds belonging to the top 1-3 prevalent varieties, i.e. $(p_{(1)sdr})$, $(p_{(1)sdr}, p_{(2)sdr})$, or $(p_{(1)sdr}, p_{(2)sdr}, p_{(3)sdr})$ where $p_{(v)sdr}$ is the proportion corresponding to the v-th top abundant classified or unclassified ("Others") variety such that $p_{(v)sdr} > p_{(v+1)sdr}$. Note that $p_{(3)sdr}$ appears to always correspond to "Others" (see Figure 2.1).

3.2. Estimating the predicted and expected seed numbers

The predicted seed number of evaluation (s, d, r) for variety v is simply calculated as the product of the number of classified seeds and the corresponding predicted proportion p_{vsdr}, i.e. $n_{vsdr} = n_{\cdot sdr} p_{vsdr}$. Similarly the expected seed number of evaluation (s, d, r) for variety v is $\mu_{vsdr} = n_{\cdot sdr} \pi_{vsdr}$.

3.3. Testing efficacy of varietal predictions

Whether the device works or not could be formulated as a simple goodness-of-fit test with

$$H_0 : p_{vsdr} = \pi_{vs} \text{ for all } v \in V_s, s \in S, d \in D, r \in R_{sw} \tag{3.1}$$

and

$$H_A : p_{vsdr} \neq \pi_{vs} \text{ for some } v \in V_s, s \in S, d \in D, r \in R_{sw}.$$

Under this formulation, rejection of the H_0 implies that the device is performing poorly.

From Figure 2.3, we can tell that there is a substantial difference by sample, thus instead of conducting an evaluation of overall device performance, we consider its performance on two levels as described next.

3.3.1. By evaluation

For one evaluation, we fix $s \in S, d \in D$ and $r \in R_{sd}$ to say, s_0, d_0 and r_0, respectively, and modify hypotheses (3.1) to $H_0 : p_{vs_0d_0r_0} = \pi_{vs_0}$ for all $v \in V_{s_0}$.

We let $E_v = n_{\cdot s_0d_0r_0} \pi_{vs_0}$ and $O_v = n_{\cdot s_0d_0r_0} p_{vs_0d_0r_0}$ for $v \in V_{s_0}$ and perform a χ^2 test using the test statistic

$$\mathrm{X}^2_{s_0d_0r_0} = \sum_{v \in V_{s_0}} \frac{(E_{vs_0d_0r_0} - O_{vs_0d_0r_0})^2}{E_{vs_0d_0r_0}}. \tag{3.2}$$

We assume that $\mathrm{X}^2_{s_0d_0r_0} \sim \chi^2_{\nu_{s_0}}$ under H_0, where the degrees of freedom is given as $\nu_{s_0} = \max\{|V_{s_0}| - 1, 1\}$, and the corresponding p-value is $P(\chi^2_{\nu_{s_0}} \geq \mathrm{X}^2_{s_0d_0r_0})$. Given that $|V_{s_0}| = 1$ for pure samples and $|V_{s_0}| = 2$ for admixed samples, $\nu_{s_0} = 1$ for all evaluations. It should be cautioned for pure samples that the degrees of freedom calculation is technically incorrect and thus the resulting p-value should not be interpreted as the technical definition in Wasserstein and Lazar (2016), however, it can be used reasonably as a measure of performance. We also note that that $\sum_{v \in V_{s_0}} p_{vs_0d_0r_0} < 1$ when a predicted variety is either unclassified or incorrect. In general, $\sum_{v \in V_{s_0}} p_{vs_0d_0r_0} \approx 1$ so we consider this effect to be negligible.

3.3.2. By sample

The performance assessment in Section 3.3.1 results in n assessments. We would normally prefer to have an overall performance assessment, however, due to obvious performance variability by sample (see Figure 2.3), we consider studying the efficacy by sample, i.e. test $H_0 : p_{vs_0dr}$ for all $v \in V_{s_0}, d \in D$ and $r \in R_{s_0d}$. This can be done by aggregating the test statistic in Equation 3.2 over all the evaluations for a sample, forming a new test statistic:

$$\mathrm{X}^2_{s_0} = \sum_{v \in V_{s_0}} \sum_{d \in D} \sum_{r \in R_{s_0d}} \mathrm{X}^2_{s_0 d_0 r_0}.$$

Under H_0, $\mathrm{X}^2_{s_0}$ is approximately distributed as $\chi^2_{\tau_{s_0}}$ where the degrees of freedom is estimated as $\tau_{s_0} = \sum_{w \in W} \sum_{r \in R_{s_0w}} \max\{|V_{s_0}| - 1, 1\}$. Then the corresponding p-value is calculated from $P(\chi^2_{\tau_{s_0}} \geq \mathrm{X}^2_{s_0})$. Same caution as Section 3.3.1 for the interpretation of this p-value hold here.

3.4. Testing prediction consistency across devices

To evaluate the consistency of the device performance, we assumed that adequate performance is defined by low between device variance and variety predictions across devices for the same samples should be similar. For $in V, s \in S, d \in D$ and $r \in R_{sd}$, we model the expectation of the predicted variety seed count, n_{vdsr}, using a Poisson regression:

$$\log(\mathbb{E}(n_{vsdr})) = \log(n_{\cdot sdr}) + \alpha_{vsdr} \tag{3.3}$$

where $\log(n_{\cdot sdr})$ is an offset of log of the number of classified seeds for evaluation (s, d, r) and α_{vsdr} is the combined effect of v, s, d and r. Under the model (3.3), we have a fully saturated model such that $\hat{\alpha}_{vsdr} = n_{vsdr}$.

We consider the alternative model

$$\log(\mathbb{E}(n_{vsdr})) = \log(n_{\cdot sdr}) + \alpha_{vsr} \tag{3.4}$$

where α_{vsr} is the combined effect of v, s and r (i.e. not dependent on the device). We compute the residual deviance of this model from the fully saturated model. We assume that the residual deviance has a χ^2 distribution with the degrees of freedom computed from the residual degrees of freedom for model (3.4).

3.5. Testing consistency of the classified seeds across devices

If we assume that each evaluation consisted of roughly similar number of seeds, then any difference in the classified seeds are attributed to the differences in the number of discarded seeds. We consider modelling the number of classified seeds, $n_{\cdot sdr}$, using a Poisson regression:

$$\log(\mathbb{E}(n_{\cdot sdr})) = \mu + \alpha_d \tag{3.5}$$

where μ is the overall intercept and α_d is the effect of device. We compute the difference in the deviance for this model and the null model $H_0 : \alpha_d = 0$ for all $d \in D$. The statistical significance of the device effect is calculated by assuming a χ^2 distribution with the degrees of freedom computed from the difference in residual degrees of freedom for model (3.5) and the null model.

4. Key Results

Throughout this report, the p-values are coded using the **performance code**: *** for $p < 0.001$, ** for $p < 0.01$, * for $p < 0.05$, and · for $p < 0.1$. It should be noted that the degrees of freedom for the χ^2 test for pure samples in determining the prediction efficacy is not technically correct, thus p-values should be taken as a measure of performance, where small values < 0.05 are generally poor and values > 0.05 are generally acceptable.

4.1. Prediction efficacy by sample

Table 4.1 shows the results for the assessment of prediction efficacy by sample as described in Section 3.3.2. Pure samples show acceptable predictive performance ($p \geq 0.05$) except for GTAPure6. This could be either the result of potential sample contamination or poor performance for the Westminster variety. Some admixed samples (specifically, GTAMix1, GTAMix2, GTAMix5, GTAMix10, and GTAMix12) exhibit clear two-variety predictions without significant issues, while remaining admix samples (e.g., GTAMix4, GTAMix11) show systematic deviation at the sample level.

Table 4.1. Results for testing for efficacy by sample. The table show the sample code, variety mixture, the test statistic, degrees of freedom, p-value and performance code.

Sample code	Variety mixture	X^2	df	p-value	Performance code
GTAMix1	Bass (85%), Maximus CL (15%)	13.91	10	0.177	
GTAMix2	Planet (85%), Bass (15%)	4.52	10	0.921	
GTAMix3	Spartacus CL (80%), Bass (20%)	111.14	10	<0.001	***
GTAMix4	La Trobe (80%), Maximus CL (20%)	58.59	10	<0.001	***
GTAMix5	La Trobe (75%), Planet (25%)	6.82	10	0.742	
GTAMix6	Maximus CL (75%), Planet (25%)	38.82	10	<0.001	***
GTAMix7	Spartacus CL (70%), Planet (30%)	31.12	10	<0.001	***
GTAMix8	Planet (70%), Maximus CL (30%)	33.10	10	<0.001	***
GTAMix9	Bass (65%), La Trobe (35%)	23.30	10	0.010	**
GTAMix10	Maximus CL (65%), Bass (35%)	13.00	10	0.223	
GTAMix11	Maximus CL (60%), La Trobe (40%)	88.67	10	<0.001	***
GTAMix12	Planet (50%), Spartacus CL (50%)	8.40	10	0.590	
GTAPure1	Bass (100%)	0.15	10	>0.999	
GTAPure2	La Trobe (100%)	0.26	10	>0.999	
GTAPure3	Maximus CL (100%)	0.09	10	>0.999	
GTAPure4	Commodus CL (100%)	0.00	10	>0.999	
GTAPure5	Spartacus CL (100%)	0.00	10	>0.999	
GTAPure6	Westminster (100%)	252.96	10	<0.001	***
GTAPure7	Planet (100%)	0.00	10	>0.999	
GTAPure8	Bass (100%)	0.03	10	>0.999	
GTAPure9	La Trobe (100%)	4.17	10	0.939	
GTAPure10	Maximus CL (100%)	0.00	13	>0.999	
GTAPure11	Commodus CL (100%)	0.15	10	>0.999	
GTAPure12	Spartacus CL (100%)	0.00	10	>0.999	

For prediction efficacy results by evaluation see Section A.

4.2. Prediction consistency

The test from Section 3.4 yields a test statistic of 70.1 with degrees of freedom of 324 and p-value of >0.999.

4.3. Consistency of classified seed across devices

The test from Section 3.5 yields a test statistic of 77.2 with degrees of freedom of 9 and p-value of <0.001.

5. Conclusion

As shown in Section 4, we generally find that the device is consistent in performance with acceptable predictive performance for pure samples (except for Westminster variety, which may be due to contamination), however, has questionable performance for admix samples. However, Section 3.4 shows that the variety prediction is generally biased towards one direction, e.g. GTAMix7 consistently under predicts the Sparatacus CL variety and over predicts the Planet variety. This suggests that some admix samples may not have a uniform mixture and poor predictive performance of admix samples may be explained by assuming an incorrect expected percentage of mixtures.

It is important to also note that the device goes through a process of sorting where seeds that are not deemed adequate for analysis are discarded. The extent to which this process can introduce further sampling error that alters the expected proportion of varietals in an admixed sample is unknown and may require further investigation.

We were unable to fully explore the effects of prolonged use of the devices on the prediction efficacy and consistency of variety determination from the data supplied. Future experiments may like to consider whether the device performance is consistent over time and ensure a uniform mixture of varieties that reflect the expected percentage of varieties.

A. Appendix

Table A1. Results for testing for efficacy by evaluation. The table show the device name, sample code, number of classified seeds, the test statistic, p-value and performance code.

Device	Sample code	Classified seeds	X^2	p-value	Performance code
Beulah	GTAPure1	285	0.00	>0.999	
Deniliquin	GTAPure1	303	0.00	>0.999	
Donald Bunker	GTAPure1	347	0.03	0.872	
Dunolly	GTAPure1	302	0.00	>0.999	
Manangatang	GTAPure1	283	0.03	0.858	
Murtoa South	GTAPure1	286	0.03	0.859	
Piangil	GTAPure1	287	0.03	0.859	
Quambatook	GTAPure1	317	0.03	0.866	
Ultima	GTAPure1	300	0.00	>0.999	
Warracknabeal	GTAPure1	251	0.00	>0.999	
Beulah	GTAPure2	324	0.00	>0.999	
Deniliquin	GTAPure2	347	0.00	>0.999	
Donald Bunker	GTAPure2	340	0.00	>0.999	
Dunolly	GTAPure2	342	0.00	>0.999	
Manangatang	GTAPure2	343	0.00	>0.999	

Murtoa South	GTAPure2	328	0.15	0.699	
Piangil	GTAPure2	318	0.11	0.737	
Quambatook	GTAPure2	314	0.00	>0.999	
Ultima	GTAPure2	310	0.00	>0.999	
Warracknabeal	GTAPure2	328	0.00	>0.999	
Beulah	GTAPure3	230	0.02	0.895	
Deniliquin	GTAPure3	256	0.00	>0.999	
Donald Bunker	GTAPure3	250	0.02	0.899	
Dunolly	GTAPure3	246	0.02	0.899	
Manangatang	GTAPure3	217	0.02	0.892	
Murtoa South	GTAPure3	223	0.00	>0.999	
Piangil	GTAPure3	222	0.00	>0.999	
Quambatook	GTAPure3	240	0.00	>0.999	
Ultima	GTAPure3	224	0.02	0.894	
Warracknabeal	GTAPure3	206	0.00	>0.999	
Beulah	GTAPure4	237	0.00	>0.999	
Deniliquin	GTAPure4	255	0.00	>0.999	
Donald Bunker	GTAPure4	263	0.00	>0.999	
Dunolly	GTAPure4	232	0.00	>0.999	
Manangatang	GTAPure4	249	0.00	>0.999	
Murtoa South	GTAPure4	204	0.00	>0.999	
Piangil	GTAPure4	221	0.00	>0.999	
Quambatook	GTAPure4	256	0.00	>0.999	
Ultima	GTAPure4	225	0.00	>0.999	
Warracknabeal	GTAPure4	206	0.00	>0.999	
Beulah	GTAPure5	289	0.00	>0.999	
Deniliquin	GTAPure5	304	0.00	>0.999	
Donald Bunker	GTAPure5	321	0.00	>0.999	
Dunolly	GTAPure5	320	0.00	>0.999	
Manangatang	GTAPure5	266	0.00	>0.999	
Murtoa South	GTAPure5	290	0.00	>0.999	
Piangil	GTAPure5	283	0.00	>0.999	
Quambatook	GTAPure5	286	0.00	>0.999	
Ultima	GTAPure5	307	0.00	>0.999	
Warracknabeal	GTAPure5	245	0.00	>0.999	
Beulah	GTAPure6	335	22.59	<0.001	***
Deniliquin	GTAPure6	337	28.50	<0.001	***
Donald Bunker	GTAPure6	347	42.19	<0.001	***
Dunolly	GTAPure6	322	21.91	<0.001	***
Manangatang	GTAPure6	307	40.13	<0.001	***
Murtoa South	GTAPure6	326	15.90	<0.001	***
Piangil	GTAPure6	284	6.51	0.011	*
Quambatook	GTAPure6	321	33.05	<0.001	***
Ultima	GTAPure6	340	34.94	<0.001	***
Warracknabeal	GTAPure6	319	7.22	0.007	**
Beulah	GTAPure7	286	0.00	>0.999	

Deniliquin	GTAPure7	295	0.00	>0.999
Donald Bunker	GTAPure7	294	0.00	>0.999
Dunolly	GTAPure7	275	0.00	>0.999
Manangatang	GTAPure7	293	0.00	>0.999
Murtoa South	GTAPure7	284	0.00	>0.999
Piangil	GTAPure7	249	0.00	>0.999
Quambatook	GTAPure7	339	0.00	>0.999
Ultima	GTAPure7	290	0.00	>0.999
Warracknabeal	GTAPure7	239	0.00	>0.999
Beulah	GTAPure8	294	0.00	>0.999
Deniliquin	GTAPure8	301	0.00	>0.999
Donald Bunker	GTAPure8	310	0.00	>0.999
Dunolly	GTAPure8	308	0.00	>0.999
Manangatang	GTAPure8	290	0.00	>0.999
Murtoa South	GTAPure8	321	0.00	>0.999
Piangil	GTAPure8	282	0.00	>0.999
Quambatook	GTAPure8	298	0.03	0.862
Ultima	GTAPure8	294	0.00	>0.999
Warracknabeal	GTAPure8	270	0.00	>0.999
Beulah	GTAPure9	311	0.00	>0.999
Deniliquin	GTAPure9	295	1.95	0.162
Donald Bunker	GTAPure9	316	0.00	>0.999
Dunolly	GTAPure9	308	0.00	>0.999
Manangatang	GTAPure9	337	0.03	0.870
Murtoa South	GTAPure9	328	0.15	0.699
Piangil	GTAPure9	335	1.58	0.209
Quambatook	GTAPure9	275	0.00	>0.999
Ultima	GTAPure9	323	0.03	0.867
Warracknabeal	GTAPure9	277	0.44	0.509
Beulah	GTAPure10	263	0.00	>0.999
Deniliquin	GTAPure10	284	0.00	>0.999
Donald Bunker	GTAPure10	272	0.00	>0.999
Dunolly	GTAPure10	286	0.00	>0.999
Manangatang	GTAPure10	264	0.00	>0.999
Murtoa South	GTAPure10	262	0.00	>0.999
Piangil	GTAPure10	266	0.00	>0.999
Quambatook	GTAPure10	281	0.00	>0.999
Quambatook	GTAPure10	317	0.00	>0.999
Quambatook	GTAPure10	295	0.00	>0.999
Quambatook	GTAPure10	309	0.00	>0.999
Ultima	GTAPure10	263	0.00	>0.999
Warracknabeal	GTAPure10	252	0.00	>0.999
Beulah	GTAPure11	252	0.04	0.850
Deniliquin	GTAPure11	259	0.03	0.852
Donald Bunker	GTAPure11	275	0.03	0.856
Dunolly	GTAPure11	259	0.00	>0.999

Manangatang	GTAPure11	239	0.02	0.897	
Murtoa South	GTAPure11	271	0.03	0.855	
Piangil	GTAPure11	256	0.00	>0.999	
Quambatook	GTAPure11	284	0.00	>0.999	
Ultima	GTAPure11	285	0.00	>0.999	
Warracknabeal	GTAPure11	297	0.00	>0.999	
Beulah	GTAPure12	241	0.00	>0.999	
Deniliquin	GTAPure12	241	0.00	>0.999	
Donald Bunker	GTAPure12	257	0.00	>0.999	
Dunolly	GTAPure12	211	0.00	>0.999	
Manangatang	GTAPure12	221	0.00	>0.999	
Murtoa South	GTAPure12	226	0.00	>0.999	
Piangil	GTAPure12	227	0.00	>0.999	
Quambatook	GTAPure12	253	0.00	>0.999	
Ultima	GTAPure12	243	0.00	>0.999	
Warracknabeal	GTAPure12	250	0.00	>0.999	
Beulah	GTAMix1	303	0.23	0.628	
Deniliquin	GTAMix1	340	0.99	0.319	
Donald Bunker	GTAMix1	331	2.17	0.140	
Dunolly	GTAMix1	334	1.15	0.283	
Manangatang	GTAMix1	275	1.67	0.197	
Murtoa South	GTAMix1	302	1.94	0.164	
Piangil	GTAMix1	316	0.90	0.343	
Quambatook	GTAMix1	322	0.39	0.531	
Ultima	GTAMix1	294	0.85	0.355	
Warracknabeal	GTAMix1	314	3.60	0.058	.
Beulah	GTAMix2	343	0.37	0.544	
Deniliquin	GTAMix2	344	0.36	0.547	
Donald Bunker	GTAMix2	346	1.11	0.292	
Dunolly	GTAMix2	338	1.13	0.287	
Manangatang	GTAMix2	324	0.00	>0.999	
Murtoa South	GTAMix2	327	0.22	0.642	
Piangil	GTAMix2	320	0.22	0.639	
Quambatook	GTAMix2	346	1.11	0.292	
Ultima	GTAMix2	331	0.00	>0.999	
Warracknabeal	GTAMix2	351	0.00	>0.999	
Beulah	GTAMix3	236	12.11	<0.001	***
Deniliquin	GTAMix3	273	13.89	<0.001	***
Donald Bunker	GTAMix3	304	18.67	<0.001	***
Dunolly	GTAMix3	267	10.62	0.001	**
Manangatang	GTAMix3	234	6.58	0.010	*
Murtoa South	GTAMix3	246	7.16	0.007	**
Piangil	GTAMix3	239	14.42	<0.001	***
Quambatook	GTAMix3	269	7.52	0.006	**
Ultima	GTAMix3	251	8.96	0.003	**
Warracknabeal	GTAMix3	227	11.21	<0.001	***

Beulah	GTAMix4	323	9.69	0.002	**
Deniliquin	GTAMix4	352	3.50	0.062	.
Donald Bunker	GTAMix4	314	1.99	0.159	
Dunolly	GTAMix4	332	4.84	0.028	*
Manangatang	GTAMix4	340	3.60	0.058	.
Murtoa South	GTAMix4	314	7.17	0.007	**
Piangil	GTAMix4	344	9.06	0.003	**
Quambatook	GTAMix4	334	1.87	0.172	
Ultima	GTAMix4	353	8.53	0.003	**
Warracknabeal	GTAMix4	271	8.35	0.004	**
Beulah	GTAMix5	311	0.62	0.433	
Deniliquin	GTAMix5	324	0.15	0.700	
Donald Bunker	GTAMix5	309	0.16	0.693	
Dunolly	GTAMix5	322	0.60	0.439	
Manangatang	GTAMix5	301	0.64	0.424	
Murtoa South	GTAMix5	310	0.84	0.360	
Piangil	GTAMix5	265	0.32	0.573	
Quambatook	GTAMix5	341	0.00	>0.999	
Ultima	GTAMix5	315	2.86	0.091	.
Warracknabeal	GTAMix5	331	0.65	0.418	
Beulah	GTAMix6	246	6.23	0.013	*
Deniliquin	GTAMix6	268	3.80	0.051	.
Donald Bunker	GTAMix6	286	1.19	0.276	
Dunolly	GTAMix6	284	1.52	0.217	
Manangatang	GTAMix6	264	3.41	0.065	.
Murtoa South	GTAMix6	267	2.34	0.126	
Piangil	GTAMix6	246	6.62	0.010	*
Quambatook	GTAMix6	280	1.22	0.270	
Ultima	GTAMix6	277	7.72	0.005	**
Warracknabeal	GTAMix6	251	4.77	0.029	*
Beulah	GTAMix7	273	6.29	0.012	*
Deniliquin	GTAMix7	307	0.77	0.381	
Donald Bunker	GTAMix7	310	0.26	0.608	
Dunolly	GTAMix7	291	2.36	0.124	
Manangatang	GTAMix7	247	1.93	0.165	
Murtoa South	GTAMix7	231	2.34	0.126	
Piangil	GTAMix7	247	2.78	0.096	.
Quambatook	GTAMix7	300	0.57	0.450	
Ultima	GTAMix7	281	3.80	0.051	.
Warracknabeal	GTAMix7	243	10.02	0.002	**
Beulah	GTAMix8	308	6.84	0.009	**
Deniliquin	GTAMix8	319	3.42	0.064	.
Donald Bunker	GTAMix8	319	2.52	0.113	
Dunolly	GTAMix8	293	2.34	0.126	
Manangatang	GTAMix8	276	2.08	0.149	
Murtoa South	GTAMix8	292	5.27	0.022	*

Piangil	GTAMix8	281	2.95	0.086	.
Quambatook	GTAMix8	315	2.30	0.129	
Ultima	GTAMix8	315	3.56	0.059	.
Warracknabeal	GTAMix8	245	1.82	0.177	
Beulah	GTAMix9	302	2.51	0.113	
Deniliquin	GTAMix9	345	1.28	0.258	
Donald Bunker	GTAMix9	319	2.52	0.112	
Dunolly	GTAMix9	343	2.93	0.087	.
Manangatang	GTAMix9	312	1.75	0.186	
Murtoa South	GTAMix9	308	2.54	0.111	
Piangil	GTAMix9	311	2.53	0.112	
Quambatook	GTAMix9	334	2.78	0.096	.
Ultima	GTAMix9	331	3.40	0.065	.
Warracknabeal	GTAMix9	303	1.07	0.301	
Beulah	GTAMix10	283	0.00	>0.999	
Deniliquin	GTAMix10	309	1.15	0.283	
Donald Bunker	GTAMix10	318	0.12	0.724	
Dunolly	GTAMix10	276	1.92	0.165	
Manangatang	GTAMix10	264	0.05	0.819	
Murtoa South	GTAMix10	285	3.02	0.082	.
Piangil	GTAMix10	269	0.15	0.701	
Quambatook	GTAMix10	277	3.11	0.078	.
Ultima	GTAMix10	259	0.42	0.515	
Warracknabeal	GTAMix10	282	3.05	0.081	.
Beulah	GTAMix11	312	12.83	<0.001	***
Deniliquin	GTAMix11	314	6.42	0.011	*
Donald Bunker	GTAMix11	316	19.06	<0.001	***
Dunolly	GTAMix11	321	5.20	0.023	*
Manangatang	GTAMix11	293	8.20	0.004	**
Murtoa South	GTAMix11	314	4.29	0.038	*
Piangil	GTAMix11	312	6.46	0.011	*
Quambatook	GTAMix11	318	3.36	0.067	.
Ultima	GTAMix11	317	10.30	0.001	**
Warracknabeal	GTAMix11	262	12.56	<0.001	***
Beulah	GTAMix12	273	1.62	0.202	
Deniliquin	GTAMix12	299	0.41	0.524	
Donald Bunker	GTAMix12	293	1.82	0.178	
Dunolly	GTAMix12	272	0.37	0.544	
Manangatang	GTAMix12	238	0.42	0.517	
Murtoa South	GTAMix12	298	0.12	0.728	
Piangil	GTAMix12	246	1.63	0.202	
Quambatook	GTAMix12	320	0.11	0.737	
Ultima	GTAMix12	287	0.28	0.598	
Warracknabeal	GTAMix12	273	1.62	0.202	

A.1. Computational details

All analysis were conducted using the programming language, R (R Core Team 2025). The data wrangling are predominately done using Wickham et al. (2019). All figures in this report are reproducibly drawn using Wickham (2016). The report is reproducibly written using Quarto. All model fitting and significance calculation are done using the `stats` package.

```
- Session info -------------------------------------------------------------
 setting  value
 version  R version 4.5.0 (2025-04-11)
 os       macOS Sequoia 15.5
 system   aarch64, darwin20
 ui       X11
 language (EN)
 collate  en_US.UTF-8
 ctype    en_US.UTF-8
 tz       Australia/Sydney
 date     2025-08-07
 pandoc   3.4 @ /Applications/RStudio.app/Contents/Resources/app/quarto/bin/tools/aarch64/ (
 quarto   1.7.30 @ /usr/local/bin/quarto

- Packages -----------------------------------------------------------------
 package      * version date (UTC) lib source
 cli            3.6.5   2025-04-23 [1] CRAN (R 4.5.0)
 colorspace     2.1-1   2024-07-26 [1] CRAN (R 4.5.0)
 commonmark     2.0.0   2025-07-07 [1] CRAN (R 4.5.0)
 digest         0.6.37  2024-08-19 [1] CRAN (R 4.5.0)
 dplyr        * 1.1.4   2023-11-17 [1] CRAN (R 4.5.0)
 evaluate       1.0.3   2025-01-10 [1] CRAN (R 4.5.0)
 farver         2.1.2   2024-05-13 [1] CRAN (R 4.5.0)
 fastmap        1.2.0   2024-05-15 [1] CRAN (R 4.5.0)
 forcats      * 1.0.0   2023-01-29 [1] CRAN (R 4.5.0)
 generics       0.1.4   2025-05-09 [1] CRAN (R 4.5.0)
 ggplot2      * 3.5.2   2025-04-09 [1] CRAN (R 4.5.0)
 ggtext         0.1.2   2022-09-16 [1] CRAN (R 4.5.0)
 glue           1.8.0   2024-09-30 [1] CRAN (R 4.5.0)
 gridtext       0.1.5   2025-07-19 [1] Github (wilkelab/gridtext@391684c)
 gtable         0.3.6   2024-10-25 [1] CRAN (R 4.5.0)
 here           1.0.1   2020-12-13 [1] CRAN (R 4.5.0)
 hms            1.1.3   2023-03-21 [1] CRAN (R 4.5.0)
 htmltools      0.5.8.1 2024-04-04 [1] CRAN (R 4.5.0)
 jsonlite       2.0.0   2025-03-27 [1] CRAN (R 4.5.0)
 kableExtra   * 1.4.0   2024-01-24 [1] CRAN (R 4.5.0)
 knitr          1.50    2025-03-16 [1] CRAN (R 4.5.0)
 labeling       0.4.3   2023-08-29 [1] CRAN (R 4.5.0)
 lifecycle      1.0.4   2023-11-07 [1] CRAN (R 4.5.0)
 litedown       0.7     2025-04-08 [1] CRAN (R 4.5.0)
 lubridate    * 1.9.4   2024-12-08 [1] CRAN (R 4.5.0)
 magrittr       2.0.3   2022-03-30 [1] CRAN (R 4.5.0)
```

```
markdown       2.0     2025-03-23 [1] CRAN (R 4.5.0)
pillar         1.11.0  2025-07-04 [1] CRAN (R 4.5.0)
pkgconfig      2.0.3   2019-09-22 [1] CRAN (R 4.5.0)
purrr        * 1.0.4   2025-02-05 [1] CRAN (R 4.5.0)
R6             2.6.1   2025-02-15 [1] CRAN (R 4.5.0)
RColorBrewer   1.1-3   2022-04-03 [1] CRAN (R 4.5.0)
Rcpp           1.1.0   2025-07-02 [1] CRAN (R 4.5.0)
readr        * 2.1.5   2024-01-10 [1] CRAN (R 4.5.0)
rlang          1.1.6   2025-04-11 [1] CRAN (R 4.5.0)
rmarkdown      2.29    2024-11-04 [1] CRAN (R 4.5.0)
rprojroot      2.0.4   2023-11-05 [1] CRAN (R 4.5.0)
rstudioapi     0.17.1  2024-10-22 [1] CRAN (R 4.5.0)
scales         1.4.0   2025-04-24 [1] CRAN (R 4.5.0)
sessioninfo    1.2.3   2025-02-05 [1] CRAN (R 4.5.0)
stringi        1.8.7   2025-03-27 [1] CRAN (R 4.5.0)
stringr      * 1.5.1   2023-11-14 [1] CRAN (R 4.5.0)
svglite        2.2.1   2025-05-12 [1] CRAN (R 4.5.0)
systemfonts    1.2.3   2025-04-30 [1] CRAN (R 4.5.0)
textshaping    1.0.1   2025-05-01 [1] CRAN (R 4.5.0)
tibble       * 3.3.0   2025-06-08 [1] CRAN (R 4.5.0)
tidyr        * 1.3.1   2024-01-24 [1] CRAN (R 4.5.0)
tidyselect     1.2.1   2024-03-11 [1] CRAN (R 4.5.0)
tidyverse    * 2.0.0   2023-02-22 [1] CRAN (R 4.5.0)
timechange     0.3.0   2024-01-18 [1] CRAN (R 4.5.0)
tinytex        0.57    2025-04-15 [1] CRAN (R 4.5.0)
tzdb           0.5.0   2025-03-15 [1] CRAN (R 4.5.0)
vctrs          0.6.5   2023-12-01 [1] CRAN (R 4.5.0)
viridisLite    0.4.2   2023-05-02 [1] CRAN (R 4.5.0)
withr          3.0.2   2024-10-28 [1] CRAN (R 4.5.0)
xfun           0.52    2025-04-02 [1] CRAN (R 4.5.0)
xml2           1.3.8   2025-03-14 [1] CRAN (R 4.5.0)
yaml           2.3.10  2024-07-26 [1] CRAN (R 4.5.0)

[1] /Library/Frameworks/R.framework/Versions/4.5-arm64/Resources/library
* -- Packages attached to the search path.

------------------------------------------------------------------------
```

References

R Core Team. 2025. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wasserstein, Ronald L., and Nicole A. Lazar. 2016. "The ASA's Statement on p-Values: Context, Process, and Purpose." *The American Statistician* 70 (2): 129–33. https://doi.org/10.

1080/00031305.2016.1154108.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.