Journal of
Caribbean Ornithology

RESEARCH ARTICLE

Vol. 36:45–61. 2023

Effects of observer skill and survey method on forest bird abundance data: recommendations for citizen science conservation monitoring in the Caribbean

Nicholas Bergen Christopher C. De Ruyck Nicola Koper

Photo: Christopher De Ruyck

BirdsCaribbean

Journal of Caribbean Ornithology

Research Article Vol. 36:45–61. 2023

jco.birdscaribbean.org ISSN 1544-4953 https://doi.org/10.55431/jco.2023.36.45-61

BirdsCaribbean
birdscaribbean.org

Effects of observer skill and survey method on forest bird abundance data: recommendations for citizen science conservation monitoring in the Caribbean

Nicholas Bergen[1,2], Christopher C. De Ruyck*[1,3], and Nicola Koper[1,4]

Associate Editor: Clint Boal

Cover Page: Flowering immortelle tree (*Erythrina poeppigiana*) at Belmont Estate, St. Patrick's, Grenada. Introduced to Grenada by Amerindians, the species is frequently retained throughout agro-forest and semi-natural habitats as the deciduous green leaves provide a nitrogen-rich mulch. Trees typically flower over the dry season when they provide an important source of nectar and associated insects for birds, with mixed flocks congregating around trees. Photograph by Christopher De Ruyck in February 2019.

Published: 9 August 2023

[1] Natural Resources Institute, 220-70 Dysart Road, University of Manitoba, Winnipeg, MB R3T 2M6, Canada

[2] e-mail: nbergen94@gmail.com

*[3] Corresponding author: e-mail: umderuy2@myumanitoba.ca

[4] Faculty of Environment, University of Northern British Columbia, Prince George, BC, Canada, V29 4Z9; e-mail: nicola.koper@unbc.ca

Cite this article as:

Bergen, N., C.C. De Ruyck, and N. Koper. 2023. Effects of observer skill and survey method on forest-bird abundance data: recommendations for citizen science conservation monitoring in the Caribbean. Journal of Caribbean Ornithology 36:45–61. https://doi.org/10.55431/jco.2023.36.45-61

Abstract

Citizen science is critical for monitoring bird populations in the Caribbean, where logistic and resource challenges significantly limit data availability. However, observer experience and environmental characteristics can influence results of surveys, including those that make use of citizen science. To evaluate how surveys could be designed to integrate volunteer citizen scientists with different levels of experience into bird population monitoring programs in the Caribbean, we used double-observer methods with pairings of variously skilled observers to compare observer perceptibility and relative abundance estimates among potential sampling protocols. In total, we conducted 265 point counts and 140 100-m line transects with 25-m truncations on the island of Grenada in 2016–2017. Our results clearly indicated that more individuals and more species were detected per survey area during point-count surveys than transect surveys. Novice observers detected as many easy-to-identify individuals as intermediate and expert observers. However, intermediate and expert observers detected significantly more difficult-to-identify species, and there was a significant correlation between observer skill and the number of unique species identified. To obtain reliable monitoring data while increasing survey skill and capacity, we recommend that surveys should: (1) implement point count rather than transect surveys, (2) be conducted during the local peak breeding period (e.g., the rainy season), (3) be conducted within 2 hrs after sunrise or 2 hrs before sunset, (4) be conducted in pairs using standardized double-observer protocols where one member of each pair has at least intermediate experience, (5) include in-person, in-situ training, (6) collect a measure of observer experience that can later be implemented in statistical analyses, and (7) include both English and local names on data collection sheets, to increase accessibility and out of respect for local knowledge of island ecosystems. Our results highlight the importance of evaluating protocols specific to Caribbean bird surveys rather than adapting protocols from other locations.

Keywords

bird abundance, Caribbean, conservation, double observer, environmental monitoring, forest, Grenada

Resumen

Efectos de la habilidad del observador y el método de muestreo en los datos de abundancia de aves de bosque: recomendaciones para el monitoreo de la conservación utilizando ciencia ciudadana en el Caribe • La ciencia ciudadana es fundamental para monitorear las poblaciones de aves en el Caribe, donde los problemas logísticos y de recursos limitan considerablemente la disponibilidad de datos. Sin embargo, la experiencia del observador y las características ambientales pueden influir en los resultados de los muestreos, incluidos lo que hacen uso de la ciencia ciudadana. Para evaluar cómo se podrían diseñar los muestreos para integrar a científicos ciudadanos voluntarios, con diferentes niveles de experiencia en los programas de monitoreo de poblaciones de aves en el Caribe, se utilizaron métodos de doble observador con parejas con diversas habilidades para comparar la perceptibilidad de los mismos y las estimaciones de abundancia relativa entre los potenciales protocolos de muestreo. En total realizamos 265 puntos de conteo y 140 transectos lineales de 100 m con ancho de banda de 25 m en la isla de Granada en 2016–2017. Nuestros resultados indicaron claramente que se detectaron más individuos y especies por área de estudio durante los muestreos de puntos de conteo que en los de transectos. Los observadores novatos detectaron la misma cantidad de individuos fáciles de identificar que los observadores intermedios y expertos; sin embargo, los observadores intermedios y expertos detectaron, de manera significativa, más especies difíciles de identificar. Hubo una correlación significativa entre la habilidad del observador y el número de especies únicas identificadas. Para obtener datos de monitoreo confiables y al mismo tiempo aumentar las capacidades/habilidades del muestreo, recomendamos que los mismos: (1) implemen-ten puntos de conteo en lugar de transectos, (2) se realicen durante el pico local de la etapa reproductiva (p. ej., la temporada de lluvias), (3) se realicen dentro de las 2 horas posteriores a la salida del sol o 2 horas antes de la puesta del mismo, (4) se lleven a cabo en parejas utilizando protocolos estandarizados de doble observador donde un miembro de cada pareja tenga, al menos, una experiencia intermedia, (5) incluyan capacitación en persona e in situ, (6) recopilen una medida de la experiencia del observador que luego pueda implementarse en análisis estadísticos, e (7) incluyan tanto los nombres en inglés como los locales en las hojas de colecta de datos, para aumentar la accesibilidad y por respeto al conocimiento local de los ecosistemas insulares. Nuestros resultados resaltan la importancia de evaluar protocolos específicos para los censos de aves del Caribe en lugar de adaptar protocolos de otros lugares.

Palabras clave

abundancia de aves, bosque, Caribe, conservación, doble observador, Granada, monitoreo ambiental

Résumé

Effets des compétences de l'observateur et de la méthode de relevé sur les données d'abondance des oiseaux forestiers : recommandations pour les suivis faisant appel aux sciences participatives dans la Caraïbe • Les sciences participatives sont essentielles pour le suivi des populations d'oiseaux dans la Caraïbe où les problèmes de logistique et de ressources limitent considérablement la disponibilité des données. Cependant, l'expérience de l'observateur et les caractéristiques environnementales peuvent avoir une influence sur les résultats des relevés, notamment lorsque l'on fait appel aux sciences participatives. Pour évaluer comment intégrer la participation d'amateurs bénévoles ayant différents niveaux d'expérience dans les programmes de suivi des populations d'oiseaux de la Caraïbe, nous avons utilisé des méthodes à deux observateurs avec des binômes de compétences différentes. Cela a permis de comparer les perceptions des observateurs et les estimations de l'abondance relative en fonction des différents protocoles d'échantillonnage potentiels. En 2016-2017, nous avons effectué au total sur l'île de Grenade 265 points fixes de comptage et 140 transects linéaires de 100 m avec des tronçons de 25 m. Nos résultats indiquent clairement que les comptages par points fixes ont permis de détecter plus d'individus et plus d'espèces par zone de suivi que les relevés par transects. Les observateurs débutants ont détecté autant d'individus faciles à identifier que les observateurs intermédiaires et expérimentés ; cependant, ces derniers ont détecté beaucoup plus d'espèces difficiles à identifier, et il existe une corrélation significative entre les compétences de l'observateur et le nombre d'espèces identifiées. Pour obtenir des données de suivi fiables tout en augmentant les compétences/capacités de suivi, nous recommandons que les suivis soient réalisés : 1) par points fixes plutôt que par transects ; 2) au cours du pic local de la saison de reproduction (p. ex. la saison des pluies) ; 3) dans les deux heures suivant le lever du soleil ou les deux heures précédant son coucher ; 4) en binôme en utilisant des protocoles standardisés à deux observateurs où l'un des membres de chaque binôme a au moins un niveau d'expérience intermédiaire ; 5) en incluant une formation en présentiel et sur place ; 6) en notant le niveau d'expérience de l'observateur de manière à pouvoir utiliser ultérieurement cette mesure dans des analyses statistiques ; et 7) en incluant les noms anglais et les noms locaux des espèces sur les feuilles de collecte de données afin de faciliter l'accessibilité et par respect pour les connaissances locales des écosystèmes insulaires. Nos résultats soulignent l'importance d'évaluer des protocoles spécifiques au suivi des oiseaux de la Caraïbe plutôt que d'adapter des protocoles provenant d'ailleurs.

Mots clés

abondance des oiseaux, Caraïbe, conservation, forêt, Grenade, méthode à deux observateurs, suivi environnemental

Citizen science is quickly becoming a key tool for avian research and monitoring (e.g., Cooper *et al.* 2014), because it can facilitate surveys of areas that would be too expensive or otherwise challenging to monitor (Tulloch *et al.* 2013). This is particularly true of small island states such as Grenada, which lack the resources and capacity to effectively manage all conservation concerns (Lack and Lack 1973, Wunderle 1985, Blockstein 1991, Watts and Wandesforde-Smith 2006, Rusk 2017). This creates opportunities for community-driven conservation efforts to provide meaningful contributions to wildlife monitoring (e.g., Caribbean Waterbird Census). Numerous studies have shown that under the correct circumstances, volunteers can contribute reliable species-specific data for a wide range of ecological monitoring projects (Gardiner *et al.* 2012, Moyer-Horner *et al.* 2012). However, different studies have drawn different conclusions about the efficacy or accuracy of citizen science data compared to data collected by trained experts (e.g., Acharya *et al.* 2009, Farmer 2012), suggesting that potential citizen science monitoring programs must be studied carefully and optimized to ensure the collection of high-quality, trustworthy data. Such data are currently absent for land-based citizen science surveys on Caribbean islands, and this is a problem because it is likely that the different biogeographic, seasonal, and socioeconomic conditions here are expected to alter optimal protocols and procedures for collecting data on birds, compared with the protocols that have been established for temperate, high-income countries, in which most citizen science programs have been evaluated (e.g., Greenwood 2007, McCaffrey 2015). Citizen science programs also play somewhat different roles in high-income compared with low-income countries. Whereas citizen science programs are often introduced to high-income countries to increase engagement of the public with their environment (e.g., Ng *et al.* 2018, West *et al.* 2021), citizen science programs may represent the only option for monitoring wildlife in low-income countries that have few financial resources dedicated to conservation. Concurrently, such programs can provide educational and skill development opportunities for participants with limited access to other professional development options. Given that citizen science program activities should be designed to achieve the specific goals of each program, our research objectives were therefore to evaluate how volunteers could best contribute to monitoring landbirds in the Caribbean country of Grenada, and better understand the unique needs and optimal design of citizen science monitoring programs on Caribbean islands broadly.

There are often trade-offs between the quantity and quality of monitoring data collected by citizen scientists (Silvertown *et al.* 2013), but appropriate survey and analysis design can help mitigate some of the costs. Citizen science projects draw on volunteers with a wide variety of experience and skill, which introduces monitoring challenges as observer effects due to differences in expertise, age, hearing acuity, and first time participant bias have all been shown to have effects on detectability (Emlen and Dejong 1992, Kendall *et al.* 1996, Diefenbach *et al.* 2003, Simons *et al.* 2009, Farmer *et al.* 2014, Leonard *et al.* 2014), and are one of the primary sources of variation when estimating detection probability during avian surveys (Rosenstock *et al.* 2002). The most effective ways to limit the effects of observer differences are to ensure high levels of training (Gallo and Waitt 2011) and use repeat sampling efforts with different observers; however, this can be restricted in citizen science projects because many of these projects arise precisely because resources are limited. Fortunately, assessment of observer expertise combined with detectability models can improve estimates of species distribution and abundance from citizen science data (Johnston *et al.* 2018). Knowledge of observer expertise also helps researchers assess which data are reliable and focus the efforts of observers in ways that suit their expertise.

Detectability models incorporate three different probabilities in wildlife surveys: the probability that the organism is currently in the survey area, p_p; the probability that the organism is in some way detectable (visible or auditory cues), called availability, p_a; and the probability that the observer actually detects the organism, or perceptibility, p_d. Detection probability can then be determined by $p = p_p * p_a * p_d$ (Nichols *et al.* 2009). However, most survey methods can only account for one or two of these aspects of detectability, while assuming the other one or two hold constant. Thus, multiple methods requiring further time and resources are required to estimate detectability across all species or locations to improve the accuracy of bird counts, which is the primary trade-off between data quality, quantity, and available resources. The suite of variables affecting counts can be split into those that affect availability and those that affect perceptibility. Duration of survey period, time of year, reproductive status, time of day, presence of observer, distance from observer, weather conditions, and density of conspecifics all affect availability (Johnson 2008). Distance from observer, attenuation of signals, habitat features, masking of cues by ambient noise, observer skills, weather conditions, conspicuousness of bird cues, and density of birds affect perceptibility (Johnson 2008). The effects of these different components can be evaluated to optimize survey procedures for specific monitoring needs, conditions, and available resources.

Different components of detectability can be assessed using double-observer surveys, distance sampling (Buckland 2006), and removal sampling (based on repeated measures; Farnsworth *et al.* 2002). Distance sampling can be prohibitively challenging in dense forest or with volunteers (e.g., Alldredge *et al.* 2007, Alldredge *et al.* 2008). Removal sampling is also difficult to implement with volunteers and requires longer census periods; it is also possible that the required assumption of a closed survey area (Farnsworth *et al.* 2002) is likely to be violated among dense populations of forest birds over longer census periods. Further, it is challenging to apply distance sampling or removal sampling in an equivalent way to both point counts and line transects (e.g., Diefenbach *et al.* 2007). As a result, for this study we chose to use double-observer sampling to assess perceptibility, given the conditions and limitations we faced that are typical of those on Caribbean islands, such as limited trained volunteers, resources, and infrastructure to travel and carry out surveys. In addition, double-observer sampling has other properties that are desirable in a volunteer program, such as increasing safety, mentorship opportunities, and opportunities for social interactions, which increases retention of volunteers (Ng *et al.* 2018). Thus, we judged double-observer methods to have the greatest scope for simultaneously producing reliable data, retaining volunteers and increasing survey skill and capacity, and examining the relationship between observer skill, detectability, and species richness and abundance estimates across different survey methods.

Point counts and line transects are the two most commonly used methods for collecting avian richness, diversity, and density data, but there is considerable habitat-specific variability in the efficiency and efficacy of one relative to the other (e.g., Edwards 1981, Verner and Ritter 1988, Dobkin and Rich 1998, Wilson *et al.* 2000, Rosenstock *et al.* 2002, Thomas *et al.* 2013). Differences in observed species richness and density estimates between transects and point counts are variable among studies and have been attributed to differences in detection probability among bird communities relative to habitat structure, distribution, and many other factors. However, few studies have directly compared detection probability of transects and point counts (Golding and Dreitz 2016, Cummings and Henry 2019), and we were unable to find any studies that compared transects and point-counts using a double-observer approach. Therefore, we designed our study so that we could compare the effects of observer skill on perceptibility within point counts and line transects, and evaluate the relative abundance indices and species counts produced.

Our overall goals were to understand how best to incorporate variably skilled observers into monitoring programs and account for this variation in analyzing trends, as well as comparing the efficacy of point counts and line transects to optimize citizen-science based programs on Caribbean islands that require broad, multi-species population trends and distribution data to help inform land-use decision making. We note that there are numerous other factors that are critical to consider in designing any wildlife monitoring survey. For example, survey locations must normally be selected to allow for repeated sampling among years, be surveyed at the same phenological period among years, be selected using stratified random sampling methods, and with consideration of independence among sample units over space and time. It was beyond the scope of our study to explore all of these principles, but we urge survey planners to consider all such factors in the design phase of monitoring programs.

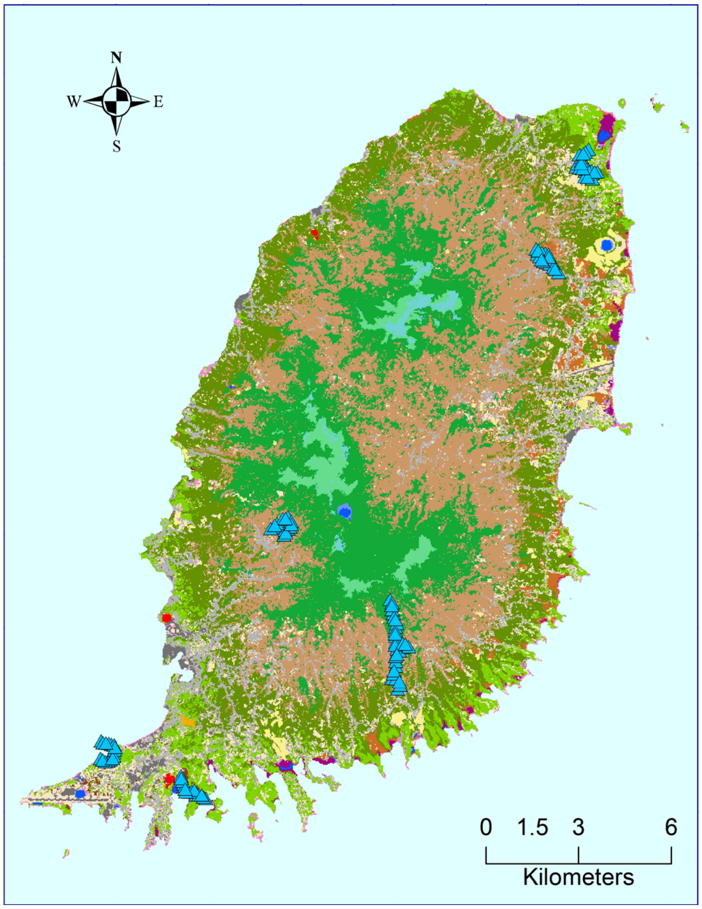

Fig. 1. Survey route locations (blue triangles) across the island of Grenada, 2016–2017. Major Land Cover classes: brown = woody agriculture, yellow = pasture/grass, light greens = coastal/dry forests, dark greens = wet tropical forests, blue green = cloud/elfin forest.

To accomplish our goals, we trained volunteers in bird identification (hereafter, "ID") and distance estimation, and conducted independent assessments of observer skill using sight and sound ID tests of easy- and difficult-to-detect birds. We then paired variously skilled observers together to conduct double observer surveys using both point counts and 100-m line transects at different times of day (dawn and dusk surveys), and over different seasons (wet and dry seasons). We modelled the various abundance counts and species richness data with their associated survey covariates to examine: (1) how observer skill affects detectability of different species and species groups; (2) how observer skill, survey method, and time of day and year affect counts of abundance and the number of species detected; (3) how observer skill influences counts of easy- and difficult-to-detect species; and (4) whether skill level of the primary observer affects the counts recorded by the secondary observer.

Methods

Study Site

Grenada is a small Caribbean nation made up of the islands Grenada, Carriacou, Petite Martinique, and numerous islets (total area ~348 km²), and is the most southerly nation in the Lesser Antilles. The Lesser Antilles were volcanically formed over the last 20 million yr and, therefore, all species colonization occurred by over-water dispersal (Ricklefs and Bermingham 2008). This has resulted in a relatively depauperate avifauna compared to nearby continental islands like Tobago (Lack 1973, Wunderle Jr. 1985), and this relatively low diversity makes it feasible to train volunteers for bird surveys in a timely manner. Grenada's bird community has one endemic species (Grenada Dove, *Leptotila wellsi*), one endemic subspecies (Grenada Hook-billed Kite, *Chondrohierax uncinatus mirus*), and several other species restricted to the Lesser Antilles (e.g., Lesser Antillean Tanager, *Stilpnia cucullata*; Grenada Flycatcher, *Myiarchus nugator*; and Lesser Antillean Bullfinch, *Loxigilla noctis*).

We conducted bird surveys at seven sites spread throughout the main island of Grenada (12°06'59"N, 61°40'44"W; Fig. 1). Elevation was 8–336 m above sea level, reflecting habitat and environmental variability found across the island. Survey routes reflected the diversity of the upland habitats found across the island, particularly secondary and dry lowland forests, which are the most extensive habitat types across the island. Habitats included dry thorny trees and bushes (e.g., *Haematoxylum campechianum*, *Bursera simaruba*, *Pisonia fragrans*, *Bourreria succulenta*, and *Leucaena leucocephala*; Rusk 2017) and closed canopy trees, usually in secondary growth forest or agroecosystems (e.g., *Dacryodes excels*, *Slonea caribea*, *Nectandra antillana*, *Cecropia peltata*), ~20 m in height within an agricultural matrix interspersed with rural residences. Sunrise varied from 0541 to 0630 and sunset from 1738 to 1823 during the study. Each of the seven survey routes consisted of eight point-count plots (1,962.5 m² per point count with 50-m detection radius) and four line-transect counts (5,000 m² per fixed width transect); therefore, 15,700 m² of each route was surveyed by point count and 20,000 m² by transect count. There were 34 different observers participating in the research, and we conducted a total of 405 surveys (265 point counts and 140 transects). Point counts were spaced 100 m apart, the minimum suggested distance for Caribbean landbird surveys (Wunderle Jr. 1994) to reduce the risk of double counting while maximizing the number of sites that volunteers could survey on foot.

Survey routes had to meet several criteria to be surveyed: (1) routes needed to be accessible and safe for volunteers, which meant that most routes were within walking distance of a bus route and took place on established trails and secondary roads that had little or no traffic during the survey; (2) routes were at least 1.2 km long so that the transect could contain 8 points, spaced 100 m apart, with a 200-m buffer before and after the first and last points respectively, and following one general direction of travel with little to no change in habitat type; (3) routes were in areas with small amounts of anthropogenic disturbance, to minimize risks to surveyors from traffic and to reduce interference from noise. While including off-road counts is usually considered an important part of any sampling design to avoid statistical problems resulting from convenience sampling (Ralph and Scott 1981, Anderson 2001), the typical definition of off-road sites being 200–400 m away from any road (Keller and Scallan 1999) proved to be too restrictive in small, densely populated Grenada. However, surveying small roads and trails where canopy cover is mostly intact mitigates some bias caused by road surveys (Keller and Fuller 1995). We controlled for variation among sites by doing all surveys on small secondary roads, with two observers, on resident bird species only, and in habitats with similar vertical structure. We also included variables to represent other factors that might influence detectability in our statistical models (such as season, time of day, and observer).

Field Methods

Training period.—Volunteers came to this project with varying degrees of proficiency and experience; no prior bird identification experience was required to participate. Volunteers were trained for a minimum of six hrs in visual and aural identification of Grenadian landbird species, distance estimation, and survey methods. Training began by demonstrating binocular use. Next, we led bird walks to work on identification skills in the field using laminated identification training cards (produced by BirdsCaribbean). Appropriate websites and training tools were provided for volunteers to study bird sounds at home. Volunteers were trained to be able to estimate a distance of 25 m, the fixed radius or strip width for the surveys. Following this training, volunteers were trained in double-observer methods and survey protocols, including filling out survey results cards that allowed for either English or Grenadian local names for each species. Volunteer observers then practiced point count and transect surveys. After completing the training requirements, observers took a visual and aural identification test (Miller *et al.* 2012) of a sample of Grenadian landbirds (Appendix 1). Volunteers were classified as either novice, intermediate, or expert based upon their results from the 29-point identification test. Those that scored less than 20 out of 29 were classified as novice, 21–26 out of 29 were classified as intermediate, and 27–29 of 29 as expert. Based on this training work and initial surveys, we also classed species as either easy-to-detect or difficult-to-detect (Appendix 1).

Surveys.—Surveys followed dependent double-observer protocols (Nichols *et al.* 2000). Two observers completed each count. One observer was designated as the "primary observer" and the other as the "secondary observer." Observers switched roles between counts so that each observer did the same number of surveys as primary and secondary observer. The primary observer identified birds and was responsible for communicating to the secondary observer the species, number of individuals, direction, whether the birds were closer or further than 25 m, and whether the detection was by sight, sound, or both. The secondary observer recorded the primary observer's identifications as well as any observations that the primary observer missed. To maintain independence between observers, secondary observers were instructed to position themselves behind the primary observer so that the primary observer could not take visual cues from the secondary observer, and observers were instructed to face away from each other, unless it was necessary to turn to make a visual ID or facilitate listening to distant calls.

Point counts were 5 min long (Dettemer *et al.* 1999), and observers recorded birds as either within 25 m of the observer (Wunderle 1994) or at a greater distance. Transects were 100 m long and travelled over 10 min, with observations categorized as within 25 m of the center of the transect or further than 25 m. We chose the 25-m threshold because detectability of our focal species in Grenada is high within 25 m (Williams 2020) and because this distance is recommended for Caribbean landbird surveys (Wunderle 1994). For this program we chose not to train observers to estimate distances to each bird (as is common for surveys conducted by professional observers) because most observers estimate distances incorrectly unless they have extensive (i.e., weeks of) training and the use of digital rangefinders to increase accuracy (Alldredge *et al.* 2007), neither of which is likely to be practical for most citizen science programs. Inaccurate distance estimates would be problematic for surveys, as inaccurate estimates of distances to birds decrease rather than increase precision and accuracy (e.g., Dawson and Efford 2009). Requiring observers to estimate distances also adds extra complexity that would make it more challenging for volunteer participants to conduct surveys, and it may also decrease participation by reducing participant's confidence in their performance. Therefore, while we recognize the value of distance sampling for many avian surveys, given the specific conditions of the citizen-science based surveys we addressed in this study, it was important to simplify distance estimates by only training observers to identify 25-m distances.

Unidentified birds were recorded as unknown to help us determine the extent to which novice observers were observing birds they did not know how to identify or simply not seeing or hearing individuals. Surveys were either completed as early morning surveys (within two hrs of sunrise), late morning (before 1300), or late afternoon (within two hrs of sunset). Surveys were postponed if it rained or if the wind speed was greater than 15 km/h.

Statistical Analysis

Effects of Observer Skill on Detection Probability.—We used the program DOBSERV (Hines 2000) to compare detectability of birds within observer pairs grouped by skill level (e.g., surveys done by a novice and an expert, a novice and intermediate, and an intermediate and expert) using data both collected only within the 25-m fixed-radius distance and with data collected using an unlimited radius, in separate analyses. Data were also analyzed using DOBSERV to determine whether detection probability differed by species or species group. We included only resident landbirds, excluding shorebirds and seabirds. Species were pooled into two groups *a priori* similar to Nichols *et al.* (2000), wherein we classified species based on plumage, song, or behavioral criteria as either easy-to-detect or difficult-to-detect. Groupings were determined from initial surveys in Grenada, and from the training period identification test, which demonstrated that observers struggled to identify certain species. Species groups contained more than 10 observations for all models (DOBSERV models listed in Appendix 2). Models were ranked by Akaike's Information Criterion for small sample sizes (AIC_C; Burnham and Anderson 2004). We estimated detection probabilities that are specific to observers at each skill level, and we also estimated joint detection probabilities, defined as the probability that at least one of the two observers in a pair will observe any given individual.

Modelling Observed Abundance and Species Richness using GLMs and GLMMs.—We used generalized linear models (GLM) and generalized linear mixed models (GLMM) to investigate how observer skill and methodological differences affected the number of individuals of each species observed. All data analyses and model fitting were conducted using R (R Core Team 2017). Generalized linear mixed models were constructed and fitted using the package *lme4* (Bates *et al.* 2014), and graphs and tables were created using *sjPlot* (Ludecke 2018) and *ggplot2* (Wickham 2016). We evaluated the effects of several predictor variables—including observer skill, time of year, time of day, and survey method used (transect/point count)—on several response variables—including the number of individual birds and species observed by primary observer, and the number of individuals observed only by the secondary observer. We had insufficient data to evaluate effects of the predictor variables on the number of species observed only by the secondary observer. We included survey route as a random effect in the GLMMs.

Diagnostic tests and plots indicated that, for most species, abundance data were best fit using a negative binomial distribution (Zurr 2009), and species richness data were best fit using a Poisson distribution. For two of the species (Lesser Antillean Tanager and Lesser Antillean Bullfinch), we fitted a binomial GLMM using presence-absence data because a negative binomial GLMM underfit the data due to the number of zeros in the count data. For the Grenada House Wren (*Troglodytes aedon grenadensis*), the inclusion of a random variable caused a lack of convergence in all models we tried, probably because it accounted for little or no variability in the data, so those data were fit using a negative binomial GLM. Similarly, we used a negative binomial GLM to model the number of individuals observed only by the secondary observer.

We used volunteers' scores on the pre-survey identification test as an index of volunteer skill (continuous variable with range 0–29). We used a categorical measure of season (dry/wet), with the first field season as the reference level (25 April–20 May 2016; rainy season). Other survey periods were 29 October–12 November 2016 (rainy season) and 30 January–3 March 2017 (dry season). We also divided time of day into three periods, with early morning (less than two hrs after sunrise) as the reference period, which was compared to late morning (before 1300), and late afternoon (within two hrs of sunset) periods. We also compared abundance of birds detected using point-count plots and transect surveys. We included survey area as an offset for fixed-radius analyses so that differences between point counts and transects could be attributed to the method and not the area surveyed, and we used sample duration as an offset for unlimited radius surveys, as survey area is undefined for unlimited-radius surveys (Simons *et al.* 2007). All offsets used in the GLM/GLMMs were log-transformed.

To ensure model fit, we used quantile-quantile diagnostic plots of scaled residuals using DHARMa (Hartig 2019). We also calculated the ratio of deviance to residual degrees of freedom to examine dispersion (Zuur *et al.* 2009, Bolker 2019a), and plotted residuals against each level of the random variable for GLMMs to examine among-site variance (Bolker 2019b). We also simulated the number of zeros expected by the models and compared them to the data to make sure that the true number of zeros was similar to that predicted by the model (Bolker 2019b).

Effects of Observer Skill and Survey Methods on Observed Species Richness.—We evaluated the effects of primary observer skill, season, time of day, and survey method (point count or transect) on observed species richness using a Poisson GLM. Survey route was used as a random effect. Because of the asymptotic nature of species richness curves, an offset could not be used to account for the non-linear relationship between survey effort and species richness. Instead, we aggregated richness data from consecutive point counts so that every two point-count plots were compared to one transect, to ensure that every data point represented the same temporal sampling effort (10 min). As a result, only the data from 206 of 265 total point counts could be used.

Effects of Observer Skill and Survey Methods on Observed Abundance.—We used a negative binomial GLMM to evaluate the effects of our predictor variables on observed abundance of easy-to-detect and difficult-to-detect species at both a 25-m fixed radius and an unlimited radius. For these analyses, abundance of each species group as identified by the primary observer was the response variable, and predictor variables were those in the saturated model, including skill of the primary observer, season, time of day, and survey method. Survey route was included as a random effect. We included survey area as an offset for fixed-radius analyses so that differences between point counts and transects could be attributed to the method and not the area surveyed, and survey duration as an offset for unlimited radius counts, as above. In addition, we tested whether there was a significant interaction between observer skill and survey method to determine whether optimal survey methods varied with observer expertise, for both easy-to-detect and difficult-to-detect species.

We also conducted similar analyses for observed abundance of individual species. We had sufficient data to assess the effects of our predictor variables on three species from the easy-to-detect group in our analyses: Bananaquit (*Coereba flaveola*), Tropical Mockingbird (*Mimus gilvus*), and Lesser Antillean Tanager; and two species from the difficult-to-detect group: Grenada House Wren and Lesser Antillean Bullfinch.

Effects of Observer Skill on Number of Birds Reported by Secondary Observer.—We also evaluated the number of individuals that were only detected by the secondary observer, to better understand how including a second observer in surveys added to data quality. We used a negative binomial GLM to test the effects of primary observer skill, secondary observer skill, survey method, and an interaction between primary and secondary observer skill, on number of easy-to-detect and hard-to-detect species that were detected only by the secondary observer. During preliminary data explorations, we found that the number of birds recorded by the primary observer was not a significant predictor of the number of birds recorded by the secondary observer as expected. Therefore, we wished to focus the model on assessing what was driving the number of birds counted by secondary observers. To simplify the model, we did not include time of year and season because the number of birds recorded by the secondary observer did not seem to rely on the number of birds present on any given survey. The form of this model followed the negative binomial described above, with the addition of an interaction term.

Results

Over three field seasons (25 April–20 May 2016, 29 October–12 November 2016, and 30 January–3 March 2017), 34 volunteers

Table 1. Detection probability (*p*) and confidence intervals (CI) sorted by each pairing of observer skill category at both a fixed radius and unlimited radius. The first four rows show the detection probability for each observer for both species groups (easy- and difficult-to-detect) when novices are paired with intermediates. The middle grouping of rows shows the same information for when novices are paired with expert observers. The last two rows show the observer-specific probability when intermediates are paired with experts (no differentiation between easy- and difficult-to-detect species). The "Best Model" column indicates the factors included in the best-fitting model for each observer pairing based on ΔAICc, S = species, G = species group, and I = observer.

		25-m Fixed Radius		Unlimited Radius		
Observer Category	**Species Group**	***p***	**95% CI**	***p***	**95% CI**	**Best Model**
Novice	Easy	0.7749	0.710–0.839	0.7105	0.656–0.765	P(G,I)
Intermediate	Easy	0.7691	0.707–0.830	0.7997	0.751–0.845	P(G,I)
Novice	Difficult	0.4847	0.349–0.621	0.3945	0.252–0.534	P(G,I)
Intermediate	Difficult	0.8101	0.710–0.910	0.7964	0.673–0.920	P(G,I)
Novice	Easy	0.8060	0.752–0.860	0.7137	0.666–0.761	P(G,I)
Expert	Easy	0.8233	0.766–0.881	0.8026	0.753–0.853	P(G,I)
Novice	Difficult	0.5726	0.445–0.696	0.4463	0.333–0.560	P(S,I)
Expert	Difficult	0.9030	0.795–1.011	0.8218	0.681–0.962	P(S,I)
Intermediate	All	0.6947	0.639–0.750	0.6838	0.643–0.725	P(.,I)
Expert	All	0.8451	0.802–0.888	0.8222	0.788–0.857	P(.,I)

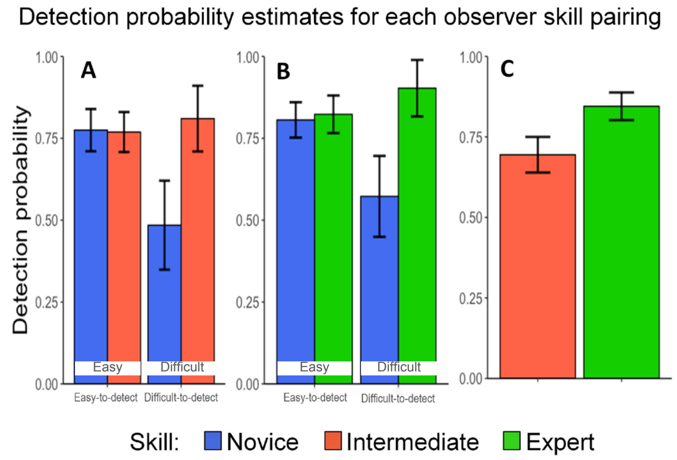

Fig. 2. Detection probability differed between observer skill levels for difficult-to-detect bird species but not for easy-to-detect species across (A) Novice–Intermediate pairs and (B) Novice–Expert pairs. (C) Detection probability was significantly higher for expert observers than intermediate observers, but this was not dependent on species group. Error bars denote 95% confidence intervals. *n* = 34 (13 novice, 15 intermediate, and 6 expert).

completed 265 point counts and 140 transect counts. A total of 5,990 detections of 39 species were recorded during these counts. In the species identification quiz used to quantify observer skill, novice observers scored an average of 3.9/13 on song identification (*n* = 13), while intermediates scored an average of 8.6/13 (*n* = 15) and experts scored 12.2/13 (*n* = 6). Differences were far less pronounced on the visual portion of the quiz, within which novice observers scored an average of 12.9/16, compared to intermediates at 15.3/16 and experts at 16/16.

Effects of Observer Skill on Detection Probability

Novice observers detected significantly fewer individuals of difficult-to-detect species than their intermediate or expert partners, but there was no evidence of effects of observer skill on detection of easy-to-detect species (Table 1; Fig. 2). Expert observers detected significantly more individuals than intermediate partners, regardless of detection difficulty of species. Model selection using Akaike's Information Criterion for a small sample size (AIC$_c$; Burnham and Anderson 2004) to select the best model for estimating detection probability was similar between fixed-radius and unlimited-radius analyses except for novice/expert pairings, where at the unlimited radius, detectability differed by individual species rather than species group (Table 1).

Adding a novice secondary observer increased detectability of easy-to-detect species, and all species combined, by either intermediate or expert observers, and this effect size was quite significant. For example, detectability of easy-to-detect species by intermediate observers increased from 0.769 to 0.954 when a novice partner was added, representing a 24% increase in detectability (Table 2). Similarly, adding an intermediate partner to a survey led by an expert increased detectability from 0.845 to 0.953 (13% increase). However, novice partners did not significantly increase detectability of difficult-to-detect species (Table 2).

Effects of Season, Time of Day, Observer Experience, and Survey Method

More birds were detected per unit area during point-count plots than on transect counts (Fig. 3A–B). Surprisingly, time of day had little effect on the number of birds observed overall; however, more easy-to-detect species were found at the unlimited radius when surveys were conducted in the evening than in the morning (Incidence rate ratio = 1.20, *p* = 0.014, *n* = 353).

Table 2. Joint detection probability relative to the detection probability of the most skilled observer for each observer pairing and species group for fixed radius surveys. CI = Confidence Interval. † denotes significantly higher joint detection probability at the 0.05 α level.

Observer Pairing	Species Group	Most Skilled Observer Detection Probability	95% CI	Joint Detection Probability	95% CI
Novice–Intermediate	Easy	0.769	0.707–0.830	0.954	0.932–0.976†
Novice–Intermediate	Difficult	0.810	0.710–0.910	0.893	0.799–0.986
Novice–Expert	Easy	0.823	0.766–0.881	0.962	0.942–0.982†
Novice–Expert	Difficult	0.903	0.795–1.000	0.960	0.909–1.000
Intermediate–Expert	All species	0.845	0.788–0.857	0.953	0.935–0.971†

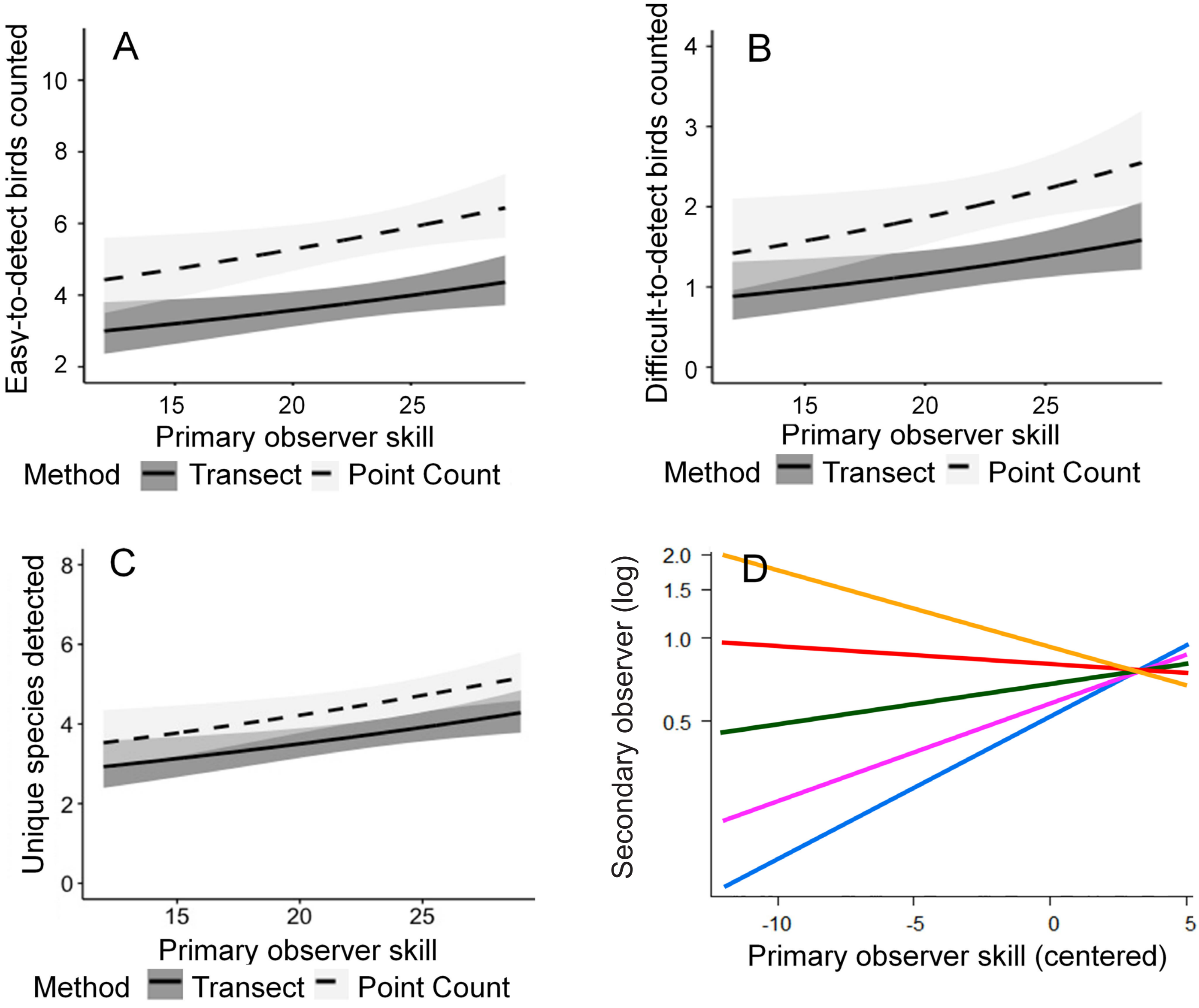

Fig. 3. Number of species detected relative to observer skill, controlling for survey effort. (A) The number of easy-to-detect species increased with primary observer skill and was higher for point counts than transects. (B) the number of difficult-to-detect species increased with primary observer skill and was higher for point counts than transects. Intervals are 95% confidence intervals. (C) The number of species detected increased with primary observer skill and was higher for point counts than transects per unit effort. (D) Interactions between detections of individuals by the secondary observer relative to primary observer skill. Each colored line shows a different skill level for secondary observers, centered around the mean to reduce collinearity. Slopes represent secondary observer skill (centered): yellow = 5, pink = 1, green = -3, magenta = -7, blue = -10.

Table 3. Abundances detected for Bananaquit, Tropical Mockingbird, and Lesser Antillean Bullfinch. Probability of presence for Lesser Antillean Tanager and Grenada House Wren. All estimates are on the log scale. † denotes significance at the 0.05 α level. Each data set was fitted according to the most descriptive model possible (GLMM or GLM). These models include only the observations made by the primary observer. Season is compared to the first field season in April-May, time of day is compared to early morning counts, and point count is compared to the reference level, transects. "Estimate" = parameter estimate, "SE" = standard error, and "*p*" = *p*-value.

	Bananaquit			**Tropical Mockingbird**			**Lesser Antillean Tanager**			**Lesser Antillean Bullfinch**			**Grenada House Wren**		
	Estimate	**SE**	***p***	**Estimate**	**SE**	***p***	**Estimate**	**SE**	***p***	**Estimate**	**SE**	***p***	**Estimate**	**SE**	***p***
Observer skill index	0.0158	0.012	0.196	0.00130	0.0200	0.948	–0.0218	0.0455	0.630	0.0404	0.0358	0.259	0.273	0.0709	< 0.005†
October/ November	–0.137	0.171	0.424	0.566	0.295	0.055	–2.0797	1.092	0.057	0.952	0.478	0.172	–0.531	0.840	0.527
January/February	–0.384	0.150	0.011†	0.348	0.261	0.182	0.454	0.564	0.421	–1.388	0.527	0.009†	1.113	0.472	0.018†
Late morning	–0.122	0.120	0.312	0.120	0.213	0.572	0.0844	0.485	0.861	–0.576	0.388	0.138	0.0449	0.431	0.916
Evening	0.256	0.136	0.0598	–0.351	0.243	0.148	0.997	0.587	0.09	0.0458	0.477	0.923	–1.516	0.592	0.011†
Point count	0.345	0.096	< 0.005†	0.354	0.168	0.035†	0.087	0.377	0.817	0.471	0.307	0.124	0.230	0.379	0.543
Model	negative binomial GLMM			negative binomial GLMM			binomial GLMM			binomial GLMM			negative binomial GLM		

Table 4. Influence of observer skill and count method on observed abundance for difficult-to-detect species at fixed and unlimited radius. This includes only the observations made by the primary observer. Season is compared to the first field season in April/May, time of day is compared to early morning counts, and point count is compared to line transect. Incidence rate ratios are the ratio at which birds are expected to be observed relative to the reference level. † denotes significance at the 0.05 α level.

Predictors	**Incidence Rate Ratios**	**CI**	***p***
Intercept	2.67	1.80–3.96	< 0.001†
Primary Observer Skill Index	1.03	1.01–1.04	0.001†
October/November	0.83	0.67–1.04	0.101
January/February	0.67	0.58–0.77	< 0.001†
Late Morning	0.95	0.81–1.11	0.514
Evening	0.93	0.78–1.10	0.378
Point Count (Fixed Radius)	1.20	1.06–1.37	0.005†
Observations	226		
Cox & Snell's R^2 / Nagelkerke's R^2	0.209 / 0.290		

For both unlimited and fixed-width point counts, the number of both easy- and difficult-to-detect birds increased with observer skill (easy-to-detect species: Incidence rate ratio = 1.02, $p = 0.036$ fixed width, Incidence rate ratio = 1.03, $p < 0.01$ unlimited radius; difficult-to-detect species: Incidence rate ratio = 1.03, $p = 0.023$ fixed width, Incidence rate ratio = 1.04, $p = 0.01$ unlimited radius, $n = 353$; Appendices 3–4). For easy- and difficult-to-detect species, significantly fewer individuals were detected during the dry season (easy-to-detect-species: Incidence rate ratio = 0.65, $p < 0.001$, fixed width, Incidence rate ratio = 0.74, $p < 0.001$, unlimited radius, $n = 353$; difficult-to-detect species: Incidence rate ratio = 0.64, $p < 0.015$, fixed width, Incidence rate ratio = 0.71, $p < 0.041$, unlimited radius, $n = 353$).

Detection of individual species (Bananaquit, Tropical Mockingbird, Lesser Antillean Tanager, Lesser Antillean Bullfinch, and Grenada House Wren) varied with different predictors (Table 3). Bananaquits and Tropical Mockingbirds were recorded at a significantly higher rate in point counts compared with transects. Detection of Lesser Antillean Tanagers was independent of the predictors. Lesser Antillean Bullfinches were observed in fewer surveys during the dry season. The number of Grenada House Wrens recorded positively correlated with observer skill, and fewer Grenada House Wrens were detected in the dry season and during evening compared with morning surveys (Table 3).

Observers with greater skill detected more species, and fewer species were identified in the dry season (Table 4). Point counts resulted in detection of more species than transects (Fig. 3C; Table 4). Observer skill was positively correlated with the number of individuals detected for both easy- and difficult-to-detect species, such that numbers of detections by experts were ~30–60% higher than novices (Appendices 3–4).

Effects of Observer Skill on Number of Birds Reported by Secondary Observer

At a fixed radius, skilled secondary observers recorded relatively more birds ($p = 0.019$) when they were paired with less skilled primary observers (Table 2; Fig. 3D). Unexpectedly, however, less skilled secondary observers detected more additional individuals when they were paired with a highly skilled primary observer. For difficult-to-detect species, more skilled secondary observers detected more individuals regardless of the experience of the primary observer (Table 2).

Discussion

Our results have practical implications for designing citizen science monitoring programs and provide important insights into the efficacy and suitability of using double-observer survey designs in Grenada and other similar small tropical islands. For most species, and cumulatively, similar numbers of individuals were detected during surveys conducted in the early morning or early evening, suggesting that surveys could be conducted in either period. This differs from many temperate-zone survey protocols for passerines, which require that surveys be conducted in the morning, but this difference offers important scheduling flexibility that would enable greater opportunities to engage citizen scientists who are dependent on public transportation or who work early in the day, as is common in Grenada's farming communities. Our results suggest that landbird surveys should be conducted during the rainy season, corresponding with the principal breeding period for most species on Grenada. There are numerous benefits of implementing double-observer protocols for citizen science surveys, both in terms of improving robustness of data collected, and for mentorship, social, and safety reasons.

Our results clearly indicated that more individuals and more species were detected on point-count surveys than transect surveys, for every metric that we tested. This result differs somewhat from the variable results of other studies that have compared point counts and transects (e.g., Ratkowsky and Ratkowsky 1979, Anderson and Ohmart 1981, Edwards in Ralph and Scott 1981, Verner and Ritter 1988, Wilson *et al.* 2000). We speculate that one reason for this discrepancy is that our point-count radii and transect distances from center were 25 m, while many temperate surveys use 100 m for the corresponding distances. It is likely that this difference in spatial extent of each plot significantly alters optimal survey protocols, particularly because detectability of birds is high within 25 m of the observer but drops considerably between 25 and 100 m (e.g., Buckland 2006). We suggest that more birds may be available for detection during 25-m fixed-radius point counts because the whole area of a point-count plot is observable for the entire duration of the count, whereas only a portion of the overall area of an equivalently sized transect is observable at any given time. In contrast, with 100-m radius plots in temperate systems, detectability throughout most of the plot is significantly less than 1.0, but individuals can often be detected at a much greater distance than is typical for tropical surveys (Waide and Narins 1988, Diefenbach *et al.* 2003, Shankar Raman 2003), so transects may allow for opportunities for detection from a greater distance and over a longer period in temperate systems. This highlights the importance of evaluating protocols for applicability to bird surveys in the Caribbean, rather than adapting protocols from other locations.

From a practical perspective, there are several other advantages of point-count plots over transects for citizen science programs. First, non-expert observers detected a higher proportion of observations by sight than by sound. Point-count plots give non-experts a better chance to see individuals and a better chance at encountering multiple stimuli from the same individual. Second, observers can focus entirely on observing and recording birds during a point count, whereas on a transect, observers must pay attention to the speed that they are walking, invest some focus on the path they are travelling, and spend less time per unit of count area (Verner and Ritter 1988).

As expected, the number of individuals and species detected increased with observer skill. However, this effect varied among species. Skill did not affect how many Bananaquits, Tropical Mockingbirds, and Lesser Antillean Tanagers were recorded. These species are common, vocal, visually distinctive, and usually seen as well as heard. This demonstrates that data collected by novice observers can be useful for long-term monitoring of certain common species. This contribution is not trivial; for example, Lesser Antillean Tanagers are near-endemics, but few studies have documented their distributions or populations trends, and more research is needed. We note that while novice observers detected as many easy-to-identify individuals as intermediate or experienced observers, on average, (Fig. 2;

McLaren and Cadman 1996, Johnston *et al.* 2018), there was still a positive correlation between experience and number of individuals recorded overall (Fig. 3A–3C; Kelling *et al.* 2015). Thus, decisions about the contributions of novices to data used to assess population trends should be made on a species-by-species basis. We found that more skilled observers identified more Grenada House Wrens and more difficult-to-detect species in general. This is not surprising, as the Grenada House Wren is brown and cryptic but has a unique song and is much more commonly detected by ear than by sight. Therefore, it is important to have observer pairs with at least one observer who is skilled at aural detections when surveying Grenada House Wrens, and this is likely true for many other species as well. Novice observers are also more likely to confuse rare with common species (Farmer *et al.* 2012), which could falsely inflate detection probability of easy-to-detect species while decreasing detection of difficult-to-detect species. Thus, data collected by novice observers should be applied to monitoring analyses with caution; this also emphasizes the importance of associating some measure of observer experience with each observer submitting data.

Surprisingly, there was a significant interaction between skill level of the primary and secondary observer when surveying easy-to-detect species. When paired with more experienced primary observers, experienced secondary observers detected fewer additional individuals not detected by the primary observer, as expected. However, less skilled secondary observers detected more individuals when paired with more skilled primary observers (Fig. 3D). We know of no previous studies that have assessed the contributions of secondary observers relative to their experience, so we consider this result important to explore further. There are a number of plausible social-psychological mechanisms that may explain these results. Our data are consistent with results predicted by *equality bias*, in which people behave as if they are as skilled as their partner, and which is particularly apparent when the skill divide between partners is large (Pareschi *et al.* 2017). This suggests that novice observers might report more detections in an effort to keep pace with their expert partner. Alternatively, novice secondary observers may experience *observer-expectancy bias* (Balph and Balph 1983), in which they equate an unknown stimulus with what they expect they should see or hear. If observers cannot identify an individual, they may identify it as another species they are familiar with. Either mechanism suggests that our estimates of the detection probability of easy-to-detect species may be higher for expert observers than our results showed, because novice secondary observers may be adding false positive observations that bias the expert's probability downwards. This theory would help to reconcile why our results show that novice and expert observers recorded similar proportions of easy-to-detect species (Fig. 2), but expert primary observers still detected more easy-to-detect individuals than less skilled primary observers (Fig. 2). We suggest that further research into behavior of citizen science participants using a social psychology lens would be informative for improving the design of monitoring programs.

For many reasons, we argue that including non-expert observers in regions that are building monitoring capacity is extremely important. Detectability of easy-to-detect species increased significantly with the addition of a novice partner relative to detectability by an intermediate or expert observer surveying alone, demonstrating that novice participants improve the quality of data collected despite their lack of experience. Research also shows that participants in citizen science programs gain expertise in a variety of skills and begin to diffuse knowledge and skills through an 'environmental advocacy network' (Johnson *et al.* 2014), resulting in benefits to both the individual and to the community. Citizen scientists have a range of different motivations for participation: to gain knowledge and skills, to meet like-minded people, to improve the areas where that they live and recreate (Bruyere and Rappe 2007), and to contribute to evidence-based governance (Greenwood 2007). Finding ways for this environmental advocacy network to grow requires inclusion of new and inexperienced citizen scientists and finding ways for them to contribute and learn. Further, in Grenada and other countries with relatively low avian diversity, it is possible for volunteers to develop from being a novice to an intermediate observer relatively quickly. Many paths for developing ecological skills and knowledge for new participants could be explored, such as shadowing more experienced observers, apprenticeships with local birdwatching tourism operators, and involvement in local conservation organizations or university programs.

Involving novice observers in monitoring programs does not come without challenges. As primary observers, novices generally recorded less than half of the difficult-to-detect individuals, and as secondary observers, they may have influenced the apparent detection probability of others through a combination of non-detections and false positive detections. Intermediate observers were much more consistent in their detections. As such, we should ask ourselves how to best incorporate observations from novice participants into data analyses. One option would be to design monitoring schemes that enable novices to build skills while contributing limited but accurate data on species detections/counts. This could mean limiting the target species to those that are easily identifiable or using the data only from the primary observer so that the novice can shadow more skilled observers without influencing the results. Ultimately, the decision to fully integrate novice observers into monitoring programs must be approached with caution and is dependent on the short and long-term goals of the monitoring project, as well as the larger social context. For example, short-term research requiring accurate abundance estimates of difficult-to-detect species (e.g., assessing impacts of recent hurricanes) requires great caution in using data collected by novice observers, and rare species monitoring may require additional bespoke approaches. In contrast, the benefits accrued by incorporating novice volunteers into long-term double-observer survey programs will likely outweigh the short-term problems of biased abundance estimates from temporarily novice participants.

We make the following recommendations for landbird surveys in Grenada and other Caribbean landbird communities. Surveys should: (1) implement point counts rather than transects, (2) be conducted during the local peak breeding period (in Grenada, the rainy season), (3) be conducted within 2 hrs after sunrise or 2 hrs before sunset, (4) be conducted in pairs using standardized double-observer protocols and where one member of each pair has at least intermediate experience, (5) include in-person, in-situ training, (6) collect a measure of observer experience that can later be implemented in statistical analyses, and (7) include both English and local names on data collection sheets, to increase accessibility and out of respect for local knowledge of island ecosystems. We note that these recommendations are in addition to other factors that must be considered in designing monitoring programs, such as the need to ensure reproducibility among years, ensure independence among sampling sites (or model its impact on results if necessary), and record (and then model) impacts of factors that can influence both abundance and detectability, such as habitat type, distance to road, number of observers, season, time of day, and observer identity (Wunderle 1994). Our results show that there are some species that can be surveyed consistently and predictably with volunteers, even those with little training, and therefore these species can be surveyed over time by many different observers to provide reasonable indices of population trends. Survey programs such as these will result not just in data to contribute towards informed decision-making and wildlife monitoring but will concurrently increase local capacity and environmental engagement by helping participants to gain experience and skills.

Acknowledgments

We are grateful to everyone in Grenada who played a role in this project. We thank Claudette David for providing housing and helping us liaise with St. Patrick's Environmental and Community Tourism Organization (SPECTO). We thank all of the volunteers who contributed towards data collection, in particular Valdon Paul and Theodore Alexander, who participated in more surveys than any other volunteers. Miya Warrington, Leon Radix, and members of Education Conservation Outreach (ECO) at Saint George's University (SGU) provided additional support. We thank research interns Quincy Augustine and Jenese Roberts from SGU. We offer a special thanks to Shadel Nyack Compton and the Belmont Estate family, who have shown tremendous generosity over the course of this research and contributed to all of our research activities, including extensive access to lands on which to conduct this research. This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada and the Queen Elizabeth II Diamond Jubilee Scholarship Program (Association of Universities and Colleges of Canada, The Rideau Hall Foundation, and Community Foundations of Canada). Author Contributions: NB and NK conceived the idea and experiment (supervised research, formulated question or hypothesis); NB and NK developed or designed methods; NB collected and analyzed data; NB and CD wrote the paper (or substantially edited the paper) and designed figures; NK contributed substantial materials, resources, and funding.

Literature Cited

Acharya, B.K., B. Chettri, and L. Vijayan. 2009. Indigenous knowledge of Lepcha community for monitoring and conservation of birds. Indian Journal of Traditional Knowledge 8:65–69.

Alldredge, M.W., K. Pacifici, T.R. Simons, and K.H. Pollock. 2008. A novel field evaluation of the effectiveness of distance and independent observer sampling to estimate aural avian detection probabilities. Journal of Applied Ecology 45:1349–1356.

Alldredge, M.W., T.R. Simons, and K.H. Pollock. 2007. A field evaluation of distance measurement error in auditory avian point count surveys. The Journal of Wildlife Management 71:2759–2766.

Anderson, B.W., and R.D. Ohmart. 1981. Comparisons of avian census results using variable distance transect and variable circular plot techniques. Pp. 186–192 *in* Estimating numbers of terrestrial birds (C.J. Ralph and J.M. Scott, eds.). Studies in Avian Biology 6. Allen Press Inc., Lawrence, KS.

Anderson, D.R. 2001. The need to get the basics right in wildlife field studies. Wildlife Society Bulletin 29:1294–1297.

Balph, D.F., and M.H. Balph. 1983. On the psychology of watching birds: the problem of observer-expectancy bias. Auk 100:755–757.

Blockstein, D.E. 1991. Population declines of the endangered endemic birds on Grenada, West Indies. Bird Conservation International 1:83–91.

Bolker, B. 2019a. GLMM FAQ. bbolker.github.io/mixedmodels-misc/glmmFAQ.html.

Bolker, B. 2019b. GLMM worked examples. ms.mcmaster.ca/~bolker/R/misc/foxchapter/bolker_chap.html.

Bruyere, B., and S. Rappe. 2007. Identifying the motivations of environmental volunteers. Journal of Environmental Planning and Management 50:503–516.

Buckland, S.T. 2006. Point-transect survey for songbirds: robust methodologies. Auk 123:345–357.

Burnham, K.P., and D.R. Anderson. 2004. Multimodel inference: understanding AIC and BIC in model selection. Sociological Methods and Research 33:261–304.

Cooper, C.B., J. Shirk, and B. Zuckerberg. 2014. The invisible prevalence of citizen science in global research: migratory birds and climate change. PLoS ONE 9:e106508.

Cumming, G.S., and D.A.W. Henry. 2019. Point counts outperform line transects when sampling birds along routes in South African protected areas. African Zoology 54:187–198.

Dawson, D.K., and M.G. Efford. 2009. Bird population density estimated from acoustic signals. Journal of Applied Ecology 46:1201–1209.

Diefenbach, D.R., D.W. Brauning, and J.A. Mattice. 2003. Variability in grassland bird counts related to observer differences and species detection rates. Auk 120:1168–1179.

Diefenbach, D.R., M.R. Marshall, J.A. Mattice, and D.W. Brauning. 2007. Incorporating availability for detection in estimates of bird abundance. Auk 124:96–106.

Dobkin, D.S., and A.C. Rich. 1998. Comparison of line-transect, spot-map, and point-count surveys for birds in riparian habitats of the great basin. Journal of Field Ornithology 69:430–443.

Emlen, J.T., and M.J. Dejong. 1992. Counting birds: the problem of variable hearing abilities. Journal of Field Ornithology 63:26–31.

Farmer, R.G. 2012. Observer error in citizen ornithology. Ph.D. Thesis. Dalhousie University, Halifax, NS, Canada.

Farmer, R.G., M.L. Leonard, J.E.M. Flemming, and S.C. Anderson. 2014. Observer aging and long-term avian survey data quality. Ecology and Evolution 4:2563–2576.

Farmer, R.G., M.L. Leonard, and A.G. Horn. 2012. Observer effects and avian-call-count survey quality: rare-species biases and overconfidence. Auk 129:76–86.

Farnsworth, G.L., K.H. Pollock, J.D. Nichols, T.R. Simons, J.E. Hines, and J.R. Sauer. 2002. A removal model for estimating detection probabilities from point-count surveys. Auk 119:414–425.

Gallo, T., and D. Waitt. 2011. Creating a successful citizen science model to detect and report invasive species. BioScience 61:459–465.

Gardiner, M.M., L.L. Allee, P.M.J. Brown, J.E. Losey, H.E. Roy, and R.R. Smyth. 2012. Lessons from lady beetles: accuracy of monitoring data from US and UK citizen-science programs. Frontiers in Ecology and the Environment 10:471–476.

Golding, J.D., and V.J. Dreitz. 2016. Comparison of removal-based methods for estimating abundance of five species of prairie songbirds. Journal of Field Ornithology 87:417–426.

Greenwood, J.J.D. 2007. Citizens, science and bird conservation. Journal of Ornithology 148:77–124.

Johnson, D. 2008. In defense of indices: the case of bird surveys. Journal of Wildlife Management 72:857–868.

Johnson, M.F., C. Hannah, L. Acton, R. Popovici, K.K. Karanth, and E. Weinthal. 2014. Network environmentalism: citizen scientists as agents for environmental advocacy. Global Environmental Change 29:235–245.

Johnston, A., D. Fink, W.M. Hochachka, and S. Kelling. 2018. Estimates of observer expertise improve species distributions from citizen science data. Methods in Ecology and Evolution 9:88–97.

Keller, C.M.E., and M.R. Fuller. 1995. Comparison of birds detected from roadside and off-road point counts in the Shenandoah National Park. Pp. 111–116 *in* Monitoring Bird Populations by Point Counts (C.J. Ralph, J.R. Sauer, and S. Droege, eds.). USDA Forest Service General Technical Report no. PSW-GTR-149. Pacific Southwest Research Station, USDA Forest Service, Albany, CA.

Keller, C.M.E., and J.T. Scallan. 1999. Potential roadside biases due to habitat changes along breeding bird survey routes. Condor 101:50–57.

Kelling, S., A. Johnston, W.M. Hochachka, M. Iliff, D. Fink, J. Gerbracht, C. Lagoze, F.A. La Sorte, T. Moore, A. Wiggins, W.-K. Wong, C. Wood, and J. Yu. 2015. Can observation skills of citizen scientists be estimated using species accumulation curves? PloS ONE 10:e0139600.

Kendall, W.L., B.G. Peterjohn, and J.R. Sauer. 1996. First-time observer effects in the North American breeding bird survey. Auk 113:823–829.

Lack, D., and A. Lack. 1973. Birds on Grenada. Ibis 115:53–59.

McCaffrey, R.E. 2015. Using citizen science in urban bird studies. Urban Habitats 3:70–86.

McLaren, M.A., and M.D. Cadman. 1996. Can novice volunteers provide credible data for bird surveys requiring song identification? Journal of Field Ornithology 70:481–490.

Miller, D.A.W., L.A. Weir, B.T. McClintock, E.H. Campbell Grant, L.L. Bailey, and T.R. Simons. 2012. Experimental investigation of false positive errors in auditory species occurrence surveys. Ecological Applications 22:1665–1674.

Moyer-Horner, L., M.M. Smith, and J. Belt. 2012. Citizen science and observer variability during American pika surveys. Journal of Wildlife Management 76:1472–1479.

Ng, C.S., J.R. Duncan, and N. Koper. 2018. Who's "hooting"? Motivations and scientific attitudes of Manitoban citizen science owl surveyors. Avian Conservation and Ecology 13:9.

Nichols, J.D., J.E. Hines, J.R. Sauer, F.W. Fallon, J.E. Fallon, and P.J. Heglund. 2000. A double-observer approach for estimating detection probability and abundance from avian point counts. Auk 117:393–408.

Nichols, J.D., L. Thomas, and P.B. Conn. 2009. Inferences about landbird abundance from count data: recent advances and future directions. Pp. 201–235 *in* Modeling Demographic Processes in Marked Populations (G.P. Patil, ed.). Springer, New York.

Pareschi, L., P. Vellucci, and M. Zanella. 2017. Kinetic models of collective decision-making in the presence of equality bias. Physica A: Statistical Mechanics and its Applications 467:201–217.

Ralph, C.J., and J.M. Scott. 1981. Estimating numbers of terrestrial birds. Studies in Avian Biology 6:1–630.

Ratkowsky, A., and D. Ratkowsky. 1979. A comparison of counting methods to obtain bird species numbers. Notornis 26:53–61.

Ricklefs, R., and E. Bermingham. 2008. The West Indies as a laboratory of biogeography and evolution. Philosophical Transactions of the Royal Society B: Biological Sciences 363:2393–2413.

Rosenstock, S.S., D.R. Anderson, K.M. Giesen, T. Leukering, and M.F. Carter. 2002. Landbird counting techniques: current practices and an alternative. Auk 119:46–53.

Rusk, B.L. 2017. Long-term population monitoring of the critically endangered Grenada Dove (*Leptotila wellsi*) on Grenada, West Indies. Journal of Caribbean Ornithology 30:49–56.

Shankar Raman, T.R. 2003. Assessment of census techniques for interspecific comparisons of tropical rainforest bird densities: a field evaluation in the Western Ghats, India. Ibis 145:9–21.

Silvertown, J., C.D. Buesching, S.K. Jacobson, and T. Rebelo. 2013. Citizen science and nature conservation. Pp. 127–142 *in* Key Topics in Conservation Biology 2 (D.W. Macdonald and K.J. Willis, eds.). Wiley-Blackwell, West Sussex, UK.

Simons, T.R., M.W. Alldredge, K.H. Pollock, and J.M. Wettroth. 2007. Experimental analysis of the auditory detection process on avian point counts. Auk 124:986–999.

Simons, T.R., K.H. Pollock, J.M. Wettroth, M.W. Alldredge, K. Pacifici, and J. Brewster. 2009. Sources of measurement error, misclassification error, and bias in auditory avian point count data. Pp. 237–254 *in* Modeling Demographic Processes in Marked Populations (T.R. Simons, ed.). Springer, Boston, MA.

Thomas, L., S.T. Buckland, K.P. Burnham, D.R. Anderson, J.L. Laake, D.L. Borchers, and S. Strindberg. 2013. Distance Sampling. Pp. 687–697 *in* Encyclopedia of Environmetrics. Wiley and Sons, New York.

Tulloch, A.I.T., H.P. Possingham, L.N. Joseph, J. Szabo, and T.G. Martin. 2013. Realising the full potential of citizen science monitoring programs. Biological Conservation 165:128–138.

Verner, J., and L.V. Ritter. 1988. A comparison of transects and spot mapping in oak-pine woodlands of California. Condor 90:401–419.

Waide, R.B., and P.M. Narins. 1988. Tropical forest bird counts and the effect of sound attenuation. Auk 105:296–302.

Watts, N.S.J., and G. Wandesforde-Smith. 2006. The law and policy of biodiversity conservation in the Caribbean: cutting a gordian knot. Journal of International Wildlife Law and Policy 9:209–221.

West, S., A. Dyke, and R. Pateman. 2021. Variations in the motivations of environmental citizen scientists. Citizen Science: Theory and Practice 6:1–18.

Williams, R.J.T. 2020. Distribution, diversity, abundance, and richness of Grenadian terrestrial birds, including endemic and restricted-range species. Master of Natural Resources Management Thesis. University of Manitoba, Winnipeg, MB, Canada.

Wilson, R.R., D.J. Twedt, and A.B. Elliott. 2000. Comparison of line transects and point counts for monitoring spring migration in forested wetlands. Journal of Field Ornithology 71:345–355.

Wunderle J.M., Jr., 1994. Census methods for Caribbean land birds. USDA Forest Service General Technical Report no. SO-98. Southern Forest Experiment Station, USDA Forest Service, New Orleans, LA.

Wunderle J.M., Jr., 1985. An ecological comparison of the avifaunas of Grenada and Tobago, West Indies. Wilson Bulletin 97:356–365.

Zuur, A.F., E.N. Ieno, N.J. Walker, A.A. Saveliev, and G.M. Smith. 2009. Mixed effects models and extensions in ecology with R. Journal of Statistical Software 32(b01).

Appendix 1. Identification tests and easy- and difficult-to-detect species categories

Species	Test	Detection
Scaly-naped Pigeon (*Patagioenas squamosa*)		easy
Common Ground Dove (*Columbina passerina*)	visual, aural	easy
Ruddy Quail Dove (*Geotrygon montana*)		difficult
Grenada Dove (*Leptotila wellsi*)		difficult
Zenaida Dove (*Zenaida aurita*)		easy
Eared Dove (*Zenaida auriculata*)	visual, aural	easy
Smooth-billed Ani (*Crotophaga ani*)	visual	easy
Mangrove Cuckoo (*Coccyzus minor*)		easy
Gray-rumped Swift (*Chaetura cinereiventris*)		difficult
Rufous-breasted Hermit (*Glaucis hirsutus*)		difficult
Green-throated Carib (*Eulampis holosericeus*)		difficult
Antillean Crested Hummingbird (*Orthorhyncus cristatus*)	visual	easy
Grenada Hook-billed Kite (*Chondrohierax uncinatus mirus*)		difficult
Broad-winged Hawk (*Buteo platypterus*)	visual, aural	easy
Barn Owl (*Tyto alba*)		difficult
Orange-winged Parrot (*Amazona amazonica*)		easy
Yellow-bellied Elaenia (*Elaenia flavogaster*)	visual, aural	easy
Caribbean Elaenia (*Elaenia martinica*)		difficult
Grenada Flycatcher (*Myiarchus nugator*)	visual, aural	easy
Gray Kingbird (*Tyrannus dominicensis*)	visual, aural	easy
Black-whiskered Vireo (*Vireo altiloquus*)		difficult
Caribbean Martin (*Progne dominicensis*)		difficult
Barn Swallow (*Hirundo rustica*)		difficult
Grenada House Wren (*Troglodytes aedon grenadensis*)	visual, aural	difficult
Tropical Mockingbird (*Mimus gilvus*)	visual, aural	easy
Cocoa Thrush (*Turdus fumigatus*)		difficult
Spectacled Thrush (*Turdus nudigenis*)	visual, aural	easy
Shiny Cowbird (*Molothrus bonariensis*)		difficult
Carib Grackle (*Quiscalus lugubris*)	visual	easy
Lesser Antillean Tanager (*Stilpnia cucullata*)	visual, aural	easy
Blue-black Grassquit (*Volatinia jacarina*)		difficult
Yellow-bellied Seedeater (*Sporophila nigricollis*)		difficult
Bananaquit (*Coereba flaveola*)	visual, aural	easy
Lesser Antillean Bullfinch (*Loxigilla noctis*)	visual, aural	difficult
Black-faced Grassquit (*Melanospiza bicolor*)	visual, aural	difficult

Breakdown of easy-to-detect and difficult-to-detect species

Species groups refer to the *a priori* pooling of two or more species that are assumed to have similar or identical detection probability (Nichols *et al.* 2000). Because we included non-expert observers in this analysis, the biggest factors influencing detectability were observer familiarity with particular species and defining characteristics that make birds easily identifiable. This differs slightly from other studies that grouped species based primarily upon how singing behavior impacts ease of detection (e.g., Leston *et al.* 2015, Nichols *et al.* 2000). Similar to the grouping by Nichols *et al.* (2000), our groups were determined by how easily the average observer was able to detect a species. Easy-to-detect species were defined as those both common to Grenada and that have characteristics that make them easy to identify. Difficult-to-detect species were less common or have characteristics that make them more difficult to identify. This could include songs that are similar to other species or species that are easily misidentified. Because there was no abundance data for Grenadian landbirds available, it was difficult to quantify rarity before the study began. Our *a priori* groupings were based upon initial surveys in Grenada, and data from training activities that demonstrated that observers struggled with particular species. For example, the Lesser Antillean Bullfinch is relatively common and males have distinctive markings, but under field conditions many volunteers were unable to differentiate this species from other, similar-sized birds, like Black–faced Grassquits or Bananaquits, so it was classified as a difficult-to-detect species. Because it is difficult to estimate detection probabilities of birds with only a few observations, species that had fewer than 10 observations were grouped together according to the species group that they had been assigned to *a priori*.

Appendix 2. DOBSERV Modelling

DOBSERV compares six models derived from the Cook and Jacobson maximum likelihood estimator (Cook and Jacobson 1979):

$$\hat{p}_1 = \frac{x_{11}\,x_{22} - x_{12}\,x_{21}}{x_{11}\,x_{22} + x_{22}\,x_{21}} \qquad \hat{p}_2 = \frac{x_{11}\,x_{22} - x_{12}\,x_{21}}{x_{11}\,x_{22} + x_{11}\,x_{12}} \qquad \hat{p} = 1 - \frac{x_{12}\,x_{21}}{x_{22}x_{11}}$$

Where $\hat{p}_I$ is defined as the detection probability for observer i ($i = 1,2$) and $\hat{p}$ is the probability that at least one observer makes the detection, called joint detection probability. x_{ij} is defined as the number of individuals counted by observer i ($i = 1,2$) when observer j ($j = 1,2$) was primary observer. The most general model (p_{is}) uses the following product binomial multiplied together over all species (Nichols *et al.* 2000):

$$B\,(x_{11}{}^{s} + x_{21}{}^{s},\ p_1{}^{s} / p^{s})\ B\,(x_{22}{}^{s} + x_{12}{}^{s},\ p_1{}^{s} / p^{s})$$

The competing models follow the same form but constrain certain parameters to equal each other. Here the "*s*" superscript indicates the x_{ij} values for each species ($s = 1, n$). The null model, P(.,.) constrains detection probability to be the same regardless of species or observer ($p_1 = p_2$ and $s_1 = s_2 = s_n$). P(S,.) allows detection probability to differ for each species, but not between observers ($p_1 = p_2$ and $s_1 \neq s_2 \neq s_n$). P(G,.) allows detection probability to differ for each species group, but not between observers. In this case $p_1 = p_2$, $s_1 = s_2$ and $s_3 = s_4$ but $s_1 \neq s_3$. P(.,I) constrains detection probability to be the same for every species, but different between the two observers ($p_1 \neq p_1$ and $s_1 = s_2 = s_n$). Finally, P(G,I) allows detection probability to differ among species groups and between observers ($p_1 \neq p_2$, $s_1 = s_2$ and $s_3 = s_4$ but $s_1 \neq s_3$).

Models were ranked by Akaike's Information Criterion for a small sample size (AIC$_c$) (Burnham and Anderson 2004). The models estimated detection probabilities that are specific to each observer (in the case of this study, each skill level) and joint detection probabilities, defined as the probability that at least one of the two observers in a pair will observe any given individual. Joint detection probability is defined as:

$$p = 1 - (1 - p_a) * (1 - p_b)$$

where p_a is the detection probability of one observer and p_b is the detection probability of the other. If species group or species were included in the best model, detection probability was estimated for each group or species.

Appendix 3. Influence of observer skill and count method on observed abundance for easy-to-detect species at fixed and unlimited radius modelled with a negative binomial GLMM. All observations were made by the primary observer and exclude any results from the secondary observer. Fixed effects are and all base levels are fully explained in the methods. Season is compared to the first field season in April-May, time of day is compared to early morning counts and point count is compared to the base level, transects. Incidence rate ratios are the ratio at which birds are expected to be observed relative to the reference level. † denotes significance at the 0.05 α level. CI = confidence interval, *p* = *p*-value, σ2 = sum of random effects variance, τ_{00} = random intercept variance (between groups/subjects variance), ICC = interclass correlation coefficient (random effect variance/sum of random effects variance).

	Easy-to-Detect Species Abundance, Fixed Radius			**Easy-to-Detect Species, Unlimited Radius**		
Predictors	**Incidence Rate Ratios**	**CI**	***p***	**Incidence Rate Ratios**	**CI**	***p***
Intercept	0.00	0.00–0.00	<0.001†	0.63	0.44–0.89	0.008†
Primary Observer Skill Index	1.02	1.00–1.04	0.036†	1.03	1.02–1.04	<0.001†
October-November	0.81	0.62–1.06	0.124	0.94	0.77–1.14	0.522
January-February	0.65	0.52–0.81	<0.001†	0.74	0.63–0.87	<0.001†
Late Morning	0.93	0.77–1.11	0.419	0.90	0.78–1.03	0.129
Evening	1.11	0.90–1.37	0.350	1.20	1.04–1.39	0.014
Point Count	1.44	1.24–1.67	<0.001†	1.28	1.15–1.42	<0.001†
Random Effects						
σ^2	0.37	0.21				
τ_{00}	0.02 $_{\text{Location}}$	0.02 $_{\text{Location}}$				
ICC	0.04 $_{\text{Location}}$	0.07 $_{\text{Location}}$				
Observations	353	353				
Marginal R^2 / Conditional R^2	0.161 / 0.195	0.202 / 0.261				

Appendix 4. Influence of observer skill and count method on observed abundance for difficult-to-detect species at fixed and unlimited radius modelled with a negative binomial GLMM. This includes only the observations made by the primary observer. Season is compared to the first field season in April/May, time of day is compared to early morning counts and point count is compared to the base level, transects. Incidence rate ratios are the ratio at which birds are expected to be observed relative to the reference level. † denotes significance at the 0.05 α level.

	Easy-to-Detect Species Abundance, Fixed Radius			**Easy-to-Detect Species, Unlimited Radius**		
Predictors	**Incidence Rate Ratios**	**CI**	***p***	**Incidence Rate Ratios**	**CI**	***p***
Intercept	0.00	0.00–0.00	<0.001†	0.15	0.07–0.30	<0.001†
Primary Observer Skill Index	1.03	1.00–1.06	0.023†	1.04	1.02–1.07	0.001†
October/November	0.71	0.48–1.07	0.101	0.68	0.46–0.99	0.046†
January/February	0.64	0.45–0.92	0.015†	0.71	0.51–0.99	0.041†
Late Morning	0.89	0.68–1.18	0.426	0.80	0.62–1.04	0.093
Evening	0.98	0.70–1.38	0.927	0.94	0.69–1.27	0.679
Point Count	1.51	1.20–1.88	<0.001†	1.31	1.07–1.62	0.010†
Random Effects						
σ^2	0.69	0.61				
τ_{00}	0.32 $_{\text{Location}}$	0.20 $_{\text{Location}}$				
ICC	0.32 $_{\text{Location}}$	0.24 $_{\text{Location}}$				
Observations	353	353				
Marginal R^2 / Conditional R^2	0.086 / 0.378	0.091 / 0.311				