# Analysis of Bird Species Detection and Accuracy

James Simmill

2024-07-18

Agriculture Data Science Research Centre

## Introduction:

This report presents findings from running various transects to detect bird species using audio recordings. The analysis covers the accuracy of these detections, the reliability of rare bird detections, and a summary of biodiversity in terms of species richness. The goal is to provide insights into the accuracy of the detection method and make recommendations to reduce false positives.

## Methods:

The data consisted of audio recordings processed to identify bird species, with each detection assigned a confidence score. Transects 5, 6, 7, and 8 were analysed, focusing on the top 3 and bottom three species in terms of average detection counts across the four transects. The following analyses were performed: Analysis of Variance (ANOVA) to identify if there are significant differences in confidence scores among species; Pearson Correlation to explore the relationship between detection counts and confidence scores; Histogram to visualise the distribution of confidence scores; and Biodiversity Summary to calculate the average species richness per day.

## Results:

The ANOVA results indicated significant differences in confidence scores among the top and bottom three species. The F-statistic was a high value of 285.153, indicating substantial differences between group means. On the other hand, the p-value was very small, approximately 2.625e-298, indicating that the observed differences in confidence scores among species are statistically significant and not due to random chance. The Pearson correlation between detection count and average confidence was 0.972, indicating a strong positive correlation. This suggests that species with higher detection counts tend to have higher confidence scores, indicating that the more frequently a species is detected, the more accurate the detection tends to be.

The average species richness was calculated to be approximately 72 unique species per day. It is important to note that this average could be incorrect depending on the number of false positives detected per transect. One way to combat this would be to consistently measure the same areas over time, as this would provide a more accurate picture of biodiversity and help identify trends in species richness. Implementing consistent measurement strategies can also help keep track of biodiversity net gain by ensuring that data collection is robust and comparable across different time periods.

## Conclusion:

The ANOVA test confirms significant differences in confidence scores among the top and bottom three detected species groups analysed. This implies that species detected more frequently have higher confidence scores compared to those detected less frequently. The high positive Pearson correlation further supports this, indicating that more frequent detections correlate with higher confidence scores and greater detection accuracy.

The histogram shows a bimodal distribution of confidence scores, suggesting that the detection model tends to be either very confident or not confident about its predictions. Lower confidence in detections of rare birds suggests that these are likely false positives, while higher count species tend to have more accurate detections. Implementing an accuracy cut-off of 0.5-0.6 to reduce false positives could help ensure that only reliable detections are considered, especially for rare species.

In summary, significant differences in confidence scores among species were identified, with a strong positive correlation between detection counts and confidence scores. The analysis indicated that rare bird detections are more likely to be false positives, while species with higher detection counts were found to be more accurately detected. The species richness averaged 72 unique species per day, indicating good biodiversity. Continuing data collection could provide a more comprehensive understanding of species presence and assess if conservation strategies are allowing for biodiversity net gain.