Significance of Sighting Rate in Inferring Extinction and Threat

GREG J. McINERNY,* DAVID L. ROBERTS,†‡ ANTHONY J. DAVY,* AND PHILLIP J. CRIBB†

*Centre for Ecology, Evolution and Conservation, School of Biological Sciences, University of East Anglia, Norwich, Norfolk, NR4 7TJ, United Kingdom

†Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3AB, United Kingdom

Abstract: *We are now entering a time of immense environmental upheaval in which, increasingly, experts are required to provide conservation assessments. Quantitative assessment of trends in species' range and abundance is costly, requiring extensive field studies over a long period of time. Unfortunately, many species are only known through a few "chance" sightings or a handful of specimens, and therefore extinction may be even harder to ascertain. Several methods have been proposed for estimating the probability of extinction. However, comparison within and between species is difficult because of variations in sighting rates. We applied a probabilistic method that incorporates sighting rate to the sighting record of Vietnamese slipper orchids (*Paphiopedilum*). The method generates a probability that another sighting will occur given the previous sighting rate and the time since last observation. This allows greater comparability between species discovered at different times. Its predictions were more highly correlated with the World Conservation Union criteria than previous methods. Trends in data collection and the political climate of a country, which affects access to material, are important potential sources of variation that affect sighting rates. A lack of understanding of the process by which data are generated makes inferring extinction from sighting records difficult because extinction status depends on how the sighting rate varies. However, such methods allow rapid conservation prioritization of taxa that are poorly known and would otherwise go unassessed.*

Key Words: collection data, herbarium specimens, *Paphiopedilum*, red-list criteria, species decline

Significado de la Tasa de Avistamiento en la Inferencia de Extinción y Amenaza

Resumen: *Estamos comenzando un tiempo de gran conmoción por el ambiente en el cual se requiere cada vez más que los expertos proporcionen evaluaciones de conservación. La evaluación cuantitativa de las tendencias en la distribución y abundancia de las especies es muy costosa, ya que requiere de estudios de campo extensivos durante un largo período de tiempo. Desafortunadamente, muchas especies solo son conocidas por medio de unos cuantos avistamientos "fortuitos" o un puñado de especimenes, y por lo tanto la verificación de extinciones es aun más difícil. Se han propuesto varios métodos para estimar la probabilidad de extinción, Sin embargo, la comparación dentro y entre especies es difícil por las variaciones en las tasas de avistamiento. Aplicamos un método probabilístico que incorpora la tasa de avistamiento al registro de avistamientos de orquídeas vietnamitas (*Paphiopedilum*). El método genera la probabilidad de que ocurra otro avistamiento en función de la tasa de avistamiento previa y el tiempo desde la última observación. Esto permite mayor compatibilidad entre especies descubiertas en tiempos diferentes. Sus predicciones estuvieron más correlacionadas con los criterios de la Unión Mundial de la Conservación que con métodos previos. Las tendencias en los datos de colección y el clima político de un país, que dificulta el acceso al material, son importantes fuentes potenciales de variación que afectan a las tasas de avistamiento. La falta de entendimiento*

‡*Address correspondence to D. Roberts, email d.roberts@rbgkew.org.uk*
Paper submitted August 19, 2004; revised manuscript accepted August 18, 2005.

DOI: 10.1111/j.1523-1739.2006.00377.x

del proceso de generación de datos dificulta la inferencia de la extinción a partir de registros de avistamiento porque el estatus de extinción depende de la variación en la tasa de avistamiento. Sin embargo, tales métodos permiten la definición de prioridades de conservación de taxa que son poco conocidos y que de otra manera no serían evaluados.

Palabras Clave: criterios de lista roja, datos de colección, declinación de especies, especimenes de herbario, *Paphiopedilum*

Introduction

Extinctions are most frequently asserted after subsequent investigation, so uncertainty often surrounds the classification of a species as extinct. This necessitates acknowledgment of the probabilistic nature of any extinction statement. Collections and censuses are often subject to limited resources, so exhaustive searches of the known habitats of a species may be impractical (McArdle 1990) and in some cases the range of a taxon may be unknown (Gaston 1995). Without extensive fieldwork the status of a species may only be inferred from sighting records (Solow & Roberts 2003) (i.e., mainly the information contained within the specimen-based collections of museums and herbaria). These records provide information on the distribution of taxa through time and space (Ponder et al. 2001) and represent primary, verifiable observations (Bachman et al. 2004).

Several methods provide a probabilistic basis for the extinction hypothesis based on such sighting records (Solow 1993*a*, 1993*b*; Burgman et al. 1995; McCarthy 1998; Solow & Roberts 2003). These methods provide the probability of another sighting given the characteristics of a sighting record. However, because species are not all discovered at a single time, inequalities arise in the number of initial sightings and in the length of the entire period. Consequently, species discovered earlier may be recognized as extinct before species that have been discovered more recently, despite the same period of absence of sighting. McCarthy (1998) suggests that relative levels of threat may be inferred from the magnitude of the probability generated, with smaller probabilities implying an increased threat of extinction through changes in range and/or abundance. This is potentially of great use because conservationists are often primarily concerned with threatened species rather than those already lost (Burgman et al. 1995), so increased comparability of sighting records will aid accuracy and application of such methods.

We present a method that uses previous sighting rate to infer extinction that is unaffected by the magnitude of the initial sighting period. We applied this method to the collection records of Vietnamese slipper orchids (*Paphiopedilum*).

Methods

Statistical Model

Sightings of a species may be arranged as a binary record, with multiple sightings recorded as a single sighting for any single time unit because the methods assume collections are independent of one another (McCarthy 1998). The sightings occur within the entire period, T, at ordered times, $t_1 < t_2 < \ldots < t_n$. The start date of collections may be known or the first sighting may be used as a starting point.

The Solow equation (1993*a*) considers the hypothesis of extinction in relation to number of sightings, n, within the period $0 \leq t \leq t_n$ given that sightings are equally likely to occur during the whole sighting period, T:

$$p = \left(\frac{t_n}{T}\right)^n. \tag{1}$$

High probability values are produced if the lack of sightings at the end of the record could happen by chance, implying that extinction or population decline has not occurred. Low probabilities indicate extinction or population decline has occurred because the sightings are unlikely to have occurred in the time period $0 \leq t \leq t_n$ given the magnitude of T and/or n.

Because the time to detection of extinction by the Solow equation is dependent on the initial period of observation, t_n, if the number of initial sightings varies then species with a smaller number of initial sightings will be inferred as extinct much more slowly, even if the previous sighting rate is exactly the same (Fig. 1a). Therefore for recently discovered species, the length of the period since the last sighting may be more informative than the p value generated. Using the sighting rate would therefore be an advantage and lead to greater comparability between taxa if there is disparity in the period of initial sightings or discovery. The probability that another sighting will occur can be generated given the previous sighting rate, n/t_n, and the time since last observation, $T - t_n$, by the following:

$$p = \left(1 - \left(\frac{n}{t_n}\right)\right)^{(T - t_n)}. \tag{2}$$

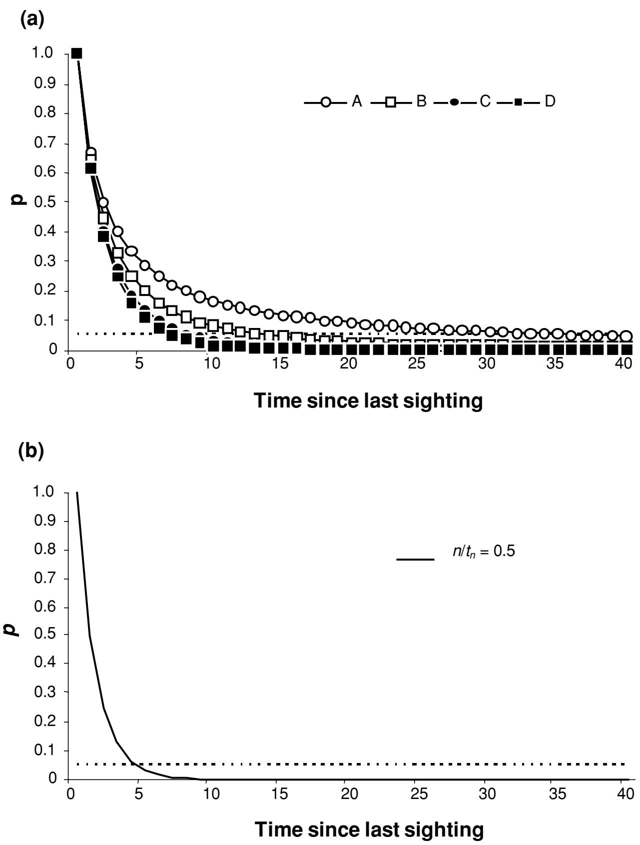

Figure 1. The p *values generated by (a) Solow equation (Eq. 1) (Solow 1993a) and (b) sighting rate equation (Eq. 2) following an initial period of sightings. In (a)* p *values are shown for four theoretical species (A–D) with differing numbers of initial sightings* (n) *(species A,* $n = 1$*; species B,* $n = 2$*; species C,* $n = 5$*; species D,* $n = 10$*). All species are seen every other year after their initial discovery, resulting in a sighting rate of 0.5. Species A was discovered earliest. As the number of initial sightings increases the period of absence required to infer extinction decreases. At the extremes, species D may be recognized as extinct after 7 years, whereas species A may be recognized as extinct after 38 years. In (b) the same species would all be inferred as extinct at the same point in time, after a 5-year absence, because they all have the same rate of sightings before the absence,* $n/t_n = 0.5$*. Dashed horizontal line represents* $\alpha = 0.05$*.*

Because the asymptotic reduction in probability occurs at the same rate when sighting rates are equal (Fig. 1b), comparability between species discovered at different times will be increased.

As mentioned above, in the case of, say, annual surveys, it is possible to select a start date for the period T. However, this is rarely possible (Solow & Roberts 2003), and because we use the first sighting as the start date, the number of sightings, n, reduces to $n - 1$.

Application to the Slipper Orchids

We applied the Solow equation and sighting rate equation to the sighting records of Vietnamese slipper orchids of the genus *Paphiopedilum*. The slipper orchids of Vietnam were chosen for the study because of the recent completion of a monograph on the group (Averyanov et al. 2003): their taxonomic status is stable, conservation status has recently been assessed, and most herbarium specimens are known. We recorded sightings for 23 taxa from data used by Averyanov et al. (2003). Because we used the first sighting as the starting point, records with a single sighting could not be used because subsequent sightings would be reduced to zero. Endpoints for each taxa were recorded as 2002. We used World Conservation Union (IUCN) categories of threat (from Averyanov et al. [2003]) as the "true" status of each species to validate the equations. Following McCarthy (1998), we carried out a Spearman rank correlation between the prediction of each equation and the IUCN category.

Results

Fifteen of the 23 species of *Paphiopedilum* had sighting rates >0 (Fig. 2), and these ranged from 1 (*P. barbigerum* var. *lockianum*) to approximately 0.1 (*P. villosum* var. *annamense*). The probabilities produced by the sighting rate equation ranged from <0.001 (suggesting considerable threat) to 0.889 (relatively low threat) for the same two taxa (Table 1). *P. delenatii* also emerged as particularly threatened ($p = 0.031$). The Solow equation consistently produced a higher probability than the sighting rate equation for every species and in some cases (particularly the more threatened species) the differences were large (e.g., *P. barbigerum* var. *lockianum* and *P. delenatii*; Table 1). A rank correlation between the p values generated by the sighting rate equation and the IUCN classifications was significant ($r_s = 0.575, p < 0.05$), whereas the equivalent correlation based on the Solow equation was not significant ($r_s = 0.492$).

Discussion

Communities often have a significant component of rare species or those that are cryptic and/or problematic to observe. Therefore any survey, however thorough, carries a chance that individuals have been overlooked (Reed 1996) or exist in refugia outside the study area (Kéry 2002). McArdle (1990) used the relationship among the number of units sampled from an area, the rarity of the

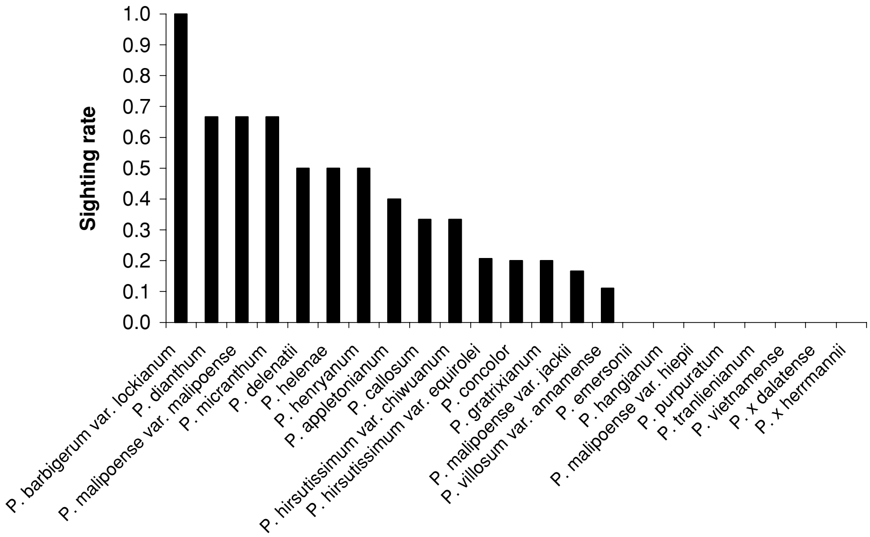

Figure 2. Ranked distribution profile of sighting rate (n/t_n) *for* Paphiopedilum *within Vietnamese sightings. Values of* n *were reduced by 1 to yield individual start dates.*

species, and the probability of detecting the species to predict the extent of sampling required to establish statistical confidence that species are extinct. The number of units required to be sampled can be calculated as

$$N = \frac{\log(1-\alpha)}{\log(1-p)}, \tag{3}$$

where p is the level of rarity of the species and a is the level of confidence required that the species is extinct (McArdle 1990). By applying this to the *Paphiopedilum* data, with the sightings of each species as proportion of the total sightings as a measure of its rarity, the statistical certainty of extinction was hard to achieve. Based on the data from Averyanov et al. (2003), rare species such as *P. purpuratum* would require over 125 surveys to provide 95% certainty of a suspected extinction. Feasibility of a survey with enough sample units to attain the confidence level may be low (McArdle 1990), due to the time and resources required. This is problematic for the IUCN categorization, which requires an exhaustive search of

Table 1. Probability values generated by the Solow and sighting rate indices based on the sighting records of *Paphiopedilum* species occurring in Vietnam.

Species	n	t_n	T	*Solow*[a]	*Sighting rate*[a]	*IUCN*[b]
P. appletonianum (Gower) Rolfe	3	5	7	0.510	0.360	VU
P. barbigerum var. *lockianum* Aver.	2	1	2	0.500	<0.001	EN
P. callosum (Rchb.f.) Stein	3	6	7	0.735	0.667	EN
P. concolor (Lindl. ex Bateman) Pfitzer	4	15	16	0.824	0.800	VU
P. delenatii Guillaumin	2	2	7	0.286	0.031	CR
P. dianthum Tang & F. T. Wang	3	3	4	0.563	0.333	EN
P. emersonii Koop. & P. J. Cribb	1	0	1	—	—	EN
P. gratrixianum Rolfe	3	10	14	0.510	0.410	EN
P. hangianum Perner & O. Gruss	1	0	1	—	—	EN
P. helenae Aver.	3	4	7	0.327	0.125	EN
P. henryanum Braem	2	2	3	0.667	0.500	EN
P. hirsutissimum var. *chiwuanum* (Tang & F. T. Wang) P. J. Cribb	3	6	7	0.735	0.667	VU
P. hirsutissimum var. *esquirolei* (Schltr.) K. Karas. & K. Saito	7	29	30	0.816	0.793	VU
P. malipoense var. *malipoense* S. C. Chen & Z. H. Tsi	5	6	7	0.540	0.333	EN
P. malipoense var. *hiepii* (Aver.) P. J. Cribb	1	0	7	—	—	EX
P. malipoense var. *jackii* (H. S. Hua) Aver.	2	6	7	0.857	0.833	EN
P. micranthum Tang & F. T. Wang	3	3	4	0.563	0.333	EN
P. purpuratum (Lindl.) Stein	1	0	3	—	—	EN
P. tranlienianum O. Gruss & Perner	1	0	1	—	—	EN
P. vietnamense O. Gruss & Perner	1	0	1	—	—	EW
P. villosum var. *annamense* Rolfe	3	18	19	0.898	0.889	VU
P. × *dalatense* Aver.	1	0	6	—	—	DD
P. × *herrmannii* F. Fuchs & H. Reisinger	1	0	5	—	—	EN

[a] *Values <0.05 imply extinction.*
[b] *World Conservation Union (IUCN) Red List categories: EX, extinct; EW, extinct in the wild; CR, critically endangered; EN, endangered; VU, vulnerable; DD, data deficient (IUCN 2001).*

the historical range of a species over an appropriate time period (IUCN 2001), and illustrates that certainty may be hard to achieve.

The fact that greater statistical confidence is not readily achievable for any extinction statement highlights the difficulties in assessing the level of threat to recently discovered species, but an indicator of confidence would still be of use. One problem with any approach based on the initial sighting period is illustrated by *P. barbigerum* var. *lockianum*. This species was sighted continuously until the end of the initial sighting period, resulting in an estimated sighting rate of 1. The sighting rate equation produced a p value of <0.001 (Table 1), suggesting a very high threat to the species, in contrast to the p value of 0.500 from the Solow equation. This sensitivity may seem extreme, for example, if an observer saw a species on three consecutive occasions, but did not see it on the fourth, they may not conclude that it was severely threatened without good reason. However, in the case of *P. barbigerum* var. *lockianum*, there is good reason to believe the species is under the threat of extinction in Vietnam (Averyanov et al. 2003). Although this case seems to illustrate a shortcoming of the sighting rate approach, it also highlights the limitations of the Solow equation.

The problem as illustrated by *P. barbigerum* var. *lockianum*, is a symptom of quantifying sighting information into a binary data, in which three circumstances are potentially recordable: a species is present = 1, a species is absent = 0, or no observation is made = 0. The latter two circumstances are indistinguishable in the binary record. In such situations, expert knowledge or records of collection effort are vital to differentiate between the sensitivity of individuals methods. When this information is unavailable, comparison between individual methods may indicate the sensitivity of individual methods to the characteristics of individual sighting records. As is shown in the example in Fig. 1, differences in the initial period of sightings are not distinguished by the sighting rate equation. Inspection of individual results in combination with the other available information, even the values of t_n or T, may help interpret individual results as demonstrated by *P. barbigerum* var. *lockianum*.

The significant correlation ($r_s = 0.575$) between the IUCN categories and the p values generated by the sighting rate equation supports the suggestion of McCarthy (1998) that such indices may be used in the inference of threat. The improvement in predictive power over the Solow equation ($r_s = 0.492$) may be in part due to the increased comparability that the sighting rate equation allows. *Paphiopedilum* had a high contemporary sighting rate, with 93% of all sightings occurring after 1994. Seven of the species had sighting records that started after 1998, resulting in T values that were 50% or less than 12 other species. This supports the suggestion that comparability between species may be increased by use of the sighting rate equation.

Sighting rate can vary in two main ways: (1) within a sighting record and (2) between sighting records. However, both types of variation can occur for the same reasons. The variation in sighting rate can vary considerably between taxa and geographical regions due to herbarium trends and the political climate of the countries (which affects access to material). Following the inclusion of all orchids on Convention on International Trade in Endangered Species (CITES) it has become increasingly difficult to obtain material for scientific studies, and this has increased variation within sighting records. Because orchids are on Appendix I or II of CITES, variation between sighting records occurs when comparing Orchidaceae records with, say, Poaceae, which are not controlled by CITES. Due to the requirements for permits this would result in a lower rate of collection in the Orchidaceae. However, with the introduction of the Convention on Biological Diversity (CBD), wherein countries exert their sovereign rights over their biodiversity, collection of material and thus collection rate may become equally as low due to the difficulties associated with access.

Access to Vietnamese sites was limited previously. Considerable contemporary interest was roused in the status of *Paphiopedilum*, because of its exploitation for the horticultural trade, once access to Vietnam was renewed (Averyanov et al. 2003). Thus, changes in the sighting rate may highlight trends unrelated to population decline. For the inference of threat, sighting rate should be a key consideration because the asymptotic decline of the p values occurs at different rates with different sighting rates. Although the rate of decline is a product of the sighting record and is assumed to be representative of the population decline, care should be taken in comparisons between groups for this reason.

The use of such probabilistic equations (Solow 1993*a*, 1993*b*; Burgman et al. 1995; McCarthy 1998; Solow & Roberts 2003) as models to "screen" herbarium and museum collections of otherwise poorly studied groups has been discussed in detail by Burgman et al. (1995), McCarthy (1998), Solow and Helser (2000), Burgman et al. (2000), and Burgman (2002). Although the behavior of the equations and their suitability to individual cases has received attention, the appropriateness of each equation may not always be obvious or consistent for all species within a grouping. For example, knowing whether species are subject to sudden extinction or that arising from a population decline is essential for accuracy in the choice of equation to be used (e.g., Solow [1993*a*] equation or Solow [1993*b*] equation for a declining population). Yet to gain an insight into threat status, more general equations, such as the sighting rate equation, may be of greater benefit because they require initial analysis of groupings to understand the scale of threat inferred by p values. Further testing of the sighting rate and previously described equations will aid the appropriate utilization of such tools as a means of rapid conservation prioritization.

Acknowledgments

We thank P. Kindlmann, I. Roberts, and A.R. Solow for useful discussions on various aspects of statistical methodology and an anonymous reviewer for useful comments.

Literature Cited

Averyanov, L., P. Cribb, P. K. Loc, and N. T. Hiep. 2003. Slipper orchids of Vietnam. Royal Botanic Gardens, Kew, United Kingdom.

Bachman, S. P., W. J. Baker, N. A. Brummitt, J. Dransfield, and J. F. Moat. 2004. Elevational gradients, area and tropical island diversity: an example from the palms of New Guinea. Ecography **27:**299–310.

Burgman, M. A. 2002. Turner review no. 5: are listed threatened plant species actually at risk? Australian Journal of Botany **50:**1–13.

Burgman, M. A., R. C. Grimson, and S. Ferson. 1995. Inferring threat from scientific collections. Conservation Biology **9:**923–928.

Burgman, M., B. R. Maslin, D. Andrewartha, M. R. Keatley, C. Boek, and M. McCarthy. 2000. Inferring threat from scientific collections: power tests and an application to Western Australian *Acacia* species. Pages 7–26 in S. Ferson and M. Burgman, editors. Quantitative methods for conservation biology. Springer-Verlag, New York.

Gaston, K. J. 1995. Biodiversity: loss. Progress in Physical Geography **19:**253–264.

IUCN (World Conservation Union). 2001. IUCN Red List categories and criteria: version 3.1. IUCN Species Survival Commission, IUCN, Gland, Switzerland, and Cambridge, United Kingdom.

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

McArdle, B. H. 1990. When are rare species not there? Oikos **57:**276–277.

McCarthy, M. A. 1998. Identifying declining and threatened species with museum data. Biological Conservation **83:**9–17.

Ponder, W. F., G. A. Carter, P. Flemons, and R. R. Chapman. 2001. Evaluation of museum collection data for use in biodiversity assessment. Conservation Biology **15:**648–657.

Reed, J. M. 1996. Using statistical probability to increase confidence of inferring species extinction. Conservation Biology **10:**1283–1285.

Solow, A. R. 1993*a*. Inferring extinction from sighting data. Ecology **74:**962–964.

Solow, A. R. 1993*b*. Inferring extinction in a declining population. Journal of Mathematical Biology **32:**79–82.

Solow, A., and T. Helser. 2000. Detecting extinction in sighting data. Pages 1–6 in S. Ferson and M. Burgman, editors. Quantitative methods for conservation biology. Springer-Verlag, New York.

Solow, A. R., and D. L. Roberts. 2003. A nonparametric test for extinction based on a sighting record. Ecology **84:**1329–1332.