Optimizing inventories of diverse sites: insights from Barro Colorado Island birds

David M. Watson*

Institute for Land, Water and Society, Charles Sturt University, PO Box 789, Albury, NSW 2640, Australia

Summary

1. Diverse sites have long-attracted ecologists, yet the overwhelming variety of species can confound attempts to enumerate species richness. Various predictive methods estimate species richness by comparing the rate at which species are first detected with the rate at which they are detected again, yielding richness estimates of known precision without exhaustive sampling.

2. While frequently used for arthropods, predictive methods are rarely applied to vertebrate surveys where species identity is often a priority. Expressing observed richness as a function of estimated richness, an estimate of survey completeness can be derived, offering the potential for inventories of standardized precision for comparison and further analysis.

3. To realize this potential, I conducted 402 h of bird surveys on Barro Colorado Island (Panama) and performed a series of retrospective analyses to address three questions: (i) How much effort is required to achieve complete inventories (maximum completeness)? (ii) What is the least amount of effort required to yield robust richness estimates (maximum efficiency)? and (iii) How much effort is required to optimize sampling, balancing completeness and efficiency?

4. Whereas the richness estimate for all species required thirty 6-h samples to attain maximum completeness, once migrants, waterbirds and non-forest-dependent species were excluded, the richness of forest-dependent residents could be estimated to the same precision with fifteen samples and to 80% completeness with four samples.

5. Of the 186 bird species detected, 70 represented unique or duplicate records, seen in only one or two sampling periods. These low detectability species were dominated by migrants (28) and raptors (14) and also included seven waterbirds, five nocturnal species and four aerial foragers, justifying the widespread practice of excluding these groups from surveys of forest assemblages.

6. In addition to demonstrating the reliability of predictive approaches, this study demonstrates the practicality of results-based stopping rules for sampling diverse sites, especially for targeted groups of species. Combining predictive methods with targeted sampling represents an efficient and rigorous design, increasing the number of sites that can be sampled and enhancing the overall power and value of the study.

Key-words: Barro Colorado island, completeness, rainforest, richness estimation, standardized search, stopping rule, turnover

Introduction

The starting point for most ecological studies of communities and assemblages are inventories – lists of species recorded at particular locations. While it may seem straightforward to establish which species are present and which are absent, a multitude of confounding factors can reduce the accuracy of biological inventories and compromise the validity of associated inferences (Hawkesworth *et al.* 1995). In addition to differences between observers, sampling methods, habitats and times-of-day, species exhibit a wide range of detectabilities that, in turn, vary with season, habitat structure and climate *inter alia* (Boulinier *et al.* 1998; Gotelli & Colwell 2001). Rather than just affecting whether or not particular species are detected, these factors also confound abundance measures (Verner 1985; Buckland *et al.* 2004) and increase sampling efforts required to attain reliable diversity estimates or descriptions of the community.

While increased sampling effort reduces the chances of missed species, more sampling does not necessarily yield more accurate inventories. With greater sampling effort per site, the likelihood that species will be detected that were not initially present increases (oversampling, *sensu* Lobo & Martín-Piera

*Correspondence author. E-mail: dwatson@csu.edu.au
Correspondence site: http://www.respond2articles.com/MEE/

2002) leading to inflated alpha diversity estimates, especially in dynamic systems with high beta diversity. Furthermore, as sampling effort for each inventory increases, the number of sites that can be surveyed within a given time period decreases, diminishing the explanatory power and broader relevance of the study. Optimal sampling balances these opposing factors: allocating sufficient effort to ensure inventories are accurate while minimizing sampling effort per site to avoid inflation and maximize the number of sites sampled. Achieving this balance becomes more difficult as alpha and beta diversities increase, presenting numerous challenges to ecologists studying diverse systems. Most inventories from tropical regions and other areas with very high alpha diversity are either exhaustive single-site studies of high accuracy (e.g. Longino, Coddington, & Colwell 2002) or preliminary lists from multiple sites of variable precision and unknown accuracy (e.g. Oliver & Beattie 1996). Neither approach generates the consistently accurate inventory data from multiple sites required to undertake meaningful ecological analysis (e.g. Hortal, Garcia-Pereira, & García-Barros 2004), limiting our understanding of these systems and constraining our ability to manage them effectively.

A key innovation for compiling inventories of diverse sites was the application of accumulation-based methods – using the rate at which species are detected to estimate the number of species present, but overlooked (Colwell & Coddington 1994). As well as yielding predictions of total species richness (Soberón & Llorente 1993), these methods have also been used to determine when sufficient sampling effort has been expended (i.e. results-based stopping rules; Peterson & Slade 1998; Scharff *et al.* 2003; Hortal, Garcia-Pereira, & García-Barros 2004). Pioneered with arthropods (Longino, Coddington, & Colwell 2002; Scharff *et al.* 2003; Hortal, Garcia-Pereira, & García-Barros 2004; Hortal & Lobo 2005) and useful for a wide range of other groups, these predictive approaches have proven less popular among vertebrate ecologists, for whom *which* species is of greater interest than *how many* species. A solution to this impasse was devised by Peterson & Slade (1998): dividing observed richness by predicted richness provides a robust estimate of inventory completeness, quantifying how well the list of recorded species represents total richness. In addition to yielding inventories with associated confidence intervals, inventories can be standardized to specified levels of completeness, maximizing efficiency and maintaining consistent accuracy (Watson 2003, 2004), to ensure comparability within and between studies.

In previous work, I used iterative completeness estimates to demonstrate the utility of results-based stopping rules for standardizing inventories of medium-diversity bird assemblages (15–40 spp.) in temperate woodlands of southern Australia (Watson 2003, 2004). Here, I evaluate the applicability of these methods to high diversity sites, using Barro Colorado Island as the test arena. With a well-studied avifauna of over 200 species (Robinson 2001), this rainforest remnant represents an ideal system to explore trade-offs between inventory completeness and sampling efficiency. Rather than comparing multiple estimators or different sampling techniques (see Walther & Moore 2005), the primary aim of this study was to determine how much sampling effort was required to generate meaningful richness estimates. Can accurate inventories with associated completeness estimates be collected in high diversity sites without exhaustive sampling and, if so, how robust are the estimates? Using two 96-h land-based surveys followed by a 210-h survey from the surrounding lake, I conducted retrospective analyses to assess how many samples were required to estimate the richness of selected groups to pre-determined levels of completeness. Although restricted to one taxonomic group on a small island sampled by one observer for a single season, this data set is used to develop a series of recommendations for collecting inventory data in a range of settings. As the most exhaustive bird survey conducted in this reserve in recent decades, it provides a useful opportunity to evaluate differences with previous inventories of the island and discuss recent colonists, local extinctions and other changes. These comparisons are briefly mentioned, but a more thorough treatment of ongoing changes to community composition and a community-level analysis of determinants of detectability will be presented elsewhere. Having determined the sampling efforts required to generate reliable richness estimates, quantitative stopping rules (based on completeness estimates) are then defined to yield inventories of maximum completeness, maximum efficiency and a compromise between the two (optimal sampling).

Materials and methods

STUDY SITE

I collected all data presented here during a 5-month period (August–December 2006) on Barro Colorado Island, central Panama. Comprising 1562 ha of lowland rainforest isolated in 1914 during construction of the Panama Canal, the island is sufficiently small and homogeneous to be sampled as a single unit. Comprehensive terrestrial inventories were conducted 15 years after isolation of the island (Chapman 1929), with successive research over the past 80 years making it the most studied rainforest remnant world-wide (Leigh, Rand, & Windsor 1982). Birds were a particular focus of this early work (Eisenmann 1952) and, despite diminishing ornithological research in recent decades, the avifauna of the island is among the best studied in the tropics. Of the 366 species known from the island (Willis & Eisenmann 1979), 217 are resident, of which 173 have been recorded during the past 15 years (Robinson 1999, 2001). Being an island of closed-canopy rainforest surrounded by the open waters of Gatun Lake (intersected by a high-traffic shipping channel), Barro Colorado Island's avifauna represents a discrete and relatively stable community: while there are seasonal and stochastic movements of individuals to and from the island, there is less dynamism than in otherwise similar habitat elsewhere (Sieving & Carr 1997).

BIRD SURVEYS

Prior to commencing the inventory, a 16-day period was used to acclimatize, familiarize myself with vocalizations and refine sampling protocols. With a 35-km network of maintained trails, canopy access towers and *c*. 65 km of shoreline easily reached by boat, access is excellent and permits thorough coverage of the island. Given prior fieldwork in Mesoamerica and northern South America (*c*. 13 months, mostly in rainforest habitats), I was experienced with

the avifauna – less than 15% of the species encountered on the island were unfamiliar. In designing the sampling protocol, I sought to balance representation of microhabitats and forest strata with temporal coverage, recognizing diel variation in calling behaviour and likelihood of detection. After trial periods spanning 2–18 h, sample duration was set at 6 h, comprising 4 h of walking within the forest (on- and off-marked trails) and 2 h of stationary watches atop canopy towers and from balconies overlooking the laboratory clearing and adjacent cove. Samples commenced within 30 min of dawn and included mid-morning, late afternoon and dusk; all on the same day. Since I was living on the island throughout this period, I also included nocturnal species encountered incidentally: detections before midnight added to the sample from the preceding day, detections after midnight added to the sample for the following day. This combination enabled forest interior species, species associated with forest edges and clearings, understorey and canopy species, forest floor and aerial species, diurnal, nocturnal and crepuscular species to be encountered. Motion-triggered cameras (Faunatech, Bairnsdale, Victoria, Australia) were deployed along trails, beside creek-lines, atop fallen logs and in treefall gaps to detect *Crypturellus soui*, *Crax rubra* and other secretive ground-dwelling birds. While successful at recording several cryptic mammals, the only birds photographed using these cameras were *Coragyps atratus*, so the cameras will not be discussed further.

Having established the sampling protocol, the avifauna of Barro Colorado Island was surveyed using two separate inventories of 16 samples (i.e. 96 h each), 2 weeks apart to quantify species turnover. The first inventory (10–25 September 2006, inclusive) was conducted during the early wet season; the second (16 days with less than 3 h of heavy rain between 10 and 30 October 2006, inclusive) in the mid-wet season. These 51 days spanned the period of peak Autumn migration – both from North America to the Neotropics and from Mesoamerica to South America. During each 6 h sample, all species detected were recorded as present, whereas species seen previously but not detected were recorded as absent. This included birds flying over the island and adjacent waters, waterbirds, nocturnal species, raptors and aerial foragers – groups often excluded from terrestrial bird surveys. For subsequent analysis, some of these groups were removed *a posteriori* yielding four species groups: all species (every bird species detected); forest species (all species minus waterbirds); island residents (forest species minus migrants, after Robinson 1999); and forest dependents (island residents minus raptors, aerial foragers, nocturnal species and species restricted to clearings and shorelines; see Appendix S1, Supporting information for species classifications).

The accuracy of these inventories was evaluated using a small dingy to survey the canopy and fringing vegetation from the waters surrounding Barro Colorado Island (totalling 210 h over 35 days with less than 3 h of heavy rain between 2 November and 19 December 2006). In addition to travelling around the island repeatedly (including numerous navigable inlets and coves), 70 h of stationary watches were conducted while anchored in areas of high avian activity (e.g. adjacent to fruiting plants). Rather than a complete inventory, this shoreline survey was to validate the two island inventories, especially the richness estimates for the four species groups. Species that were previously not recorded were noted, while species previously recorded in only one sampling period (uniques) or in only two sampling periods (duplicates) were also noted.

ANALYSIS

Incidence data were analyzed with EstimateS 8·0·0 (Colwell 2006), using the Chao 2 estimator to generate richness predictions – an equation that yields robust estimates while being sufficiently simple to cal-

culate in the field with presence/absence data (Chao 1984, 1987; Herzog, Kessler, & Cahill 2002; Watson 2003):

$$S_{est} = S_{obs} + a^2/2b,$$

where S_{est} is estimated species richness; S_{obs} is the observed species richness; a is the number of unique species and b is the number of duplicate species (Chao 1987). This equation uses the ratio of species seen once to species seen twice to estimate the number of species missed altogether. With greater sampling effort, more uniques will become duplicates and the number of undetected species will decrease: once all species have been seen in two or more samples, estimated richness equals observed richness (Colwell 2006). Prior to calculating richness predictions, I quantified within- and between-inventory differences to measure the effect of observer learning and quantify species turnover between the two inventories. To remove any influence of sample order and generate variance estimates, mean predicted richness was calculated from 50 randomizations of the sample accumulation. The primary aim of these calculations was to evaluate the effect of increasing sample number on richness estimates and associated completeness values. This was achieved in three ways. First, for each data set (two 16 sample inventories individually and the combined 32 sample inventory for four species groups, i.e. 12 data sets), observed richness was divided by estimated richness (mean) to generate mean completeness estimates for the full survey (either 16 or 32 samples), expressed as a percentage. Secondly, the point at which predicted richness stabilized was identified for each data set, defined as the point when the standard deviation of the Chao 2 estimator began decreasing consistently, and the completeness estimates and associated sample number were noted. Finally, the numbers of samples required for observed richness to exceed pre-determined thresholds (70%, 80% and 90% of estimated richness) were calculated – scenarios reflecting various compromises between inventory completeness and sampling efficiency. Comparisons were conducted between inventories and across the four species groups using Tukey's *post hoc* tests (alpha set at $P = 0.05$). In addition to using the Chao 2 estimator, richness estimates were also calculated using two other incidence-based estimators: ICE and Jacknife 2 (Colwell 2006). Although both consistently predicted slightly greater richnesses, these differences declined with sample number and final estimates did not differ qualitatively from the Chao 2 estimates depicted here (Appendix S2, Supporting information).

Finally, to estimate the actual richness of birds on Barro Colorado Island during the sampling period, all analyses were repeated using the two island inventories plus additional records from the shoreline survey. These additional records were classified into three groups: species seen on only 1 day, species seen on two separate days and species seen on three or more separate days, allowing recalculation of richness estimates for the four species groups.

Results

A total of 125 bird species were detected in the first inventory, compared with 148 for the second inventory. Eighteen species were seen only in the first inventory, 22 species observed only in the second inventory (Appendix S1, Supporting information) and a total of 166 species detected overall. While most species were identified using auditory or visual cues, eggs (*Tinamus major*), window-strike casualties (*Catharus minimus*) and recent skeletal remains (*Cairina moschata*) were used as

evidence of occurrence (in all cases documented with voucher specimens or confirmed by subsequent observations).

Although the initial 16-day familiarization period was excluded, the inventory data may still have been confounded by observer learning – progressively more species detected in later samples if my acuity and familiarity with species improved with time. To quantify any learning effect, the two inventories were divided in half to yield four 8-sample inventories for which I calculated the mean number of species recorded in a single sample (Table 1). Significantly fewer species were seen in samples of the first inventory in two species groups: all species and forest species (Tukey's *post hoc* test; $P < 0.05$). Both of these groups include migrant species that are more likely to exhibit temporal dynamism and, once removed, there were no differences in the number of species detected per sample between the second inventory and the second half of the first inventory. In terms of within-inventory differences, two of eight pair-wise comparisons within inventories were significant (island residents and forest dependents, both in the first inventory; Tukey's *post hoc* test; $P < 0.05$). Neither of these two groups included migrants, suggesting that the lower numbers of detections per sample in early sampling periods resulted from learning effects (an estimated nine island residents including six forest-dependent species were present but missed in each of the first eight samples; Table 1).

To quantify species turnover within and between the two 16-day inventories, I calculated richness estimates for the same four 8-sample groups (Table 1). Mean richness estimates were significantly lower in the first inventory than the second ($P < 0.05$, Tukey's *post hoc* tests) for all groups except forest dependents, suggesting some components of the avifauna were not consistent between the two sampling periods. The fact that

these patterns were concordant across species groups indicates they were not driven by changes in the proportion of migrants and suggests changes in composition during the intervening fortnight between inventories. In terms of within-inventory differences, both halves of the second inventory were comparable for all species groups, whereas there were significant differences for all groups between the two halves of the first inventory ($P < 0.05$, Tukey's *post hoc* tests). This finding – that the lowest richness estimates for all four species groups were in the first half of the first inventory – reinforces the finding that early samples were confounded by missed species.

Given this within- and between-inventory variation, no further analyses were conducted using the data in the order in which they were collected. Rather, sample order was randomized 50 times for all sample numbers between 1 and 16, and richness estimated with the Chao 2 equation, these 50 values summarized with a mean and standard deviation for each sample number for both inventories (Table 2; comparable predictions generated from other estimators summarized in Appendix S2, Supporting information). In inventory 1, richness of all species was estimated to be 152 species after 16 samples (SD = 11·76), the 125 species observed representing 82% estimated completeness. Excluding waterbirds, 122 forest species were estimated (SD = 10·74), giving the 104 species detected a completeness value of 85%. Once migrant species were excluded, the richness of island residents was estimated to be 102 (SD = 6·78), the 91 species observed representing 89% completeness. Finally, once raptors, aerial foragers, nocturnal species and non-forest-dependent species were excluded, a richness of 80 forest-dependent species (SD = 3·78) was estimated after 16 samples, giving the 75 species detected a completeness value of 94%.

Table 1. Summary of two 16 sample inventories of the birds of Barro Colorado Island, Panama

	First inventory		Second inventory	
	Periods 1–8	Periods 9–16	Periods 1–8	Periods 9–16
$\hat{S}_{obs}$ per period (SD)				
All species	33·9 (7·16)a	43·6 (7·23)a	54·9 (8·94)b	55 (4·87)b
Forest species	29·4 (5·99)a	37·1 (6·53)a	42·1 (7·64)b	43·0 (2·56)b
Island residents	30·6 (6·69)a	39·3 (6·34)b	45·1 (7·53)b	45·88 (3·04)b
Forest dependents	25·8 (4·8)a	32·3 (6·09)b	33·8 (6·07)b	36·4 (2·45)b
Total S_{obs}				
All species	89	103	129	123
Forest species	74	91	99	96
Island residents	69	81	87	88
Forest dependents	62	66	66	72
$\hat{S}_{est}$(SD)				
All species	105·3 (8·42)a	130·8 (12·76)b	166·4 (15·24)c	172·4 (21·4)c
Forest species	83·6 (6·03)a	113·2 (11·08)b	130 (14·79)c	133·8 (19·44)c
Island residents	77 (5·39)a	97·5 (9·29)b	102·4 (8·68)c	115·6 (15·12)c
Forest dependents	69·5 (5·22)a	73·58 (5·44)b	75·1 (6·48)b	85·6 (8·74)b

$\hat{S}_{obs}$ is the mean number of species observed in a single sampling period in each half of the two inventories ($n = 8$, with associated standard deviation), whereas total S_{obs} is the total number of species recorded in all eight sampling periods and $\hat{S}_{est}$ is the mean estimated richness when applying the Chao 2 estimator to those eight samples based on 50 randomizations (with associated standard deviations). Where a, b, c denote groups that are significantly different from one another ($P < 0.05$, Tukey's *post hoc* test).

Table 2. Summary of observed and predicted species richness under different sampling strategies designed for maximum efficiency, maximum completeness and balancing both (optimal)

Data set	S_{obs}	$\acute{S}_{est}$ (SD)	C_{final} (%)	C_s, N_s	N_{70}	N_{80}	N_{90}
Single inventories (16 samples)							
All species (first)	125	152 (11·76)	82	73%, 12	4	15	>16
All species (second)	148	171 (10·74)	87	76%, 6	2	9	>16
Forest species (first)	104	122 (9·12)	85	74%, 9	3	14	>16
Forest species (second)	115	137 (11·73)	84	78%, 8	2	10	>16
Island residents (first)	91	102 (6·78)	89	79%, 10	3	11	>16
Island residents (second)	102	117 (9·03)	87	79%, 7	2	8	>16
Forest dependents (first)	75	80 (3·78)	94	83%, 8	3	5	13
Forest dependents (second)	80	93 (9·01)	86	82%, 7	2	3	>16
Combined inventories (32 samples)							
All species	166	182 (8·00)	91	73%, 8	4	13	30
Forest species	128	140 (7·22)	91	77%, 9	3	12	28
Island residents	111	117 (4·58)	95	77%, 5	3	10	21
Forest dependents	86	91 (4·49)	95	82%, 7	3	4	15
All data (32 samples + shoreline survey)							
All species	186	203 (8·67)	92	69%, 8	9	14	23
Forest species	162	172 (5·92)	94	71%, 7	7	13	22
Island residents	122	129 (4·77)	95	74%, 8	4	12	22
Forest dependents	87	88 (1·01)	99	80%, 8	3	8	20

S_{obs} is the observed number of species for each inventory; S_{est} is the mean estimated species richness predicted using the Chao 2 estimator with the standard deviation (of 50 randomizations); C_{final} is the completeness estimate derived from all samples for each inventory ($S_{obs}/S_{est} \times 100$), whereas C_s is the completeness estimate when predicted richness begins to stabilize (when the variance begins to decrease consistently), with associated sample number N_s; and N_{70}, N_{80} and N_{90} are the number of samples required to attain completeness estimates of at least 70%, 80% and 90%, respectively.

The same patterns emerged from inventory 2, with variance of richness estimates narrowing and completeness estimates of recorded species increasing as successive groups were excluded (Table 2). It was noteworthy that the number of detected species for the four groups in inventory 2 closely matched the predictions from inventory 1; in two cases (island residents and forests dependents), the values matched the predictions exactly. All four richness estimates from inventory 2 exceeded those from the first inventory, but variance estimates were also higher for all groups except all species (Table 2).

These richness estimates and completeness values all relate to the two complete 16-sample inventories but, to determine the minimum number of samples required to generate reliable richness predictions, estimates were calculated for all sample numbers, with means and standard deviations calculated from 50 randomizations to remove any effect of sample order. The point at which richness estimates stabilized was identified statistically as the sample number where the standard deviation of estimated richness first began to decrease consistently – equating to the sample number when observed richness and predicted richness began to converge (Fig. 1). For the four species groups in both inventories, this inflection point occurred consistently between 73% and 83% estimated completeness (highest values for forest-dependent species: 83% and 82%), and took between 5 and 12 samples to attain (Table 2, Fig. 1).

To determine which results-based stopping rule to use for the four species groups, thresholds corresponding to completeness estimates lower than these inflection points are unreliable. Accordingly, estimates for the number of samples required to

attain 70% completeness were deemed unreliable for all groups (as with the 80% threshold for forest dependents) and likely represent local minima tripping the stopping rule prematurely. The optimal threshold for the other three groups was 80%, with progressively fewer samples required to achieve 80% estimated completeness as waterbirds and migrants were excluded, from a high of 15 and 9 samples for all species, down to 11 and 8 samples for island residents (for inventories 1 and 2, respectively; Table 2). While necessarily greater numbers of samples were required to achieve estimates of higher completeness, the values for 90% completeness were unreliable, with only one of eight estimates (forest-dependent species in inventory 1) below the maximum of 16.

To evaluate how many samples were required to attain 90% completeness estimates, the two inventories were merged to yield a single combined inventory of 32 samples. While the total species observed increased to 166, the estimated richness for all four species groups from the combined inventory were comparable with estimates derived from inventory 2 (Table 2). Sample numbers required to achieve 90% completeness showed the same pattern as for 80%, decreasing from a high of 30 samples for all species, down to 15 for forest-dependent species (Fig. 2).

Subtracting observed richness from estimated richness yields estimates of how many species were overlooked in the combined 32 sample inventory: 16 species, comprising four waterbirds, six migrants, six island residents of which five were forest-dependent species. To verify these predictions and identify these 'missed' species, a shoreline survey was undertaken,
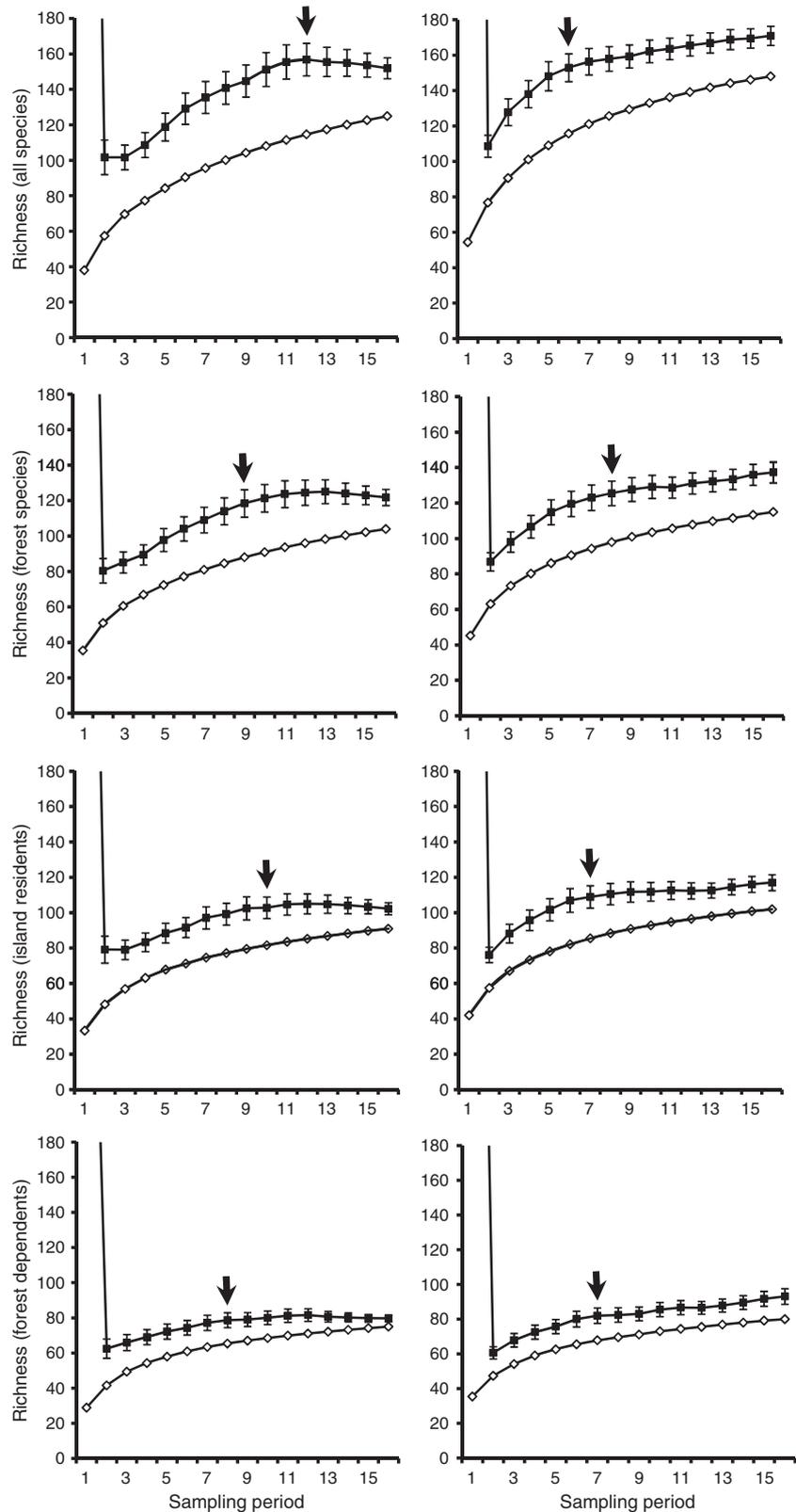
Fig. 1. Species accumulation curves for the four species groups for the first (left) and second (right) inventories (both consisting of sixteen 6-h samples). Open diamonds denote mean observed species richness, whereas closed diamonds denote mean predicted richness using the Chao 2 estimator (EstimateS 8·0·0; Colwell 2006). Error bars quantify the standard deviation of predicted richness and the arrows identify the point when the standard deviation began decreasing consistently – representing the minimum number of samples required for reliable predictions. All values were calculated from 50 randomizations of *n* samples to remove any effect of sample order. Full details of species summarized in Appendix S1 (Supporting information).

using complementary methods to sample the avifauna of the entire island. After 210 h of water-based surveys, 20 previously undetected species were observed: seven waterbirds, six migrants and seven island residents including one forest-dependent species. Of 30 species previously represented as uniques (seen in only one of 32 sampling periods), 8 were seen on 1 day and 12 seen on two or more days of the shoreline survey (Appendix S1, Supporting information). Of 23 species
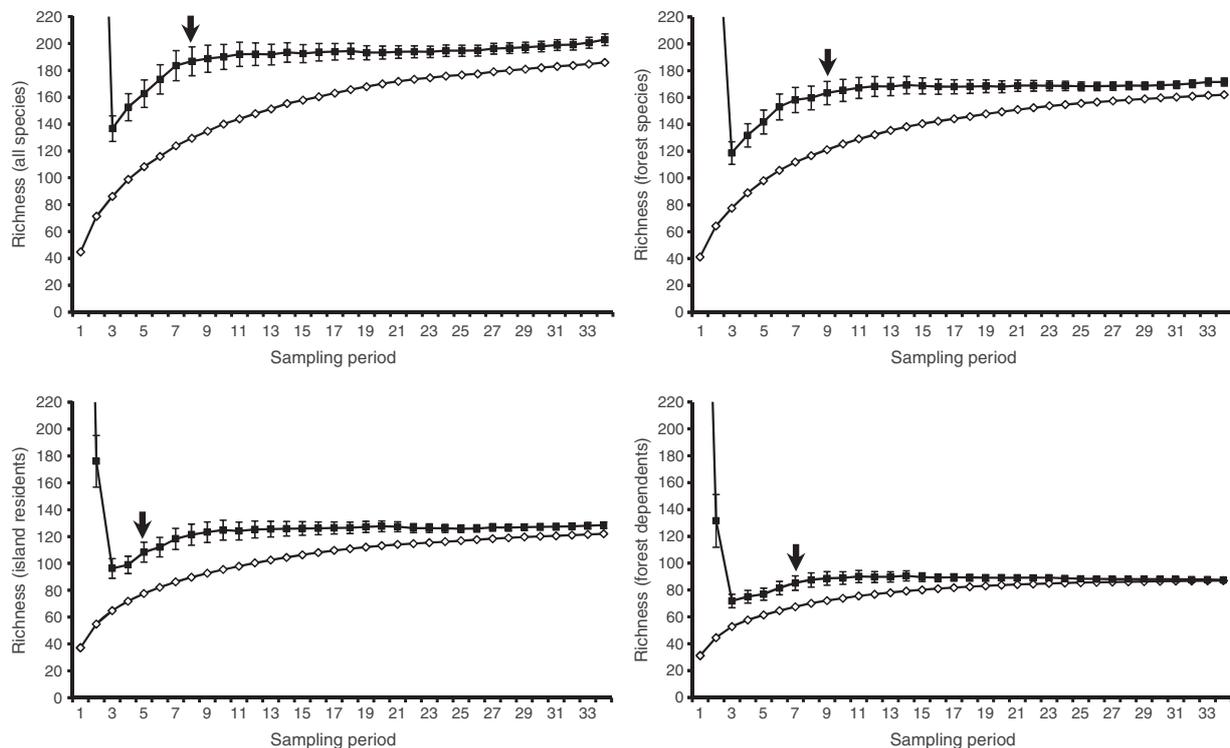
Fig. 2. Species accumulation curves for the four species groups for both inventories combined (i.e. thirty-two 6 h samples). Open diamonds denote mean observed species richness, whereas closed diamonds denote mean predicted richness using the Chao 2 estimator (EstimateS 8·0·0; Colwell 2006). Error bars quantify the standard deviation of predicted richness and the arrows identify the point at which the standard deviation began decreasing consistently – representing the number of samples required for predicted richness to stabilize. All values were calculated from 50 randomizations of *n* samples to remove any effect of sample order. Predictions using other estimators are summarized in Appendix S2 (Supporting information).

previously represented as duplicates (seen in only 2 of 32 sampling periods), 12 were seen on one or more days of the shoreline survey (Appendix S1, Supporting information). With the revised number of unique and duplicate species (18 and 25, respectively), a final set of predictions was calculated from the entire data set (Table 2). The final estimate for all species was 202·9 (SD = 8·67), which accords closely to the 217 species recorded from Barro Colorado Island since 1995 (Robinson 1999; in litt.). Final completeness estimates for all four species groups exceeded 90%, with the final prediction for forest dependents less than one species greater than observed richness (Table 2).

Of the 217 species recorded from Barro Colorado Island since 1995, 67 were not detected during the terrestrial surveys and 54 not detected during all sampling (although 4 species were seen incidentally). These 54 species comprised 8 waterbirds, 5 migrants, 17 non-forest-dependent species and 24 forest-dependent island residents (*sensu* Robinson 2001), many of which are known to leave the island during the wet season (Robinson, pers. comm.). A total of 23 species recorded during this study represented additions to the current bird list for Barro Colorado Island, 16 from the island inventories and 7 from the shoreline survey: 3 waterbirds, 14 migrants, 4 raptors and 2 low detectability species (a swift and a potoo). Of these, 18 have been reported from Barro Colorado Island historically by

Willis & Eisenmann (1979), while five represent genuine additions to the recorded avifauna (a resident duck, two migratory raptors and two resident raptors) of which at least one (*Milvago chimachima*) likely represents a recent colonist.

Discussion

Establishing how many species occur at a particular location is the fundamental purpose of biological inventories, and they form the basis for most community-level studies and conservation applications. In addition to accuracy and representation, efficiency is important in designing a sampling strategy, with results-based stopping rules offering a repeatable and objective means of achieving these combined objectives (Peterson & Slade 1998; Watson 2003). In this study, previously undetected species were still being recorded after 400 h of systematic sampling, but richness predictions for all species groups stabilized long before then (after 5–12 samples). Not only does this mean the richness estimates based on the entire data set can be considered robust, it also means accurate richness estimates can be generated without the exhaustive efforts typically associated with inventories of high diversity sites. While previous research has evaluated the robustness of various richness estimators, most studies lack exhaustive and/or independent baseline data with which to compare richness estimates and few have used

richness estimates to inform quantitative results-based stopping rules (Hortal & Lobo 2005; see also Peterson & Slade 1998 for an elegant case study using state license plates as 'species').

In evaluating these findings, I start by exploring the underlying drivers affecting species richness estimates and associated completeness measures: (i) Why did it take so long to attain a complete inventory? (ii) What were the effects of observer learning and species turnover? and (iii) Which components of the fauna required the greatest sampling effort? Secondly, I examine the detailed patterns underlying these broader findings: (i) which particular species and groups of species were best suited to the sampling approach and which were not? and (ii) How do these results compare with other work? While I refer to the recent changes and ongoing dynamics in the avifauna of Barro Colorado Island and adjoining regions, thorough treatments of these site- and taxon-specific matters will be presented elsewhere. Thirdly, I examine the balance between inventory completeness and sampling effort, proposing some of the key elements required to achieve an optimal sampling protocol. Finally, I explore the implications of this work for further research on sampling diverse assemblages, in terms of methodological approach and overall study design.

TEMPORAL VARIATION AND SPECIES TURNOVER

Although the data presented here were all collected from the same area, they relate to different sampling periods of various durations: two 16 sample inventories spanning 16 and 21 days (14 days apart); the combined 32 sample inventory spanning 51 days; the 32 sample inventory and 35 day of shoreline survey over 101 days; and the sum of all records drawn from the 126 full days spent on the island. Species turnover during this 5-month period mean that species recorded during later samples may not have been present or detectable on the island during earlier samples. As such, the 186 species do not represent the number of species occurring on the island at any one time but, rather, the pool from which the composition of the island's avifauna is derived within this 5-month period. For most comparisons, therefore, 186 species are not treated as the absolute benchmark against which all other estimates are compared. Instead, the predicted species richness estimates are used for the different data sets relating to different time periods. These predictions are unaffected by sample order and all of them stabilized with partial sample numbers of all three incidence-based estimators (Chao 2, ICE and Jacknife 2; Appendix S2, Supporting information), yielding rigorous numeric benchmarks of how many species were present during the relevant sampling period but overlooked. These estimates can be contrasted with one another to evaluate drivers of species turnover, and can be compared against the overall species list to determine which species and groups were overlooked (after Longino, Coddington, & Colwell 2002).

With any diverse assemblage, observer experience and acuity is important and must be demonstrated before inventory data can be considered. Although the first 16 days of sampling were set aside for familiarization with the local fauna, learning effects were still evident, depressing species numbers detected in the first eight sampling periods and likely underestimating richness during that period. This was exemplified by *Ornithion brunneicapillum* and *Camptostoma obsoletum*, two small-bodied tyrannids with non-descript plumages that spend most of their time high in canopy actively glean for insects. While rarely seen and even less frequently identified by sight, both have distinctive calls which, once learned, alert the observer to their presence. Rather than making the first inventory untrustworthy, however, this learning effect was effectively corrected by using mean values from 50 resamples of the data for all analyses. Although these richness estimates stabilized, more samples were required to achieve the same completeness for all four species groups than the second inventory. Thus, while an initial period of familiarization and the use of experienced observers are essential, incorporating part of the learning phase into the data collection period can maximize efficiency, with the use of resampling methods and results-based stopping rules removing patterns in sample order and ensuring data quality is maintained.

Having established comparability, evaluating differences between the two complete 16-sample inventories enabled the magnitude of species turnover on the island to be quantified. Of the 166 species recorded during the 32 sampling periods, 18 species were seen only in the first inventory and 22 species were observed only in the second inventory, with a further 20 species detected only in the subsequent shoreline survey (Appendix S1, Supporting information). Taken together, these 60 species represent approximately a third of all species recorded: a major component of the avifauna that would be under-sampled or completely missed in most conventional surveys. The component of the Barro Colorado Island avifauna most likely to exhibit high species turnover is the migrants – both neotropical migrants that leave their north American breeding range to winter in the region and Mesoamerican species moving through the region to their wintering grounds further south. Of the 50 migrants recorded overall, 20 were recorded only during one of the two inventories, with a further 10 recorded only during the shoreline survey, accounting for half of the temporal variation noted earlier. This group includes birds of prey, waterbirds and aerial foragers, most of which are not restricted to rainforest and range beyond the shores of the island. Of the 24 waterbirds recorded (including raptors and migrants), nine species were seen in only one of the two inventories with an additional eight species seen only during the shoreline survey. The sampling methods used for the two inventories were designed primarily for forest-dependent species, with restricted access to fringing vegetation and associated coves and creek-lines explaining the lower reliability for groups dependent primarily on other habitats.

Waterbirds and migrants accounted for 47 of the 60 species that were not detected in either one or both of the 96-h inventories, justifying the common practice of excluding them from bird surveys. This adjustment represents a refining of the question: rather than simply asking *how many species of bird occur on the island*? but asking *how many resident terrestrial birds*

occur on the island?, sampling to the same degree of completeness can be achieved far more efficiently – completeness stabilizing at 77% after five sampling periods, as compared with 73% completeness in eight periods. By focusing the question on the habitat and temporal scale most relevant, sampling effort is almost halved, while yielding data of greater accuracy.

LOW DETECTABILITY SPECIES

For all richness predictions discussed here, estimates are based on infrequently detected species – the predictive equation uses the ratio of species recorded twice (duplicates) to species recorded once (uniques) to calculate the number of species present but not recorded. In addition to these taxa, there is a third group of low detectability species: those species missed during the 32 sample inventory but recorded in the subsequent shoreline survey. What do these rarities have in common, and why were they encountered so infrequently? Of all these rarities, which were incidentals passing through the forested island and which were island residents that were sufficiently cryptic to make detection challenging?

Of the 70 species in this group, 28 are migrants and 7 are resident waterbirds, and have been discussed already. Of the remaining 35 species, several ecological groups are over-represented, including aerial species, nocturnal species and more than half of all raptors detected (Appendix S1, Supporting information). Occurring in low density, calling infrequently and flushing at great distances, raptors are notoriously difficult to sample (Jullien & Thiollay 1996) and illustrate many of the difficulties associated with sampling low detectability species. Comparing my data with the working list of birds recorded during systematic fixed-effort surveys of the island since 1995 (Robinson 1999; in litt.), there are several noteworthy differences. Only one species recorded during previous surveys was not detected in this study (*Elanus leucurus*, a resident associated with clearings and open habitats), while another (*Elanoides forficatus,* a forest species that migrates through the region) was seen flying over the island on several occasions, but not during any of the sampling periods. While two species were 'missed', six additional species were recorded, including two migrants (*Falco columbarius* and *Buteo jamaicensis*) and two large-bodied residents (*Leucopternus princeps* and *Morphnus guianensis*). Three of these species have not been recorded previously from the island (Willis & Eisenmann 1979), and the latter was not detected in more than 10 years of systematic surveys of the island (Robinson 2001). The final two species *Milvago chimachima* and *Buteo nitidus* are both common lowland raptors in the canal zone, and the records from the island pertain to individuals observed in fringing vegetation during the shoreline survey. Rather than vagrants, it is more likely that these individuals represent colonists from the adjoining second-growth habitats, with the falcon recorded as first colonizing the region in the 1980s (Ridgely & Gwynne 1989). Given the history of Barro Colorado Island and the detailed documentation of local extinctions and colonizations, the addition of six raptor species to those 18 recorded since 1995 (and four species to those species recorded since 1929) is unprecedented,

and demonstrates the level of sampling required to sample these low detectability species adequately: the more effort the better, and an absence of records need not indicate actual absence (Kery 2002).

Rather than an idiosyncratic finding, similar patterns were also exhibited by aerial species and nocturnal species – more than half the species representing uniques, duplicates or new records (Appendix S1, Supporting information). Most community-level bird survey methods ignore flyovers and are restricted to day-time (Verner 1985; Buckland *et al.* 2004), so these groups are typically intentionally excluded from bird surveys. Most records for these groups were incidental or from canopy-level watches: non-standard elements to most bird surveys and impossible in most sites. The issue of access becomes increasingly problematic as structural complexity and habitat heterogeneity increase – additional microhabitats and substrates become effectively inaccessible to observers and, for substrate specialists, decreasing the likelihood of detection (the 'methodological edge' of Longino, Coddington, & Colwell 2002). Accordingly, the widespread practice of excluding raptors, aerial species and nocturnal groups may be pragmatic, but must be explicitly acknowledged.

EVALUATING ABSENCES

In terms of the proportion of species represented by uniques and duplicates, one of the least represented groups is of greatest interest from a conservation and management perspective: residents that depend on intact forest. Five species were seen in only one or two periods and all but one of these were seen during subsequent shoreline surveys, along with a single previously undetected forest resident (Appendix S1, Supporting information). In addition to explaining the close match between observed and predicted richness, this explains the relatively modest effort required to attain robust richness estimates: exceeding 80% completeness after four samples, less than one third of the effort required to sample all species to the same degree (Table 2).

While no new species of diurnal, forest-dependent resident were added, 21 of the 94 forest-dependent residents previously recorded from the island since 1995 were not detected. With the possible exception of one tyrannid (*Platyrinchus coronatus*), I am confident that these species were not on the island during the 5-month study period. For some species (e.g. *Crax rubra*, *Geotrygon violacea*, *Notharcus pectoralis*, *Notharcus tectus*, *Celeus loricatus*, *Tityra inquisitor*, *Schiffornis turdinus, Vireolanius pulchellus*, *Tangara gyrola* and *Pitylus grossus*), I suggest these species no longer reside on the island, and dedicated follow-up surveys are recommended to confirm their absence. Interestingly, four of these species were additions to the island in 1995 so, rather than local extinctions, their current absence suggests these previous records related to individuals moving through the island habitats or new arrivals that did not establish island populations. For other species known to undertake seasonal movements (including *Laniocera rufescens*, *Rhytipterna holyerythra*, *Lipaugus unirufus*, *Myiopagis gaimardii* and *Dacnis venusta*), these absences may be due to

movements off the island during the wet season, and systematic surveys in other seasons are needed before local extinction can be considered. Comparing extinction rates of species that undergo seasonal movements with year-round residents will shed further light on the drivers of faunal relaxation, with those movements likely buffering short-term changes in resource availability and enhancing local persistence.

TOWARDS OPTIMAL SAMPLING

As with all samples, inventories are necessarily incomplete (Colwell & Coddington 1994), but optimal sampling involves matching the level of completeness with the scale and resolution of the question being addressed (Hortal & Lobo 2005). Using predictive methods to estimate the number of species missed yields richness estimates of known completeness enabling comparisons of standardized samples, but this approach necessarily generates species of unknown identities. As demonstrated with this study, it is often these low detectability species that are of greatest interest or conservation concern: low-density habitat specialists, wary predators or long-distance migrants. Rather than excluding these groups, a hybrid approach using successive iterations of dividing observed richness by estimated richness is a useful compromise, yielding lists of known species with associated completeness estimates. Completeness can be monitored continuously to maximize efficiency (i.e. stop sampling once predictions stabilize), comparability (stop when completeness exceeds a pre-determined threshold) or completeness (stop once completeness shows no sign of increasing). While results-based stopping rules are generally defined operationally, using this quantitative approach ensures data of maximum comparability, even with less complete data that can be standardized *a posteriori* through rarefaction or extrapolation.

In this study, the maximum completeness of 91% was attained after 192 h of intensive sampling on 32 days over a 51-day period and, even when that effort was more than doubled (using complementary methods), less than 92% completeness was achieved (Table 2; Appendix S1, Supporting information). If 210 h of sampling can achieve less than a 1% improvement, 90% can be considered to signify an upper limit to meaningful completeness estimates. Hence, if estimating total richness is the objective, attaining very high completeness is not just difficult and time-consuming, it may simply be unattainable. As demonstrated here, however, much of this noise relates to particular groups within the community, enabling the research question to be refined and the sampling approach optimized. By simply removing the migrants and waterbirds, variance measures of richness estimates were halved at all levels – from the four 8-sample half inventories to the entire 402 h data set – enabling 80% estimated completeness to be attained with 30% less sampling.

This improved efficiency notwithstanding, this refined approach still involves 10 6-h samples, or 2 weeks of intensive fieldwork to conduct an inventory of a single site – a prohibitive effort for most ecological studies. By going further and removing those species not wholly dependent on the rainforest

(i.e. second-growth and edge species), those species poorly suited to the sampling approach (nocturnal species), and low density species that range beyond the study area (raptors, aerial species), sampling efficiency can be further optimized, taking less than half the time to achieve inventories of the same completeness.

Reconciling these findings with previous work validates many of the conventions already in widespread use, quantifying their benefit to sampling efficiency, albeit at the expense of information about these low detectability species (Hawkesworth 1995). Non-breeding migrants are typically excluded from analyses of avian occurrence, often justified by the fact that, being only seasonal visitors they would not be as sensitive to subtle differences in habitats, disturbance regimes or resource availability as species residing in the area year-round. Although reasonable, this leads to far less knowledge about occurrence patterns, habitat preferences and resource requirements of migrants in their non-breeding range, information considered critical for their longer term conservation management. While raptors, owls, nightjars, swifts, swallows and species restricted to second-growth habitats are components of avian assemblages world-wide, does their inclusion justify a doubling of sampling effort to achieve data of the same accuracy? This is an open question, and depends primarily on the motivation driving the study. Importantly, the scope of the study also determines what those data can be used for, both initially by the researcher and subsequently by others interpreting the findings. For example, using community-level surveys of woodland birds to discern diversity patterns of raptors is inappropriate and would underestimate the actual number of species present. If raptors or aerial species are the priority, water-based surveys and canopy watches proved far more effective than land-based searches, highlighting the need to tailor sampling to the group(s) of interest.

Having determined the scope of the question, how much sampling is enough? In addition to allowing continuous real-time monitoring of completeness, combining completeness estimation with a quantitative results-based stopping rule (e.g. the standardized search; Watson 2003) yields inventories of fixed completeness. In this study, richness estimates of 70% completeness or less were unstable, reflecting local minima where predicted richness increased more rapidly than observed richness and variance was fluctuating or increasing. Once variance of the predicted richness stabilized and began to decrease consistently, predictions and associated completeness estimates became reliable. For these data, this threshold occurred consistently at the 75% mark, with values for all species groups except forest dependents occurring between 73% and 79%, although the sample number when this occurred ranged from 5 to 12. The point where completeness estimates stabilized for forest dependents was similar (seven and eight sampling periods), but the associated values were higher – 82% and 83%. In addition to reinforcing the thesis that more focused sampling yields more complete samples, this also suggests that predictions based on 80% completeness are unreliable for all groups, and 85% would be a more robust benchmark (comparable with Lobo 2008). Regardless, the fact that all 12 sets of

analyses converged to within 10% of one another (Table 2) lends considerable confidence to the broader applicability of these methods and suggests that estimates of 75–85% completeness should be regarded as maximally efficient. While estimates with lower completeness estimates could be used to make comparisons of relative numbers, absolute values would be unreliable and limited time would be better invested increasing sampling effort per site that adding additional sites.

PROSPECT

Unlike inventories in entomology, ornithological inventories are composed exclusively of taxa identified to species – a practice that warrants objective scrutiny. Do we really need to know the name of every species occurring within a given area to make meaningful inferences, draw reliable conclusions and inform useful management recommendations? For island residents and forest-dependent species, accurate richness estimates were attained after 46–60 h of effort, yet a further 132–146 h of effort were required to ascribe names to all of these species: time that could easily have been devoted to sampling two or three additional sites to the same degree. In tropical areas and many other diverse systems, there is far less autecological knowledge than in temperate systems, so knowing the name of a species need not confer much additional understanding. For many studies, knowing the diet, mass, nesting ecology or phylogenetic affinities of taxa are needed, and necessitate exhaustive sampling. If so, and if these components are central to the aims of the study, then sampling for maximum completeness is required. But, if species numbers are needed, proportion of residents vs. migrants, or habitat specialists vs. generalists, then using results-based stopping rules to fix completeness estimates of all surveys (nominally, 80–85%) would be adequate and represent the optimal approach. Alternatively, if the aim of the study was simply to evaluate determinants of diversity (e.g. typical habitat fragmentation or island biogeographical questions), then only relative estimates of species richness are required to rank sites from most to least diverse. The optimal strategy here would be to maximize the number of sites sampled, sampling each site until richness estimates stabilized, regardless of completeness values.

Having demonstrated that accurate inventories of known completeness can be achieved, what is the best way to use them to sample high diversity sites, both for birds and other groups? In this study, the minimum effort required to sample the group of most interest (forest-dependent species) with confidence was seven 6-h samples – i.e. one full week of comprehensive sampling. Given the documented effects of temporal variation, this required effort precludes studying more than seven or eight sites by a single observer, limiting the number of comparisons possible and decreasing the statistical power of associated analyses. Using multiple observers working simultaneously is one potential solution that is rarely employed, especially in tropical forests where there are so few experienced observers able to identify all individuals detected to species. While training observers to a minimum level of competency is essential, inter-observer effects could then be quantified as the data were

being collected. Rather than have all observers collecting the same number of samples, iterative application of a results-based stopping rule would be conducted by each observer, ensuring completeness is consistent across surveys and maintained at the desired level. Coupling these data with contemporaneous sampling by a highly experienced observer for at least one survey for each observer would enable comparability to be quantified rather than assumed, and allow observer effects to be minimized. While this approach would be most useful for generating relative richness estimates, it would also identify those groups of greatest and least variance, those sites of highest and lowest diversity, those observers of greatest and least ability and those species of highest and lowest detectabilities; all valuable foundations for further and more focused research.

Acknowledgements

References

Boulinier, T., Nichols, J.D., Sauer, J.R., Hines, J.E. & Pollock, K.H. (1998) Estimating species richness: the importance of heterogeneity in species detectability. *Ecology*, **79**, 1018–1028.

Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L. & Thomas, L. (2004) *Advanced Distance Sampling: Estimating Abundance of Biological Populations*. Oxford University Press, New York.

Chao, A. (1984) Non-parametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics*, **11**, 265–270.

Chao, A. (1987) Estimating the population size for capture-recapture data with unequal catchability. *Biometrics*, **43**, 783–791.

Chapman, F.M. (1929) *My Tropical air Castle*. Appleton-Century, New York.

Colwell, R.K. (2006) *EstimateS: Statistical estimation of species richness and shared species from samples*. *Version 8*. User's Guide and application published at: http://purl.occlc.org/estimates. Department of Ecology & Evolutionary Biology, University of Connecticut, Storrs.

Colwell, R.K. & Coddington, J.A. (1994) Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society, London B*, **345**, 101–118.

Eisenmann, E. (1952) Annotated list of birds of Barro Colorado Island, Panama Canal Zone. *Smithsonian Miscellaneous Collections*, **117**, 1–62.

Gotelli, N.J. & Colwell, R.K. (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, **4**, 379–391.

Hawkesworth, D.L. (1995) *Biodiversity: Measurement and Estimation*. Royal Society, UK.

Herzog, S.K., Kessler, M. & Cahill, T.M. (2002) Estimating species richness of tropical bird communities from rapid assessment data. *Auk*, **119**, 749–769.

Hortal, J., Garcia-Pereira, P. & García-Barros, E. (2004) Butterfly species richness in mainland Portugal: predictive models of geographic distribution patterns. *Ecography*, **27**, 68–82.

Hortal, J. & Lobo, J.M. (2005) An ED-based protocol for the optimal sampling of biodiversity. *Biodiversity and Conservation*, **14**, 2913–2947.

Jullien, M. & Thiollay, J.-M. (1996) Effects of rain forest disturbance and fragmentation: comparative changes of the raptor community along natural and human-made gradients in French Guiana. *Journal of Biogeography*, **23**, 7–25.

Kery, M. (2002) Inferring the absence of a species: a case study of snakes. *Journal of Wildlife Management*, **66**, 330–338.

Leigh, E.G. Jr, Rand, A.S. & Windsor, D.M. (1982) *The Ecology of a Tropical Forest: Seasonal Rhythms and Long-Term Changes.* Smithsonian Institution Press, Washington DC.

Lobo, J.M. & Martin-Piera, F. (2002) Searching for a predictive model for species richness of Iberian dung beetle based on spatial and environment variables. *Conservation Biology*, **16**, 158–173.

Lobo, J.M. (2008) Database records as a surrogate for sampling effort provide higher species richness estimations. *Biodiversity and Conservation*, **17**, 873–881.

Longino, J.I., Coddington, J.A. & Colwell, R.K. (2002) The ant fauna of a tropical rainforest: estimating species richness three different ways. *Ecology*, **83**, 689–702.

Oliver, I. & Beattie, A.J. (1996) Designing a cost-effective invertebrate survey: a test of methods for rapid assessment of biodiversity. *Ecological Applications*, **6**, 594–607.

Peterson, A.T. & Slade, N.A. (1998) Extrapolating inventory results into biodiversity estimates and the importance of stopping rules. *Journal of Biogeography*, **4**, 95–105.

Ridgely, R.S. & Gwynne, J.A. Jr (1989) *A Guide to the Birds of Panama With Costa Rica, Nicaragua and Honduras*, 2nd edn. Princeton University Press, Princeton NJ.

Robinson, W.D. (1999) Long-term changes in the avifauna of a tropical forest isolate: Barro Colorado Island, Panama. *Conservation Biology*, **13**, 85–97.

Robinson, W.D. (2001) Changes in abundance of birds in a Neotropical forest fragment over 25 years: a review. *Animal Biodiversity and Conservation*, **24**, 51–65.

Scharff, N., Coddington, J.A., Griswold, C.E., Hormiga, G. & Per de Place, B. (2003) When to quit? Estimating spider species richness in a northern European deciduous forest *Journal of Arachnology*, **31**, 246–273.

Sieving, K.E. & Carr, J.P. (1997) Avian extinction and persistence mechanisms in lowland Panama. *Tropical Forest Remnants: Ecology, Management, and Conservation of Fragmented Communities* (eds W.F. Laurance & R.O. Bierregaard Jr), pp. 156–170. University of Chicago Press, Chicago, IL.

Soberon, J.M. & Llorente, J.B. (1993) The use of species accumulation functions for the prediction of species richness. *Conservation Biology*, **7**, 480–88.

Verner, J. (1985) Assessment of counting techniques. *Current Ornithology* (ed. R.F. Johnston), pp. 247–302. Plenum Press, New York.

Walther, B.A. & Moore, J.L. (2005) The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. *Ecography*, **28**, 815–829.

Watson, D.M. (2003) The "standardized search": an improved way to conduct bird surveys. *Austral Ecology*, **28**, 515–525.

Watson, D.M. (2004) Comparative evaluation of new approaches to survey birds. *Wildlife Research*, **31**, 1–11.

Willis, E.O. & Eisenmann, E. (1979) *A Revised List of Birds of Barro Colorado Island, Panama.* Smithsonian Institution Press, Washington DC.

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Appendix S1. List of low detectability species used in the calculation of richness estimations.

Appendix S2. Species richness predictions: complete outputs from EstimateS (Version 8.0.0; Colwell 2006).

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.