

**Escuela Agrícola Panamericana, Zamorano**  
**Department of Environment and Development**  
**B.S. in Environmental Science and Development**



Special Graduation Project  
**Evaluating the Influence of Sample Size and Taxonomic  
Group on Chao's Estimator of Species Richness Using iNaturalist  
Photographic Observations in Honduras**

Student

Fernanda Lucía Alvarado Vásquez

Advisors

Oliver Komar, Ph.D.

Eric van den Berghe, Ph.D.

Honduras, August 2024

**Authorities**

**SERGIO ANDRÉS RODRÍGUEZ ROYO**

President

**ANA M. MAIER ACOSTA**

Vice President and Academic Dean

**VICTORIA CORTÉS MATAMOROS**

Director of the Environment and Development Department

**HUGO ZAVALA MEMBREÑO**

Secretary General

**Table of Contents**

List of Figures .....4

Abstract .....6

Resumen .....7

Introduction .....8

Methods .....11

Results and Discussion .....14

Reptiles .....14

Birds .....19

Bats .....23

Amphibians .....28

Conclusions .....34

Recommendations .....36

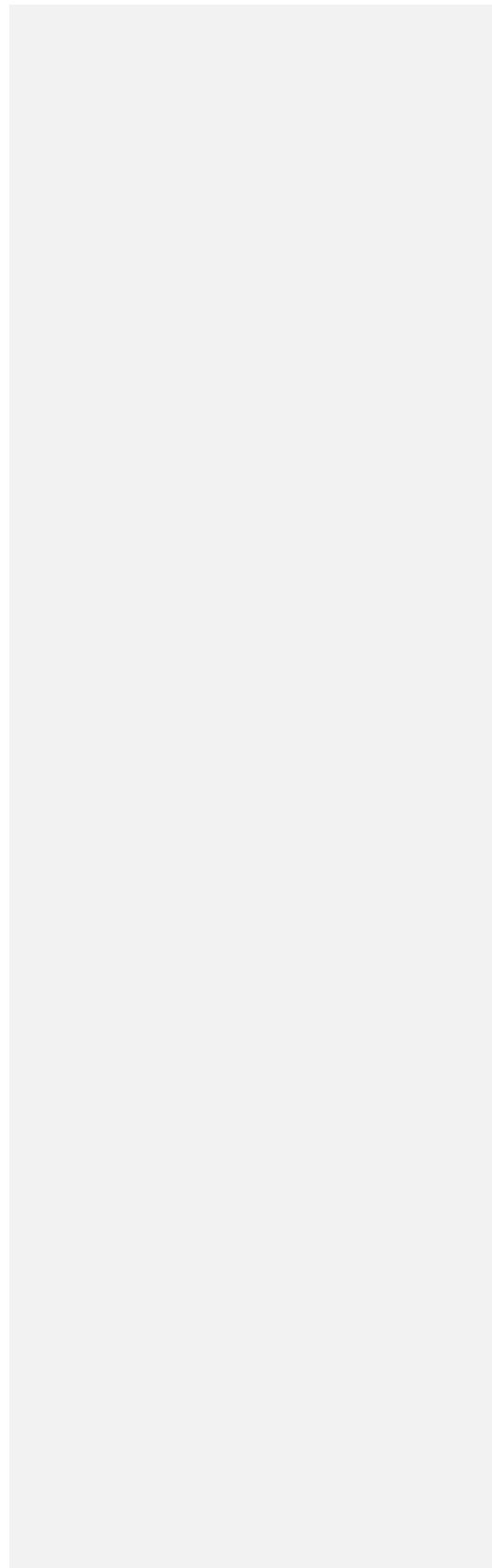
References .....37

Comentado [GU1]: Fernanda, hay errores en los números de página. Tienes que actualizarlas. Es automatico si sigues las opciones de menú. Esto es una de las cosas que te señalé hacer ayer!!!

### List of Figures

Figure 1 Species richness in Honduras for reptiles, all observations included .....	14
Figure 2 Species Richness in Honduras for reptiles, research grade observations only .....	15
Figure 3 Species richness in Francisco Morazán for reptiles, all observations included .....	16
Figure 4 Species richness in Francisco Morazán for reptiles, research grade observations only .....	17
Figure 5 Species richness in San Antonio de Oriente for reptiles, all observations included .....	18
Figure 6 Species richness in San Antonio de Oriente for reptiles, research grade observations only .....	19
Figure 7 Species richness in Honduras for birds, all observations included .....	20
Figure 8 Species richness in Honduras for birds, research grade observations only .....	20
Figure 9 Species richness in Francisco Morazán for birds, all observations included .....	21
Figure 10 Species richness in Francisco Morazán for birds, research grade observations only .....	21
Figure 11 Species richness in San Antonio de Oriente for birds, all observations included .....	22
Figure 12 Species richness in San Antonio de Oriente for birds, research grade observations only .....	23
Figure 13 Species richness in Honduras for bats, all observations included .....	24
Figure 14 Species richness in Honduras for bats, research grade observations only .....	24
Figure 15 Species richness in Francisco Morazán for bats, all observations included .....	25
Figure 16 Species richness in Francisco Morazán for bats, research grade observations only .....	26
Figure 17 Species richness in San Antonio de Oriente for bats, all observations included .....	27
Figure 18 Species richness in San Antonio de Oriente for bats, research grade observations only .....	27
Figure 19 Species richness in Honduras for amphibians, all observations included .....	28
Figure 20 Species richness in Honduras for amphibians, research grade observations only .....	29
Figure 21 Species richness in Francisco Morazán for amphibians, all observations included .....	30
Figure 22 Species richness in Francisco Morazán for amphibians, research grade observations only .....	30
Figure 23 Species richness in San Antonio de Oriente for amphibians, all observations included .....	31

Figure 24 Species richness in San Antonio de Oriente for amphibians, research grade observations  
only.....32



### **Abstract**

Species richness estimation is an important tool for biodiversity assessment and conservation. However, numerous factors can affect the accuracy of these estimates, such as sample size and the taxonomic group studied. This study evaluated the applicability of Chao's species richness estimator in different taxonomic contexts using photographic observations from the iNaturalist platform in Honduras. Counts of low abundance individuals (observed 1 or 2 times) were conducted, and data from reptiles, birds, bats, and amphibians were analyzed based on photographic observations from the iNaturalist citizen science platform. Species presence was monitored in different sampling areas, varying in size and sampling effort. Additionally, the effectiveness of Chao's estimator was compared in these areas to determine its reliability at different sample sizes. The results highlight the importance of adequate sampling effort to obtain accurate estimates, especially for underrepresented taxa. This study provides a baseline for future research to improve biodiversity estimation methods from incomplete inventories and their application in conservation strategies.

*Keywords:* Accumulation curve, citizen science, estimating biodiversity, sampling effort

### Resumen

La estimación de la riqueza de especies es una herramienta clave para la evaluación de la biodiversidad y la conservación. Sin embargo, la precisión de estas estimaciones puede verse afectada por diversos factores, como el tamaño de la muestra y el grupo taxonómico estudiado. En este estudio, se evaluó la aplicabilidad del estimador de riqueza de especies de Chao en diferentes contextos taxonómicos, utilizando observaciones fotográficas de la plataforma iNaturalist en Honduras. Se realizaron conteos de individuos con baja ocurrencia (observados 1 y 2 veces) y se analizaron datos de reptiles, aves, murciélagos y anfibios a partir de observaciones fotográficas de la plataforma de ciencia ciudadana de iNaturalist. Se monitoreó la presencia de especies en diferentes áreas de muestreo, variando en tamaño y esfuerzo de muestreo. Además, se comparó la efectividad del estimador de Chao en estas áreas para determinar su fiabilidad con distintos tamaños de muestra. Los resultados resaltan la importancia de un esfuerzo de muestreo adecuado para obtener estimaciones precisas, especialmente en taxones subrepresentados. Este estudio proporciona una línea base para futuras investigaciones que buscan mejorar las metodologías de estimación de la biodiversidad a partir de inventarios incompletos y su aplicación en estrategias de conservación.

*Palabras clave:* Ciencia ciudadana, curva de acumulación, esfuerzo de muestreo, estimación de la biodiversidad

### Introduction

Honduras, located in Central America, is renowned for its rich biodiversity, which includes 14,849 species reported in iNaturalist and many more not yet reported (*Observaciones*). However, this biodiversity is increasingly threatened by habitat destruction, climate change, and other anthropogenic pressures. The environment in Honduras is particularly characterized by agricultural practices that have led to significant environmental deterioration, especially in mountainous areas (Shmelev, 1999). Land use for agriculture, including practices such as burning fields to clear them, has contributed to widespread environmental degradation. Population growth and increasing land pressure have exacerbated this degradation. Understanding the biodiversity of this region provides valuable insights for both ecological research and practical conservation strategies, as we intend to mitigate the adverse effects of human activities on this biodiversity hotspot.

In this context, species richness, defined as the total number of unique species within a given biological community (Rafferty, 2024), becomes a fundamental measure of biodiversity that provides valuable insights into the complexity and health of ecosystems. Its simplicity and intuitive nature make it an accessible and widely used metric for both scientific research and conservation efforts. Firstly, it allows researchers to assess the biodiversity of an area and identify potential threats to specific species, thus informing conservation efforts. Additionally, accurate estimations of species richness provide a baseline for monitoring changes in biodiversity over time, aiding in ecosystem management and restoration. Understanding the number of species in a population is also essential for ecological research, as it offers insights into the functioning of ecosystems and the relationships between species. Overall, accurate estimates of species richness are vital for making informed decisions regarding the conservation and management of natural resources (Chiu et al., 2014).

Citizen science platforms like iNaturalist have revolutionized biodiversity monitoring by enabling the collection of extensive observational data from a wide range of contributors. The photographic observations collected through iNaturalist provide a valuable, although often

incomplete, inventory of species. This highlights the need for precise estimation methods to derive meaningful insights from such data. Researchers have begun integrating iNaturalist data into their studies, particularly concerning the presence and distribution of species, illustrating the platform's growing significance in biodiversity research (Rode & Torkar, 2023). iNaturalist can significantly contribute to assessing species richness in biodiversity hotspots by providing a platform for students and other participants to observe and identify a wide range of organisms, much like how it was utilized in research assessing biodiversity responses and recovery to the Australian megafires (Gorta et al., 2023). Additionally, the platform offers opportunities to enhance data literacy among its users, including geography and taxonomic classification, graphical interpretation, and the open discussion of species identification.

Complete species inventories are nearly impossible to achieve in practical applications due to the presence of rare, reclusive, or undescribed taxa (Chao & Colwell, 2017). This challenge underscores the importance of estimation methods in biodiversity studies. Chao's estimator is a widely used statistical tool designed to estimate species richness, particularly effective in handling incomplete datasets by accounting for the frequency of rare species. Despite its utility, the accuracy and reliability of Chao's estimator can be influenced by factors such as sample size, taxonomic differences, and the completeness of the inventory. Understanding these influences is crucial for improving species richness estimates and making informed conservation decisions.

Studies have shown that Chao's estimator performs optimally at intermediate sampling efforts (Iknayan et al., 2014). This finding supports the use of incomplete inventories. In the present study, we present iNaturalist as a potential source of data for species richness estimation.

The primary objective of this research is to evaluate the effectiveness and reliability of Chao's estimator for species richness in Honduras, specifically focusing on the impact of sample size, taxonomic differences, and the precision of the estimator in incomplete inventories, using photographic data collected via the iNaturalist platform. To achieve this, the study aims to compare

the estimator's performance across different taxa and geographic areas, and to assess its behavior with incomplete iNaturalist inventories. This research is justified by the need to improve biodiversity assessments and conservation strategies in Honduras, especially given the challenges of incomplete data in citizen science platforms such as iNaturalist.

The research questions guiding this study are: Does the amount of sampling affect Chao's estimator performance? How effective is Chao's estimator for estimating species richness based on incomplete iNaturalist photographic inventories? Does Chao's estimator behave differently for different taxa? By answering these questions, this study seeks to enhance our understanding of species richness estimation methods and support effective biodiversity conservation in Honduras.

### Methods

The study employed a retrospective analysis of observational data from iNaturalist, focusing on four taxonomic groups: bats, birds, reptiles, and amphibians. These taxa were selected due to the well documented species count, allowing a comparison between these documented counts and the results obtained by using Chao's estimator based on the iNaturalist observations. For bats, the total count recognized for the country was 114 species (Mora et al., 2018); for amphibians and reptiles, 135 and 261 respectively (Solís, J. M., Wilson, L. D., Townsend, J. H., 2014); and for birds it is 798 species (Dubón et al., 2024).

In the iNaturalist platform observations from Honduras for the 10 years up to June 4<sup>th</sup>, 2024, was sorted by observation date. Observations labeled as "wild" were retained, while those identified as captive (referring to those in zoos or collections *ex situ*) were excluded.

The data collection involved the counting and recording of singleton and doubleton observations, representing species observed only once or twice, respectively, within the specified time frame. This process was performed year by year, over the 10-year study period. Subsequently, Chao's estimator was then applied to the recorded singleton and doubleton counts. The estimated number of missing species was then added to the actual number of species observed in the iNaturalist platform. In Equation 1, the singletons were represented by the variable  $a$ , and doubletons by the variable  $b$ .  $S$  represents the actual number of recorded species observed and identified.

$$Chao = S + \frac{a^2}{2b} \quad [1]$$

The resulting data was graphed, showing the curve of species richness over time, the estimation using the Equation 1, and the actual number of species according to literature (Chao & Chiu, 2014). This arrangement allowed the comparison of the curves and an analysis of the efficiency

of the estimation method. A previous study pointed out three a priori criteria (Foggo et al., 2003) for using a species estimation curve to determine if the estimate is viable:

The dataset had to show a clear asymptote with at least the final two values being identical.

The variance of the mean estimate of the observed number of species at the asymptotic point defined above must be zero.

The final 20% of the values of the curve preceding the first asymptotic sample had to be within 5% of the asymptotic value.

These criteria were employed to assess whether the curves generated in this study for estimated species richness met the standards observed in comparable research. Ensuring the final two values of each curve being identical indicates stabilization of species accumulation over time, which is named as the point of stabilization. This point is a critical aspect because it helps in understanding when sampling efforts have adequately captured the diversity present in the environment or dataset under study.

Following the initial process, the research proceeded to repeat the entire process, this time filtering the observations to include only those marked as "Research Grade" in the platform's filter. This step aimed to ensure higher data quality and reliability in the dataset by comparing the effect of the formula on those observations that were not reviewed and those that were. However, it eliminates some observations and reduces the sample size. The entire procedure was repeated for each of the selected taxa, allowing observation of Chao's estimator across different taxonomic groups. The data obtained were analyzed and compared with results obtained from previous studies, to assess the accuracy of Chao's estimator tool for species richness with the selected taxa in Honduras. Finally, the process was repeated twice more at smaller geographical scales. The platform has search filters that allow the selection of geographical limits. This study was repeated for the department of Francisco Morazán, and again for the municipality of San Antonio de Oriente where the Zamorano University campus is located.

For birds, reference data at the department and municipality levels were obtained through the eBird platform. The process began by accessing the eBird platform and navigating to the "Explore" section, where the "Bar Charts" option was selected. In the "Bar Charts" section, the geographical limits under study were selected. For the municipality, it was necessary to individually select hotspots within the department that corresponded to the geographical limits of the municipality. Reference data (Total species richness) at the department and municipal level of San Antonio de Oriente were not available for bats, amphibians, and reptiles.

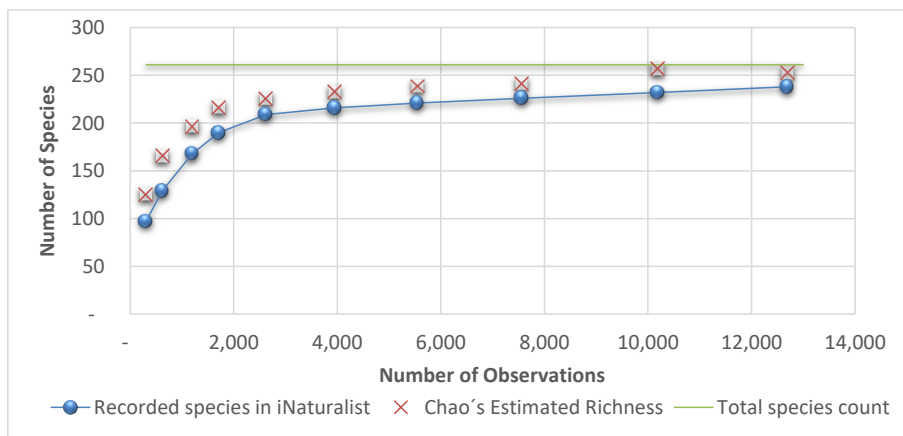
## Results and Discussion

### Reptiles

For reptiles (Figure 1), Chao's estimator calculated closely the species richness for Honduras (261 according to other authors). The formula estimated 257 species, which represents 98% of the total species, but considering the number of observations needed to estimate this percentage, 10,000 observations, it was not as practical as desired.

#### Figure 1

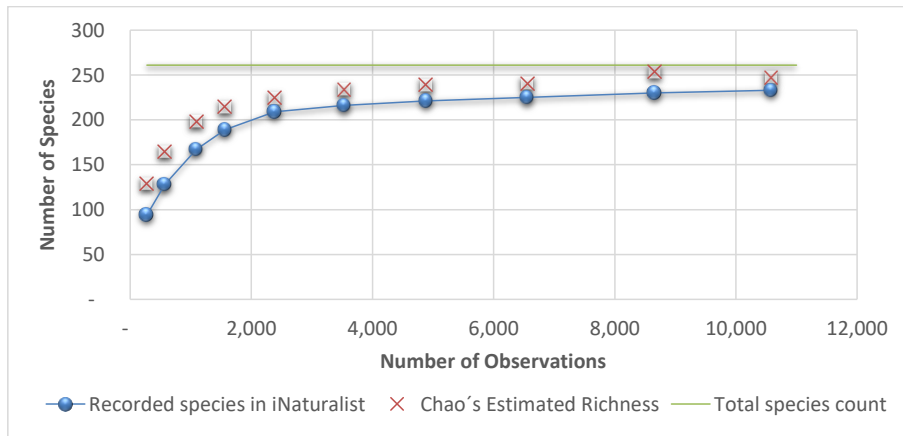
*Species richness in Honduras for reptiles, all observations included*



In the case of "only research grade observations", 8,662 observations were needed to closely reach the reference data (Figure 2), estimating 253 species (96%), which is less than the number of observations needed in Figure 1, and a similar percentage of species estimated.

**Figure 2**

*Species Richness in Honduras for reptiles, research grade observations only*

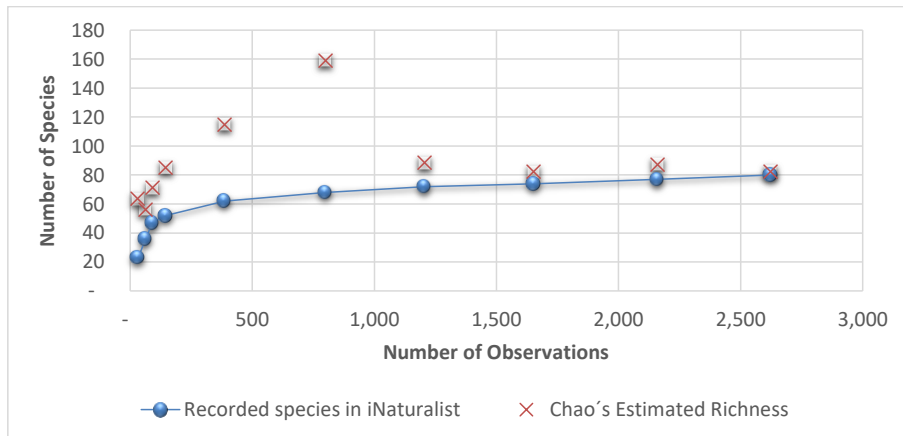


At the department level of Francisco Morazán, the accumulation curve seems to reach the asymptote (Figure 3) but does not accurately meet the three standards proposed by Foggo et al. (2003). At the end of the curve, the estimator approaches the accumulation curve, so the observed species and the estimator curve both indicate that the total number of species has been observed by year 10 with 2,624 observations. This convergence point indicates that 83 species is the total number of species. However, by year 8 with 1,653 observations, the formula estimated 98% of the expected total number of species for Francisco Morazán, which is an adequate percentage of estimation, and needed a smaller number of observations to be precise.

Comentado [VC2]: Explain a little better when this type of behavior is reached in the curves (# of observations, level, etc).

**Figure 3**

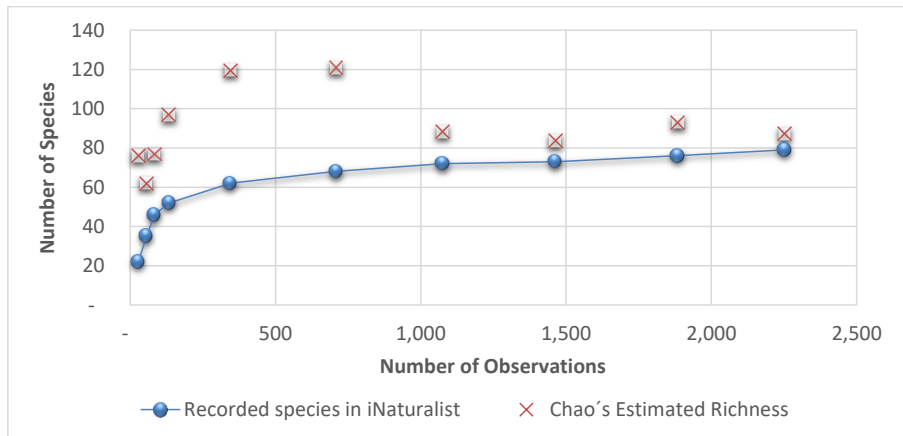
*Species richness in Francisco Morazán for reptiles, all observations included*



The distribution of estimates Figure 3 for the department was different when only research observations were considered (Figure 4). The outliers were less extreme compared to those in Figure 3. There is no point where the estimate and the accumulation curve meet, but both curves are close, meaning that the asymptote is likely to be close. There were 1,464 observations needed to accurately estimate this point where the curves meet (or closely meet), estimating the 101% of the species (considering the convergence point as the total number of species). For both Figure 3 and Figure 4, there was no reference data to which to compare the estimate to. Within the departmental level, the use of only research-grade observations is justified, as fewer observations were needed to estimate the point where the two curves converge.

**Figure 4**

*Species richness in Francisco Morazán for reptiles, research grade observations only*

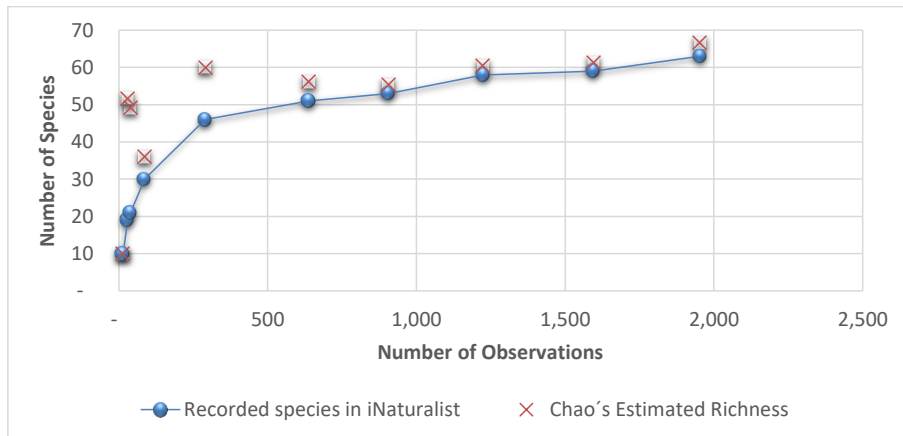


For the municipality limits as seen in Figure 5, the estimator did not perform well, as it simply followed the tendency of the accumulation curve for observed species. The three standards by Foggo et al. (2003) were not met. There is no convergence point within the curves, so it is not possible to predict the total number of species, and there is no reference data for reptile species count in San Antonio de Oriente. It is important to point out that reptiles include species of snakes, lizards, iguanas, geckos and more. These animals are usually not easy to observe or identify through photos, which is how iNaturalist works, and might lead to misidentification of species. It seems unlikely to ever have a complete inventory for reptiles, according to the curves.

Comentado [VC3]: Why??? What is happening in this case and how can this performance be changed?

Figure 5

*Species richness in San Antonio de Oriente for reptiles, all observations included*



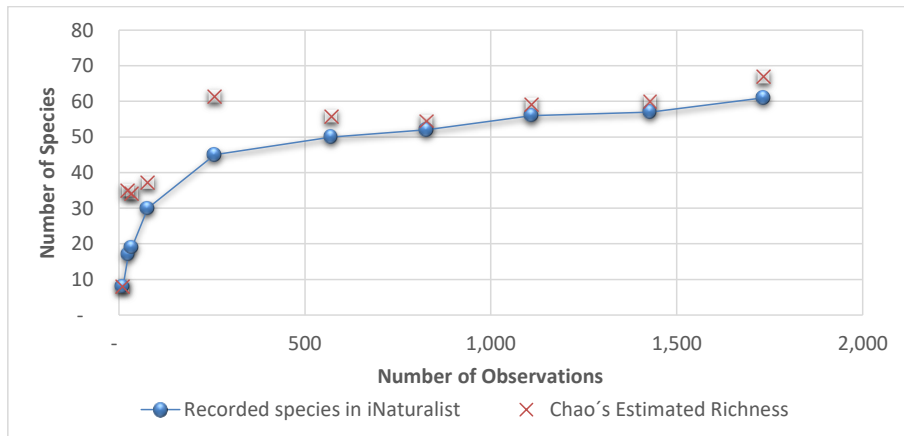
The estimator had a similar performance with research grade observations only (Figure 6) in San Antonio de Oriente, and for this case it is not justified to filter the observations. As with Figure 5, it is not possible to predict the total number of species. This might be caused by misidentification of species through photos in the platform, interfering in the formula's precision. Also, there might be species that are hidden under a genus. In other words, individuals which genus was identified, but never the specific species, so the individual is assigned to a group which it does not accurately belong to.

Comentado [VC4]: Same question... Is it because of the limited number of observations?

Comentado [FA5R4]: I don't think so, it is possible that there is a lot of misidentification of individuals.

Figure 6

*Species richness in San Antonio de Oriente for reptiles, research grade observations only*



## Birds

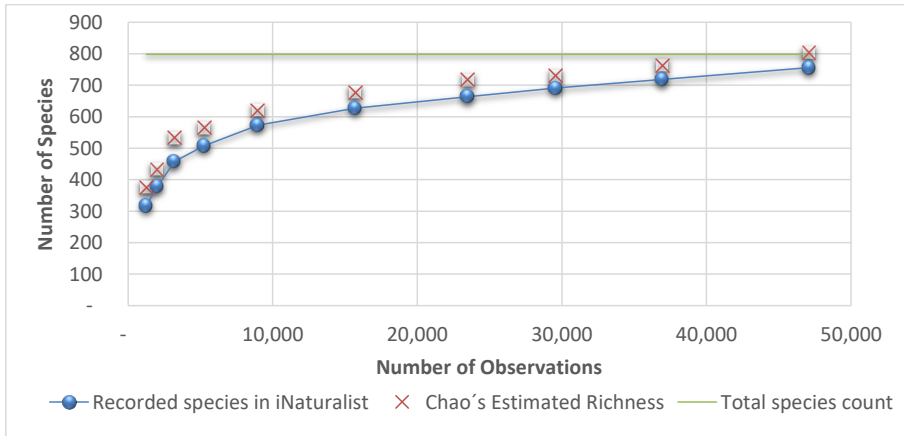
At the country level of Honduras (Figure 7 and Figure 8), the accumulation curve for bird observations demonstrated a steady increase in the number of species identified over time. By the 10th year, Chao's estimator closely approximated the actual number of species recognized in the literature (Figure 7), estimating 803 species (100%) with 47,112 observations. Figure 8 demonstrates a proximity between the reference and the estimate, estimating 778 species (97%) also by year 10, for research grade observations only. This indicates an elevated level of accuracy for Chao's estimator in predicting species richness for birds at a national scale, yet more than 40,000 observations were needed, and the accumulation curves do not reach the asymptote. This may be due to the high mobility characteristic of birds, and the reports of marginal and vagrant species that are not part of the regional bird community. Comparable results were obtained for this taxon, even when changing to department level (see Figure 9 and Figure 10), or to municipal level (see Figure 11 and Figure 12). The estimator was accurate but required many observations to reach the reference number of species.

Comentado [VC6]: Closure?

Comentado [GU7]: Dra. Cortes solicitó que se hace uniforme el uso de Chao o Chao's. Pero en este párrafo, aparecen ambas formas todavía. En mi opinión, "the Chao's estimator" es pobre gramática, pues lo correcto será utilizar Chao como adjetivo, y siempre usar "the Chao estimator".

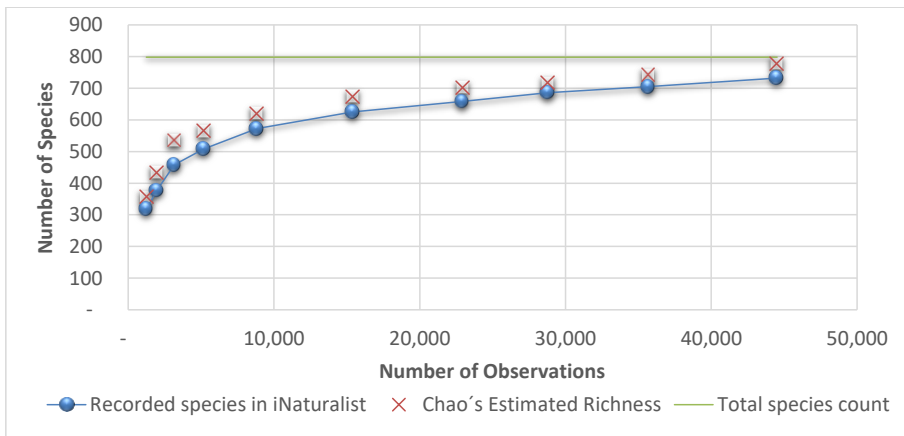
**Figure 7**

*Species richness in Honduras for birds, all observations included*



**Figure 8**

*Species richness in Honduras for birds, research grade observations only*



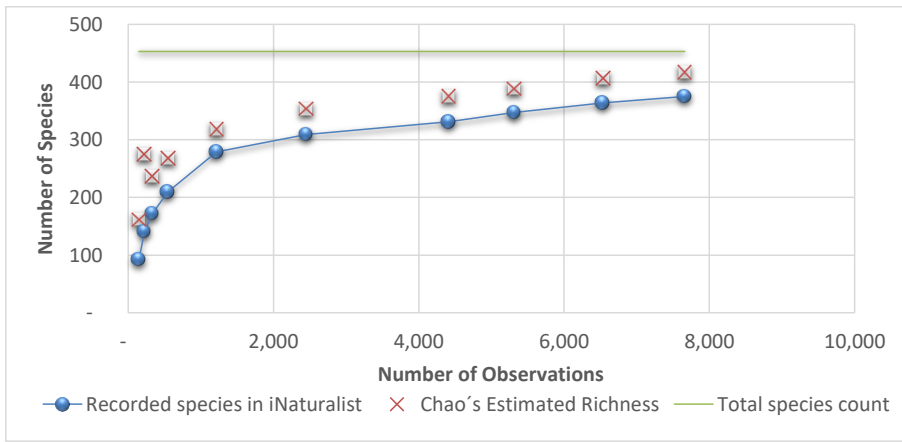
Within the limits of Francisco Morazán (Figure 9 and Figure 10), Chao's estimator estimated 92% and 90% of the bird species (all observations included and research grade observations only,

Comentado [VC8]: Verify that Chao's estimator is included throughout the document. There are sections where only Chao is used. Ensure consistency

respectively), with more than 7,000 observations needed. The use of research grade filter was not justified in this case, as the estimator performed better with all observations included.

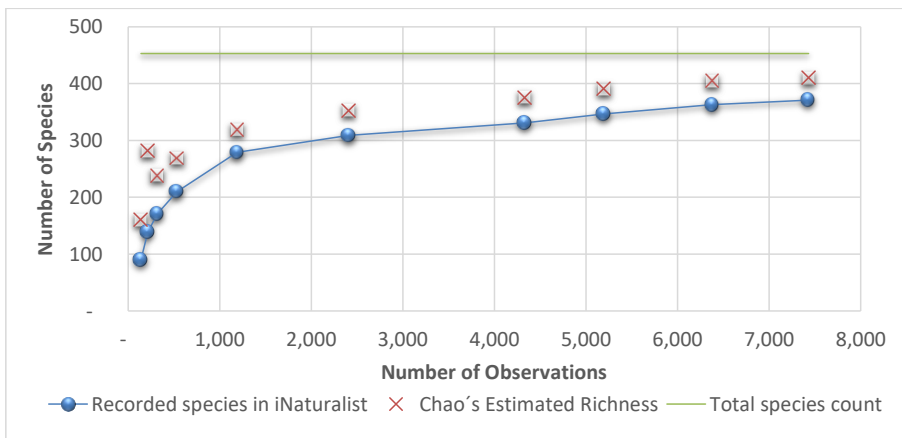
**Figure 9**

*Species richness in Francisco Morazán for birds, all observations included*



**Figure 10**

*Species richness in Francisco Morazán for birds, research grade observations only*



At the municipality level (San Antonio de Oriente), with every observation included (Figure 11), 89% of species were estimated with 5,756 observations. However, 88% of species were estimated with only 4,000 observations, which is precise and more practical. In the case of research-grade observations only (Figure 12), 88% of species were estimated, with 3,929 observations. In this context, the use of higher quality data is justified, as the estimation was precise for both cases, but with research grade observations only it required less observations to estimate the 88% of species.

Comentado [VC9]: Final comment is required.... What is the main finding? - Is the estimator adjusted? yes/no? why?

**Figure 11**

*Species richness in San Antonio de Oriente for birds, all observations included*

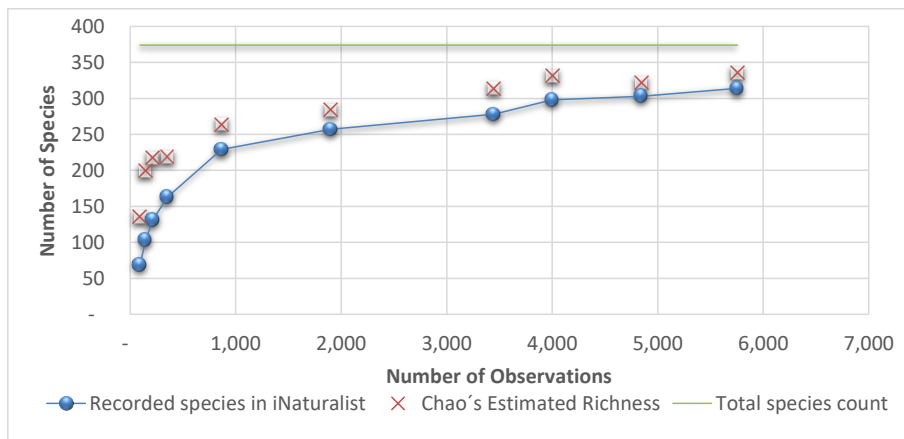
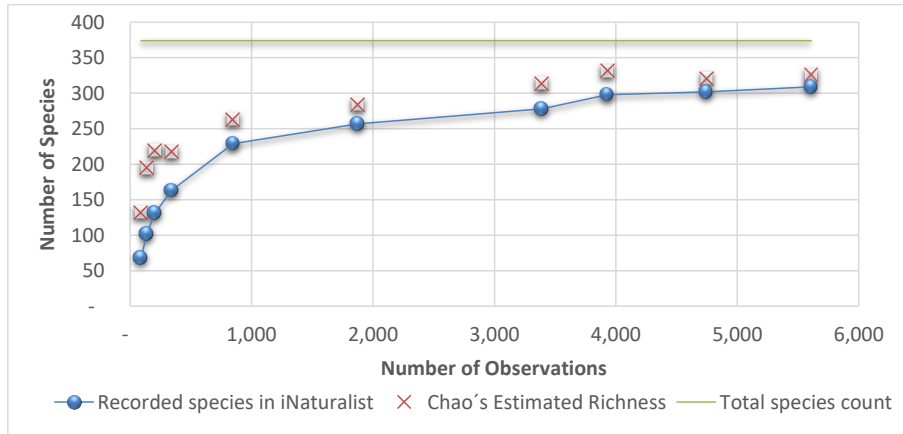


Figure 12

*Species richness in San Antonio de Oriente for birds, research grade observations only*



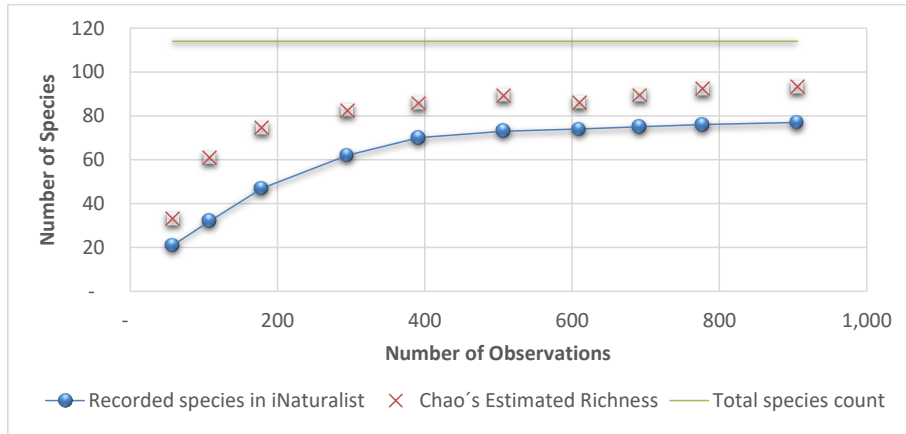
## Bats

At the national level (Honduras), Chao's estimator consistently fell short of reaching the literature reference for species richness estimates for bats, irrespective of whether all observations or only research-grade observations were considered. This similarity in performance suggests that the quality of observations (research grade observations only, or all the observations) did not significantly affect the estimator's ability to approximate species numbers. At the country level (Honduras), with all observations included (Figure 13), 81% of the total number of species was estimated with 906 observations. Figure 14 shows that 71% of species were observed after 446 observations in case of only research grade observations included. This underestimation might result from the difficulty in identifying bats in photographs. It is important to highlight that like birds, bats are highly mobile, so that the literature estimate for overall richness may include some marginal or irregular species.

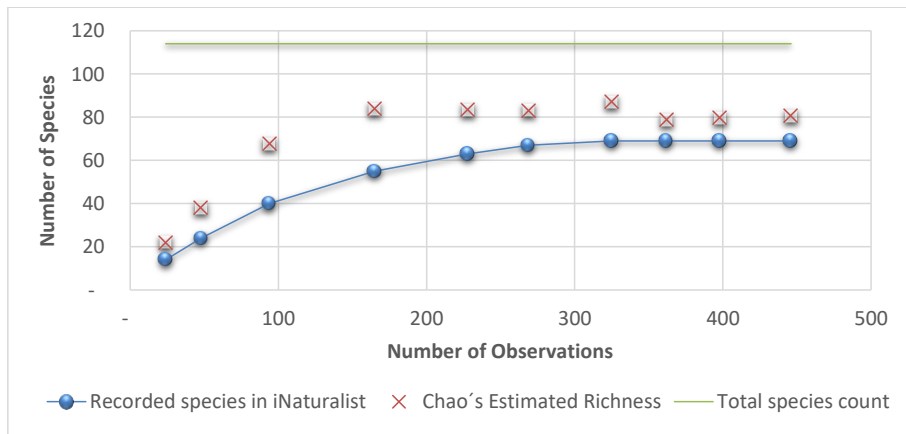
Comentado [VC10]: Explain "quality" of observations (quality or number?)

**Figure 13**

*Species richness in Honduras for bats, all observations included*

**Figure 14**

*Species richness in Honduras for bats, research grade observations only*

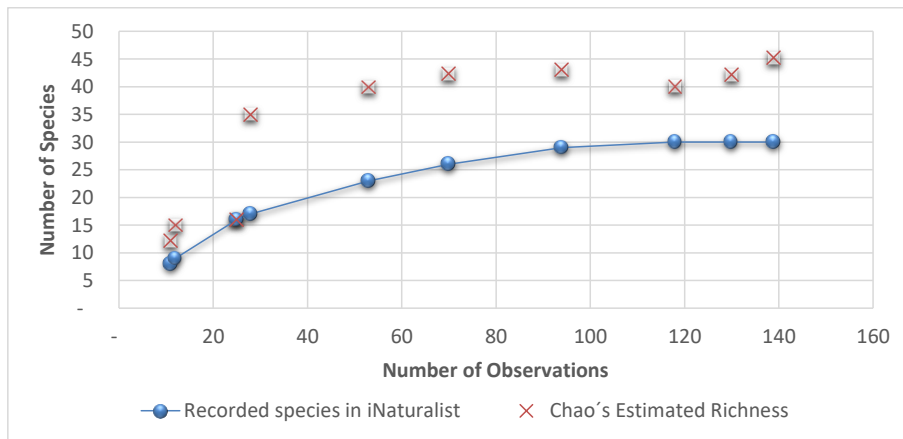


Similarly, in Francisco Morazán (Figure 15), the accumulation curve reaches the asymptote, meaning that for the municipality the total number of species is 30. In this case, the three a priori criteria (Foggo et al., 2003) are achieved, but the estimation curve does not converge with the

accumulation curve. For the municipality limits, the use of higher quality data is not justified as the curves in Figure 16 failed to reach the asymptote, and did not conform to expected standards, posing challenges for predictive accuracy. There is no reference data to evaluate the accuracy of the estimator for this case.

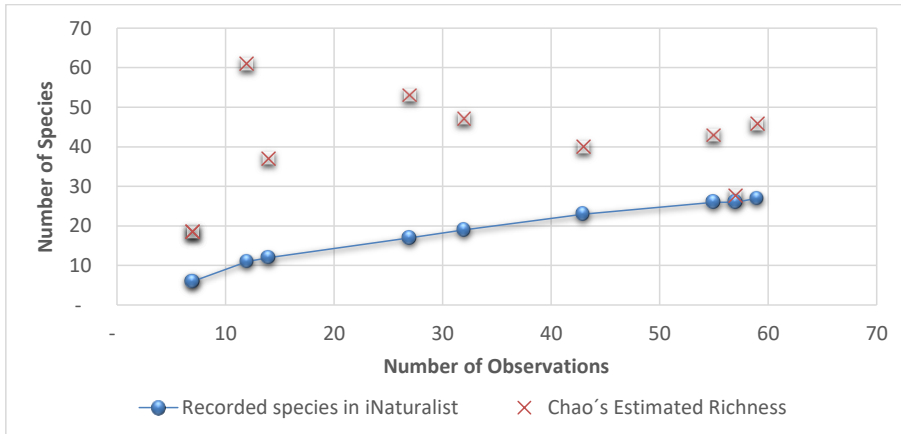
**Figure 15**

*Species richness in Francisco Morazán for bats, all observations included*



**Figure 16**

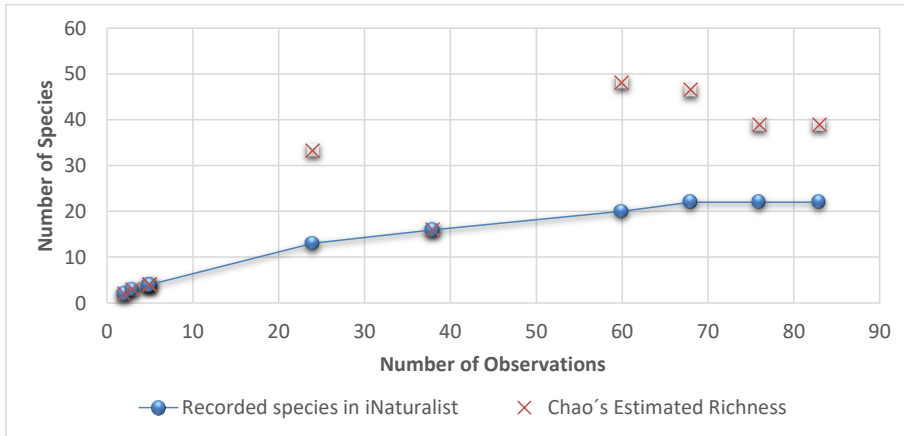
*Species richness in Francisco Morazán for bats, research grade observations only*



Furthermore, at the municipality level of San Antonio de Oriente (Figure 17 and Figure 18), limited data availability hindered Chao's estimator predictive capacity, indicating that more data is needed. In addition, in Figure 18 there is no visible accumulation curve due to the lack of historical data.

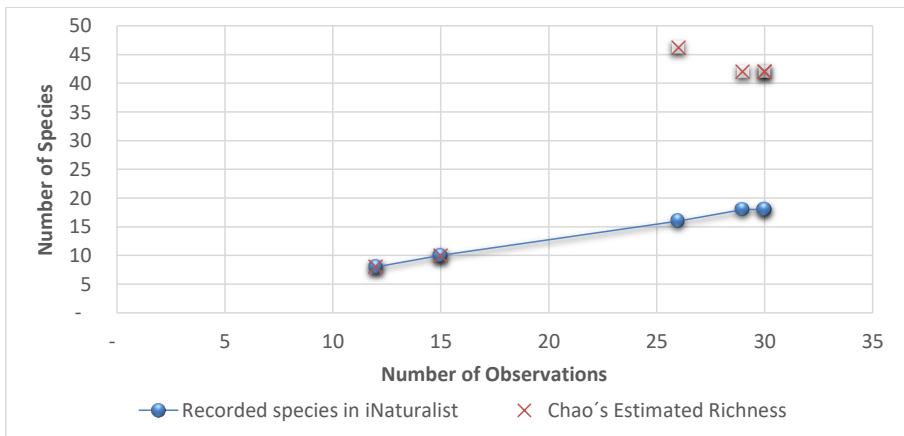
**Figure 17**

*Species richness in San Antonio de Oriente for bats, all observations included*



**Figure 18**

*Species richness in San Antonio de Oriente for bats, research grade observations only*

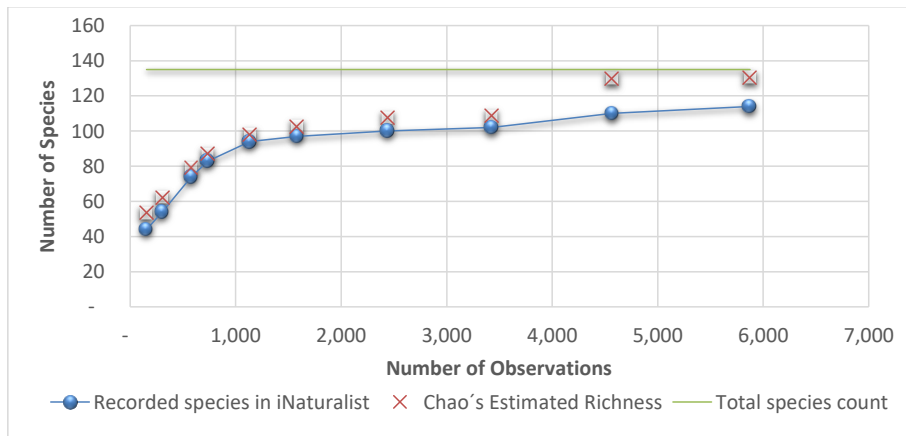


## Amphibians

For the national level (Honduras), Chao's estimator approached predicting total amphibian species richness (96%) with around 4,500 observations when all data were considered (Figure 19), demonstrating precision. Filtering for research-grade observations resulted in less precise estimates (Figure 20), with 83% of the species estimated, after 4,423 observations. Neither curve met the established standards (Figure 19 and Figure 20).

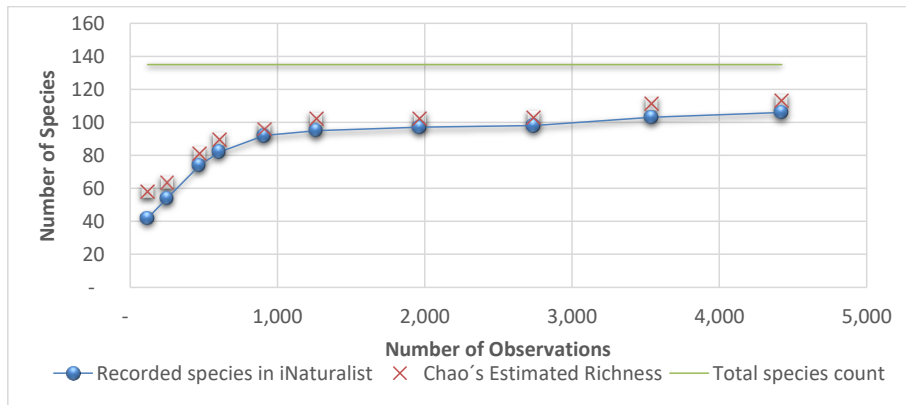
**Figure 19**

*Species richness in Honduras for amphibians, all observations included*



**Figure 20**

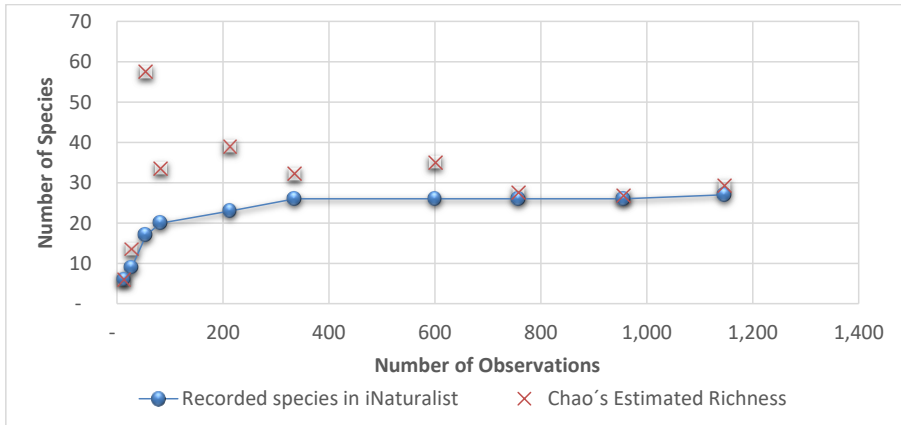
*Species richness in Honduras for amphibians, research grade observations only*



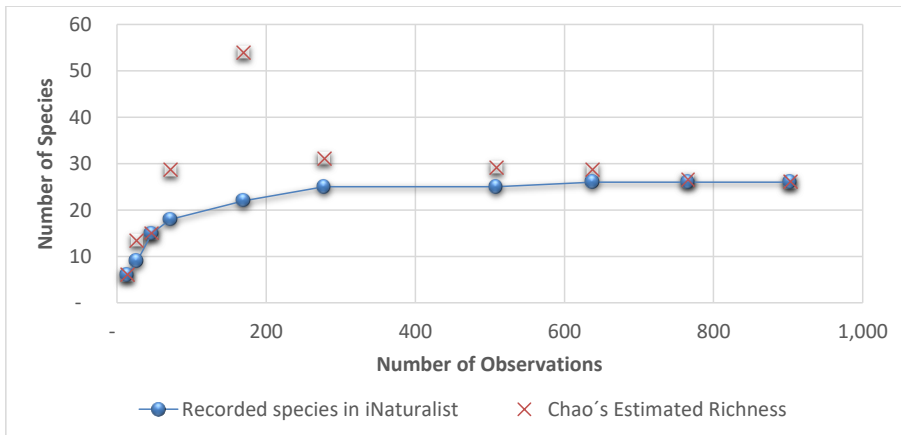
At the departmental level of Francisco Morazán, the accumulation curve that included all observations (Figure 21) produced a higher prediction from Chao's estimator compared to Figure 22. This overestimation was likely caused by misidentified non-research-grade observations. Thus, the higher estimates observed in Figure 21 can be attributed to the inclusion of lower-quality data, leading to an inflated species count. In contrast, when using research-grade data, the accumulation curve stabilized (Figure 22), indicating that 26 species represent the true richness in Francisco Morazán.

**Figure 21**

*Species richness in Francisco Morazán for amphibians, all observations included*

**Figure 22**

*Species richness in Francisco Morazán for amphibians, research grade observations only*



In San Antonio de Oriente, within municipal limits, accumulation curves failed to reach asymptotes (Figure 23), and the estimator did not exhibit stabilizing trends, thus proving ineffective in estimation (Figure 24). This discrepancy may arise from the smaller sample size and lower species

detectability at the municipal level (San Antonio de Oriente), which often results in fewer observed species and a higher proportion of rare species, making Chao's estimator less reliable. This outcome aligns with expectations, as the estimator's accuracy typically improves with increased sample size and coverage.

**Figure 23**

*Species richness in San Antonio de Oriente for amphibians, all observations included*

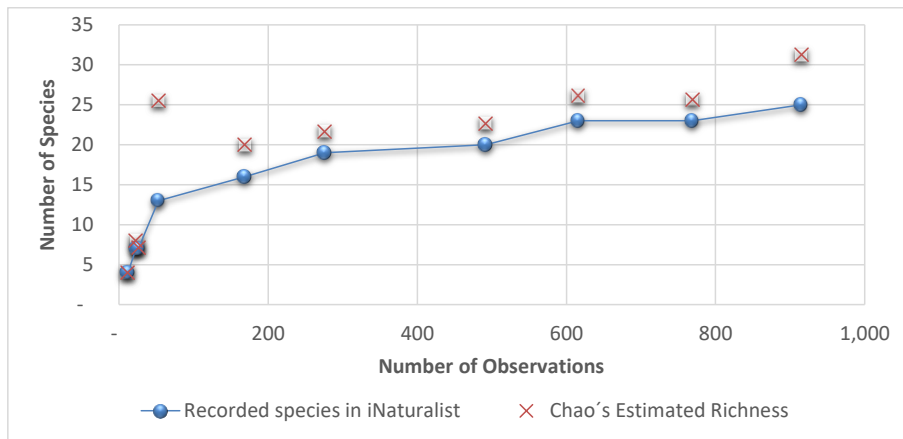
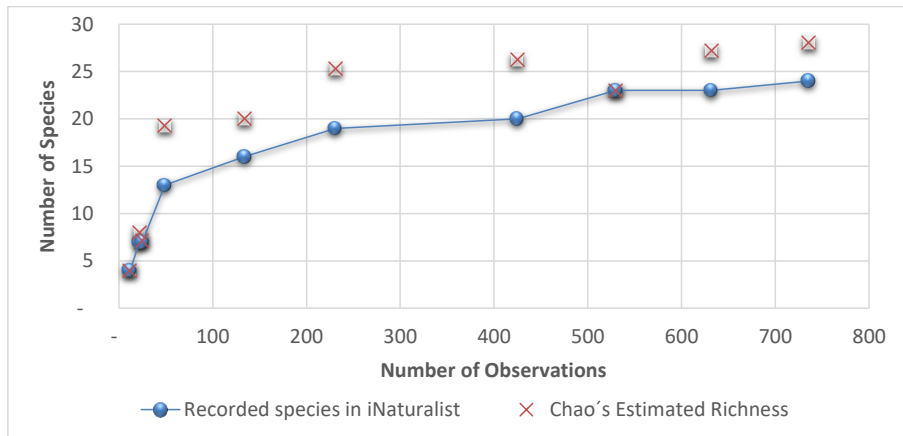


Figure 24

*Species richness in San Antonio de Oriente for amphibians, research grade observations only*



The use of Chao's estimator at the national level of Honduras showed promising accuracy in predicting bird species richness over a long-term observation period, but substantial data requirements and incomplete accumulation curves suggest limitations in capturing (photographing) species with lower detectability or sporadic occurrence. This may be due to a lower number of unique species by the end of the time frame. According to Foggo et al. (2003), Chao's estimator works better when there are fewer singletons and doubletons. Similarly, for reptiles, the inability of the estimator to reach an asymptote after a considerable number of observations, despite approaching the literature reference number of species indicates potential underestimation and the need for more samples. At finer spatial scales, such as departmental and municipal level, varying degrees of estimator performance were observed. Departmental analyses showed closer alignment between estimator predictions and actual species counts, especially when focusing on research-grade observations, which required fewer data points to converge. Similarly, for reptiles, the inability of the estimator to reach an asymptote despite approaching the literature reference species count after a substantial number of observations, indicates potential underestimation and the need for more

Comentado [GU11]: This is a comma that violates a grammar rule, so should be eliminated. The rule is "never separate a subject from its verb with a comma" (there is an exception but it doesn't apply here).

samples. However, municipal-level analyses often suffered from insufficient data, interfering in the estimator's ability to stabilize and accurately estimate species richness. In general, Chao's estimator tends to underestimate, similar to other studies (Ugland & Gray, 2004).

In most of the figures, the estimation curve followed the accumulation curve, similar to the figures in previous studies (Rocha et al., 2014). However, with amphibians and bats the estimator was different. The results from the formula tend to overestimate in the first years that were considered. This may be due to the proportion of singletons and doubletons in relation to the total inventory from iNaturalist. These taxa are the two that have the smallest inventories.

It is important to highlight the fact that there might be observations that were misidentified in iNaturalist, or that are hidden under a genus. In other words, there might be observations identified only to the genus level, but not the species, so the species was not counted in the inventory, or used to generate the estimate. Also, for groups with more than 3,000 species, the platform hides singletons and doubletons, so this type of study wouldn't be possible with flowering plants (*angiosperm* subphylum) with 3,685 reported species in iNaturalist for Honduras), or the class *insectae* (with 6,615 reported species in iNaturalist for Honduras).

### Conclusions

The study found that the effectiveness of Chao's estimator improves with larger sample size and with a reduced number of unique species. For instance, the estimator closely approximated the reference species counts for birds and reptiles at the country level of Honduras when the number of observations were sufficiently high. This indicates that larger datasets enhance the precision of species richness estimates, affirming that the amount of sampling affects the estimations.

The performance of Chao's estimator varied across different taxonomic groups. For birds, the estimator provided accurate species richness estimates when observations were available. In the case of reptiles, the estimator also showed high accuracy with a large dataset. However, for bats and amphibians, the estimator was less reliable, particularly with smaller sample sizes. With 4,567 observations for amphibians, the estimator was close to the reference data, but for bats, it did not approximate the reference counts effectively. This variability underscores the importance of taxonomic differences in the performance of the estimator.

While iNaturalist provided a valuable platform for collecting observational data, the study found that incomplete inventories and smaller sample sizes significantly impacted the accuracy of species richness estimates using Chao's estimator. These findings suggest that while iNaturalist is a powerful tool for citizen science and biodiversity monitoring, its effectiveness for species richness estimation is sensitive to the quantity and completeness of the data collected. This encourages users of iNaturalist to improve its data quality and quantity, by being active observers and identifiers to increase the reliability of the data.

The analysis revealed an important finding: the number of observations, singletons, and doubletons has a significant impact on the estimate, regardless of observation quality. Comparable results were obtained when comparing research-grade observations to the entire dataset, indicating that the estimator's performance is not affected by potential inaccuracies associated with non-

research grade data. This suggests that citizen science data can be reliable for biodiversity assessments, even when including non-expert observations.

### **Recommendations**

To further validate and expand upon the findings of this study, it is recommended to repeat the research in various locations that contain a higher number of observations. By diversifying the geographical limits, it will be possible to determine if the patterns observed in Honduras are consistent across other regions with varying levels of biodiversity. Additionally, conducting the study in larger areas could provide insights into whether the size of the study area influences the effectiveness of Chao's estimator and the overall assessment of species richness. Given that Honduras is a relatively small country, replicating the study in more extensive areas could reveal important trends and correlations that might not be evident in smaller-scale studies.

Including additional species richness estimators for comparison is another critical recommendation. By evaluating the performance of other estimators alongside Chao's estimator, the study can offer a more comprehensive understanding of the most effective tools for different taxa and geographical contexts. This comparative analysis would be instrumental in identifying the strengths and limitations of each method. Also, extending the time frame of the study is advisable. Data before 2014 was not available for Francisco Morazán, so the study was limited by start. A longer duration of data collection would account for seasonal variations and provide a more detailed picture of the species richness trend over time.

Finally, conducting a statistical analysis is essential to determine the significance of the variations observed in this research. This type of tests will quantify accuracy and reliability of Chao's estimator, providing a numeric understanding of its performance.

## References

- Chao, A., & Chiu, C.-H. (2014). Species Richness: Estimation and Comparison. In R. S. Kenett, N. T. Longford, W. W. Piegorisch, & F. Ruggeri (Eds.), *Wiley StatsRef: Statistics Reference Online* (pp. 1–26). Wiley. <https://doi.org/10.1002/9781118445112.stat03432.pub2>
- Chao, A., & Colwell, R. K. (2017). Thirty years of progeny from Chao's inequality: Estimating and comparing richness with incidence data and incomplete sampling. *SORT (Statistics and Operations Research Transactions)*, *41*(1), 3–54. <https://doi.org/10.2436/20.8080.02.49>
- Chiu, C.-H., Wang, Y.-T., Walther, B. A., & Chao, A. (2014). An improved nonparametric lower bound of species richness via a modified good-turing frequency formula. *Biometrics*, *70*(3), 671–682. <https://doi.org/10.1111/biom.12200>
- Dubón, F., Herrera, L., Juárez, R., Komar, O., Mejía, M., Rodríguez, F., & Dort, J. (2024). *Lista de Aves de Honduras*. Comité Científico de Aves. [https://aveshonduras.org/wp-content/uploads/2024/03/LISTA\\_AvesHonduras\\_2024\\_DIGITAL.pdf](https://aveshonduras.org/wp-content/uploads/2024/03/LISTA_AvesHonduras_2024_DIGITAL.pdf)
- Foggo, A., Attrill, M. J., Frost, M. T., & Rowden, A. A. (2003). Estimating marine species richness: an evaluation of six extrapolative techniques. *Marine Ecology Progress Series*, *248*, 15–26. <https://doi.org/10.3354/meps248015>
- Gorta, S. B. Z., Callaghan, C. T., Samonte, F., Ooi, M. K. J., Mesaglio, T., Laffan, S. W., & Cornwell, W. K. (2023). Multi-taxon biodiversity responses to the 2019–2020 Australian megafires. *Global Change Biology*, *29*(23), 6727–6740. <https://doi.org/10.1111/gcb.16955>
- Iknyan, K. J., Tingley, M. W., Furnas, B. J., & Beissinger, S. R. (2014). Detecting diversity: Emerging methods to estimate species diversity. *Trends in Ecology & Evolution*, *29*(2), 97–106. <https://doi.org/10.1016/j.tree.2013.10.012>
- Mora, J. M., López, L. I., Espinal, M. R., Marineros, L., & Ruedas, L. A. (2018). *Diversidad y conservación de los murciélagos de Honduras*. Master Print S. de R.L.
- Observaciones.*
- Rafferty, J. P. (2024). *Species richness | Definition, Examples, & Facts*. Encyclopedia Britannica. <https://www.britannica.com/science/species-richness>
- Rocha, R. G., Ferreira, E., Martins, I. C. M., Costa, L. P., & Fonseca, C. (2014). Seasonally flooded stepping stones: Emerging diversity of small mammal assemblage in the Amazonia-Cerrado ecotone, central Brazil. *Zoological Studies*, *53*(1), 1–10. <https://doi.org/10.1186/s40555-014-0060-0>
- Rode, Ž., & Torkar, G. (2023). The iNaturalist Application in Biology Education: A Systematic Review. *International Journal of Educational Methodology*, *9*(4), 725–744. <https://doi.org/10.12973/ijem.9.4.725>
- Shmelev, S. (1999). Political Ecology, Mountain Agriculture, and Knowledge in Honduras, by Kees Jansen (1998). Reviewed by Michael Paolisso. *Journal of Political Ecology*, *6*(1). <https://doi.org/10.2458/v6i1.21436>
- Solís, J. M., Wilson, L. D., Townsend, J. H. (2014). An updated list of the amphibians and reptiles of Honduras, with comments on their nomenclature, *1*(1), 123–144.

Ugland, K. I., & Gray, J. S. (2004). Estimation of species richness: analysis of the methods developed by Chao and Karakassis. *Marine Ecology Progress Series*, 284, 1–8. <https://doi.org/10.3354/meps284001>