AVIAN COMMUNITIES OF LUNDY 2008-2016

by
THOMAS E. DICKINS[1] AND LAYLA TWIGGER[2]
[1]Faculty of Science and Technology, Middlesex University, London, NW4 4BT
[2]Independent Researcher
[1]*Corresponding author, e-mail: t.dickins@mdx.ac.uk*

ABSTRACT
Nine years of citizen science data about avian abundance on Lundy are analysed for diversity and similarity. Diversity is relatively stable across the nine-year period, as is similarity in year-to-year comparisons, but there is approximately 30% dissimilarity between years. This appears to be caused by migrant and vagrant birds, and passerines predominate. We make comment on the nature of the data collected, its limitations and also suggest how to improve data collection. We see this as a first step, and it is our hope that this article will stimulate further analyses of existing data, and encourage current and future researchers to further develop data collection and analytic techniques to better understand Lundy's ecology. Such activity will be of direct use to management and conservation efforts on the island but also of intrinsic scientific interest.

Keywords: *Lundy, avian community, diversity, dominance, citizen science*

INTRODUCTION

The Lundy Field Society (LFS) has reported annual bird lists since 1948, compiled from the logged observations of visitors and residents including the wardens. Various individuals have compiled these lists across the decades and the level of detail and manner of recording have differed accordingly. The bird lists are published in the Annual Report and are available online as PDF documents (http://tinyurl.com/yxv5xmov). This kind of citizen science is far from new, as the timeline attests, and it represents a venerable part of the history of ecology (Silvertown, 2009).

From 2008 Tim Davis and Tim Jones have compiled the LFS bird lists in a systematic fashion, making these citizen science data more readily processed. This came after their excellent overview of prior Lundy bird reports, published in book form (Davis & Jones, 2007). Their reports give maximum monthly counts of birds across the year, but also detailed notes of rarities, migrants and vagrants. Commentary is also made where systematic survey work has been conducted in a given year, such as the Royal Society for the Protection of Birds (RSPB) and Joint Nature Conservancy Council (JNCC) seabird monitoring work. Nonetheless, all counts recorded come from informal observational work.

Whilst count data collected by volunteers are of considerable use, it does have weaknesses. Notably, the LFS observations have not been systematically derived using standard techniques, such as set transects, point surveys or territory mapping (Bibby *et al.*, 2000). The data are also compiled by multiple observers, a majority of whom are most likely regular visitors and others who are not. Within this group of observers, abilities to observe and identify will be varied. Furthermore, it is likely that survey effort will be unequal across all observation bouts. This is in part because of the highly seasonal nature of the observations on Lundy. More expert observers tend to visit in spring and autumn, leaving the rest of the year to less expert visitors and the warden. The current warden is an extremely experienced observer, but this has not always been the case, and in the early years when the LFS employed a warden, that person was more often not present between October and March (we are indebted to Tim Davis and Tim Jones for this detail). Gaps in data have necessarily emerged as a consequence.

All of these factors will introduce considerable inconsistency and an unknown degree of error into the data, leading to wide confidence intervals about any point estimate (Snäll *et al.*, 2011). As a consequence any weak population trends will go unnoticed. As Snäll *et al.* point out, the solution to this is to focus on common species where detection is more assured, leaving scarcer birds to formally designed and executed studies. This then makes detailed monitoring an issue of resources, as formal studies are time intensive and rely on a specific level of expertise.

It is possible to design citizen science monitoring programmes that reduce some of the problems listed above (Conrad & Hilchey, 2011) and both the British Trust for Ornithology (BTO) and the RSPB rely on such well considered schemes, and we shall make some recommendations for Lundy in the Discussion. For now, however, we would note that the activity of watching and counting any species on Lundy is useful for reasons beyond formal monitoring. As Conrad and Hilchey make clear, such activities help to educate the visiting population about conservation issues but also help to cement relevant social bonds among those who are well placed to care for the site, with hope such sentiment and action will generalise beyond the island.

Davis and Jones have made a remarkable contribution through their work, one that has made bird reporting on the island more accessible and, as a consequence, opens the data to systematic scientific enquiry, with due caution around data distribution and error. Tracking avian biodiversity and abundance can help to understand the community ecology of these species. As avian species largely consist of mesotrophic predators, with many seed eaters relying upon insects during the breeding season, the community ecology of birds is indicative of the general biodiversity and ecological health of a site (Bibby *et al.*, 2000). In this article we present a series of basic analyses of bird community structure on Lundy drawn from data reported between 2008 and 2016. The principal analyses we present are basic diversity, dominance and similarity indices for breeding and non-breeding birds on Lundy across nine years of data. These indices are effectively tracking alpha diversity of the whole site. In order to drill into beta diversity we would require observations to be tied to habitat and related data, which are not currently available. We then go on to make some comments about the avian community structure of the island before moving onto concluding comments and recommendations.

The aim of our work is to indicate what can be done with such data and to encourage further, and more sophisticated, analyses, by current and future researchers. We note that the Annual Reports also include data on many other aspects of the flora and fauna of Lundy, which presents not only the possibility of similar analyses for other taxonomic groups, but also for integrated data set creation. Our choice of birds, to illustrate the potential of all LFS data, is perhaps born of our other interests, but in no sense should be read as an exclusive recommendation. Thus, we hope that the analyses we present are of intrinsic interest, but more importantly, that they present a starting point for analysing specific scientific questions around, for example, the impact of island management and conservation schemes. Moreover, such data could be usefully related to secondary data on weather and climate, as well as data from other neighbouring sites, such as the island of Skomer. To this end, we hope to initiate a concerted scientific effort. We believe collecting data is only one half of the custodial responsibility we have for sites such as Lundy. The other half is the use of it.

METHODS

Raw data preparation

Our first task was to record the data reported in the LFS Annual Report bird lists using Microsoft Excel 2011™. We used this format as Excel files and comma separated variable (CSV) files could be readily shared between the two authors, but also because Excel enabled the construction of key formulae for later calculations. All analyses reported were conducted on MS Excel 2011™, R (R Development Core Team, 2009) and IBM SPSS v24 on an iMac OS10.13.

The basic data we recorded for all 214 species observed between 2008 and 2016 were: maximum count per month within each year; their British status (from the British List 2017 found at: https://www.bou.org.uk/); and their Lundy status following Davis and Jones. This last amounted to whether or not they were resident breeders, Lundy vagrants, British vagrants, Devon rarities, migrants, etc. We then added Latin nomenclature for each species, as well as information about Order, Family and Genus. Additionally we recorded basic foraging data following descriptions from *The Birds of the Western Palearctic* (BWP) (Snow & Perrins, 1997). The non-exclusive categories for food or dietary choice can be seen in Table 1.

Table 1: Foraging sub-category of food choice: each option was recorded in terms of presence or absence for each species

Category	Options
Food choice	Vertebrates
	Invertebrates
	Plants

RESULTS

Highest count data and species richness

Table 2: Species and bird counts with mean bird count and lower (LCI) and upper 95% confidence intervals (UCI), and Menhinick's Index (MI) of species richness across nine years of citizen science data collection on Lundy

Year	Total Species	Total Count	Mean Count	LCI	UCI	MI
2008	143	47577	332.71	124.06	541.36	0.66
2009	132	71926	544.89	120.15	969.63	0.49
2010	147	61919	421.22	162.00	680.44	0.59
2011	140	42998	307.13	148.39	465.87	0.68
2012	150	78623	524.15	167.74	880.57	0.53
2013	128	38085	383.87	176.01	591.72	0.58
2014	143	35492	248.20	115.12	381.27	0.76
2015	136	60534	445.10	83.38	806.82	0.55
2016	139	48490	348.85	110.25	587.45	0.63

As can be seen from Table 2 the confidence intervals around the count means are large, which introduces a great deal of uncertainty when attempting to derive point estimates of abundance across the nine years of data collected. This is in keeping with the observations made in the introduction (Snäll *et al.*, 2011) but a portion of that may well be driven by fluctuations in migrant numbers. We shall return to this point.

Whilst the total number of species recorded gives basic species richness data, this does not account for sampling effort. Put simply, the more sampling done, the greater the number of species one would expect to record. Given this, Menhinick's Index was calculated (Gardner, 2014), as a form of standardisation, by dividing the total number of species in each annual sample by the square root of the total number of individual birds in each annual sample. The mean for Menhinick's Index=0.61, with a lower 95% confidence interval of 0.56 and an upper 95% confidence interval of 0.66. The coefficient of variation for Menhinick's Index is 0.14. The coefficient of variation should be read as a percentage, thus this represents 14% variation for this index. Whilst this is a descriptive statistic, with no inferential test for significance, at this point we have no reason to reject the hypothesis that richness has been stable across the sampled nine years.

Diversity

Diversity indices use the proportion of each species' contribution to the overall total count in order to produce an unbounded measure of species diversity. We calculated two forms of Shannon diversity (H-index), using a natural log and log base 2 standardisation of proportionality of representation. The natural log method is the most common, and we provide it for the purposes of comparison between studies. However, the log base 2 version is more readily interpretable as the number of yes-no questions one would have to ask to ascertain the identity of a species drawn at random from that sample (Gardner, 2014). We further calculated the exponential of the natural log index (ExpH) in order to give a measure of effective species, which is the number of equally common species within a sample and gives a more intuitive sense of true diversity (Jost, 2006) (Table 3). The mean for the log base 2 H-Index=4.04, with a lower 95% confidence interval of 3.82 and an upper 95% confidence interval of 4.25. The coefficient of variation is 0.08. This gives no reason to reject the hypothesis that diversity has been stable across the nine years. The mean for ExpH=16.78, with a lower 95% confidence interval of 14.39 and an upper 95% confidence interval of 19.17. The coefficient of variation is 0.22. This indicates relatively more variance in the effective species across nine years, and thus diversity, compared with those estimates derived from Shannon calculations. Nonetheless, we consider this indicative of core stability.

Table 3: Two measures of Shannon diversity (natural log (H-LN) and log base 2 (H-Log 2)) and effective species number (ExpH) across nine years of citizen science data collected on Lundy

Year	H-LN	H-Log 2	ExpH
2008	2.86	4.13	17.49
2009	2.46	3.55	11.74
2010	2.82	4.06	16.69
2011	3.09	4.45	21.91
2012	2.76	3.97	15.72
2013	2.99	4.28	19.48
2014	3.05	4.39	21.02
2015	2.45	3.54	11.61
2016	2.73	3.94	15.36

Similarity/Dissimilarity

We converted count data to presence-absence data for all species in all years. This allowed calculation of the similarity between annual samples in terms of species composition but also attenuated issues relating to large confidence intervals as a result of observer and observational effort differences and migration. We used both the Jaccard and Sorensen indices for this purpose, as two of the most commonly used (Gardner, 2014), and ran similarity comparisons across year pairs in a temporal sequence in order to give a sense of change in composition over the total sampling effort. These trends are displayed in Table 4. The indices are bounded as a 0-1 scale and should be read as the amount of similarity at species level across comparisons.

Table 4: Two indices of similarity across eight pairs of years derived from nine years of citizen science data collected on Lundy. A value can be read as a percentage similarity, thus using Jaccard, there was an approximately 70% similarity in species composition between 2008 and 2009; leaving an approximate 30% dissimilarity. The coefficient of variation (CV) is given for each index

	2008-2009	2009-2010	2010-2011	2011-2012	2012-2013	2013-2014	2014-2015	2015-2016	CV
Jaccard	0.70	0.74	0.74	0.73	0.70	0.69	0.78	0.76	0.04
Sorensen	0.82	0.85	0.85	0.84	0.82	0.82	0.87	0.87	0.03

Whilst similarity indices yield slightly different absolute values, they consistently track patterns in fluctuation and are suitably standardised to allow comparison to be made between communities, in this case annual populations of birds on Lundy. Given this, we focused on Jaccard's index. The mean for Jaccard's Index=0.73, with a lower 95% confidence interval of 0.71 and an upper 95% confidence interval of 0.75. The coefficient of variation for Jaccard (0.04) gives no reason to reject the hypothesis that similarity of species composition was similar across the year-on-year comparisons.

We wanted to track certain aspects of similarity across the nine years of data, and in particular we were interested in the kinds of birds that were present for all nine years, or proportions of that overall time. In other words, we wished to better understand aspects of the core community structure on Lundy. To do this we analysed the data at the Order, breeding status (yes-no) and dietary choice levels, as this packaged species into phylogenetically related groupings (to make visual analysis more tractable), those using the island for continuation, and those with some functional commonality at the dietary guild level.

Table 5: The number of species within each Order across years present. Thus Nine indicates each contributing species was present for all nine years, whilst One indicates presence for only one of the survey years. This is a heat map, such that red indicates extreme absence, and green high presence

	Number of years present								
ORDER	Nine	Eight	Seven	Six	Five	Four	Three	Two	One
Anseriformes	2	2	0	0	1	1	1	5	4
Galliformes	0	0	0	0	0	0	0	1	0
Gaviiformes	1	0	0	0	0	1	0	0	1
Procellariiformes	4	0	0	0	0	0	0	2	1
Suliformes	3	0	0	0	0	0	0	0	0
Pelecaniformes	1	1	0	0	0	0	0	0	0
Podicipediformes	0	0	0	0	0	0	0	0	2
Accipitriformes	2	0	2	0	1	0	0	4	2
Falconiformes	3	1	0	0	0	0	0	0	0
Gruiformes	1	0	0	0	0	0	0	0	2
Charadriiformes	15	9	0	2	3	1	5	7	4
Columbiformes	3	1	1	0	0	0	0	0	0
Cuculiformes	1	0	0	0	0	0	0	0	0
Strigiformes	1	0	0	0	0	1	0	0	0
Caprimulgiformes	0	0	0	0	0	0	1	0	0
Apodiformes	0	1	0	0	0	0	0	0	1
Coraciiformes	0	0	0	0	0	0	1	0	1
Bucerotiformes	0	0	0	0	0	1	0	0	0
Piciformes	0	1	1	0	0	0	0	0	0
Passeriformes	52	4	4	3	5	3	6	8	15

Table 6: The number of species within each dietary choice category across years present. Thus Nine indicates each contributing species was present for all nine years, whilst One indicates presence for only one of the survey years. Note that there were data missing for four species for dietary choice as this was not recorded in BWP. This is a heat map, such that red indicates extreme absence, and green high presence

	Number of years present								
DIET	Nine	Eight	Seven	Six	Five	Four	Three	Two	One
All foods	11	2	0	0	2	0	1	0	4
Vertebrates and Invertebrates	12	6	2	0	1	1	5	5	8
Invertebrates and Plants	38	5	4	3	4	3	4	11	5
Vertebrates and Plants	0	0	0	0	0	0	0	1	0
Vertebrates	8	0	1	0	2	2	0	0	4
Invertebrates	19	5	1	2	0	1	3	5	8
Plants	1	2	0	0	1	1	0	4	2

Tables 5 and 6 give a sense of how the bird communities on Lundy have differed across the nine-year period analysed in this paper. Breeding birds are present in all years, with one exception: an eight-year presence for one species (Chaffinch *Fringilla coelebs*). As a result we have not tabulated breeding data. More sporadic attendance is left to non-breeding birds that are migrants or vagrants. To look into Order and dietary choice more systematically, two separate hierarchical cluster dendrograms were produced based on Euclidean distance measures of dissimilarity, which use abundance data rather than mere presence-absence (Figure 1). These were created using the hclust function and the complete method in R. These diagrams show how species within particular presence categories (from one year only to all nine years) cluster in terms of similarity on Order composition and dietary choice. A visual inspection shows that the clustering patterns are very similar for categories 1,3,4 and 6 in both diagrams, but most notably that whilst 1 is a clear outlier in both analyses, the remaining categories are relatively similar to one another. This is in keeping with the presence-absence similarity analyses above.

An independent samples Kruskal-Wallis test was conducted to test the hypothesis that Order was unevenly distributed across each category of the number of years present. No significant results were returned and so we should retain the null hypothesis of even distribution. A visual inspection of Table 5 suggests that passerines dominate generally across the sample. This was tested with a series of Wilcoxon paired sample tests and found to be statistically significant (Table 7). A further independent samples Kruskal-Wallis test was conducted to test the hypothesis that dietary choice was unevenly distributed across each category of the number of years present. No significant results were returned (as above). No further analyses were conducted on dietary choice.

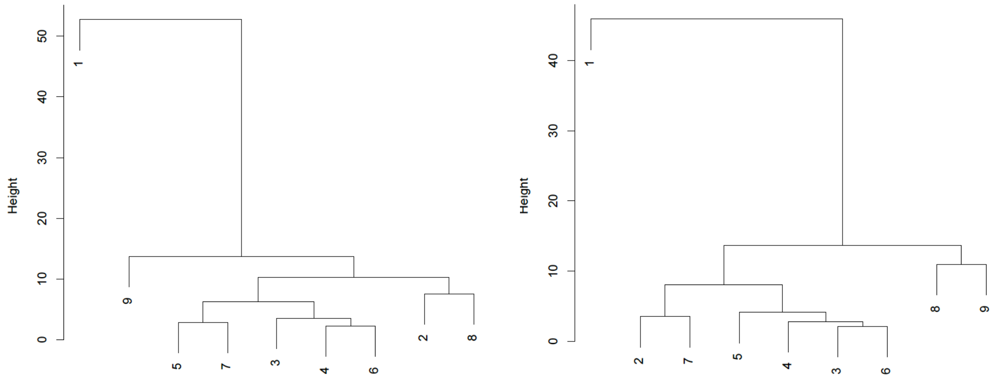


Figure 1: Hierarchical cluster dendrograms displaying dissimilarity between Order structure (left panel) and dietary choice structure (right panel) across the number of years present. The y-axis represents Euclidean distance between clusters, such that the cluster labelled 1 in each panel is in fact the most dissimilar. These clusters represent years of occurrence such that 1 means one year of occurrence for the particular species that contributed data to Order or dietary choice categories, whereas 9 means nine years of occurrence from contributing species

Table 7: Wilcoxon paired sample tests (z) comparing Passeriformes with all other Orders within the whole sample of data. Passerines clearly dominate as an Order on Lundy

	Passeriformes compared with:				
Order	Picif.	Bucerotif.	Coraciif.	Apodif.	Caprimulgif.
Z	-2.689	-2.668	-2.673	-2.675	-2.673
Asymp. Sig. (2-tailed)	0.007	0.008	0.008	0.007	0.008
Order	Strigif.	Cuculif.	Columbif.	Charadriif.	Gruif.
Z	-2.668	-2.670	-2.689	-1.844	-2.670
Asymp. Sig. (2-tailed)	0.008	0.008	0.007	0.065	0.008
Order	Falconif.	Accipitrif.	Podicipedif.	Pelecanif.	Sulif.
Z	-2.675	-2.677	-2.670	-2.675	-2.670
Asymp. Sig. (2-tailed)	0.007	0.007	0.008	0.007	0.008
Order	Procellariif.	Gaviif.	Gallif.	Anserif.	
Z	-2.673	-2.668	-2.670	-2.673	
Asymp. Sig. (2-tailed)	0.008	0.008	0.008	0.008	

Seasonal analyses

A number of the species that breed on Lundy, such as Linnet (*Linaria cannabina*), Meadow Pipit (*Anthus pratensis*), Skylark (*Alauda arvensis*) and Wheatear (*Oenanthe oenanthe*) also occur in large numbers during migratory periods. This means that the overall highest count numbers are far greater than estimates of breeding populations on the island, not least because of limited breeding habitat for these species. Whilst our focus was on the basic nature of which species are present on the island across the nine years, and therefore within each year, we also conducted a series of seasonal analyses. To do this we took highest count data reported on a monthly basis and created totals for spring (March, April, May), summer (June, July, August), autumn (September, October, November) and winter (December, January and February) across all nine years. Thus each seasonal data package consisted of 9 years × 3 months of data. Table 8 presents the basic richness and diversity analyses.

Table 8: Seasonal data calculated as total highest counts for each season across all nine years (see main text for details). The number of species registered in each season for the entire sample is given, along with the mean of the total highest counts, the 95% confidence intervals (CI) and the coefficients of variation (CV) for each compound season. Menhinick's Index (MI) of species richness and two measures of Shannon diversity (natural log (H-LN) and log base 2 (H-Log 2)) and an effective species number (ExpH) are also given for each compound season

	Species descriptives				Richness and diversity metrics			
	No. of species	Mean count	95% CI	CV	MI	Effective Species	H-LN	H-Log 2
Spring	165	1037.70	800.73	5.06	0.40	12.47	2.52	3.64
Summer	133	909.51	1001.60	6.48	0.38	7.47	2.01	2.90
Autumn	180	792.02	470.90	4.07	0.48	19.79	2.99	4.31
Winter	86	720.40	638.71	4.20	0.35	8.05	2.09	3.01

Table 8 shows that there is a large amount of variation (CV) in abundance within the compound seasonal data, but these data are not normally distributed within each season (all tests for deviation from normality returned a significant result). A non-parametric Friedman's test recommended rejecting the null hypothesis that distributions were similar across all seasons (*184.187, df=3, p=0.0001*). Multiple pair-wise comparisons, with a Bonferroni correction, revealed that only spring and autumn were not significantly different in their distribution *(-2.303, p=0.128*). Species richness was relatively stable across compound seasons, the highest values were to be found during spring and autumn. Effective species also differed between the seasons, with spring and autumn having the highest effective species numbers. This pattern is most likely a consequence of migration. The log-2 diversity indices relate to this pattern, showing that more yes-no questions would be required to identify an individual species when drawn at random from the spring and autumn samples, compared to those of the summer and winter.

Table 9: The rank abundance (RA) for the top 13 species in each season (denoted by British Ornithologists' Union Common Names (BOU)) and cumulative percentage (Cum.%) contribution to the overall count across compound seasons (see text for details)

	Spring		Summer		Autumn		Winter	
RA	**BOU**	**Cum. %**	**BOU**	**Cum. %**	**BOU**	**Cum. %**	**BOU**	**Cum. %**
1	Swallow	30.61	Manx Shearwater	54.53	Swallow	22.56	Guillemot	34.46
2	Guillemot	53.17	Guillemot	67.86	Chaffinch	37.04	Kittiwake	60.98
3	Razorbill	60.70	Razorbill	71.67	Starling	47.23	Herring Gull	75.12
4	Manx Shearwater	66.79	Kittiwake	74.57	Redwing	52.99	Starling	78.75
5	Herring Gull	70.05	Herring Gull	77.37	Meadow Pipit	58.36	Fulmar	82.30
6	Kittiwake	73.33	Linnet	79.76	Kittiwake	63.36	Razorbill	85.08
7	House Martin	76.37	Lesser Black-backed Gull	81.65	Guillemot	67.99	Great Black-backed Gull	86.85
8	Willow Warbler	79.19	Shag	83.44	Manx Shearwater	70.50	Lesser Black-backed Gull	88.58
9	Blackcap	81.63	Puffin	85.17	Linnet	72.71	Carrion Crow	90.13
10	Lesser Black-backed Gull	83.52	Gannet	86.52	Goldcrest	74.76	House Sparrow	91.47
11	Meadow Pipit	85.17	Swallow	87.72	House Martin	76.75	Redwing	92.59
12	Wheatear	86.61	Starling	88.90	Herring Gull	78.73	Skylark	93.28
13	Sand Martin	87.86	House Sparrow	90.03	Gannet	80.33	Shag	93.90

We further inspected the seasonal data by calculating rank abundance for recorded species. Table 9 displays the ranking results for species in the top 13 most numerous species during each season, where each season is a compound of 9 years × 3 months, as above. Table 9 clearly shows the summer months as dominated by breeding seabird species.

Rank abundance can obscure issues of detectability (Bibby *et al.*, 2000) meaning that those birds that are harder to detect are also registered less frequently. To inspect this, Pareto charts were constructed for each compound season (Figures 2a-d). The Pareto Principle, with reference to ornithological surveys, assumes that 80% of the birds seen belong to 20% of the species registered (Rispoli, Zeng, Green & Higbie, 2014). Looking at the number of species contributing to the 80% is another way of looking at species dominance, with due caution around detectability. Thus the top 13 ranked species are displayed in Table 9, as this number incorporates the 80% threshold for all seasons. It is worth noting that the 80% threshold varies considerably across seasons in terms of ranking, with autumn meeting it after only 13 species. The x-axis on the Pareto charts give an abridged list of birds by BOU common name in rank order to give a sense of the less abundant species.

Figure 2a: A Pareto chart for all spring bird data collected on Lundy from 2008-2016 (see main text for details). The steep decline of the blue bars (count data in rank order; left y-axis) and the sharp rise of the cumulative percentage data (red line; right y-axis) indicate relatively low diversity (see Table 8 for comparison with Shannon H indices). The list of BOU common names on the x-axis is not a complete list of all birds observed but a reduced list, in rank order, due to space constraints

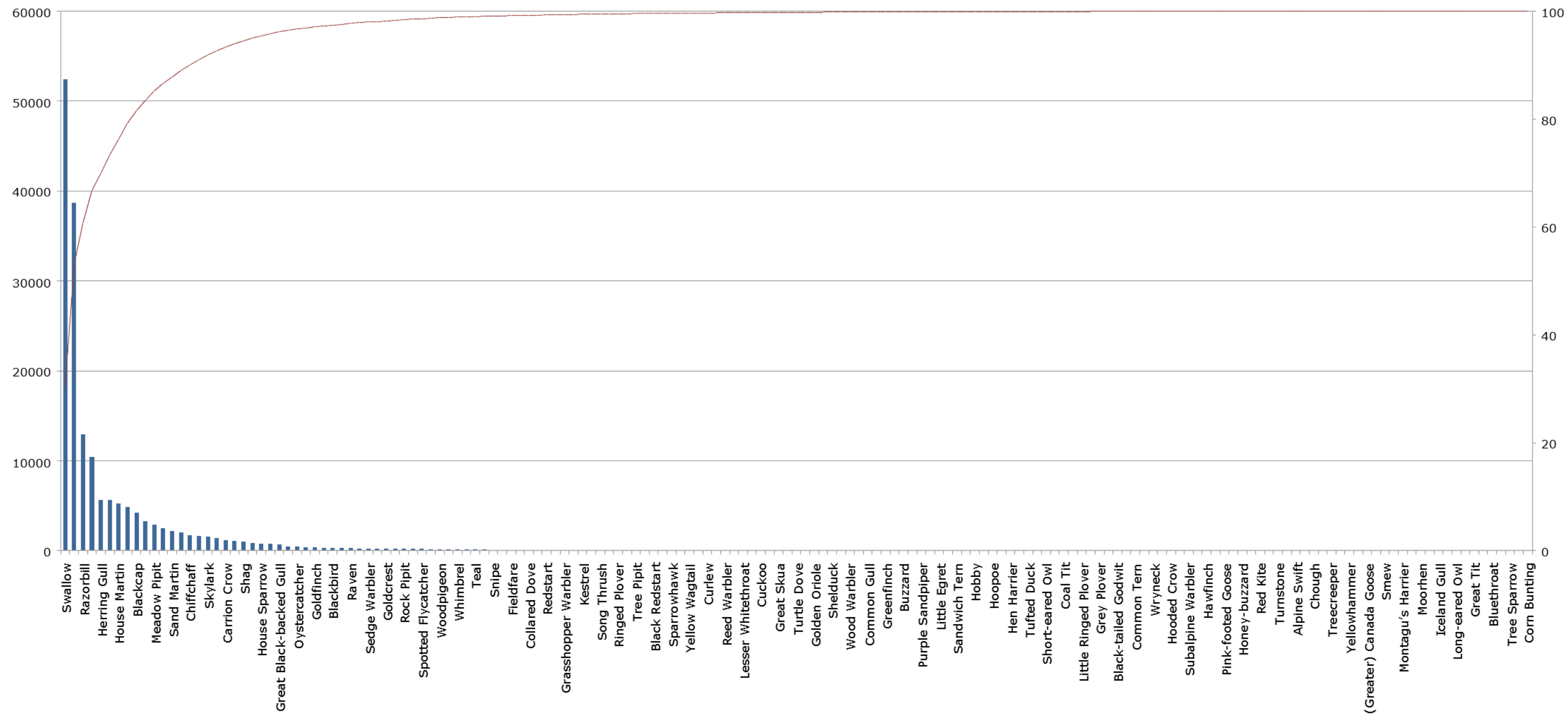


Figure 2b: A Pareto chart for all summer bird data collected on Lundy from 2008-2016 (see main text for details). The steep decline of the blue bars (count data in rank order; left y-axis) and the sharp rise of the cumulative percentage data (red line; right y-axis) indicate relatively low diversity (see Table 8 for comparison with Shannon H indices). The list of BOU common names on the x-axis is not a complete list of all birds observed but a reduced list, in rank order, due to space constraints

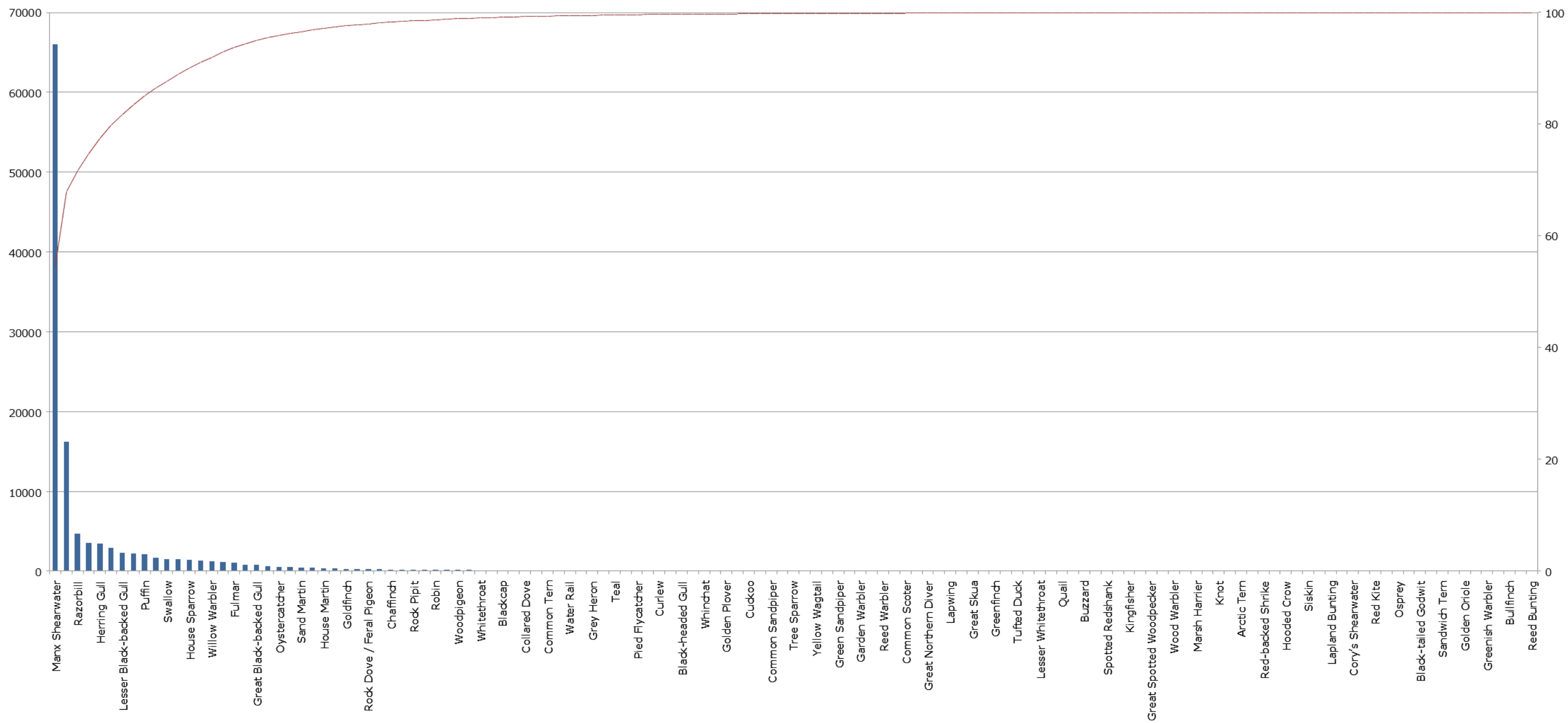
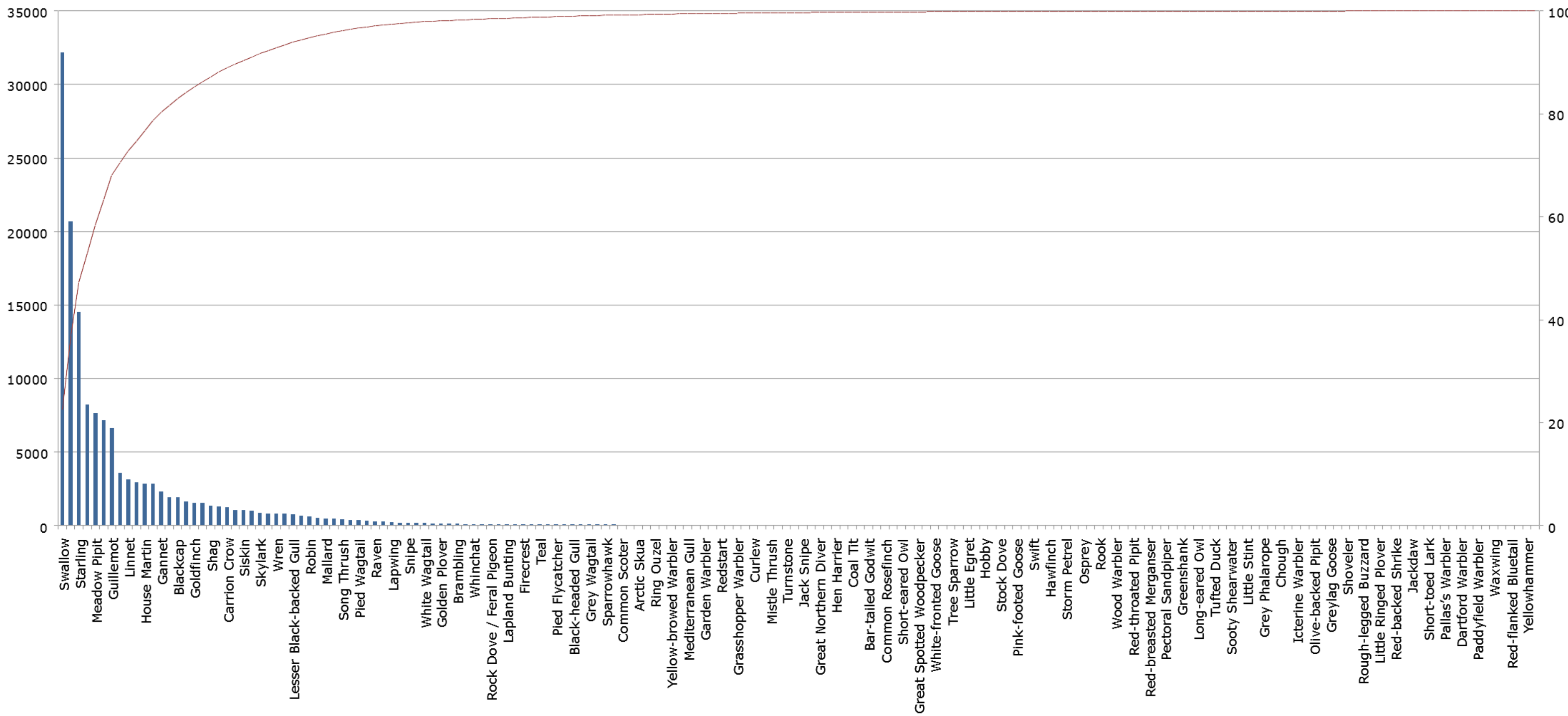


Figure 2c: A Pareto chart for all autumn bird data collected on Lundy from 2008-2016 (see main text for details). The steep decline of the blue bars (count data in rank order; left y-axis) and the sharp rise of the cumulative percentage data (red line; right y-axis) indicate relatively low diversity (see Table 8 for comparison with Shannon H indices). The list of BOU common names on the x-axis is not a complete list of all birds observed but a reduced list, in rank order, due to space constraints



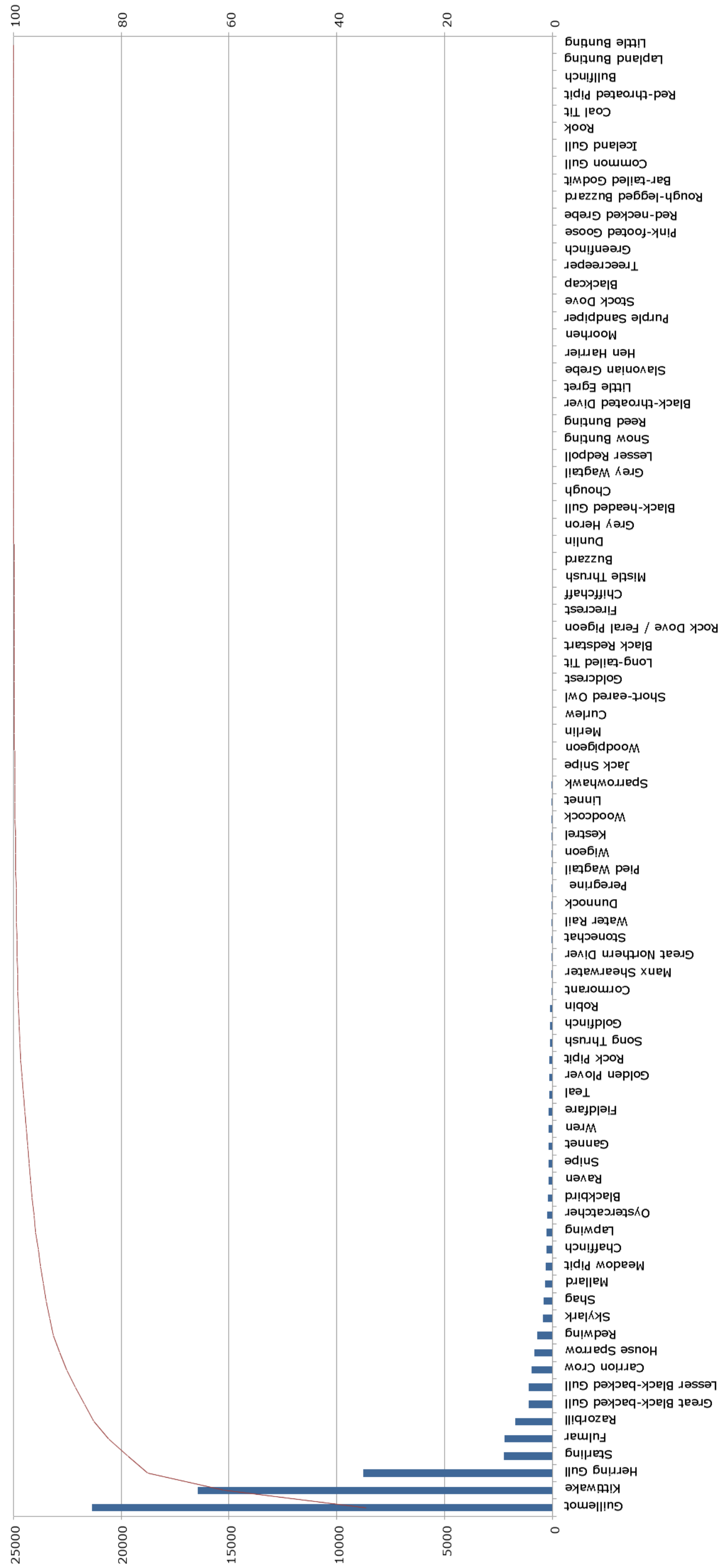


Figure 2d: A Pareto chart for all winter bird data collected on Lundy from 2008-2016 (see main text for details). The steep decline of the blue bars (count data in rank order; left y-axis) and the sharp rise of the cumulative percentage data (red line; right y-axis) indicate relatively low diversity (see Table 8 for comparison with Shannon H indices). The list of BOU common names on the x-axis is not a complete list of all birds observed but a reduced list, in rank order, due to space constraints

DISCUSSION

From our simple analyses of nine years of citizen science data about avian abundance on Lundy, we are confident that the diversity and year-to-year similarity of the population has remained fairly stable for this period. Across all nine years, 41.6% of the 214 total observed species were present at some point in time on the island. This means that more irregular avian visitors to Lundy have introduced variance into the similarity data over time. Nonetheless, a similarity comparison between 2008 and 2016 revealed a Jaccard's Index of 0.699, which is similar to the 2008-2009 comparison, but marginally lower than the lower 95% confidence interval reported in the Results section; this leaves approximately 30% dissimilarity between the species composition at the start of the sample and at the end.

In an attempt to interrogate this variance in similarity, we looked at Order and dietary choice composition across categories of the number of years a species had been present on the island. Whilst some clustering was revealed, using hierarchical dendrograms based on Euclidean distance measure of similarity, the basic distribution of Order and diet was even across all categories. If we had seen an uneven distribution, this would have been indicative of possible challenges to specific kinds of birds that might have driven dissimilarity at certain times.

At present our best hypothesis is that much of the dissimilarity across years is caused by non-breeding migrant and vagrant birds, especially as all breeding birds were present for either all nine or eight years in this sample. Our seasonal analyses support this view. Thus variance will be caused by multiple distal factors. These birds appear to be sustained in terms of diet when on the island. However, Lundy has seen a number of changes over the period of the sample. The sample began at the point that rats had been eliminated from the island, and breeding bird populations have been able to exist without the threat of that predator for the entire period. It is also the case that management practices have changed over this period, including the introduction of large herbivores that will potentially impact on invertebrate communities. Rabbit populations have risen and fallen dramatically, which could have multiple impacts on vegetation, invertebrates and possible predator-prey dynamics (Dickins *et al.*, 2018). The removal of Rhododendron on the east coast and bracken control will also have effects in terms of invertebrate communities, but also on cover and nesting habitats. Whilst all of these factors are duly noted in passing in Annual Reports and similar publications, clear data are not recorded and critically not married to the citizen science data collected in the logbooks. Efforts in this direction would yield potential reward.

Recommendations for future data collection

Re-ordering old data in order to better understand patterns in avian and other communities on Lundy is an important approach, but it does not detract from the problems of citizen science data discussed in the Introduction. We looked at all data from 1948 to 2016 before embarking on this project, and realised that the best one could do with pre-2008 data was really to note presence-absence. We decided to focus on the 2008-2016 data, as abundance data were systematically recorded (but see Tovo *et al.*, n.d., forthcoming, for ways of deriving abundance from presence-absence data).

A key problem for this project was the absence of a digital record. The ideal would be the production of data sets as comma separated variable (CSV) files for use in a number of different applications. This would allow subsequent organisation and analysis by researchers and other parties. There are many ways to do this. As an example, the BTO collate digital information through a dedicated smartphone application called BirdTrack. This enables data to be recorded at various levels of resolution, starting with just basic species identification with no count data, and this is coordinated with location data and uploaded to a central database. This is a very good system for day-to-day birdwatcher data, and becomes more useful when recruited by those who have regular patches, thus generating data both about the patch and observer learning and consistency. The BTO assiduously scan this data and will contact users when something unusual is recorded for verification. Users can look through their own data at any time.

The other system that the BTO use is online uploading of data collected in regular organised surveys at particular sites with set transects, during the breeding and winter seasons, as well as for more focused efforts, for example specific species surveys. Observers may collect data using pen and paper but this eventually finds its way to an online data set, and observers can interrogate their own findings. These surveys are supported by a large amount of online material that explains methods. This enables a certain amount of control over survey effort and also the way in which observations are collected and logged. Clearly individual differences in skill and experience can still have an influence, but the methods make each individual effort more consistent and comparable to others. The BTO is an organisation that recruits its members from dedicated bird enthusiasts who either have much experience or are prepared to gain it. This too helps to reduce variance.

Direct, digital recording of the sort used by the BTO via BirdTrack is difficult on Lundy due to inadequate mobile device coverage. This will most likely improve as 5G rolls out across the United Kingdom, but any solution relying upon mobile applications runs the risk of data loss at upload, when signal is patchy; and data loss due to observers forgetting to upload when returning to the mainland. Given this, we would not recommend developing Lundy-specific applications as a first initiative. Instead, we would recommend looking to investment in computers on the island that enable logbook entries to be made directly in digital format. Files created in this way could be stored on dedicated drives and/or uploaded to cloud storage solutions via the island management team's private internet service. Cloud storage would enable mainland researchers to access data, with due permissions, continuously. Our recommendation would be to make this data open access in keeping with current open science initiatives (see for example: https://openscience.com/ and https://cos.io/).

Computer entry logbook solutions could operate at a number of levels, as indeed the current paper logbooks do. Those visiting the island for vacations collect many observations, and whilst they may have a general interest in nature, they will not be as dedicated as those BTO surveyors tramping through fields on cold, winter days, working established transects. It is therefore important not to make logging observations effortful for these contributors. Thus the opening 'page' should simply ask the contributor to opt for the kind of observation they should like to make and those choices could include all the existing options, which are species specific count data, and more narrative based descriptions of interesting sightings and counts. Whilst this article has processed highest count data, having digitally recorded narrative accounts provides a rich source of data for other kinds of more interpretive and qualitative analyses that would greatly inform those trying to develop user friendly conservation initiatives, for example. Software packages exist in order to do this (e.g. NVIVO).

When creating a digital version of the current logbook solution, a number of modifications might be considered. First, it would be useful to ask observers to rate their own expertise and skill levels on a Likert scale from 0=novice to 3=expert. This would create a useful categorical variable for sifting data in future analyses. For example, it could be used to look at only expert collected portions of data or as a controlling variable in various forms of generalised linear modelling.

Additional levels of entry could be listed as front-page options and these could be tied to more systematic attempts at surveying the island for particular purposes. There will be many ways to do this, but any solution would need to incorporate detailed mapping of the island into portions, akin to the BTO site and transect approach, and should be accompanied by some advice on methods. Further details of systematic methods would depend upon the desired outcome of such work and this in turn would be best linked to the conservation and management interests on the island. It is perhaps at this point that more focused work pulling together observations across taxa could be implemented enabling linkage between, say, vegetative and floristic surveys, invertebrate observations and bird work. In effect a number of available projects could be offered to visitors wishing to make this level of contribution. This might also dovetail with the interests of visiting university parties looking for student projects. This leads to a final suggestion: the creation of a computer based logbook system could be done in such a manner as to coordinate with systems designed to record the professional interventions and data collection of the conservation and management teams working on the island. This would not necessarily have to be publicly available at the point of logbook entry, but data from such work could be made open access after appropriate periods of embargo.

The preceding recommendations require much fully project-managed work to design and implement, and our reviewers feel this may sit best within a formal Bird Observatory set-up as Lundy once was, and could be again. However it is managed, such a project requires a dedicated team and funding to cover both their costs, and costs associated with exploring and trialling different solutions. The manner in which we have described our recommendations reads most like a web page based, or a more formal database solution. The latter is perhaps more useful on an island with poor internet connectivity, the former is better for close to continuous updating. Hidden costs include the development of appropriate metadata for the dataset, error management procedures etc. and decisions would need to be made about just how much investment is worthwhile. It is also likely that the overall project would need to be managed in stages, with clear performance indicators and a decision process around transition between stages. We have not done this work, but we believe it is possible. We also believe it will represent a challenge to extant cultures of recording. But current practices, relying purely on pen and paper logbooks, run the risk of actively creating a non-modern archive and losing much of the richness of the amassed experiences of many dedicated visitors to Lundy.

Data note

Raw data are available at: https://figshare.com/authors/Tom_Dickins/3116847.

ACKNOWLEDGEMENT

We thank Ben Dickins for critical comments on an earlier draft of this paper, and for useful discussions about the analyses. We are also extremely grateful for the detailed commentary of Tim Davis and Tim Jones. All errors are our own.

REFERENCES

Bibby, C.J., Burgess, N.D., Hill, D.A. & Mustoe, S.H. (2000). *Bird Census Techniques* (2nd ed.). London: Academic Press

Conrad, C.C. & Hilchey, K.G. (2011). A review of citizen science and community-based environmental monitoring: Issues and opportunities. *Environmental Monitoring and Assessment*, *176* (1-4), 273-291. https://doi.org/10.1007/s10661-010-1582-5

Davis, T. & Jones, T. (2007). *The Birds of Lundy*. Exeter: Devon Bird Watching & Preservation Society and Lundy Field Society

Dickins, T.E., Neller, K. & Spencer, R. (2018). Clutch size in Kittiwakes *(Rissa tridactyla)* on Lundy. *Journal of the Lundy Field Society*, *6*, 7-26

Gardner, M. (2014). *Community Ecology: Analytical Methods using R and Excel*. Exeter: Pelagic Publishing

Jost, L. (2006). Entropy and diversity. *Oikos*, *113*(2), 363–375. https://doi.org/10.1111/j.2006.0030-1299.14714.x

R Development Core Team. (2009). R: A Language and Environment for Statistical Computing. Vienna, Austria. Retrieved from http://www.r-project.org

Rispoli, F.J., Zeng, S., Green, T. & Higbie, J. (2014). Even birds follow Pareto's 80-20 rule. *Significance*, *11*(1), 37–38. https://doi.org/10.1111/j.1740-9713.2014.00725.x

Silvertown, J. (2009). A new dawn for citizen science. *Trends in Ecology & Evolution*, *24*(9), 467–471

Snäll, T., Kindvall, O., Nilsson, J. & Pärt, T. (2011). Evaluating citizen-based presence data for bird monitoring. *Biological Conservation*, *144*(2), 804–810. https://doi.org/10.1016/j.biocon.2010.11.010

Snow, D.W. & Perrins, C.M. (1997). *The Birds of the Western Palearctic: Volumes 1 & 2* (Concise). Oxford: Oxford University Press

Tovo, A., Formentin, M., Suweis, S., Stivanello, S., Azaele, S. & Maritan, A. (n.d.). Inferring macro-ecological patterns from local species' occurrences. *Oikos*. https://doi.org/doi.org/10.1111/oik.06754