# Reliability of Citizen Science for Avian Species Identifications

#### Tyler Christian

### Abstract

The iNaturalist platform is a widespread species identification tool and acts as a long-term citizen science biodiversity database. Despite this, few studies have quantified the platform’s reliability and tested whether the data can be accurately used to calculate ecological trends. Here, all UK avian observations over 20 years were modelled to test for reliability and used to approximate species richness. Observations were most accurate if they contained high-resolution images of the birds. The data successfully approximated species richness, implying its utility in further ecological analysis. Suggestions were made on how further to improve the platform’s reliability and observation quality.

### Introduction

Citizen science is an increasingly popular data collection method in ecology, allowing large quantities of data to be collected without travelling to multiple field sites. However, this can result in spatial bias (Johnston et al., 2020) and inconsistent data quality, particularly when the study is poorly designed with basic data collection methods (Brown & Williams, 2019). One strategy to overcome this is to combine the data with pre-existing datasets (e.g. museum specimens) to track ecological trends (Pizarro, DeRaad & McCormack, 2023).

One of the most popular citizen science projects is the iNaturalist platform, on which users can upload pictures, audio recordings, and written descriptions of species that they can identify manually or by using a built-in algorithm. Observations are posted on the platform for others to discuss the identifications and reach a census. Once it reaches a set criterion, the observation is classified as ‘research grade’ and deemed fit for use in ecological analysis. Despite the scale of this database, few studies analyse it as a whole; instead, they focus on specific species and trends (Mesaglio & Callaghan, 2021). From these smaller studies, it is evident that the less well-known taxa are commonly misidentified (McMullin & Allen, 2022). Few studies have assessed the observation reliability of more charismatic and easier-to-identify taxa, such as birds. In this analysis, avian observations from iNaturalist were censused to compare their utility in measuring biodiversity to official databases and tested to determine the primary factors influencing their identification reliability.

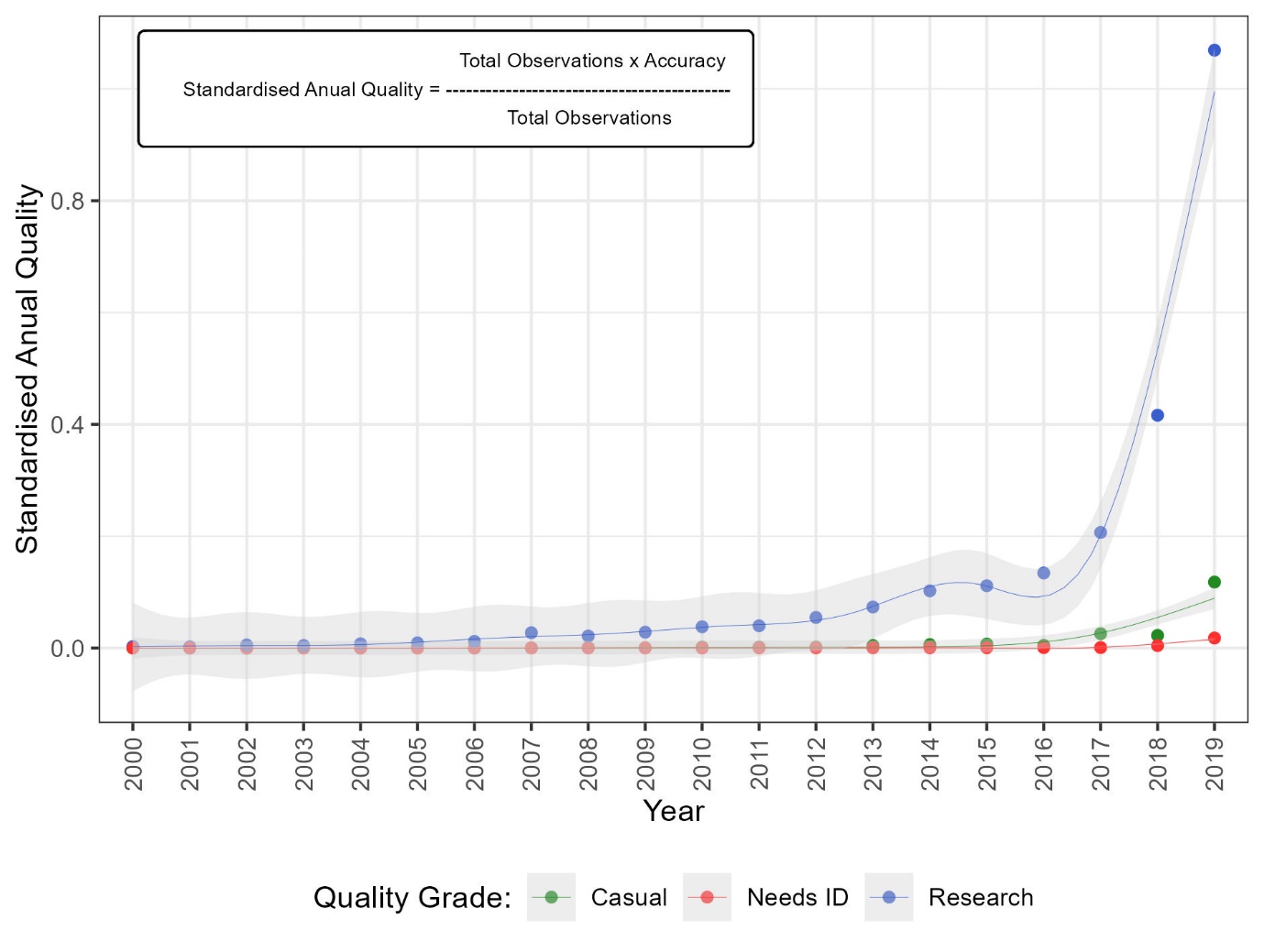
### Methods

To test for their reliability, all iNaturalist avian observations between 01/01/2000 and 31/12/2019 (iNaturalist, 2024) were censused, with forty random observations from the three quality grades (Needs ID, Casual, and Research) had their identifications verified using an ID guide (Collins, 2009) to calculate the percentage success of each grade. This metric was used as the response variable in a Linear Mixed Model using identification date and location as crossed random factors. The model tested which additive factors influenced the identification success: the ratio between identification agreements and disagreements (z-standardised), total user identifications (z-standardised), and correct initial species guess. The presence of an image, audio recording, and description were also added as interactive explanatory variables to represent how they collectively support species identification. To visualise the shift in observation quality over time, the average percentage accuracy for each quality ranking per year was plotted and overlayed with a Generalised Additive Model (GAM).

To test the data’s biological accuracy, it was compared to a UK avian species richness dataset (Dyer and Oliver, 2016). Since the species richness dataset spanned between 2000 and 2013, the iNaturalist data was filtered to match this range. The official dataset calculated species richness using the Frescalo method. This could not be applied to iNaturalist due to the lack of neighbourhood statistics, and richness was instead calculated by dividing the total species by the area of the region (ONS, 2019) it was located within. The two species richness metrics were z-standardised and compared using a Chi-Squared test to determine their statistical similarity. Data and R scripts are available on GitHub: <https://github.com/123-Tyler/miniproject_tyler_christian>.

### Results

The iNaturalist data contained 123,009 observations from 9,691 users across 190 locations in the UK. After manual identification, it was estimated that the “casual” quality grade was 10% accurate, “needs ID” was 22.5% and “research” quality was 97.5%. These approximations were used as the response variable within the linear mixed model, finding the agreement rate (β = 0.34, SE < 0.05), species guess (β = 0.02, SE < 0.05), total user avian IDs (β = 0.02, SE < 0.05), image (β = 1.09, SE < 0.05), and audio recording (β = 1.09, SE < 0.05) to all have significantly positive impacts on the observation reliability. The interaction effect between the presence of audio recordings and image (β = -1.08, SE < 0.05) was negative, suggesting the presence of both reduced observation reliability. The presence of a description had no impact on the reliability (β = -0.003, SE < 0.05), nor did the three-way interactive effect (β = 0.006, SE = 0.09). The random effects of location had minimal impact on the variance (0.004 and 0.0003, accordingly) but were retained in the model to account for potential temporal and spatial skews. All three observation quality rankings have experienced an exponential increase in reliability with time, with a massive increase after 2016 (Fig.1). When compared to an official dataset, there was no significant difference in the species richness of iNaturalist (chi-squared = 2926, d.f. = 2916, p = 0.44) but iNaturalist had a lower interquartile range (0.03) compared to the official data (1.09), suggesting there is some variance in the values.



**Fig.1 –** The quality of iNaturalist observations has exponentially increased, with research-quality observations being the most reliable, particularly since 2016. Regression lines were plotted using GAM smoothing, and standardisation of annual quality accounted for differing sample sizes.

### Discussion

Citizen science platforms such as iNaturalist play a major role in ecological data collection, and assaying their reliability is thus vital to ensure their ongoing usability. This study first quantified the estimated accuracy of the platform’s “quality grades”, finding their “research” data to be accurate in 97.5% of cases, supporting its ongoing use in scientific publications. Furthermore, when the database was used to calculate avian species richness in the UK, the results were somewhat similar to an official database, suggesting that iNaturalist data could be used for approximating ecological trends but may lack the precision of more rigorous studies.

The individual observations on iNaturalist varied in quality; when inspecting random samples, it was evident that they were easier to verify with multiple high-resolution images of the bird from various angles; the statistical model supported this. The model implied that having access to an image and recording of the birdcall decreased an observation’s reliability. Whilst this may appear counterintuitive, if not done correctly (i.e. quiet or muffled recordings), the presence of both media could confuse the identifier, skewing the results. Likewise, user descriptions tend to be vague and offer little-to-no identification support, hence their lack of significance in the model. Therefore, I propose that the implementation of stricter user guidelines on the platform and the introduction of a brief training/how-to guide could improve the overall reliability of observations. Whilst this study was limited to UK avian data and a complete taxonomic census is required to validate these trends, it can be assumed that similar faults exist across the platform, particularly in the most common taxa such as plants, insects, mammals and fungi.

### References

Brown, E.D. & Williams, B.K. (2019) The potential for citizen science to produce reliable and useful information in ecology. *Conservation Biology*. 33 (3), 561–569. doi:10.1111/cobi.13223.

Collins (2009) *Collins Bird Guide*. 2nd edition. London, Harper Collins Publishers Ltd.

Dyer, R. and Oliver, T. (2016). Estimated species richness data used in study of UK Ecological status. NERC Environmental Information Data Centre. <https://doi.org/10.5285/6c535793-034d-4c4f-8a00-497315e7d689>

iNaturalist community (2024). Observations of birds from UK observed between 01/01/2000 - 01/01/2021. iNaturalist. <https://www.inaturalist.org>

Johnston, A., Moran, N., Musgrove, A., Fink, D. & Baillie, S.R. (2020) Estimating species distributions from spatially biased citizen science data. *Ecological Modelling*. 422, 108927. doi:10.1016/j.ecolmodel.2019.108927.

McMullin, R.T. & Allen, J.L. (2022) An assessment of data accuracy and best practice recommendations for observations of lichens and other taxonomically difficult taxa on iNaturalist. *Botany*. 100 (6), 491–497. doi:10.1139/cjb-2021-0160.

Mesaglio, T. & Callaghan, C.T. (2021) An overview of the history, current contributions and future outlook of iNaturalist in Australia. *Wildlife Research*. 48 (4), 289–303. doi:10.1071/WR20154.

ONS Geography (2019). Counties and Unitary Authorities (December 2019) Boundaries UK BUC. Office For National Statistics. <https://geoportal.statistics.gov.uk/datasets/bbf13c40ca8f4931afc0a46d079af58c_0/about>

Pizarro, A.K., DeRaad, D.A. & McCormack, J.E. (2023) Temporal stability of the hybrid zone between Calocitta magpie-jays revealed through comparison of museum specimens and iNaturalist photos. *Ecology and Evolution*. 13 (3), e9863. doi:10.1002/ece3.9863.

### Reflection