Reliability of Citizen Science Data for Avian Species Identification

Tyler Christian

Abstract

The iNaturalist platform is a popular species identification tool functioning as a long-term citizen science biodiversity database. Despite its widespread use, few studies have quantified the platform's reliability and tested whether the data can accurately measure ecological trends. All UK avian observations over 20 years were modelled to test for reliability and 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 ongoing 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 to be collected without travelling to multiple field sites. However, this can result in a trade-off causing spatial bias (Johnston *et al*., 2020) and inconsistent data quality, particularly when the study is poorly designed with simplistic 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 such as temporal shifts (Pizarro, DeRaad & McCormack, 2023).

iNaturalist is a thriving ecological citizen science 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 computer vision 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. iNaturalist was released in 2008, but users can upload past sightings dating back to the 1980s. Despite the database's scale, few studies have analysed it as a whole; instead, focusing on a specific trend or species (Mesaglio & Callaghan, 2021). From these smaller studies, it is evident that the less well-known taxa are commonly misidentified (McMullin & Allen, 2022), allowing incorrect observations to be added to research databases, including GBIF. Furthermore, few studies have assessed the observation reliability of more charismatic and easier-to-identify taxa, such as birds, questioning the overall reliability of the platform. In this analysis, UK 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. 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 category. This metric was used as the response variable in a Linear Mixed Model using identification date and location as 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 plot the shift in observation quality over time, the annual quality was approximated by multiplying the annual observations with the proportion of platform identification agreements and quality grade's accuracy and dividing by the annual observations. This was plotted using 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. All data and R scripts are accessible on GitHub: https://github.com/123-Tyler/miniproject_tyler_christian.

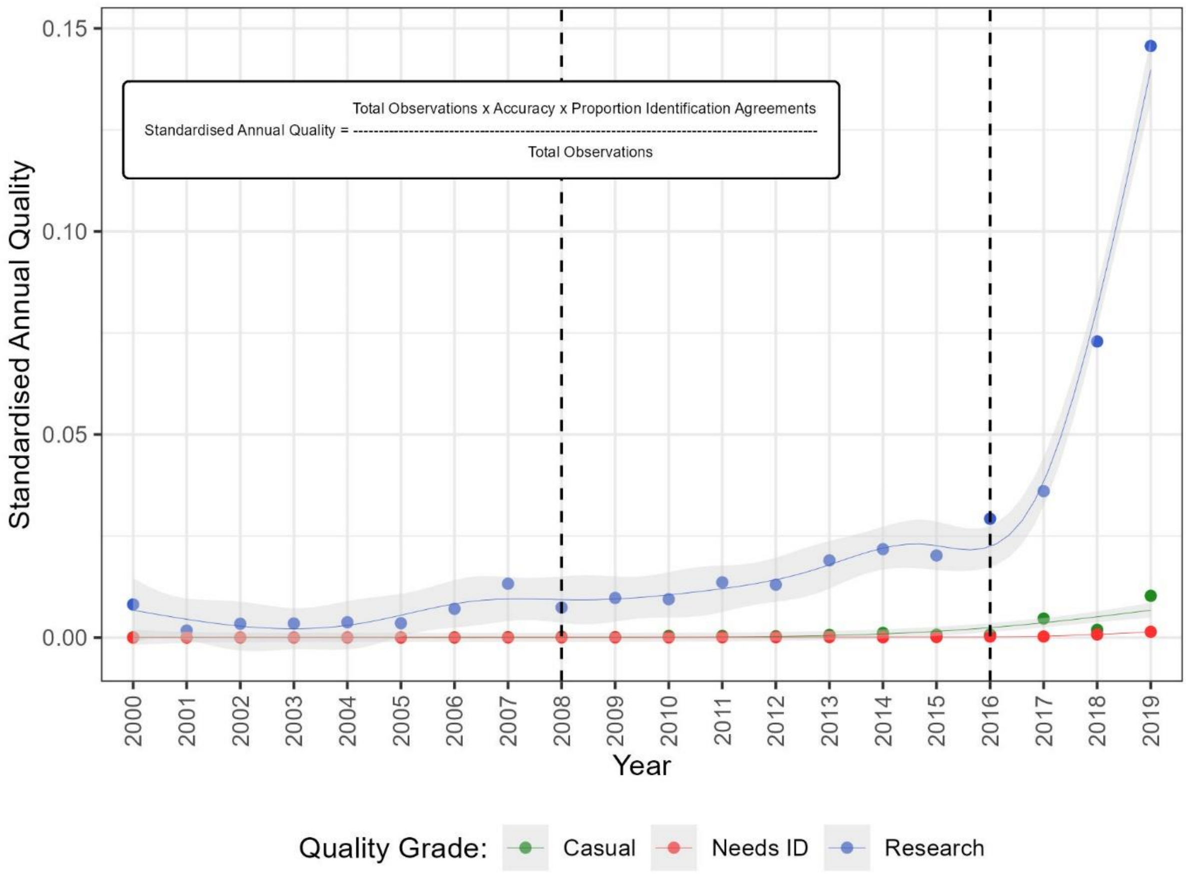

Fig.1 – The quality and accuracy of iNaturalist observations have exponentially increased, with research-quality observations being the most reliable. The increase in "casual" observation quality is debatable. Regression lines were plotted using GAM smoothing, and standardisation of annual quality accounted for differing sample sizes. The two dashed lines represent the platform's release in 2008 and the influx of media attention in 2016 which plausibly caused the rapid incline in observation quality (Iwane, 2016).

Results

The 20-year 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 ($\beta = 0.34$, SE < 0.05), species guess ($\beta = 0.02$, SE < 0.05), total user avian IDs ($\beta = 0.02$, SE < 0.05), image ($\beta = 1.09$, SE < 0.05), and audio recording ($\beta = 1.09$, SE < 0.05) to all have significantly positive impacts on the observation reliability. The interaction effect between the presence of audio recordings and image ($\beta = -1.08$, SE < 0.05) was negative, suggesting the presence of both reduced observation reliability. The presence of a description had no impact on the reliability ($\beta = -0.003$, SE < 0.05), nor did the three-way interactive effect ($\beta = 0.006$, 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.

Discussion

Citizen science platforms such as iNaturalist play a major role in ecological data collection, and assaying their reliability is 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 identifying 97.5% of bird observations, 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 resolution 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. Observations containing poor-quality images and audio recordings were the most challenging to verify. The statistical model supported this, further suggesting that having access to both an image and a 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, leading to mistakes. Likewise, user descriptions tended to be vague and offer little-to-no identification support, hence their lack of significance in the model. In 2016, iNaturalist received an influx of media attention, resulting in a rapid surge of users, explaining the incline of observations. This peak also coincided with the release of popular outdoor exploration apps. This included Pokémon Go, where users 'identified' fantasy animals in an interactive virtual environment similar to how species are identified on iNaturalist (Iwane, 2016). This plausibly inspired people to spend more time in nature and download iNaturalist as a realistic alternative to the fantasy games, supporting the rise in users and thus observation reliability.

In summary, iNaturalist and similar citizen science identification platforms (such as eBird and PlantNet) and can provide useful and somewhat reliable ecological data but rely heavily upon user input and efficient marketing to improve engagement. To support this, I propose implementing clearer user guidelines and introducing a compulsory user training page or video upon downloading the app to improve consistency and reliability between observers. The platform would also benefit from a modernised layout and the addition of a reminder before submitting observations, encouraging users to submit more reliable data. Note that this study was limited to UK avian data, and a complete taxonomic census is required to validate these trends and support potential platform changes.

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.

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 – 31/12/2019. *iNaturalist*. https://www.inaturalist.org

Iwane, T. (2016). iNat and Media Attention in 2016. *iNaturalist*. https://www.inaturalist.org/blog/7101-inat-and-media-attention-in-2016. Assessed: 20/02/2024.

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.

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.

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.

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.

Reflection

Through this mini-project, I have experienced independent research and project management skills, expanded on the working strategies developed during my undergraduate dissertation, and learnt new techniques. I have thoroughly enjoyed this process and wish I had more time to develop my ideas further, assaying the trends across multiple taxonomic groups, not just UK birds, to gain a more thorough understanding of the platform's reliability.

From this project, I have found one of my greatest strengths to be time management. When studying, I follow a strict 9-5 working day, taking only short breaks for refreshments and an extended lunch break. Whilst this strategy may appear trivial, I find it essential to maintaining a healthy work-life balance. The strict working hours also ensured I maintained focus and avoided distractions, I rely on this method not only for the mini-project but also for my general studies. I will continue to utilise it during my thesis, making scheduling adjustments for fieldwork where appropriate.

At the beginning of the fortnight, I mainly worked in my room but soon found this unideal, so switched to working in the library instead. I regularly became distracted when in my room, breaking up my workflow with mundane tasks such as tidying and laundry. I also started to frequently check my phone and take regular snack breaks (the majority of which were unhealthy). To resolve this, on the first Tuesday afternoon, I decided to move over to the library. This distraction (and snack) free environment drastically improved my concentration and overall quality of work. I maintained this habit for the remainder of the fortnight, entering the library at 9 am, taking a lunch break at 12, and finishing working at 5, after which I had time to exercise, relax and complete non-academic day-to-day jobs. Upon reflection, the reason for this was likely psychological: I associate my room with relaxing and thus struggle to concentrate in it, constantly being tempted to procrastinate.

At the end of each working day, I found writing a checklist of what I have achieved useful. By keeping a record of these checklists, I could regularly check my progress to ensure I am not falling behind. During my mini-project, I have started to add my tasks for the following day onto these checklists, planning whether they are morning or afternoon jobs. This means that I waste no time thinking of what I need to do and am able to make an immediate start the following morning. Occasionally, my time plans were not entirely accurate, but they gave me a rough guide on what needed doing, thus supporting my workflow. This has been highly successful and resulted in me saving time, finishing the mini-project a couple of days early, and allowing myself to allocate time towards revision and finalising the seminar diaries. To further improve this technique during my thesis, I will also attempt to plan a weekly working schedule on Friday evenings, using the daily checklists to ensure I am sticking to the plan and adjusting where required.

When working on the project, it didn't take long to decide upon a research question and formulate a hypothesis: I initially wanted to study bird diversity and decided to use the iNaturalist database, leading to my idea of testing for its reliability. Whilst the main reliability aspect of the project was relatively simplistic, I realised that the platform's quality rankings were not quantifiable and, thus, not an ideal response variable for the linear mixed model. From this, I decided to manually identify random observations to quantify the success of each quality ranking. While this was only a minor (and successful) detour, I also spent time developing larger ideas, which I later scrapped. The most prominent example of this was when I was struggling to visualise the data and decided to plot the spatial trends in reliability by creating a GIS heatmap of the UK. I spent over a day unsuccessfully attempting to code this before deciding to switch to plotting temporal trends in Figure 1. In my thesis, I need to learn to recognise these detours before they happen by checking for simpler alternatives and ensuring their feasibility before attempting them. If I do decide to continue with the idea, I will set a realistic deadline, at which point I will assess my progress and determine whether the idea is worth continuing, thus preventing me from falling behind on more pressing tasks.

Overall, my working style was efficient throughout the mini-project, with my major flaw being distractions. These came in two forms: casual (i.e. mundane jobs and snacks) and academic (i.e. unfeasible ideas). To overcome this during the thesis, I will continue to manage my time efficiently and work in the library (or lab where appropriate). I will also take time to thoroughly think through ideas before implementing them, maintaining a system of progress checkpoints. Furthermore, I will continue to reflect upon my working style, making further adjustments as/when issues arise to ensure efficiency.