ASSESSING iNATURALIST AS A TOOL TO ESTIMATE BIODIVERSITY RICHNESS

Maxim Velli, Tarek Nouneh, Lola Kengen, Eva Koskova - Master AIRE, Université Paris Cité

Supervisors: Ignacio Atal, Rona Aviram

INTRO

The rise of citizen science over the past few decades presents a new opportunity to measure biodiversity independent of traditional methods like field surveys which can be costly and time consuming (Kühl et al., 2020). iNaturalist has emerged as one of the most successful and widely used platforms around the world in identifying species photographed by users (Campell 2023). iNaturalist uses a crowdsourced species identification system, which means other users of the platform have to come to an agreement in identifying the species in question (iNaturalist, 2013).

In this study, we investigate the potential for using citizen science data from iNaturalist as an alternative method for estimating species richness. New Zealand was chosen as a case study because of the comprehensive efforts made to map biodiversity and the availability of public access data. Here, we take the observation data provided by iNaturalist and use it to estimate species richness in the country. The result is compared with the official numbers obtained through traditional methods.

The latest species richness estimate for New Zealand, provided by distinguished researcher Dennis P. Gordon, suggests the existence of approximately 52,000 species (Gordon, 2013; Smith et al., 2023). This number is derived through years of systematic documentation of the flora and fauna by experts in the field through traditional methods. This latest estimate will serve as our point of comparison in the study.

Biodiversity: the measure of genetic, ecosystem, and species variability in a certain location (Yadav and Mishra, 2013).

Species richness: the number of different species present in an ecosystem (Kühl et al., 2020)

METHODO

The methodology of the research focused on evaluating iNaturalist as a tool for estimating biodiversity in New Zealand, contrasting it with data derived from traditional methodologies as represented by the New Zealand Threat Classification System (NZTCS). To achieve this, the iNaturalist February 2024 snapshot data was downloaded from the Global Biodiversity Information Facility (GBIF) and compared to official New Zealand data (NZTCS, 2023). The New Zealand biodiversity data only listed observed species without individual observations, and although some 55,000 species are estimated to exist in New Zealand, the dataset only contains 15,000.

The discrepancy in data necessitated an alternative comparison approach. In this approach, we compared saturation curves derived from iNaturalist data against NZTCS species number. The ACE and Chao1 estimators were employed to extrapolate the iNaturalist saturation curve, facilitating a comparison between the datasets (Gotelli & Colewell, 2010).

Additionally, the study analyzed iNaturalist's performance across biological kingdoms and classes by comparing species counts to NZTCS, employing a binomial distribution model to evaluate the significance of differences in species distribution. This approach allowed for an in-depth comparison of citizen science versus traditional biodiversity assessment methods.

RESULTS

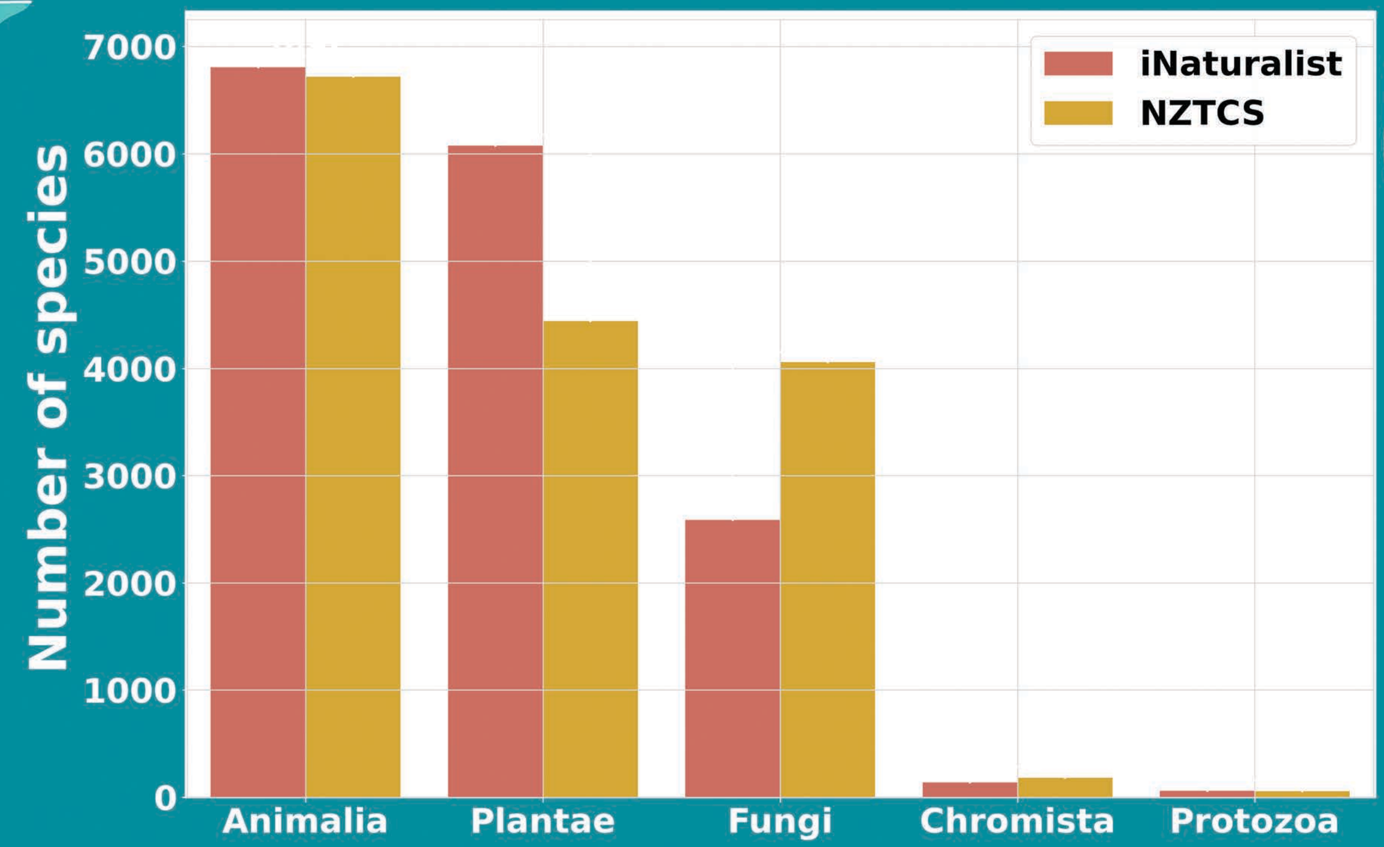


Fig. 1 | Comparison of species richness between different kingdoms for the iNaturalist and NZTCS datasets. Bacteria and Archaea not shown due to their absence in iNaturalist.

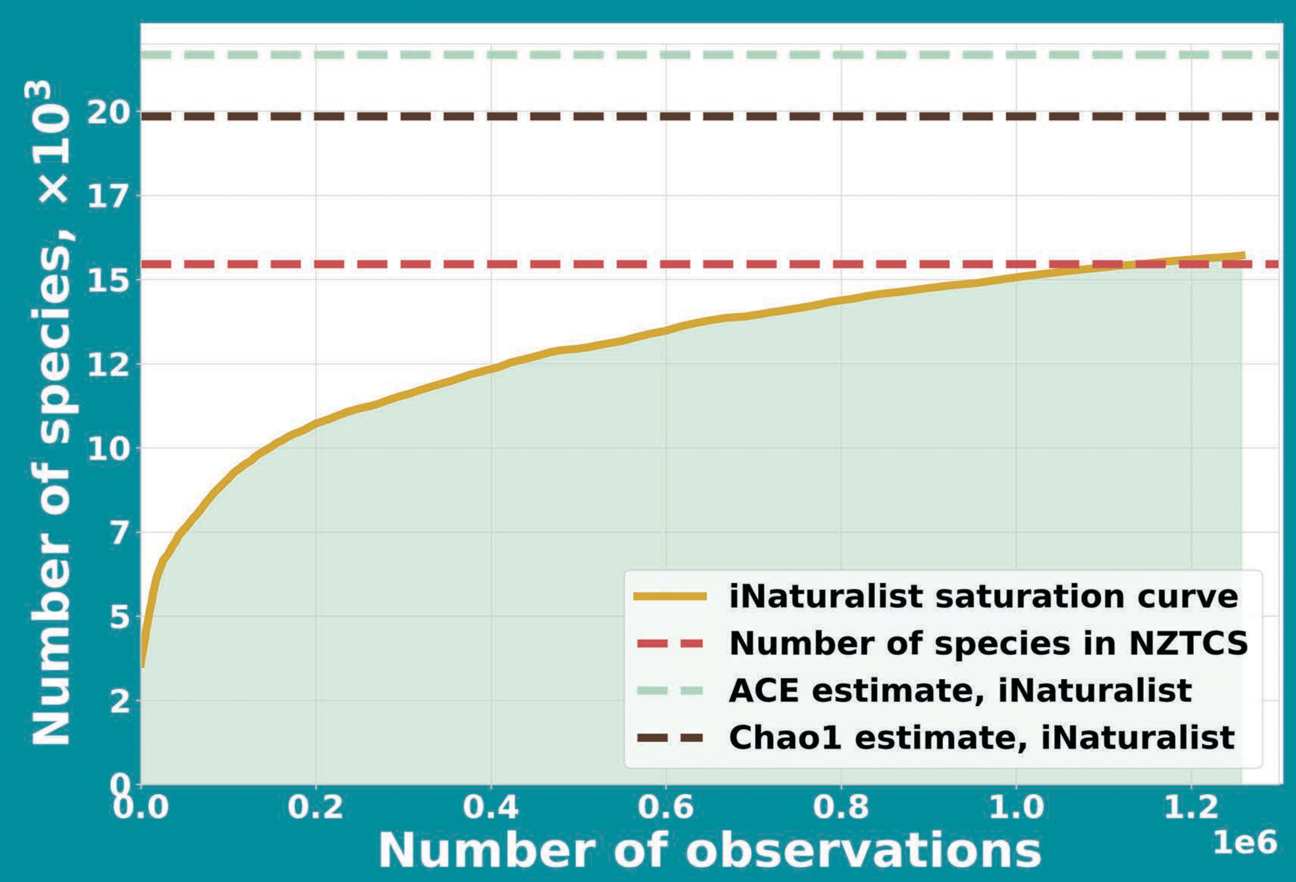


Fig. 2 | Saturation curve for iNaturalist dataset, compared with the total number of species in the NZTCS dataset. Species richness estimators Chao1 and ACE are also provided. The official NZTCS estimate is 50,000.

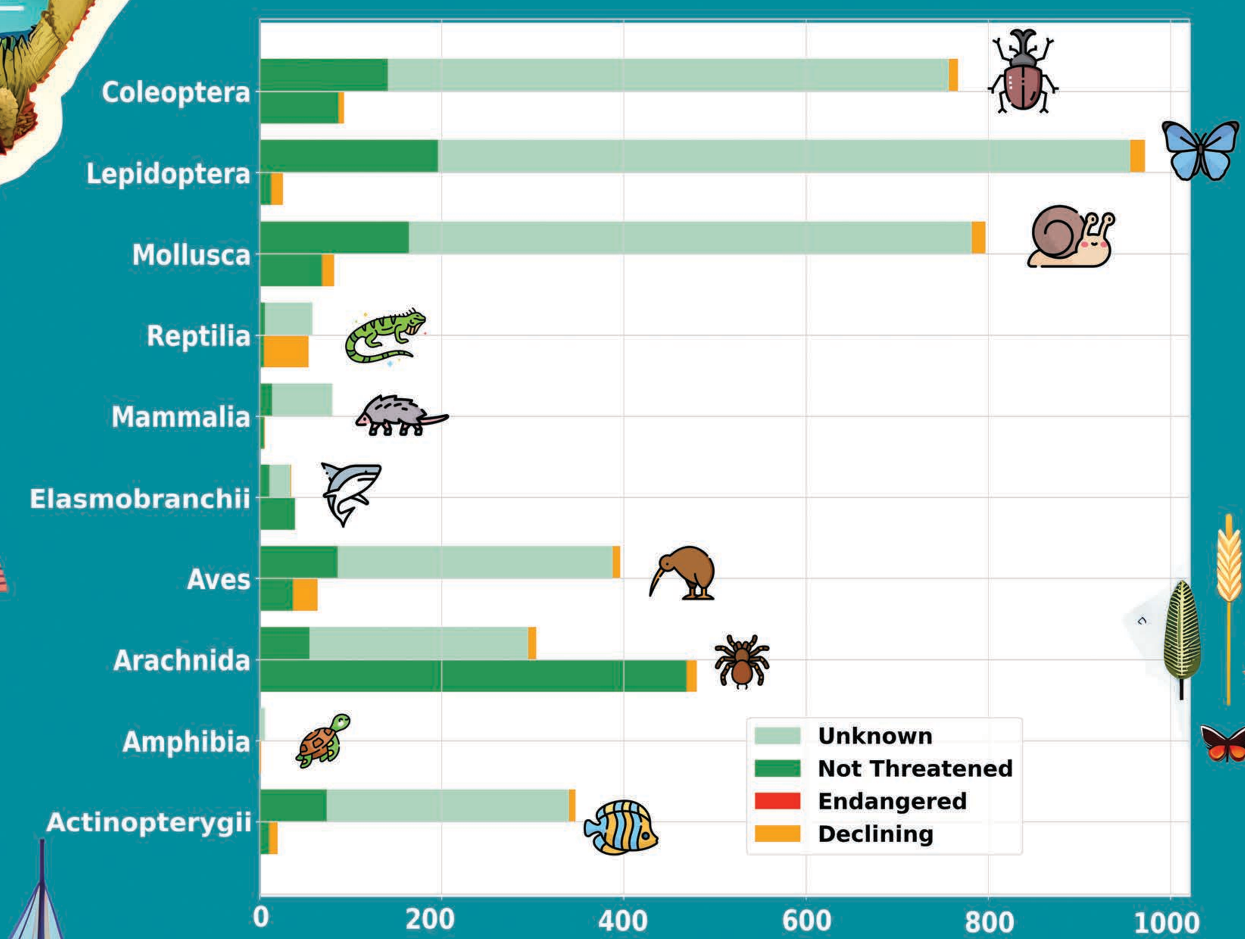
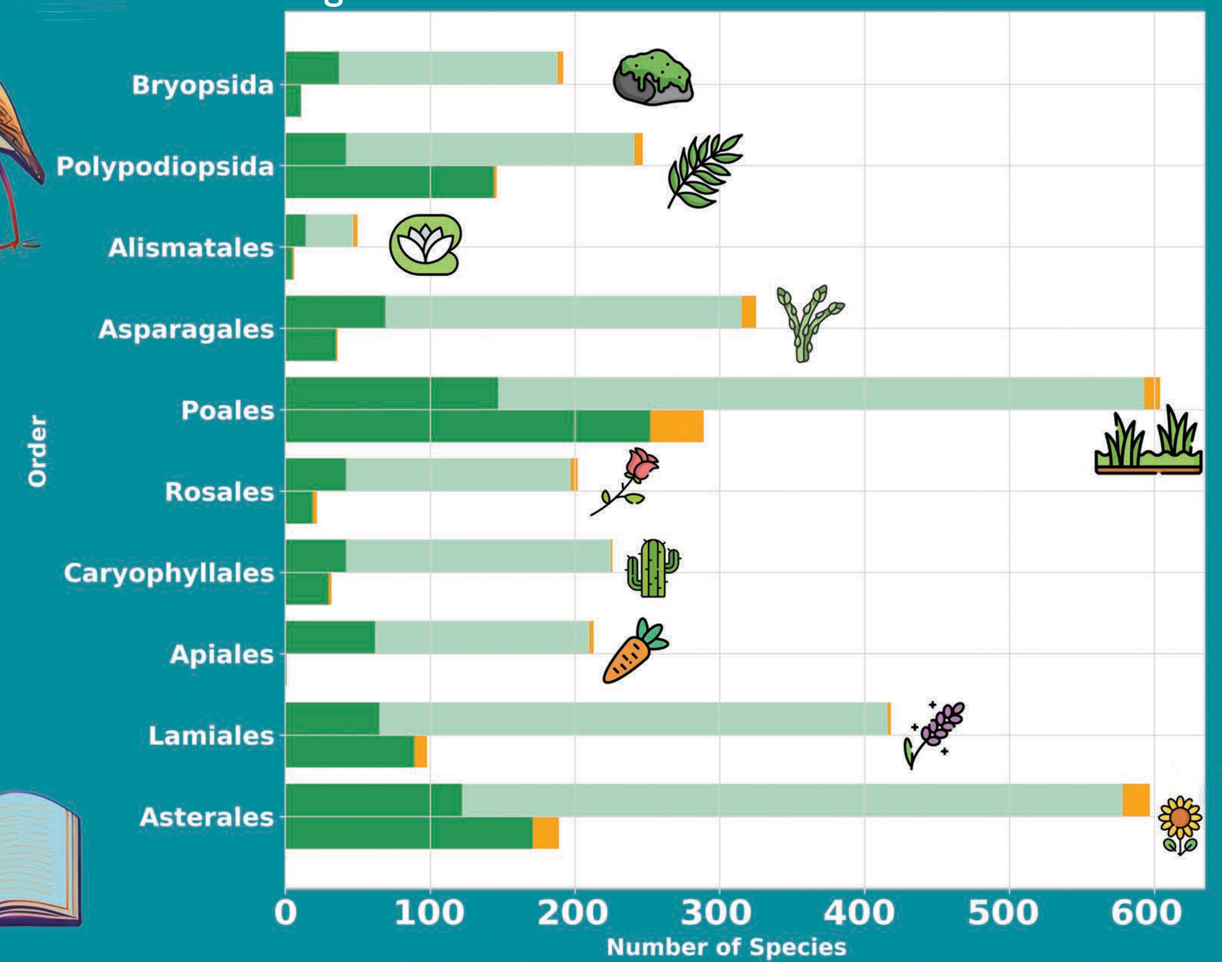


Fig. 3 & 4 | Comparison of species richness for 10 most common classifications in the Animal and Plant kingdoms between the iNaturalist and NZTCS datasets.



DISCUSSION

Main Conclusions

iNaturalist shows promise in recording species richness more effectively than traditional methods for certain datasets. Although it's difficult to specify which factors influence iNaturalist's performance, its potential as a supplementary tool to traditional methods is evident.

Biodiversity Estimation Discrepancies

A significant finding from our study is the disparity between biodiversity estimates derived from iNaturalist data, predicting approximately 20,000 species, and the official governmental estimate of 50,000 species. We believe the biggest cause of the disparity, which is also one of the major shortcomings of citizen science data, is the uneven sampling effort from different regions of the country (Reddy & Davalos, 2003; Hugo & Altwegg, 2017). Particularly, the sampling bias towards highly populated areas and trails (Hugo & Altwegg 2017; Fitzpatrick et al. 2009). The lack of normalized data from all regions can negatively impact the accuracy of estimation models (Hughes et al., 2001). It is also important to note that the Chao1 model is a minimum asymptotic estimator, which means the 20,000 projection is a lower end estimate (Chao, 1984).

Moreover, we question the accuracy of the accepted estimate of 55,000 species in New Zealand. There is no clear justification to the figure, and it has proven difficult to track the source of the data. Furthermore, the estimation models employed wevre not designed for large datasets generated from citizen science which might be bias towards certain geographic locations (source). There is an urgent need for developing new abundance estimators tailored to the expansive datasets typical of citizen science efforts.

Limitations

The study was limited by the use of the NZTCS dataset, which lacked geolocation data, and the partial incompatibility between iNaturalist and NZTCS data, particularly in standardized species naming and extracting extinction statuses. These issues underscore the importance of standardized data formats and the need for more detailed spatial information (Feldman 2021; Bullion et Bahlai 2011).

A natural limitation of iNaturalist data is the uncertainty associated with the identification process of species. For an observation to reach research grade, two users must agree on identifying the sample (Harper, 2024). Research grade label is roughly 85% accurate (Kueda, 2019). Although iNaturalist data is already being used for scientific publications, it is important to highlight this limitation and note that it might play a bigger role in New Zealand where the number of distinct species is large (Moles & Xirocostas 2022; Forty et al. 2022).

Future Perspectives

Future research should explore biases in iNaturalist's species recording, like 'photogenicity' and habitat accessibility, and investigate temporal trends in species observations to understand migration patterns, population changes, and the effects of environmental changes. This could offer valuable insights for biodiversity conservation, pest control, and disease vector management.

Overall, while iNaturalist offers significant advantages in biodiversity documentation, addressing its limitations and further exploring its capabilities will enhance our understanding and conservation of biodiversity.

OPEN SCIENCE

Our study evaluates iNaturalist's effectiveness in documenting species richness to enhance traditional biodiversity conservation strategies amidst the ongoing 6th mass species extinction (iPBES 2019; Tittensor et al. 2014). Our code is openly accessible and interoperable in other geographical zones.

As we challenge traditional scientific practices, we emphasize integrating indigenous knowledge into biodiversity monitoring and conservation (Agrawal 1995; Leach and Fairhead 2002; Stevens et al. 2014). Exploring synergies between modern citizen science technologies and indigenous knowledge remains crucial, particularly in New Zealand (Tengo et al. 2021, Department of Conservation 2022)

Science operates within heteropatriarchal, hierarchical, and capitalistic power structures (Hubard 1983). Embracing open science requires contextualizing the history of scientific practice and acknowledging the social subjectivities of scientists (Haraway 2003). Our methodology and literature selection demonstrate diversity in author gender and geographical focus, reflecting a commitment to inclusive perspectives.

Attributions

Eva: Data mining, data analysis, data visualisation

Lola: Conceptual framing, inclusivity, adherence to open science principles

Tarek: Biodiversity assessment framing, data mining

Maxim: Data mining, data analysis, data visualisation, statistical analysis

Bibliography + Github