Estimating the rarity of birds and its ecological context in the University of the Philippines Diliman Campus

Benjamin Vallejo Jr.* and Alexander B. Aloy
Institute of Environmental Science and Meteorology, College of Science
Villadolid Hall, University of the Philippines, Diliman, Quezon City

In this paper we use frequentist (classical) and Bayesian inference to estimate the rarity of birds in the University of the Philippines Diliman campus. Rare species have a detection probability of 1%. As sightings of certain species of birds are extremely rare, a frequentist approach to estimation will often result in overestimates of observation precision. Using McArdle's rarity and Bayesian inference we estimated the probability of detecting rare species as between 4 to 10%. The Bayesian estimates are lower and may be a better method for estimating rarity. The coefficient of variation (CV) of the frequentist (24%) and Bayesian (25%) estimates are similar suggesting imprecision. Rare species are likely to be detected in areas of campus with common and highly abundant species. This may be an artifact of the fast paced Jokimäki method that was used in surveys of bird abundances. The higher probability of detecting rare species from the suggested 1% cutoff and similar CV estimates are likely due to the small data set used in the estimations and the lack of prior information. The ecological context of our observations is related to the increasing fragmentation of habitat in the campus as a consequence of urbanization.

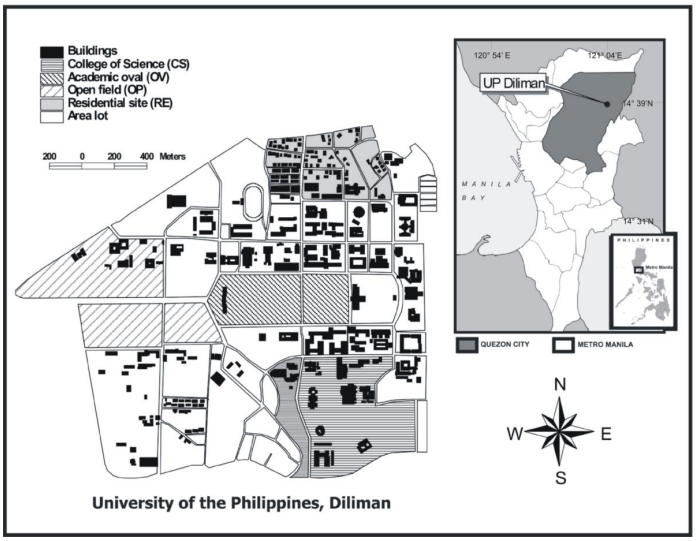

Figure 1. The University of the Philippines (UP) Diliman campus in northeast Quezon City, Metro Manila. Bird survey subplots are indicated.

Introduction

Estimating the frequency of rare events are important in planning for conservation and environmental science (Dixon, et al. 2005). As the populations of species that are vulnerable to human impacts are often low, sightings of individuals of these species may be rare. However in the faunal lists of a region, these species often contribute to an absolute estimate of species diversity (Rosenzweig 1995). Decisions on where and what to conserve are usually framed in terms of species diversity and estimates of endemism (Burgmann and Lindenmayer 1998). These estimates are often point estimates and may give unwarranted conservation importance to certain regions. As resources for protected areas management are limited, conservation planners have to perform "conservation triage" that is placing differing levels of management priorities in areas needing protection (Meffe and Carroll 1994, Catibog-Sinha and Heaney 2006).

This is one of the reasons proposed by conservation organisations on why the 4.9 km² University of the Philippines Diliman campus (14° 31' N, 121° 03' E) needs to be declared as a conservation area (Fig 1). The campus is located in rapidly urbanizing Quezon City 15 kilometers from downtown Manila. Amateur birdwatchers have listed at least 82 species while our earlier landscape ecology study has noted 53 species in the campus (Vallejo, et al. 2008). These estimates suggest that the university has conservation importance in a rapidly urbanizing region of Metropolitan Manila. This conservation importance is particular to urban areas and is different from that considered for wilderness areas. Similar importance has been given to urban parks and nature reserves in Singapore (Koh and Sodhi 2004).This region that includes the Diliman district and the nearby La Mesa Watershed in Novaliches district are the last remaining significant areas of green cover in the metropolis (Fig 2). Unlike the La Mesa Watershed, the university campus is a multi-use community that has academic, commercial, residential and government areas. It is unlikely that the whole campus can be protected but parts of the campus can receive different levels of protection.

The preponderance of rare species with small ranges in tropical regions is predicted by biogeographic theory. These rare species contrib-

*To whom the correspondence should be addressed. E-mail:bmvallejo@up.edu.ph
Received May 20, 2008; Accepted December 13, 2008.
Editor-in-charge: Porfirio M. Aliño

ute a majority of global species diversity (Brown and Lomolino 2004). Range theory predicts that range size is determined to a large part on the abundances of individuals within the distributional range (Gaston 1996).

In areas subject to human impact, rare species may be observed. The question that needs to be answered is whether these sightings are due to chance alone or not. Estimating the frequency of rare events presents a statistical challenge due to small sample size of events. If a frequentist view is adopted the true probability of a discrete rare event is denoted as π . The frequentist estimate of this probability is *p*. This is calculated as the number of observations (*n*) of an event over the total numbers of events (*N*)*p=n/N*. The standard error (SE) is given as:

$$SE_p = \sqrt{p(1-p)/N}$$

In ecological statistics an extremely rare event has a π=0.01. The frequentist estimate *p* will achieve a reasonable precision demanded by most ecological studies if the coefficient of variation CV≤ 10%. This can only occur if the sample size is sufficiently large *n*>1000. However most ecological studies are limited by logistical constraints such that sample size is usually *n*<100. The convention in most ecological studies is n ≥ 30.

In estimating the frequency of rare events, the main constraint is small sample size. It is not unusual that for a certain period of sampling in time or in space, *n* ranges from 1 to 2. Estimating the precision of such events becomes a problem. The occurrence of any large number of sightings in any single traverse will have a marked effect on the mean.

Another commonly used method of estimating species rarity include McArdle's statistic α (McArdle 1990),which estimates a probability of detection. This method is more sensitive to the detection of rare species than other classical methods.

In this paper we present frequentist and Bayesian method for evaluating the rarity of birds in the UP Diliman campus. Frequentist methods rely on empirical data from surveys while Bayesian method employs the use of prior information about the probability of a rare event π. Bayesian inference combines this prior with empirical data. If the probability estimates of the prior is similar to that of the empirical, then a great degree of precision exists for the estimate rather that from the empirical alone.

Methods

We used the Jokimäki walk (Jokimäki and Jokimäki 2003) in rapid assessment of bird abundance and diversity in four plots in the campus throughout 2005-2006. These plots are the Academic Oval (OP), College of Science (CS), Open Fields (OP), and Residential Areas (RE). (Fig.1) The plot sizes range from 20-25 ha. The method and its rationale are described in an earlier study (Vallejo, et al. 2008) and our observations span the period of 1.5 years with 20 Jokimäki walks made in each plot. In each Jokimäki walk, there is a probability that a bird species will be encountered. In many species that are abundant the probability of encounter is almost 100% as estimated using the classical frequentist approach (Eq 1). We first estimated the probability of encounter using the frequentist approach based on the number of times all bird species have been encountered and the times a rare or

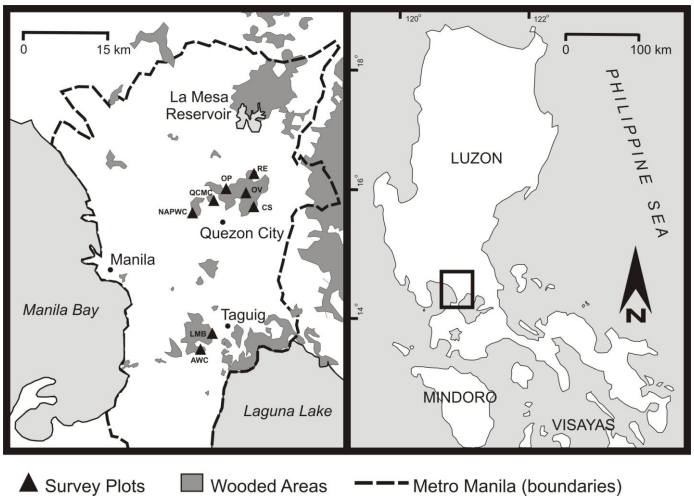

Figure 2. Map showing the location of green areas in Metro Manila. The survey plots are mapped for an earlier study of urban bird biodiversity. Plots OP, OV, RE, and CS are within the University of the Philippines Diliman campus.

uncommon species was encountered. We did not assess species whose probability of encounter is greater than 0.20. We considered these as uncommon. Species with a probability of encounter greater than 0.5 were considered common.

Classical frequentist probability estimate of rarity (Dixon, et al. 2005)

$$p=n/N, \quad (1)$$

where *p* is probability of sighting, *n* is actual sighting, and *N* equals the total number of observed events.

We defined a cut off probability estimate of rarity of 10% (Gaston 1994). We considered any species as rare if it has a probability of being detected 10% or less in a single random Jokimäki walk.

We preferred to use the more conservative frequentist estimate McArdle's statistic (Eq 2) (McArdle 1990) in our analysis. In using this estimate, we assumed that the rarest event is a single sighting. The probability of encountering a species once in the year was evaluated in each plot. McArdle's rarity statistic:

$$\alpha=1-(1-f)^N \quad (2)$$

where α=probability of detection, *f*=probability that a species will occur in a single random work, *N*=total sighting effort. We assumed that a rare species will have an *f*=0.01 as suggested by McArdle.

The standard error of the frequentist estimates was also evaluated. Because of the small sampling effort sizes, the coefficient of variation of the probability estimates is poor. To improve the probability estimate, a Bayesian approach is required (Dixon, et al. 2005).

Bayesian estimation. The frequentist estimates do not use prior information of sightings. However with the lack of previous statistical estimates of abundance and bird sightings, it is justified to use a frequentist estimate that could later be used for Bayesian estimates of rarity.

Bayesian inference deals with the probability π. This is a random variable that follows a statistical distribution. Parameters are then estimated from the statistical distribution. The distribution of a parameter

summarizes the expectation of the parameter and its variance. Bayesian inference uses empirical observations with information about the parameters from the prior distribution before the data is analyzed. A new distribution is then made that incorporates the prior. A function $f(\theta)$ is defined as the likelihood function (LF) that indicates how well the data supports the parameter θ. The value that is most supported by the data is the maximum likelihood estimate (MLE).

By combining LF with other information from earlier studies a prior probability density $f(\theta)$ is derived. Using Bayes' theorem LF and $f(\theta)$ are combined

Bayes' theorem is shown below:

$$f(\theta \mid \text{data}) = \frac{prior \times likelihood}{\int prior \times likelihood \quad d\theta} = \frac{f(\theta) f(data \mid \theta)}{\int f(\theta) f(data \mid \theta) d\theta} \tag{3}$$

The chance of encounter in any Jokimäki walk is a discrete event. It is either a rare bird is encountered (1) or is not (0). This is best modelled using a binomial distribution,

$$f(k \mid \theta) = \binom{n}{k} \theta^{k} (1-\theta)^{n-k} \tag{4}$$

where n is the number of observations in all Jokimäki walks, k is p (encounter) and $f(\theta)$ is the likelihood function (LF). This θ is the estimator for p (encounter). Since there were no earlier studies on assessing the rarity of birds in UP Diliman from where we can derive inference on the probability distribution of sightings, we assumed a uniform prior distribution $f(\theta)$. The uniform prior means that we desire an outcome based in the empirical data set alone. Using the uniform prior we assume that the probability of all values of θ are equal and that all values of θ are represented as $f(\theta)=1$ when species are encountered and $0<\theta<1$. Thus the probability of all values of θ are equal. No species is rare but all species are common since the sighting probability is 0.5.

Bayes' theorem is written as

$$f(\theta \mid k) = \frac{\theta^{k}(1-\theta)^{n-k}}{\int \theta^{k}(1-\theta)^{n-k} d\theta} = \frac{\theta^{k}(1-\theta)^{n-k}}{B(k+1, n-k+1)} = (n+1)\binom{n}{k}\theta^{k}(1-\theta)^{n-k} \tag{5}$$

$B(k+1, b-k+1)$ in Eq (5) is the beta function. The posterior in this form of the Bayes theorem is a beta density $f(\theta|k)=B(k+1, n-k+1)$ and expresses a level of certainty for values of θ. In statistical inference theory, the mode of this density is the most probable value of θ (MLE) and also occurs at the critical point, where $df(\theta|k)d\theta=0$.

We equated the density to 0 since it is one of the steps in estimating MLEs in any distribution. In order to simplify differentiation we take the logs of the beta density and get

$$d \ln f(\theta \mid k) d\theta = \frac{n-k}{1-\theta} \tag{6}$$

The main question here is whether these sightings of rare species are attributable to chance alone or not. In order to estimate the possibility of the species being observed due to chance alone, we used a uniformed flat prior. The uniform flat prior assumes that all sighting events estimates θ are equally probable such that $f(\theta)=1$ $0<\theta<1$.

Table 1. Ecological rarity probability and binomial prior probabilities.

Ecological rarity p	0.001	0.010	0.050	0.080	0.090	0.100
Prior	0.050	0.100	0.300	0.300	0.200	0.050

The relationship between the beta distribution and the binomial distribution are well known and are discussed in probability theory textbooks. The beta distribution is a convenient prior since the integral in the denominator can be easily evaluated (Lee 1989).

Detecting rare species with prior information from an assessment of ecological rarity. We used the binomial distribution as a prior to assess the possibility of detecting rare species in each plot. The observations made per Jokimäki walk are discrete events (*e.g.* the possibility of sighting of a rare bird) and are independent of each other. The prior probabilities of the events are given in Table 1

In these Bayesian estimations, we restricted our prior probabilities to represent extremely rare to rare events ($0.001<p<0.1$) as suggested by other researchers (Dixon, et al. 2005). These prior probabilities are commonly used in assessing ecological rarity. By definition a truly rare event has a probability of less than 0.01 of being detected in a survey. The results are reported as posterior probabilities.

Results

Frequentist estimation of rarity

The rare species listing is given in Table 2 using the standard frequentist probability criteria estimate ($p<0.10$). Table 3 shows classical and McArdle's rarity probability estimates

There are 25 species have only been observed once in a plot throughout the two year course of this study. There are six migrants, six endemics and one possible invasive on the list. However species in italics are rare in some plots but uncommon in others. The scaly breasted munia *Lochura punctulata* and Pied Triller *Lalage nigra* are considered rare but is observed in multiple plots. The single possible invasive is *Psittacula sp*. We were not able to verify if this species is the Alexandrine parrot *P. alexandri* that is a common cage bird. This parrot is commercially bred and hybridized with other *Psittacula* species. This parrot is a documented invasive in other Southeast Asian countries after being introduced in the late 1940s (Sodhi and Sharp 2006). In Diliman it is likely a pet bird that has escaped.

The classical probability estimates for these rare species are likely overestimates. A measure of how reliable these estimates are is in the use of coefficient of variation (CV) statistic. Since the number of Jokimäki walks is not large and considering that the rare species often exist in low abundances, there will be a large degree of imprecision of the estimates. The CV of our classical probability estimate for rare bird sightings in all plots is 30%. The CV using the McArdle's statistic is 24%. These values suggest imprecision although. the McArdle's rarity statistic tend to reduce this level. McArdle's statistic is a better frequentist estimate of rarity. McArdle's statistic assumes that each sampling event for the rare species is independent and that f is a constant.

Table 2. Rare species in this study.

Rare species		Plots where species occur
1	*Apus pacificus**	CS
2	*Amaurornis phoenicurus*	CS
3	*Collocalia esculenta*	CS
4	*Chalcochaps indica*	OP
5	*Collocalia troglodytes*[E]	OP
6	*Cacomantis variolosus*	OP, OV
7	*Centropus viridis*[E]	OV
8	*Dendrocopus maculatus*[E]	RE
9	*Ducula poliocephala*[E]	OV
10	*Dicaeum sp.*	OV
11	*Egretta intermedia**	RE
12	*Galirallus torquatus*	RE
13	*Halcyon chloris*	RE
14	*Halcyon smyrnensis*	CS
15	*Lalage nigra*	CS, OV, OP
16	*Loriculus philippensis*[E]	CS
17	*Lonchura punctulata*	CS, OV, OP
18	*Monticola solitarius*	CS
19	*Megalurus timoriensis*	CS
20	*Merops viridis*	CS
21	*Phylloscopus borealis**	OP
22	*Psittaccula sp.*[I]	OP
23	*Streptopelia chinensis*	OP
24	*Sterna/Chlidonia sp*[I]	OP
25	*Turdus sp.*	OP

Note: *migrants, [E]Philippine endemics, [I]possible invasive.

Bayesian estimation of rarity

In the College of Science plot, the posterior probabilities from Bayesian analysis is 0.042. Thus the possibility of sighting a species just once is 4.2%. The Bayesian estimates for the other plots are given in Table 4.

The chance of detecting extremely rare species sightings is less than 10%, a result that is more or less similar to McArdle's statistic, although by restricting the prior to 10% or less probability, a lower estimate results from Bayesian analysis. However since even with suggested binonimal prior probabilities, the CV in this analysis is 25%. This is close to the McArdle's CV estimate. Since the Bayesian rarity estimates are lower than the frequentist one, this is likely to be the better estimate if there is a more informative prior (Lee 1989).

The chance of detecting rare species is lowest in the College of Science plot since these have the highest species diversities as determined from our earlier study (Shannon Index CS=1.9) with only a few species predominating. The observer has a lower chance of detecting rare species in areas where there are many species at lower abundances. In the UP Residential Area there are a few predominant species that exist in high abundances. The observer is likely to detect the presence of a rare species in an ecological system with a lot of common and abundant species.

Table 3. Estimate of single sighting probability using classical probability (p) and McArdle's statistic (α).

Plot	classical probability (p) (%)	McArdle's probability (α) (%)
College of Science (CS)	11.11	7.43
Academic Oval (OV)	10.22	9.76
Open fields (OP)	11.67	7.77
Residential Area (RE)	7.89	10.05

Discussion

One of the major goals of conservation priority setting is determining which species are rare and what factors contribute to rarity (Burgmann and Lindenmayer 1998). In this study 25 species are considered rare and this accounts for 47.1% of the 53 species of birds recorded on campus. The 3 migrant species and seven endemic species represent 20% of the rare species detected in the study.

Comparison and limitations of frequentist and Bayesian analyses

Given the limited data used in analysis and the lack of prior information, the frequentist and Bayesian rarity estimates show similar trends. The probability of detecting rare species is lowest in areas with many species at low abundances but highest in areas with common and abundant species. This may be an artifact of the Jokimäki method. The fast paced survey method may not be suitable in detecting rare birds,especially in species diverse communities.

The data set is too small and observation period too short to assess whether the single sightings of rare bird species are merely due to chance. The CV of the estimates are large which suggests imprecision. It is only when a larger set of observations are made that a better estimate of species rarity can be made. Under these conditions, frequentist and Bayesian estimates of rarity will eventually converge. A larger and long observation time data set would allow for a more robust and rarity. Aggregation of observations at larger scales may reduce the CV. Dixon et al (2006) reports that aggregating the observations of prey capture in a rare carnivorous plant species reduced the CV from 76% as estimated from by frequentist methods to 36% using Bayesian analyses. However their observations involved long observation times. The researchers also used as prior capture probability data from a confamilial species. Thus with informative priors and aggregation, precision may be increased. In our study these conditions cannot be applied. We note that our McArdle's and Bayesian estimates are far from the cut off 1% limits of detection probability of rare species. The CV in both McArdle's and Bayesian analyses are similar at 24% and 25%. This is likely due to the small size of the data set. A better estimate for Bayesian rarity can made if information from past ecological studies on the UP Diliman campus are included as priors. These studies are often reported as unpublished theses.

The ecological context of rarity

In ecology rarity is defined with respect to the niche concept. In a niche a species is able to persist indefinitely. This is affected by biotic and abiotic factors that impinge on the survival and reproduction of populations. Species have a range of tolerance to these factors which comprises the realized niche.

Table 4. Bayesian posterior probability estimates of single sighting of a rare species in a plot.

Plot	Bayesian probability estimate (%)
College of Science (CS)	4.2
Academic Oval (OV)	4.8
Open fields (OP)	4.5
Residential Area (RE)	7.0

Biogeography theory represents the realized niche in spatial terms as the geographic range (Gaston 2006). Range size is positively correlated with abundance. The larger the range size the more likely the species is abundant and less rare (Hengeveld 1993, Hengeveld 1994). There are many factors that determine rarity and these may be related to habitat availability, life history traits and catastrophic environmental events (Gaston 1994). For some species that were recorded as common to moderately common in areas outside the campus *Lanius schach, Loriculus philippensis*, and *Amarornis phoenicurus* their rarity in UP Diliman College of Science is probably related to the lack of available habitat. The Pied Triller *Lalage nigra* is tolerant of human presence but may require significant areas of open woodland or "parang" in the Filipino language. The species is largely absent in high density built-up areas. Many habitat-limited species while resistant to disturbance will have to evolve behavioral responses to human impact. This may involve radical changes to life history characteristics that would allow them to efficiently colonize what remains of their original habitat in an increasingly fragmented urban landscape. If biogeographic theory is applied to this phenomenon, the ecological niche in space of these species has changed from the original "parang" to the built-up landscape now present. This is even if the spatial extent of the geographic range has remained largely unchanged.

Urbanization may contribute to some species shifting from dispersal limitation to dispersal facilitation. It is dispersal limited species that are not resilient to human disturbance and prone to extinction. These species are likely unable to recover their former population state even if the original landscape recovers. These species eventually would become rare and sightings in the urban environment are likely derived from vagrants from adjacent forest blocks. In the UP Diliman geographic context these blocks are the La Mesa Watershed (8 km from campus) and the secondary forest areas of the southern Sierra Madre range (50-100 km from campus). The urban environment, even though may still possess patches of original habitat, will then serve as a sink for these species.

A likely indicator of urbanization's effect on the original distribution of a bird is in the occurrence of *Lonchura punctulata*. It is very likely that this species was once widespread in the original landscape of Diliman. With increasing urbanization and loss of its ricefield habitat, it has become rare. Sightings of this bird are common in the grassy areas of the College of Science and the last remaining ricefields near Krus na Ligas a barrio adjacent to the campus. In Hawaii, this bird was introduced in the early 20th century and became a pest. With agriculture becoming a less important economic activity in the US state, the bird has lost its pest status and now is considered a "wayside bird"(Sodhi and Sharp 2006). Even if the bird has habitat specific requirements, it is not dispersal limited and thus is able to persist in the urban environment. If UP Diliman is allowed to revert to its original "parang" landscape, this bird will become common.

The rarity estimates we have presented may be of importance in designing statistical indices for conservation planning. Earlier data used are often from faunal lists and are usually point estimates of rarity. Our estimates are probabalistic and can be used as priors for future surveys. The estimates can also be used in a future ecological impact assessment of building developments in campus such as the National Science Complex and the North (Commonwealth Avenue) Science Park. These areas are also the areas that have been previously identified as critical bird habitats.

Acknowledgements. This work was funded by University of the Philippines NSRI grant ESM-05-2-01. We would like to thank M. Lu and A. Jensen of the Wild Bird Club of the Philippines and Prof. A. de Villa, CSSP Department of Philosophy for helping in bird identification, the Delaware Museum of Natural History for sending us species and identification catalogues and all volunteers who participated in the surveys.

References

Brown JH, Lomolino MV. Biogeography. 3nd. Sunderland, MA: Sinauer Associates, 2004.
Burgmann MA, Lindenmayer DB. Conservation Biology for the Australian Environment. Chipping Norton NSW: Surrey Beatty and Sons Pty, Ltd, 1998.
Catibog-Sinha C, Heaney LR. Philippine Biodiversity: Principles and Practice. Quezon City: Haribon Foundation for the Conservation of Nature, 2006.
Dixon P, Ellison A, Gotelli A. Improving the precision of estimates of the frequency of rare events. Ecology, 2005; 86 1114-1123.
Gaston KJ. Rarity. London: Chapman and Hall, 1994.
Gaston KJ. Species-range-size distributions: Patterns, mechanisms and implications. Trends in Ecology and Evolution, 1996; 11 (5): 197-201.
Gaston KJ. The Structure and Dynamics of Geographic Ranges. Oxford: Oxford University Press, 2006:
Hengeveld R. Biogeographical ecology. Journal of Biogeography, 1994; 21 341-351.
Hengeveld R. Ecological biogeography. Progress in Physical Geography, 1993; 17 (4): 448-460.
Jokimäki J, Jokimäki M. Spatial similarity of urban bird communities: A multiscale approach. Journal of Biogeography 2003; 30 1183-1193.
Koh LP, Sodhi NS. Importance of reserves, fragments, and parks for butterfly conservation in a tropical urban landscape. Ecological Applications, 2004; 14 (6): 1695–1708
Lee, PM. Bayesian Statistics: An Introduction. New York: Oxford University Press, 1989.
McArdle, B When are rare species not there? Oikos, 1990; 57 276-277.
Meffe GK, Carroll CR. Principles of Conservation Biology. Sunderland MA: Sinauer Associates, Inc, 1994.
Rosenzweig ML. Species Diversity in Space and Time. Cambridge UK: Cambridge University Press, 1995.
Sodhi N, Sharp I. Winged Invaders: Pest Birds of the Asia Pacific. Singapore: SNP International Publishing PTE LTD, 2006.
Vallejo B, Aloy A, Ong P, Tamino A, Villasper J. Spatial patterns of bird diversity and abundance in an urban tropical landscape: The University of the Philippines Diliman campus Science Diliman, 2008; 20:1-10