Biological Conservation 144 (2011) 1374–1381

Contents lists available at ScienceDirect

Biological Conservation

journal homepage: www.elsevier.com/locate/biocon

Rapid monitoring of species abundance for biodiversity conservation: Consistency and reliability of the MacKinnon lists technique

Ross MacLeod [a,*], Sebastian K. Herzog [b], Aidan Maccormick [a], Steven R. Ewing [a], Rosalind Bryce [a], Karl L. Evans [c]

[a] *Institute of Biodiversity, Animal Health and Comparative Medicine, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow G12 8QQ, UK*
[b] *Asociación Armonía – BirdLife International, Casilla 3566, Santa Cruz de la Sierra, Bolivia*
[c] *Department of Animal & Plant Sciences, University of Sheffield, Western Bank, Sheffield S10 2TN, UK*

ARTICLE INFO

Article history:
Received 22 November 2009
Received in revised form 8 December 2010
Accepted 11 December 2010
Available online 21 February 2011

Keywords:
Biodiversity survey methods
Rapid assessment surveys
Species lists technique
Bird surveys

ABSTRACT

Effective monitoring of biodiversity for conservation requires information on spatial and temporal variation in species' abundances. As conservation resources are limited, monitoring methods are required that enable rapid and cost effective data collection. There are many traditional methods of estimating absolute abundance, such as territory mapping and distance sampling. However, these typically require more time, expertise and finances than are available across much of the globe. This is especially so in the tropics, where high species richness, low densities of many species and structurally complex environments also make monitoring particularly challenging. The MacKinnon lists technique is a rapid assessment methodology designed for use in species rich environments. This method is typically used to estimate species richness, but it has also been suggested that it can generate consistent abundance indices, even when observer experience and environmental conditions vary. If this suggestion is correct, the MacKinnon lists method could be used to assess spatial or temporal changes in abundance using diverse survey data. Here, we provide the first detailed assessment of intra- and inter-observer consistency of the Mackinnon List method in generating species abundance indices that could be useful for conservation monitoring purposes. As a case study, we use one of the world's most diverse avifaunas, that of the forested Bolivian Andes. We show that MacKinnon lists can provide species abundance indices that are consistent between observers of markedly different experience of the focal avifauna (zero to six years), and between assessments carried out in different stages of the breeding season, between which detectability of individuals differed significantly. We believe this is the first time that a biodiversity monitoring method has been demonstrated to produce consistent abundance indices for a highly diverse avian tropical assemblage. We also suggest that the MacKinnon lists methodology has the potential to be a very useful conservation monitoring tool for many taxa in species rich environments, such as the tropics.

1. Introduction

Conservation requires effective monitoring of species distributions and abundances in order to assess the relative importance of different sites, the effectiveness of management activities and population trends (Mace and Baillie, 2007; Marsh and Trenham, 2008; Maxwell and Jennings, 2005; Pollock, 2006). There are several sophisticated methods available to generate density estimates and calculate absolute abundances, such as mark-recapture, distance sampling and territory mapping (e.g. Bibby et al., 2000; Sutherland, 1996). These methods are resource intensive as they rely on large data sets that require significant expertise to collect and interpret (Bibby et al., 2000; de Thoisy et al., 2008; Peres, 1999). Moreover, there is growing awareness that they are not infallible, and resultant density estimates can be biased by a number of methodological concerns (Alldredge et al., 2008; Newson et al., 2008). In tropical regions that are characterised by high species richness, low population densities and structurally complex vegetation, it is particularly difficult to use these traditional methods without violating their basic assumptions (Bibby et al., 2000; Herzog et al., 2002). Monitoring in many tropical areas is further hindered by a lack of experienced field workers and reduced availability of scientific support and financial resources. There is thus a pressing need to develop and test survey methods that could enable rapid assessment of spatial and temporal variation in species abundances by relatively inexperienced personnel in tropical regions. The MacKinnon lists (ML), or species lists technique, is advocated as one such method (Bibby et al., 2000; MacKinnon and

* Corresponding author. Address: Tel.: +44 (0)141 330 5340.
E-mail addresses: ross.macleod@glasgow.ac.uk (R. MacLeod), skherzog@armonia-bo.org (S.K. Herzog), karl.evans@sheffield.ac.uk (K.L. Evans).

0006-3207/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.biocon.2010.12.008

Phillips, 1993), but its use in rapid relative abundance assessments has had only minimal testing.

The ML technique was originally developed for rapid assessment of avian species richness in tropical environments (Bibby et al., 2000; MacKinnon and Phillips, 1993). In brief, the method consists of listing all individuals encountered in chronological order of detection. This master list is then broken down into lists or samples of a pre-determined number of species; ten is often recommended. Each list thus provides a sample of the species community at the study site. The ML method can then be used to derive abundance indices of individual species by calculating the proportion of samples in which each species occurs (Bibby et al., 2000; MacKinnon and Phillips, 1993). The ML samples are suggested to be independent of the amount of time needed to collect them, the spatial extent over which they are collected and observer expertise; the method may therefore be less sensitive to spatial and temporal changes in a species' detectability than many of its alternatives (Bibby et al., 2000). As a result the ML method has the potential to be useful in estimating changes in abundance over time and comparing spatial differences in abundance. The ML method does not take inter-specific variation in detectability into account, and thus cannot be used to compare abundances of species with different detectability (Bibby et al., 2000). It also cannot be used to estimate species densities or absolute abundances and it is therefore not suitable for answering ecological questions that require this type of data or for estimating population sizes. However, this is rarely of central importance in conservation monitoring, where intra-species comparisons of abundance indices between sites and time periods are usually more valuable (e.g. Bibby et al., 2000; Sutherland, 1996).

The ML method is frequently used to estimate species richness (Bibby et al., 2000; Fjeldsa, 1999; Herzog and Kessler, 2006; Herzog et al., 2002, 2005; Poulsen et al., 1997a,b; Trainor, 2002). In contrast the ML method has been used surprisingly rarely to compare species' abundances at different sites or over time (although see Herzog, 2008). This is probably largely due to a lack of awareness of its potential, not least because there have been few formal tests of its utility in this regard. The only such assessment compared the ranking of the five most common bird species in two tropical forests and although the authors drew negative implications from their limited assessment, the results presented appeared to show that abundance rankings were broadly similar (although not identical) when calculated from density estimates derived from point counts and the ML method (O'Dea et al., 2004). A potential major advantage of the ML method over other methods such as point counts for generating species abundance indices is its suggested ability to produce consistent estimates for a focal species despite variation in survey effort and observer expertise.

Here, we assess whether the ML method generates consistent relative abundance indices despite marked variation in observer experience and environmental conditions. We do so using the avifauna of the forested east Andean slope in Bolivia as a case study. This region comprises one of the most diverse terrestrial hotspot for biodiversity and supports many species of high conservation concern (Myers et al., 2000; Soria-Auza and Hennessey, 2005). We test the hypotheses that the ML sampling technique generates consistent species abundance indices: (i) during different periods of the main breeding period, during which time species' detectability will change and (ii) when observer expertise varies (from no experience to over six years experience of the focal avifauna). We also test the relative performance of three different forms of the ML methodology that vary in how species lists samples were used to construct abundance indices. Finally, in addition to assessing the utility of ML abundance indices across the entire assemblage, we assess their utility for subsets of species that varied in the primary means of their identification at the study site (visual and aural), and for a subset of species of special importance due to their conservation status (Table S1).

2. Methods

2.1. Study area

The study was carried out on the east side of the Cordillera de Cocapata, Department of Cochabamba, Bolivia between 2nd August and 19th September 2001. This mountain range, which forms part of the eastern Andean slope, is located in the Cuzco-Cochabamba section of the central Andes zoographic region (Stotz et al., 1996) and lies within the Bolivian and Peruvian Upper and Lower Yungas Endemic Bird Areas (EBAs 054 and 055; (Stattersfield et al., 1998). The study focused on the Rio Pampa Grande valley (16°40′S; 66°29′W) and was primarily conducted within undisturbed and slightly disturbed primary montane evergreen forest. Small areas of greater disturbance, including secondary forest and a few areas cleared for grazing, were also present. From a base camp at 2100 m, four transects (10 km total length) ran through forested habitat between 1900 and 2600 m so that they covered a representative selection of the microhabitats in the area. A full description of the site, where no ornithological survey work had previously been conducted, is provided by MacLeod et al. (2005).

2.2. Survey work

The survey work reported here was carried out as an independent study during an inventory of avian biodiversity that aimed to document the avian assemblage occupying the focal area during the pre-breeding and early breeding period of the main breeding season. The survey work was carried out by a team of six people. Fieldwork began at first light (~06:15 local time) and continued for on average three to four hours, with the exact survey time depending on transect length and the speed at which observers walked. The work was conducted with each observer using 8× or 10× binoculars for visual identification, and sound recording equipment to back up audio identification (either a Sony TCM 5000EV cassette tape recorder or Sharp Minidisk MT280E recorder each with a Sennheiser ME 66 shotgun style directional microphone).

Although all six observers had previously conducted ornithological fieldwork in the Neotropics and were competent ornithological observers, they varied markedly in bird identification skills, particularly with regard to knowledge and prior experience of the focal avifauna. This allowed investigation of whether the ML method provided species abundance indices that were robust to inter-observer differences, a key requirement for any data being collected for conservation assessments. Two observers had previous South American experience, but not with the Bolivian avifauna; three observers had prior experience with the avifauna of the eastern slope of the Bolivian Andes (the Yungas); the final observer had six years prior experience with the Yungas, and is one of the most experienced field ornithologists in this region.

For the purposes of this study, surveys were conducted during two periods (5th–20th August and 26th August–15th September 2001). These represented respectively the pre-breeding period (when birds started to sing and vocalise to defend territories and attract mates), and the early breeding season (when the majority of birds were singing strongly in conjunction with the onset of nesting activity). The increased singing and activity levels in the second period increased the detectability of individuals so that with the same survey effort 28% more individuals were detected in the second period than the first, and this difference was significant (paired t-test: $t_4 = 4.8$, $p = 0.017$). During each survey period each observer aimed to conduct surveys once along each of four transects so that each observer was surveying the same standardised area once. Each observer was randomly assigned a survey day for each transect and so carried out data collection, for this study, on only four of the days within the survey period. No transect was surveyed by more than one observer on any one day.

Abundance indices for each species were estimated using the following methodology. Every individual bird detected, either visually or aurally, within 50 m on each side of the transect route during the outward journey along each transect was recorded. Data were therefore recorded from a standard survey area by each observer. Registrations for return journeys were not included to avoid the possibility of prominent individuals being registered twice and biasing the sampling (Herzog et al., 2002). Observers took descriptions, or sound recordings of any bird not immediately identified but that was seen, or heard, sufficiently well for identification. These individuals were subsequently identified using standard reference works. To ensure that the study was a robust test of inter-observer differences no other special instructions were given, and the observers were free to use their own preferred methods of finding and identifying species. For each survey period, a chronologically ordered master list of all individuals registered by each observer during their four transects was then compiled. Following common practice (e.g. Poulsen et al., 1997b; Herzog et al., 2002), the master list was divided into samples of 10 species each, as constructing sample lists of more than 10 species would have reduced the sample size and variation in species relative abundance estimates. The initial sample recorded the first 10 species registered, with additional registrations of the same species being excluded so that the sample comprised 10 different species. After completing the first 10-species list the process was repeated to produce a second 10-species sample, and the process was then repeated continuously until the last complete list of 10 species was generated. Each observer thus produced a series of samples of the bird assemblage from which the relative abundance of a species could be estimated by calculating the proportion of samples (i.e. lists) in which it occurred (Poulsen et al., 1997b; Bibby et al., 2000).

2.3. Analysis

For each observer three types of species abundance indices were calculated. First, we used the exact proportion of ML samples in which each species appeared (ML exact method). Second, we categorised species into 20 abundance categories of equal width; as an example, species in 0.00–4.99% of list samples were assigned to the first abundance category, and species in 5.00–9.99% of samples to the second abundance category, etc. (ML category method). Finally, we assessed an alternative abundance measure based on the proportion of individuals of each species detected compared to the total number of individuals detected by the observer (individual frequency or IF method). This is equivalent to a ML technique where species lists are compiled with just one, rather than 10, species. Prior to analysis, an arcsine square root transformation was applied to all abundance indices so that data were normally distributed. One observer, with an intermediate level of experience, did not record numbers of individuals detected, so contributed no data to tests of the individual frequency method and as a result comparisons between the IF methods and the other methods have lower degrees of freedom. The same observer was unable to complete all four transects in each period due to inclement weather and sickness so most of the analysis in the results only includes comparisons between and within five observers (see Table 1 for observer details).

We assessed whether the MacKinnon lists technique produced comparable abundance indices when used in different circumstances. First, we tested whether the same observer produced similar species abundance indices for both survey periods (intra-observer consistency). Second, we tested whether species abundance indices were consistent between observers with different knowledge and experience, when collecting data on different days during the same survey period (inter-observer consistency). The latter was assessed across the entire assemblage and separately for three sub-groups of species: (i) those mainly identified visually because they rarely vocalise, or are difficult to identify by voice alone, (ii) those mainly identified vocally because they are more often heard than seen, or are difficult to identify on plumage alone, and (iii) those of special conservation importance, i.e. restricted-range and globally threatened species (BirdLife International, 2000, 2008; Stattersfield et al., 1998). Species were assigned to visual and vocal identification categories by the authors with previous Yungas experience prior to the start of the analysis based only on experience during the study. Species that are often identified vocally at other sites, but were not vocalising during the fieldwork, were thus classified in the visual identification category. Species that were not predominately identified vocally or predominately identified visually because they were identifiable by both methods were excluded from the visual and vocal sub-groups.

Intra-observer and inter-observer consistency were assessed using intra-class correlation coefficients (McGraw and Wong, 1996) and reliability analysis (Cronbach, 1951). These methods are not well known in ecology or conservation biology. However, they are a form of random effects model and in other disciplines, including medicine, psychology, behavioural genetics, and behavioural science, they are widely accepted techniques for comparing results obtained under varying conditions, such as by different observers or in different time periods (Howell, 1997; McGraw and Wong, 1996; Shrout and Fleiss, 1979; Siegel and Castellan, 1999). Intra-class correlation coefficients (ICCs) are used to assess the strength of association between two or more measures of the same variable, and Cronbach's alpha reliability metric (Cronbach, 1951) is used to assess the similarity between each estimate of a particular variable and the mean of all estimates of that variable. Absolute values of ICCs and Cronbach's alpha range from 0 (no association or reliability) to 1 (perfect association or reliability); values of 0.7 or greater are typically considered to indicate high consistency (Table 2; Cohen and Holliday, 1996). We used ICCs to assess the strength of association between abundance indices obtained under different conditions, and Cronbach's alpha reliability measure to assess how close each observer's abundance indices were to the average across all observers.

We used an ICC (A, 1) model to test for absolute agreement between abundance indices for each species (McGraw and Wong, 1996). In the intra-observer analysis used to test for variation over time, observer formed the class. Intra-class variation and reliability were calculated from the difference between the observer's two abundance indices for each of the 156 species detected in the study. In the inter-observer analysis used to test for variation between observers, time period formed the class. Intra-class variation and reliability for each period were calculated, for each of the 156 species detected, from the differences between the five observers who collected data in period one and the six observers who collected data in period 2. In each analysis degrees of freedom were therefore 155. In the results ICCs and Cronbach's alpha are expressed as absolute values.

Measures of consistency and reliability were normally distributed, and we thus used paired t-tests (2-tailed) to assess if these metrics were influenced by the type ML method used to calculate abundance indices. We used Spearman rank correlations to assess if intra-observer consistency was correlated with observer expertise or survey effort (as the latter two variables did not meet the assumptions of parametric tests). Analysis was carried out using the SPSS statistical programs (SPSS, 2008).

Table 1
Observer survey information: 'samples' represents mean observer survey effort recorded as the number of 10-species list samples, 'ind/sample' represents the mean number of individuals detected per list sample, and 'species' represents the mean number of species detected per observer averaged across the two periods.

Observer	Experience of focal avifauna prior to this study	Samples	Ind/sample	Species
1*	Yungas experience	50	–	86
2	Yungas experience	26	12.0	85
3	South American experience	33	12.2	88
4	Yungas experience	27	14.0	75
5	South American experience	32	11.8	89
6	Yungas expert	39	14.9	97

* Complete data were not available for this observer due to inclement weather and illness.

Table 2
Standard interpretations of correlation and reliability coefficients (Cohen and Holliday, 1996).

Coefficient	Interpretation
0.00–0.19	Very low correlation
0.20–0.39	Low correlation
0.40–0.69	Modest correlation
0.70–0.89	High correlation
0.90–1.00	Very high correlation

3. Results

3.1. Variation in survey effort and ability

The fieldwork conducted for this study took 140 person hours during eight mornings of fieldwork. Each observer took approximately three and a half hours to complete the outward leg of each transect, and thus spent approximately 14 h on data collection, and collected on average 34 10-species MacKinnon list samples in each assessment period (Table 1). The fieldwork for this study detected and calculated ML abundance indices for 156 bird species, which represents 74% of the 211 species identified in the altitudinal range of the study area by the team during the 50 days spent in the area (Macleod et al., 2005). Of these 156 species, 96 (61.5%) and 31 (19.9%) were predominately identified respectively vocally and visually, the remaining 18.6% of species were identified using both methods depending on circumstances.

Although each observer surveyed the same length of transects for a similar period of time there was marked variation in survey effort as measured by the number of ML samples collected per observer, which varied by a factor of two (Table 1). We consider number of samples to be the most appropriate measurement of survey effort because it is a direct measure of the quantity of data collected and is thus more appropriate than less direct measures such as time spent in the field. There was also considerable variation in objective measures of observer ability, i.e. the total number of species recorded and the number of individuals recorded per ML sample. The Yungas expert recorded approximately one third more species, and one quarter more individuals per sample, than the observers who recorded the fewest species and individuals per sample (Table 1). Moreover, there was a marked association between prior observer experience and the number of individuals recorded per sample suggesting that prior experience was a suitable measure of observer ability (Spearman rank correlation $r_s = 0.738$); although this correlation was not statistically significant this is due to the small sample size ($p = 0.155$, $n = 5$). There was no association between prior experience and total number of species recorded ($r_s < 0.01$, n = 6).

3.2. Intra-observer consistency

Species abundance indices calculated by the same observer in the two different survey periods were closely associated with each other (Table 3). Across the five observers who collected data in both periods mean ICC values were greater than 0.70 regardless of the method used to calculate the abundance indices (Table 3) and the highest mean ICC value (0.84) was obtained using the category method. This was not significantly greater than that obtained using the exact method (ICC = 0.82; paired t-test: $t_4 = 1.3$, $p = 0.270$), but it was significantly greater than that obtained using the individual frequency method (ICC = 0.75; paired t-test: $t_4 = 5.5$, $p = 0.005$). The intra-observer abundance indices were also highly consistent between the two survey periods as measured by Cronbach's alpha reliability (Table 3). Reliability was also greatest when abundance indices were calculated using the category method (Cronbach's alpha = 0.92). This was not significantly different from that obtained by the exact method (Cronbach's alpha = 0.91; $t_4 = 1.1$, $p = 0.335$), but it was significantly greater than reliability when the individual frequency method was used to estimate abundance indices (Cronbach's alpha = 0.86; $t_4 = 4.7$, $p = 0.009$).

Intra-observer consistency often increased with survey effort, as measured by the number of list samples the observer collected. Such correlations occurred for the ML category method with regard to both mean ICC value and Cronbach's alpha (Spearman rank correlations: in both cases $r_s = 1.00$, $p < 0.001$, $n = 5$), but only with regard to Cronbach's alpha when abundance indices were calculated using the exact method ($r_s = 0.90$, $p < 0.037$, $n = 5$). There were no other significant correlations between measures of intra-observer consistency and other potential measures of survey effort (i.e. mean number of individuals per sample or total number of species recorded; $p > 0.1$ in both cases). Prior observer experience was not correlated with intra-observer consistency measured as either mean ICC value or Cronbach's alpha ($p > 0.1$ in both cases).

3.3. Inter-observer consistency

Species abundance indices calculated by different observers generated similar results (Table 4). In the first survey period (Table 4a) inter-observer consistency was highest when abundance indices were calculated using the category method (ICC = 0.78), and this was significantly greater than that for abundance indices calculated with the exact method (ICC = 0.71; paired t-test: $t_9 = 10.5$, $p < 0.001$), and also significantly greater than that obtained using the individual frequency method (ICC = 0.68; paired t-test: $t_9 = 7.0$, $p < 0.001$). ML exact correlations where not significantly higher than those generated by the individual frequency methodology (paired t-test: $t_9 = 0.29$, $p = 0.78$). Reliability measures showed very high inter-observer consistency using all three methods (always >0.9) and showed a similar pattern to the ICCs, with the category method producing the most consistent results (Table 4a).

In the second survey period (Table 4b), inter-observer consistency measures (mean ICC and reliability values) were slightly higher than in the first survey period (cf. Table 4a), but the differences were not statistically significant (paired t-tests: $p > 0.1$ in all cases). Abundance indices based on the ML category method again

Table 3
Intra-observer consistency of species abundance indices. Results generated using three MacKinnon lists methodologies (see text) with consistency measured by intra-class correlation coefficients (ICC) and Cronbach's alpha reliability measure for the five observers who collected data in both periods ($N = 156$ species). 'Obs' represents the observer code from Table 1.

Method	Obs	ICC	L95% CI	U95% CI	Reliability	L95% CI	U95% CI
a. Mackinnon lists	2	0.690	0.598	0.764	0.816	0.748	0.866
Exact	3	0.887	0.848	0.916	0.940	0.918	0.956
	4	0.843	0.790	0.883	0.915	0.883	0.938
	5	0.840	0.787	0.881	0.913	0.881	0.937
	6	0.862	0.816	0.980	0.949	0.930	0.964
Mean		0.824	0.768	0.885	0.907	0.872	0.932
b. Mackinnon lists	2	0.759	0.682	0.817	0.862	0.811	0.899
Category	3	0.880	0.839	0.911	0.936	0.912	0.953
	4	0.833	0.774	0.873	0.907	0.872	0.932
	5	0.847	0.960	0.886	0.917	0.886	0.940
	6	0.898	0.864	0.926	0.964	0.950	0.974
Mean		0.843	0.824	0.883	0.917	0.886	0.940
c. Individual frequency	2	0.683	0.589	0.759	0.812	0.741	0.863
	3	0.733	0.646	0.800	0.846	0.784	0.889
	4	0.751	0.669	0.815	0.858	0.801	0.898
	5	0.744	0.660	0.809	0.853	0.795	0.895
	6	0.850	0.802	0.890	0.945	0.924	0.960
Mean		0.752	0.673	0.815	0.863	0.809	0.901

Table 4
Inter-observer consistency of species abundance indices. Results generated using three MacKinnon lists methodologies (see text) during the (a) first and (b) second assessment period. Consistency is measured by the mean intra-class correlation coefficient (ICC) between each pair of observers and Cronbach's alpha reliability measure. Values are calculated for the five observers who collected complete data in the first period and six observers who collected data in the second period ($N = 156$ species).

Methodology	Mean ICC	L95% CI	U95% CI	Mean reliability	L95% CI	U95% CI
(a)						
Exact	0.714	0.658	0.767	0.926	0.906	0.940
Category	0.779	0.732	0.822	0.946	0.932	0.959
Individual frequency	0.676	0.607	0.739	0.913	0.885	0.925
(b)						
Exact	0.752	0.701	0.800	0.938	0.921	0.952
Category	0.803	0.760	0.842	0.953	0.941	0.964
Individual frequency	0.696	0.633	0.754	0.901	0.873	0.925

generated the highest measures of inter-observer consistency, and the individual frequency methodology the lowest. The ML category mean ICC (0.80) was significantly higher than that of the individual frequency methodology (ICC = 0.69; paired t-test: $t_5 = 6.2$, $p = 0.002$), and also significantly higher than when using the ML exact methodology (ICC = 0.75; paired t-test: $t_9 = 6.0$, $p < 0.001$). The ML exact correlations were significantly higher than those generated by the individual frequency methodology (paired t-test: $t_5 = 3.3$, $p = 0.022$). A similar pattern was observed for the reliability values (Table 4b).

3.4. Consistency within sub-groups of species

Calculating abundance indices using the ML category method generated the highest ICC and reliability values. We thus selected this method to test how inter-observer consistency of abundance indices varied between sub-groups of species. Analysing only the abundance indices of our 12 species of special conservation concern (see Appendix Table S1) produced the highest consistency measures of any analysis in our study (Fig. 1), with a mean ICC of 0.843 (95% CI: 0.690–0.944) and a mean Cronbach's alpha of 0.964 (95% CI: 0.917–0.988). Species predominantly identified vocally also showed high consistency (Fig. 1) with a mean ICC of 0.778 (95% CI: 0.718–0.833) and a mean Cronbach's alpha of 0.946 (95% CI: 0.927–0.962). For predominantly visually identified species there was a tendency for both mean ICC and reliability to increase in the second survey period, but these values were much reduced compared to vocally detected species (Fig. 1). On average the mean ICC was 0.362 (95% CI: 0.208–0.546), and the mean Cronbach's alpha was 0.728 (95% CI: 0.545–0.854).

4. Discussion

The MacKinnon lists technique allowed species abundance indices to be rapidly estimated for a large number of species, thus in just eight mornings of fieldwork the team was able to produce species abundance indices for 74% (156 species) of the avian assemblage. Intra-class correlation coefficients and reliability metrics demonstrated that these abundance indices were highly consistent within and between observers even though samples were collected by observers with different levels of expertise and during two periods when avian activity levels and thus detectability of individuals varied. The ML technique worked particularly well for the twelve species of special conservation concern. This group comprised a diverse set of species that did not appear to share a particular suite of traits other than their conservation interest (see Appendix Table S1). It thus seems likely that the ML technique was especially successful for species of conservation importance because each observer paid particular attention to these species when familiarising themselves with the avifauna and when conducting fieldwork. This suggests that the ML method could be an effective way of rapidly monitoring abundance indices of any group of species that were of special interest, be it for conservation reasons or other purposes.

To our knowledge this is the first time that a biodiversity monitoring methodology has been shown to produce consistent abundance indices for the majority of species in a tropical forest bird

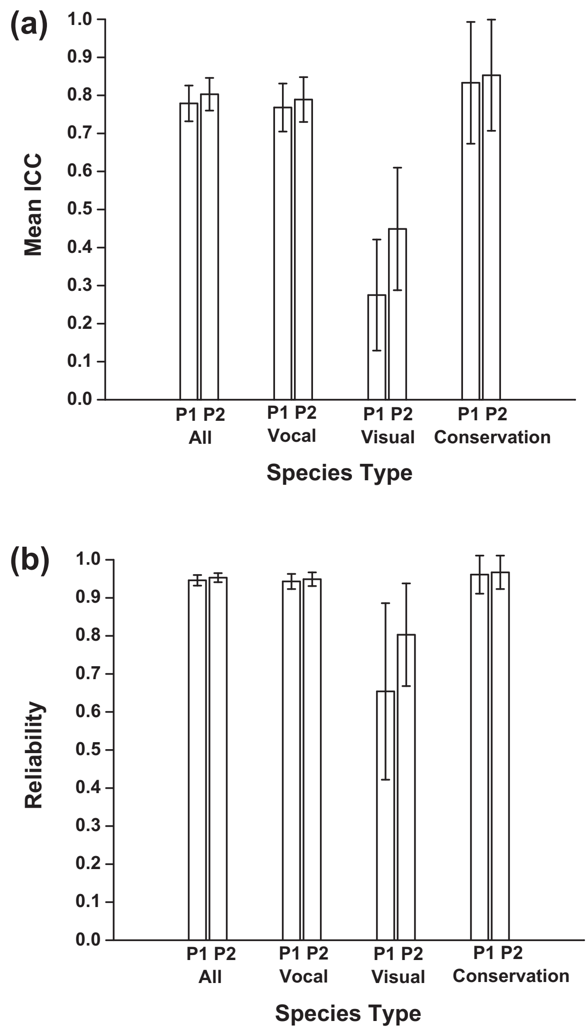

Fig. 1. Inter-observer consistency of species abundance indices. These are calculated for the entire assemblage, and for species that are mainly identified vocally, mainly identified visually and are of special conservation concern. Consistency is measured by (a) mean intra-class correlation coefficients and (b) Cronbach's alpha reliability. Error bars represent 95% CI. P1 and P2 refer respectively to the first and second observation periods.

community within a short time period. In one of the most detailed comparisons of bird census techniques in tropical forests Raman (2003) was able to estimate abundance for just 13 of the approximately 68 resident species (19%) that regularly occurred in the rainforest of the Western Ghats (India). Therefore, despite the relatively low avian diversity of the Western Ghats system, and a minimum of 18 days of fieldwork by the observer per method, each of the traditional survey methodologies used (line transects, point counts and territory mapping) were only able to generate sufficient data to estimate species abundances for the commonest species and many of these estimates had very wide confidence intervals (Raman, 2003). This is frequently the case for these standard methodologies that attempt to estimate absolute abundance rather than abundance indices (Bibby et al., 2000). In contrast, on average, each observer in our study was able to detect and produce abundance indices for 87 species from four mornings of fieldwork (Table 1). The ability of the ML method to generate consistent species abundance indices suggests that the method has excellent potential as a rapid assessment and monitoring tool. Further testing will help determine just how useful the method can be. We suggest it will be particularly important, although challenging, to study how relative abundance indices generated by the ML method compare to density estimates generated by more traditional methodologies.

The ML methodologies worked well for the avifauna as a whole and for species that were mainly identified vocally, which represented the majority of species in this study, as is commonly the case in tropical forest environments (e.g. Bibby et al., 2000). However, the method worked less well for the approximately 20% of species predominately identified visually. We believe this is because observers varied in how they combined the use of vocal and visual detection methods, with those focusing on aural identification and sound recording perhaps doing so at the expense of visually detecting non-vocalising individuals. Some evidence for this is provided by the large confidence intervals for the consistency metrics relating to visually detected species, which reflect considerable variation within different observers' assessments, with ICCs ranging from 0.167 to 0.568. It is also possible that variation between observers was more pronounced with visually identified species because, in contrast to the use of sound recording to confirm vocal identification, no physical evidence is routinely collected for later verification.

We suggest problems in combining different bird detection methods within a single survey method are unlikely to be a challenge that is unique to the ML method. In traditional survey methods such as distance sampling, there also seems to be considerable potential for variation in abundance estimates to arise from the greater difficulty in accurately estimating distance to a vocalising individual than one detected visually (Alldredge et al., 2007, 2008). We suggest that when using the ML methods it would be useful to increase the constraints placed on observer methodology and to assess whether such constraints can generate even greater consistency for the avifauna as a whole and for visually identified species in particular. For example, observers could choose to use the ML method only for vocally detected species or only for visually detected species or could aim to put equal emphasis on detecting species both vocally and visually. Further work is also required to assess how abundance indices generated by the ML methodology respond to large variation in species richness and population size.

Although intra-class correlation coefficients and reliability metrics are widely used in many scientific disciplines, they have been neglected by ecologists and conservationists. We provide, to the best of our knowledge, the first example of how these analyses can indicate whether survey methods generate consistent results under the widely varying conditions in which conservation monitoring data are typically collated. Although we have shown the ML method can produce very reliable results, we are not suggesting that the methodology is suitable for generating absolute abundance (density) estimates. Density estimates are essential for answering many interesting scientific questions and for estimating population sizes (e.g. Bibby et al., 2000; Sutherland, 1996) and in these cases the ML method will not be a suitable replacement for traditional methods that are capable of producing absolute abundances estimates. Calculating the latter can, however, only be achieved using complex and detailed survey methods such as distance sampling. Broadly speaking, these will be of limited use for rapid conservation assessments in tropical forests for two reasons. First, distance sampling rests upon a number of critical assumptions, including that all individuals along the transect route are detected, and that individuals do not move prior to detection. It seems highly likely that such assumptions are invalid in structurally complex habitats such as tropical forest (Bibby et al., 2000). Second, methods such as distance sampling require extremely intensive survey efforts to generate absolute abundance estimates, and these can thus only typically be calculated for a small proportion of the species in diverse assemblages (de Thoisy et al., 2008; O'Dea et al., 2004; Peres, 1999; Raman, 2003). Moreover, such species are often the commonest ones that are thus usually of limited conservation interest (but see Lloyd (2004)).

The ML category methodology produced the highest consistency in each test, but it was rarely significantly better than the ML exact methodology. We thus believe that both methods could be useful for conservation assessments and monitoring. The individual frequency methodology generated less consistent results, and we thus do not recommend its use. The reduced consistency is probably partly because the abundance indices generated by the individual frequency methodology are highly dependent upon whether large flocks of a single species, such as parrots, are present during a survey. The individual frequency method will also have lower consistency because it is more sensitive to variation in observer ability than alternative methods in which sample lists are comprised of a greater number of species. This is because for two observers to generate the same result, i.e. the presence of a species within a sample list, they both need to detect the focal species once during the sample. The probability of this occurring is much lower when lists comprise just a single individual, than when lists comprise a greater number of species.

The consistency of the ML sampling methodologies was not dependent on observer experience of the focal avifauna. This strongly indicates that observers which vary in expertise should be able to generate comparable abundance estimates. Further research is required, however, to assess the minimum level of experience required as all participants in this study were competent ornithologists. A key factor that was related to the consistency of abundance indices was survey effort, measured by the number of list samples generated. We thus suggest that, until consistency at lower survey efforts is assessed, survey effort should not be much smaller than the lowest used in this study (i.e. 23 list samples).

In conclusion, the ML methodology appears to be well suited for biodiversity monitoring in the tropics. It can certainly rapidly generate consistent indices of species abundance for large proportions of species rich communities, despite marked variation in observer experience and variation in avian activity. Although our study focused on birds, the method could be employed for species rich assemblages of other taxonomic groups. Potential applications could include, but are not limited to, amphibians, large terrestrial mammals, bats, coral reef fish, or some insects groups. Wider adoption of ML methodologies could enable conservation monitoring programmes to be implemented where they have previously been regarded as too complicated and time consuming to be feasible.

Acknowledgements

KLE is funded by the Leverhulme Trust and RCM by a Royal Society of Edinburgh Scottish Government Personnel Research Fellowship. We gratefully acknowledge the help of Colección Boliviana de Fauna for organizing the fieldwork permits. We thank all those who helped and supported the Yungas 2001 expedition during which the field data were collected. This expedition or preparation of this paper were financially supported by the Thriplow Charitable Trust, BP Conservation Programme, Royal Geographic Society (Rio Tinto Award), Oxford University (A.A. Paton Fund and Exploration Council), British Ecological Society, Gilchrist Educational Trust, Glasgow University, The Russell Trust, Royal Society of St George Award, Chester Zoo, Scottish Royal Geographic Society, Albert Reckitt Charitable Trust, BOU and the Darwin Initiative.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.biocon.2010.12.008.

References

Alldredge, M.W., Simons, T.R., Pollock, K.H., 2007. A field evaluation of distance measurement error in auditory avian point count surveys. Journal of Wildlife Management 71, 2759–2766.
Alldredge, M.W., Pacifici, K., Simons, T.R., Pollock, K.H., 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.
Bibby, C., Burgess, N., Hill, D., Mustoe, S., 2000. Bird Census Techniques. Academic Press, London.
BirdLife International, 2000. Threatened Birds of the World. Lynx Edicions and BirdLife International, Barcelona, Spain and Cambridge, UK.
BirdLife International, 2008. Threatened Birds of the World. CD-ROM. BirdLife International, Cambridge, UK.
Cohen, L., Holliday, M., 1996. Practical Statistics for Students. Paul Chapman Publishing, London.
Cronbach, L.J., 1951. Coefficient alpha and the internal structure of tests. Psychometrika 16, 297–334.
de Thoisy, B., Brosse, S., Dubois, M.A., 2008. Assessment of large-vertebrate species richness and relative abundance in Neotropical forest using line-transect censuses: what is the minimal effort required? Biodiversity and Conservation 17, 2627–2644.
Fjeldsa, J., 1999. The impact of human forest disturbance on the endemic avifauna of the Udzungwa Mountains, Tanzania. Bird Conservation International 9, 47–62.
Herzog, S.K., 2008. First ornithological survey of Cordillera Mosetenes, with a latitudinal comparison among lower Yungas bird communities in Bolivia. Revista Boliviana de Ecología y Conservación Ambiental 23, 59–71.
Herzog, S.K., Kessler, M., 2006. Local vs. regional control on species richness: a new approach to test for competitive exclusion at the community level. Global Ecology and Biogeography 15, 163–172.
Herzog, S.K., Kessler, M., Cahill, T.M., 2002. Estimating species richness of tropical bird communities from rapid assessment data. Auk 119, 749–769.
Herzog, S.K., Kessler, M., Bach, K., 2005. The elevational gradient in Andean bird species richness at the local scale: a foothill peak and a high-elevation plateau. Ecography 28, 209–222.
Howell, D., 1997. Statistics Methods for Psychology, 5th ed. Wadsworth Publishing, Belmont, USA.
Lloyd, H., 2004. Habitat and population estimates of some threatened lowland forest bird species in Tambopata, south-east Peru. Bird Conservation International 14, 261–277.
Mace, G.M., Baillie, J.E.M., 2007. The 2010 biodiversity indicators: challenges for science and policy. Conservation Biology 21, 1406–1413.
MacKinnon, J., Phillips, K., 1993. A Field Guide to the Birds of Borneo, Sumatra, Java and Bali, the Greater Sunda Islands. Oxford University Press, Oxford.
Macleod, R., Ewing, S.K., Herzog, S.K., Bryce, R., Evans, K.L., Maccormick, A., 2005. First ornithological inventory and conservation assessment for the Yungas forests of the Cordilleras Cocapata and Mosetenes, Cochabamba, Bolivia. Bird Conservation International 15, 361–382.
Marsh, D.M., Trenham, P.C., 2008. Current trends in plant and animal population monitoring. Conservation Biology 22, 647–655.
Maxwell, D., Jennings, S., 2005. Power of monitoring programmes to detect decline and recovery of rare and vulnerable fish. Journal of Applied Ecology 42, 25–37.
McGraw, K.O., Wong, S.P., 1996. Forming inferences about some intraclass correlation coefficients. Psychological Methods 1, 30–46.
Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B., Kent, J., 2000. Biodiversity hotspots for conservation priorities. Nature 403, 853–858.
Newson, S.E., Evans, K.L., Noble, D.G., Greenwood, J.J.D., Gaston, K.J., 2008. Use of distance sampling to improve estimates of national population sizes for common and widespread breeding birds in the UK. Journal of Applied Ecology 45, 1330–1338.
O'Dea, N., Watson, J.E.M., Whittaker, R.J., 2004. Rapid assessment in conservation research: a critique of avifaunal assessment techniques illustrated by Ecuadorian and Madagascan case study data. Diversity and Distributions 10, 55–63.
Peres, C.A., 1999. General guidelines for standardizing line-transect surveys of tropical forest primates. Neotropical Primates 7, 11–16.
Pollock, J.F., 2006. Detecting population declines over large areas with presence–absence, time-to-encounter, and count survey methods. Conservation Biology 20, 882–892.
Poulsen, B.O., Krabbe, N., Frolander, A., Hinojosa, M.B., Quiroga, C.O., 1997a. A note on 20-species lists. Bird Conservation International 7, 293.
Poulsen, B.O., Krabbe, N., Frolander, A., Hinojosa, M.B., Quiroga, C.O., 1997b. A rapid assessment of Bolivian and Ecuadorian montane avifaunas using 20-species lists: efficiency, biases and data gathered. Bird Conservation International 7, 53–67.
Raman, T.R.S., 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.
Shrout, P.E., Fleiss, J.L., 1979. Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin 2, 420–428.
Siegel, S., Castellan, N.J., 1999. Nonparametric Statistics. McGraw-Hill, New York.
Soria-Auza, R., Hennessey, A.B., 2005. Áreas importantes para la conservación de las aves en Bolivia. In: BirdLife International y Conservation International. Áreas importantes para la conservación de las aves en los Andes tropicales: sitios

prioritarios para la conservación de la biodiversidad. BirdLife International, Quito, Ecuador, pp. 57–116.
SPSS, 2008. SPSS for Windows. SPSS Inc., Chicago, USA.
Stattersfield, A.J., Crosby, M.J., Long, A.J., Wege, D.C., 1998. Endemic bird areas of the world. Priorities for biodiversity conservation. Birdlife Conservation Series 7, 1–846.
Stotz, D.F., Fitzpatrick, J.W., Parker, T.A., Moskovits, D.K., 1996. Neotropical Birds Ecology and Conservation. University of Chicago Press, Chicago, USA.
Sutherland, W.J.e., 1996. Ecological Census Techniques: A Handbook. Cambridge University Press, Cambridge.
Trainor, C.R., 2002. Status and habitat associations of birds on Lembata Island, Wallacea, Indonesia, with reference to a simple technique for avifaunal survey on small islands. Bird Conservation International 12, 365–381.