Conservation Biology

Contributed Paper

A robust nonparametric method for quantifying undetected extinctions

Ryan A. Chisholm,* Xingli Giam,† Keren R. Sadanandan,* Tak Fung,* and Frank E. Rheindt* ¶

*Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, 117558, Singapore
†School of Aquatic & Fishery Sciences, University of Washington, Seattle, WA, 98195-5020, U.S.A.

Abstract: *How many species have gone extinct in modern times before being described by science? To answer this question, and thereby get a full assessment of humanity's impact on biodiversity, statistical methods that quantify undetected extinctions are required. Such methods have been developed recently, but they are limited by their reliance on parametric assumptions; specifically, they assume the pools of extant and undetected species decay exponentially, whereas real detection rates vary temporally with survey effort and real extinction rates vary with the waxing and waning of threatening processes. We devised a new, nonparametric method for estimating undetected extinctions. As inputs, the method requires only the first and last date at which each species in an ensemble was recorded. As outputs, the method provides estimates of the proportion of species that have gone extinct, detected, or undetected and, in the special case where the number of undetected extant species in the present day is assumed close to zero, of the absolute number of undetected extinct species. The main assumption of the method is that the per-species extinction rate is independent of whether a species has been detected or not. We applied the method to the resident native bird fauna of Singapore. Of 195 recorded species, 58 (29.7%) have gone extinct in the last 200 years. Our method projected that an additional 9.6 species (95% CI 3.4, 19.8) have gone extinct without first being recorded, implying a true extinction rate of 33.0% (95% CI 31.0%, 36.2%). We provide R code for implementing our method. Because our method does not depend on strong assumptions, we expect it to be broadly useful for quantifying undetected extinctions.*

Keywords: biodiversity loss, birds, nonparametric model, Singaporean avifauna

Un Método No-Paramétrico Robusto para la Cuantificación de Extinciones No Detectadas

Resumen: *¿Cuántas especies se han extinguido en tiempos modernos antes de haber sido descritas por la ciencia? Para responder a esta pregunta, y por lo tanto tener una evaluación completa del impacto de la humanidad sobre la biodiversidad, se requieren métodos estadísticos que cuantifiquen las extinciones no detectadas. Dichos métodos han sido desarrollados recientemente y están limitados por su dependencia de las suposiciones paramétricas; específicamente, asumen que el acervo de especies existentes y no detectadas decae exponencialmente, mientras que las tasas reales de detección varían temporalmente con el esfuerzo del censo y las tasas reales de extinción varían con el incremento y decremento de los procesos amenazantes. Diseñamos un nuevo método no-paramétrico para estimar las extinciones no detectadas. Como contribuyentes, el método sólo requiere de la primera y la última fecha en las que se registró cada especie en un conjunto. Como resultados, el método aporta estimados de la proporción de especies que se han detectado o no, extinguido, y del caso especial en el que el número de las especies existentes no detectadas en la actualidad es asumido como cercano a cero, del número absoluto de especies extinguidas no detectadas. La principal suposición del método es que la tasa de extinción por especie es independiente de si una especie ha sido detectada o no. Aplicamos el método para las aves nativas y residentes de Singapur. De las 195 especies registradas, 58 (29.7%) se han extinguido en los últimos 200 años. Nuestro método proyectó que unas 9.6 especies adicionales (95% CI 3.4, 19.8) se han extinguido sin haber sido registradas primero, lo que implica una tasa de extinción verdadera de 33.0% (95% CI 31.0%, 36.2%). Proporcionamos un código R para la implementación de nuestro método. Ya que este no depende de suposiciones fuertes, esperamos que sea útil de manera general para la cuantificación de extinciones no detectadas.*

Palabras Clave: aves, avifauna de Singapur, modelo no-paramétrico, pérdida de biodiversidad

¶*email dbsrfe@nus.edu.sg*
Paper submitted February 20, 2015; revised manuscript accepted September 29, 2015.

Conservation Biology, Volume 00, No. 0, 1–8

DOI: 10.1111/cobi.12640

Introduction

Many modern species extinctions are well documented. At the time of European settlement in North America, the Passenger Pigeon (*Ectopistes migratorius*) was one of the world's most abundant birds, but it declined rapidly over the course of the 19th century and finally went extinct in 1914. In Australia the thylacine (*Thylacinus cynocephalus*) was already rare and restricted to Tasmania when Europeans arrived (Johnson 2006). It was then hunted to extinction; the last known survivor died in captivity in 1936. There have also been many documented local extirpations. The city-state of Singapore was home to tigers (*Panthera tigris*) during early colonial times, but hunting, deforestation, and urban development led to its extirpation and the extirpation of many other species (Corlett 1992). The last Singaporean tiger was killed in the 1930s (Khan 1986).

But the list of documented extinctions tells only part of the story: many species have gone extinct before they could be discovered (Ceballos & Ehrlich 2009; Tedesco et al. 2014). For example, 9 of 10 species in a radiation of helicinid land snails from the Gambier Islands, French Polynesia, went extinct in the 19th century, but this only became known in 2013 after the discovery of empty shells (Richling & Bouchet 2013). One can infer that such undocumented extinctions have occurred widely and are ongoing because human pressures on biodiversity are ubiquitous (Sodhi et al. 2009; Wilcove et al. 2013; Laurance et al. 2014) and the discovery of new species occurs in many places (Joppa et al. 2011; Giam et al. 2012; Scheffers et al. 2012). The extent of undocumented extinctions remains underexplored. How many more species might there have been to discover if they had not gone extinct first? How many species have been driven to extinction by human activities without scientific notice? These questions must be answered to make a comprehensive assessment of humanity's effects on the biosphere and to make accurate projections about future extinction rates. Accurate estimates of extinction rates are also necessary for estimating the global taxonomic effort required to discover Earth's unknown biodiversity before too many more species go extinct undetected (Costello et al. 2013).

Previously ratios (analogous to mark-recapture estimation of population size) were used to calculate the likely number of birds that have gone extinct without leaving fossilized bones (Pimm et al. 1994). However, this method can be applied only in regions where the fossil record is well documented and for taxa that fossilize well (e.g., large vertebrates) (Barnosky et al. 2011). To get around these issues, Tedesco et al. (2014) developed a parametric model for estimating undetected extinctions that requires only species description and extinction dates. The authors specified functional forms for detection and extinction rates over time and fitted the resulting model to time series of species discovery and extinction: the fitted parameters are used to infer the numbers of both undetected extant and undetected extinct species. This approach requires that detection and extinction rates be constant or follow some regular distribution. Assuming constant detection and extinction rates, Tedesco et al. (2014) estimated that undiscovered extinctions as a proportion of total extinctions are in the range of 15% to 59%, depending on the region and taxonomic group considered.

Unfortunately, the applicability of parametric models is limited because of the highly irregular extinction and detection rates often observed in empirical data. Indeed, at regional scales, detection rates show peaks and troughs associated with the activity of particular biologists, government funding priorities, or wars. At a global scale, these effects in some cases average out to a reasonably smooth function (e.g., world birds [Tedesco et al. 2014]), but in other cases they do not (e.g., world marine mammals [Tedesco et al. 2014]). At smaller scales, detection and extinction rates tend to be even more variable over time. Development of nonparametric methods that require fewer assumptions about the functional forms of temporal detection and extinction trends will allow more general and parsimonious estimates of undetected extinctions from species inventories.

At first glance, the problem of estimating undetected extinctions without making parametric assumptions may seem intractable. Our goal here was to prove otherwise. A remarkable amount of information about undetected species can be gleaned from time series of detected extant and detected extinct species without making strong assumptions. We made nonparametric estimates of the total proportion of extinct species and, for groups of species where the number of undetected extant species in the present day can be assumed close to zero, of the absolute number of undetected extinct species. We applied our new method to the resident bird fauna of Singapore (Wang & Hails 2007), for which we compiled a new comprehensive list of resident bird species and dates of first and last detection.

Methods

Nonparametric Undetected Extinctions Model

We consider four classes within a regionally or taxonomically defined group of species: detected extant species (S_t); undetected extant species (U_t); detected extinct species (E_t); and undetected extinct species (X_t). Subscripts t on each variable indicate time (years). The total number of extant and extinct species is $N = S_t + U_t + E_t + X_t$, which is constant over time. We used

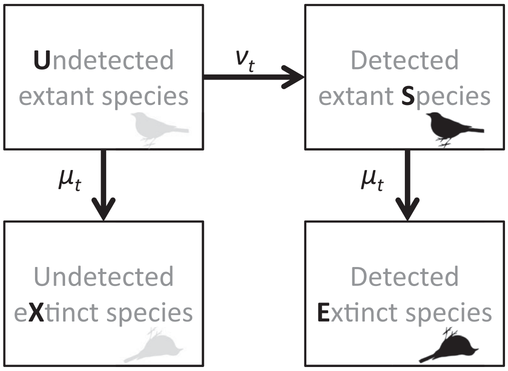

Figure 1. State-and-transition diagram corresponding to our mathematical model of the extinction and detection process (described in the text). Bold capital letters in each state rectangle are variable names. Arrows indicate transitions due to species discovery (at rate ν_t) and extinction (at rate μ_t).

$t = 0$ as a baseline value. We assumed $S_0 = E_0 = X_0 = 0$ and so $U_0 = N$; that is, we did not consider detections or extinctions that occurred before year 0 so all species are initially undetected and extant. Possible transitions between the four classes are from undetected to detected ($U \rightarrow S$) and from extant to extinct ($U \rightarrow X$, $S \rightarrow E$) (Fig. 1). Dynamical equations describing the system are

$$\begin{aligned} S_{t+1} &= (1 - \mu_t)\, S_t + \nu_t U_t \\ E_{t+1} &= E_t + \mu_t S_t \\ U_{t+1} &= (1 - \mu_t)\, U_t - \nu_t U_t \\ X_{t+1} &= X_t + \mu_t U_t \end{aligned} \tag{1}$$

where μ_t and ν_t are the per-species extinction and detection rates at time t. These rates are not constrained by the model to be constant or to follow any particular functional form, although there is an implicit assumption that while the rates may vary over time, they do not vary systematically across species (see below for an exploration of the robustness of the model to violation of this assumption).

The number of detected extant and extinct species (S_t and E_t) over time is, by definition, known. The second equation in equation set 1, which defines the deterministic model, can be rearranged to give the extinction rate $\mu_t = (E_{t+1} - E_t) S_t$ in a particular year t. But real extinction processes must be stochastic (if for no other reason than the discreteness of the number of species); thus, this formula gives in practice only an estimate $\hat{\mu}_t$ of the extinction rates:

$$\hat{\mu}_t = \frac{E_{t+1} - E_t}{S_t} \tag{2}$$

In year $t = 0$, the extinction rate μ_0 cannot be estimated because $S_0 = 0$. Therefore, we assumed in the calculations that $\mu_0 = 0$, which means we considered only extinctions that occurred after the first species was detected.

One quantity of interest is the fraction p_t of all species, including both detected and undetected species, that have gone extinct by time t. It can be proven by induction (Supporting Information) that

$$p_t = \frac{E_t + X_t}{S_t + U_t + E_t + X_t} = \frac{E_t + X_t}{N} = 1 - \prod_{i=0}^{t-1} (1 - \mu_i) \tag{3}$$

This formula has a direct interpretation: the probability of being extinct is one minus the probability of being extant, which is in turn the product of the probability of persistence over all years. Substituting in Eq. 2 for $\hat{\mu}_t$ gives

$$\hat{p}_t = 1 - \prod_{i=0}^{t-1} \left(1 - \frac{E_{i+1} - E_i}{S_i}\right) \tag{4}$$

This formula allowed us to estimate the fraction of extinct species knowing only the time series of detected extant and extinct species or, equivalently, the first and last date of record of each species.

Equation 4 assumes perfect detection (i.e., that a species goes extinct immediately after it is detected for the last time). If this assumption is violated, such that a species was last detected in year $\tau - \theta$ but actually went extinct in year τ, then the error in the estimate of $\hat{p}_t$ will be approximated by the following formula if the number of detected extant species S_τ is sufficiently large (Supporting Information):

$$\epsilon \approx \left(1 - \hat{p}_t\right) \frac{1 + \Delta E_{\tau-\theta} - \Delta E_\tau + S_\tau - S_{\tau-\theta}}{(\Delta E_\tau - S_\tau)\left(\Delta E_{\tau-\theta} - S_{\tau-\theta}\right)} \tag{5}$$

where $\Delta E_t \equiv E_{t+1} - E_t$. In most cases, such errors will be tiny; they will only be large if a huge fractional change in the number of detected extant species occurred between years $\tau - \theta$ and τ. For example, assuming the focal species was the only species that went extinct in year τ and that zero species actually went extinct in year $\tau - \theta$, then Eq. 5 simplifies to

$$\epsilon \approx \left(1 - \hat{p}_t\right) \frac{S_\tau - S_{\tau-\theta}}{(S_\tau - 1)\, S_{\tau-\theta}} \tag{6}$$

which will generally be small because the numerator of the second factor is linear in the number of species, whereas the denominator is quadratic in the number of species.

Equations 5 and 6 estimate the error in $\hat{p}_t$ assuming imperfect detection for only one species. To investigate imperfect detection in multiple species, we compared the estimated $\hat{p}_t$ in our case study (see below) to values of $\hat{p}_t$ from simulations in which the true extinction date for each species was drawn uniformly from the interval between the species' last detection date and the last detection date plus 10 years. The simulation was repeated 10,000 times. We then investigated a more extreme case of imperfect detection by repeating the simulations with the upper limit of each uniform interval being the present day.

Another quantity of interest is the absolute number of extinct species, including both detected and undetected species. One can write Eq. 3 as

$$E_t + X_t = N\, p_t = (S_T + U_T + E_T + X_T)\; p_t \qquad (7)$$

where the second step comes from noting that the total number of species (extant and extinct) is, by definition, constant over time. With $t = T$, one can solve for X_T:

$$X_T = \frac{(S_T + U_T)\, p_T}{1 - p_T} - E_T \qquad (8)$$

Substituting into Eq. 7, simplifying, and replacing p_t with its estimator (Eq. 4) gives

$$\hat{X}_t = \frac{S_T \hat{p}_t}{1 - \hat{p}_T} - E_t + \frac{U_T \hat{p}_t}{1 - \hat{p}_T}. \qquad (9)$$

If one assumes the number of undetected extant species in the present day is close to zero ($U_T \approx 0$), the approximation is

$$\hat{X}_t \approx \frac{S_T \hat{p}_t}{1 - \hat{p}_T} - E_t. \qquad (10)$$

Recalling that $\hat{p}_t$ and $\hat{p}_T$ are expressible in terms of the time series E_t and S_t (Eq. 4), Eq. 10 shows us that one can estimate the absolute number of undetected extinct species at time t ($\hat{X}_t$) as a function of the time series of detected extant and extinct species. The assumption that the number of undetected extant species in the present day is small ($U_T \approx 0$) is justifiable in circumstances where a region or particular taxonomic group has been intensively surveyed and very few new species are being detected in the present day. From Eq. 10, we estimated the number of undetected extant species at time *t* as

$$\hat{U}_t = 1 - \hat{X}_t - S_t - E_t. \qquad (11)$$

We developed an algorithm to generate confidence intervals on the estimated time series of undetected extinct species ($\hat{X}_t$) by considering a stochastic version of Eq. 1 in which the extinction process follows a binomial distribution. Specifically, the number of extinctions of detected species in time step t ($E_{t+1} - E_t$) is treated as a random realization of a binomial distribution with number of trials equal to the number of detected extant species S_t and probability parameter equal to μ_t, and the number of extinctions of undetected species ($X_{t+1} - X_t$) is treated as another binomial distribution with number of trials equal to the number of undetected extant species U_t and the same probability parameter μ_t. We called the realized extinction rate of detected species $\hat{\mu}_t$ (as in Eq. 2) and the realized extinction rate of undetected species $\hat{\mu}_t^* = (X_{t+1} - X_t)/U_t$. Our algorithm first estimates a probability distribution for each value in the intrinsic extinction rate time series μ_t by resampling from a binomial distribution with number of trials S_t and probability parameter $\hat{\mu}_t$ (Eq. 2). It then generates alternative realizations of $\hat{\mu}_t^*$ by sampling from a binomial distribution with number of trials $\hat{U}_t$ (estimated from Eq. 11) and probability parameter drawn from the sampling distribution of μ_t. These alternative time series of $\hat{\mu}_t^*$ are then used to generate alternative realizations of $\hat{X}_t$ from a more general version of Eq. 10 that allows different realized extinction rates for detected and undetected species (Supporting Information). We verified that this sampling procedure produces accurate 95% CIs by generating simulated time series of S_t, U_t, E_t, and X_t (each simulation was run until there were no more undetected extant species [i.e., $U_T = 0$]) and checking that the 95% CIs on $\hat{X}_T$ estimated from S_t and E_t contained the true value of X_T roughly 95% of the time (i.e., nominal coverage probability $\approx$ actual coverage probability) (Supporting Information).

To test the model's robustness to violation of the assumption of equal intrinsic extinction and detection rates across species, we then simulated scenarios that were biologically more realistic in which these rates varied across species according to beta distributions (Supporting Information). A beta distribution has support [0,1] and two free parameters; thus, it allows simultaneous control of the mean and variance. In the beta distribution of extinction rates, we fixed the mean at the empirical value from our application (see below) and used a range of coefficients of variation (0.0, 0.2, 0.5, and 1.0). We followed the same procedure for the detection rates. We used Eqs. 4 and 10 with just the time series of S_t and E_t to produce estimates $\hat{p}_t$ and $\hat{X}_t$ of the quantities p_t and X_t (these estimates are now biased because the assumption of equal intrinsic extinction and detection rates across species has been violated). For each parameterization, we repeated the simulation 1000 times (for a total of 16,000 simulations). Afterward, we generated CIs on the estimates using the methods described in the previous paragraph. We then measured model performance by comparing the actual number of undetected extinct species in the final time step of each simulation (X_T) with the predictions of the model ($\hat{X}_T$): we recorded the fraction of times X_T was excluded from the 95% CI for $\hat{X}_T$ (error rate = 1 – actual coverage probability) and the bias, measured as the mean of the relative error $(\hat{X}_T - X_T)/X_T$.

Application

We applied our method to the task of estimating the number of bird species that have gone extinct undetected in the last 200 years in Singapore, a tropical city state of original area 540 km^2 and current area ~720 km^2 (owing to land reclamation from the sea). Because Singapore is highly urbanized and deforested (<5% of total forest remains, most of it secondary), it has often been used as a case study of tropical biodiversity loss (e.g., Brook et al. 2003).

Previously, Brook et al. (2003) estimated that 33.9–58.5% of bird species have gone extinct in Singapore. The lower bound was based on observed extinctions and the upper on the assumption that all lowland species (excluding savanna species because there was no savanna in Singapore) found in peninsular Malaysia also inhabited Singapore in 1819. Following this assumption, they assumed that species that were never found in Singapore but were recorded from peninsular Malaysia had gone extinct in Singapore since 1819. In applying our method to the Singaporean bird inventory, we aimed to provide an alternative upper bound that is based on direct estimation of undocumented extinctions.

Based on detailed information available from past survey work (summarized by Wang and Hails [2007] and Gibson-Hill [1949], but see Chasen [1924] and Bucknill and Chasen [1927]; the full reference list is in the Supporting Information), we generated a comprehensive list of native resident Singaporean bird species from 1819, since collections began, to the present (Supporting Information). Our list differed from the one used by Brook et al. (2003) in that we did not make assumptions about the occurrence of a species in Singapore based on its presence in peninsular Malaysia. Given the exceptionally detailed documentation of Singaporean avifauna since the 1800s, we accepted species based only on actual breeding records or their regular occurrence in a fashion that is compatible with breeding activity (Supporting Information). Singapore's resident bird fauna is very well known—no new species have been detected since 1996. Thus, it is reasonable to assume that the number of undetected extant species in the present day is approximately zero (and so we can set $U_T = 0$) as required by Eq. 10.

For each bird species, we noted the first and last year of record and used these years to construct a time series of the number of detected extant species and the number of detected extinct species in each year up to the present (S_t and E_t). We then used Eq. 4 and Eq. 10 to estimate the proportion of all species that are extinct and the absolute number of undetected extinct species at each time step ($\hat{p}_t$ and $\hat{X}_t$ respectively). We generated 95% CIs on the $\hat{X}_t$ time series using the methods described in the previous subsection.

We also tested the assumption that average extinction rate is independent of whether a species has been detected or not by comparing the extinction rate of species that were detected in the first year of record (1819) with those detected later (and therefore were, on average, harder to detect).

Table 1. Error rates[a] and biases[b] in estimates of numbers of undetected extinct species ($\hat{X}_T$) in simulations with variation in intrinsic detection and extinction rates across species.

	CV[c] (% error rate, % bias)			
CV[c]	*0.0*	*0.2*	*0.5*	*1.0*
0.0	6.8, +6.9	5.7, +5.4	6.8, +4.9	12.7, -3.3
0.2	6.9, +4.5	6.9, +5.3	7.0, +2.4	9.3, -1.2
0.5	7.4, +7.0	6.3, +6.6	6.9, +6.6	8.4, -2.1
1.0	8.4, +3.9	6.3, +4.1	6.5, +7.1	6.0, -2.1

[a]*Measured as percentages of instances in which the true simulated value of the number of undetected extinct species X_T was outside the modelled 95% CI. Error rates should be close to 5% for an accurate model.*
[b]*Measured as the mean of the relative error $(\hat{X}_T - X_T)/X_T$. Biases should be close to zero for an accurate model.*
[c]*Coefficient of variation of intrinsic detection rates across species.*

All analyses were done in R. In the Supporting Information (Supporting Information), we provide code to calculate $\hat{p}_t$ and $\hat{X}_t$ from time series E_t and S_t. The application to the Singaporean bird data set is in .csv format. We also provide code to calculate CIs around $\hat{X}_t$.

Results

Simulations with synthetic data showed that our method produced accurate results, even when intrinsic detection and extinction rates varied across species (Table 1). The bias of the method was usually slightly positive (up to +7.1%) (Table 1), but negative biases arose (down to –3.3%) when variation in intrinsic detection rates across species was highest. The true number of undetected extinct species was within the 95% CIs usually 90–95% of the time (Table 1), indicating that the confidence intervals were roughly accurate but a little narrow.

Of a known resident bird fauna of 195 species in Singapore, 58 species have not been recorded recently and were assumed to be extinct. The estimated number of undetected extinct species was 9.6 (95% CI [3.4, 19.8]) (Fig. 2), which implies 67.6 total extinctions [61.4, 77.8]. The corresponding estimate for the fraction of extinct species (both detected and undetected) was $(58 + 9.6)/(195 + 9.6) = 33.0\%$ [31.0%, 36.2%].

With our main method, we assumed perfect detection. Under a scenario of imperfect detection, where the true extinction date of each species is drawn uniformly from the interval between its last detection date and 10 years thereafter, our simulations estimated the number of undetected extinctions of Singapore birds as slightly lower, 8.9 [3.4, 19.3]. Under a more extreme imperfect detection scenario, where each species' extinction date is drawn uniformly between the last detection date and the present day, our simulations estimated the number of undetected extinctions as substantially lower, 4.1 [0.9, 10.3].

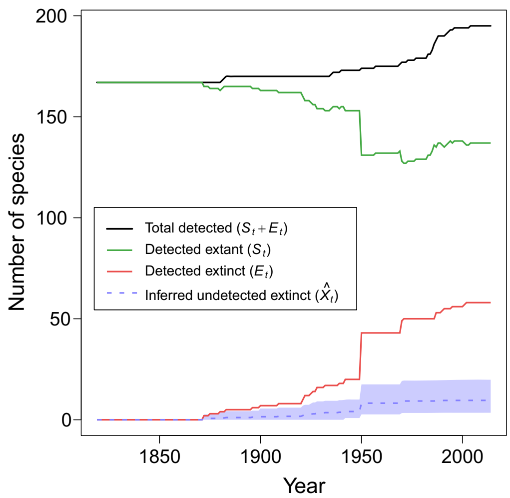

Figure 2. Time series of detected resident bird species in Singapore (solid lines) and inferred undetected extinctions (dashed line) with 95% CI (shaded region).

Annual extinction rates of early- and late-detected species were similar in the Singapore data set, providing support for the key assumption that extinction rates are independent of whether a species has been detected or not. Of the 167 species known to have occurred from the beginning of modern record keeping in 1819, 56 were extinct by 2014, which gives an overall average annual extinction rate of $1 - (1 - 56/167)^{1/(2014-1819)} = 0.0021$ species^{-1}year^{-1} for this group. Applying the same average annual rate to the 28 species detected after 1819, from each species' year of detection until 2014, there was a predicted mean of 2.56 extinctions in this group (95% CI [0,5]), whereas there were 2 observed extinctions. Thus, the annual extinction rate was slightly lower for species detected later (and, by implication, for undetected species), but the difference was minor.

Discussion

We devised the first nonparametric analytical model for estimating extinction rates that account for undetected extinctions. The model has minimal data requirements and makes few assumptions. Previously, most estimates of contemporary extinction rates have been based only on known species (Stork 2010; Barnosky et al. 2011; Pimm et al. 2014) and are therefore underestimates (Pimm et al. 2014). Our model allows contemporary extinction rates to be adjusted to account for extinctions that have escaped the scientific community's notice. It may seem surprising that undetected extinctions can be quantified without making any parametric assumptions about extinction and detection rates. It is, however, possible to intuit a reason our method works. Essentially, the method integrates two sources of information, both extractable from the time series of detected extant and extinct species. The first source of information is the detected extinction-rate time series. The second source of information concerns the past number of undetected extant species: Species that are detected and extant at time t_2 but not at time $t_1 < t_2$ were certainly undetected and extant at time t_1. This information is then enough to estimate undetected extinctions in proportional terms (Eq. 4), assuming only that extinction rates are unrelated to whether a species has been detected or not.

Estimating undetected extinctions in absolute, rather than proportional, terms from our model (Eq. 10) requires one more piece of information: the current number of undetected extant species. Intuitively, the reason for this is clear: If one has no knowledge of the current number of undetected extant species, one has no basis for making estimates of the total number of undetected species and thence the number that are extinct. For groups of species that have been studied intensively and for which the rate of new species discovery is now very low (e.g., our case study of Singaporean birds), it is reasonable to assume there are at present no undetected extant species (i.e., $U_T = 0$). But for cases involving taxa and regions where new species are frequently described, our method can be used only to estimate the total extinction rate accounting for undetected extinctions (Eq. 4), not the absolute number of undetected extinctions. In such cases, estimates of undetected extinctions may be obtained from parametric approaches such as that of Tedesco et al. (2014), with the caveat that the accuracy of the results depends on the extent to which empirical time series follow the functional forms specified by their model.

Our method is accurate and robust to variation in intrinsic variation in detection and extinction rates across species. For the simulated data sets, the true value of undetected extinctions (X_T) was within the model's 95% CI ($\hat{X}_T$) usually 90–95% of the time, and the bias ranged from -3% to +7% (Table 1; Supporting Information). The tendency toward small positive biases suggests that our method may slightly overestimate the number of undetected extinctions, which, from a precautionary perspective, may be preferable to underestimating them. Nevertheless, a goal for the future is to reduce these biases in the method even further.

In common with the parametric model introduced by Tedesco et al. (2014), we assumed in our nonparametric model that extinction rates do not vary systematically between detected and undetected species. This assumption deserves some scrutiny. On the one hand, one might expect extinction rates to be higher for detected species because they tend to be larger and more visually conspicuous than undetected species (Gaston & Blackburn 1994) and are therefore more likely to be hunted or trapped by humans. Compounding this, large species may have smaller population sizes and thus be more vulnerable to extinction (Gaston & Blackburn 1994). On the other hand, one might expect extinction rates to be higher for undetected species because many undetected species are rare and have small ranges (Blackburn & Gaston 1995) and rare species are vulnerable to extinction (Sodhi et al. 2009). In addition, undetected species may be more vulnerable because they cannot benefit from targeted conservation efforts. The net result of these countervailing effects on the extinction rate of undetected species is unclear a priori. Evidence from global data sets indicates that recently described species are more likely to be listed as threatened (Giam et al. 2012), which implies that species that are difficult to detect go extinct faster and that estimates of undetected extinctions determined using our method may therefore be biased toward lower values. One way of testing our assumption is to compare the annual extinction rate of species detected later (a subset of species that are difficult to detect) with those of species detected earlier. For our Singapore bird data set, the extinction rate of species detected later was very similar to that of species detected earlier (in fact, it was slightly lower), indicating that the assumption of equal average extinction rates, regardless of detected status, is reasonable for this case study.

For our model we also made the simplifying assumption that detection is perfect (i.e., each extinct species' extinction occurred immediately after its last detection). The bias induced by imperfect detection is positive (Eq. 6) and for a single species will generally be small unless, from the time of last detection to actual extinction, there is a huge fractional change in the number of detected species. The error can be greater if multiple species are imperfectly detected and delays between last detection and actual extinction average decades. This extreme scenario of imperfect detection does not apply to our Singaporean bird case study—there are no modern examples of rediscoveries of species thought to be extinct, despite a dramatic increase in birdwatching over the last 10 years.

Another possible complication, not considered in our model, is that species may recolonize after a local extirpation. Several species that went extinct in Singapore at some point within the last 200 years but have intermittently recolonized the island (Supporting Information) present a complicated case. Most of these are frugivores (e.g., the Pied Imperial Pigeon [*Ducula bicolor*]) whose plight in Singapore seems to have been partially reversed with maturation of the city's garden vegetation. We opted to treat these species as extant, neglecting their temporary absence in the recent past. It is possible that a limited number of additional species will recolonize the island in the future despite being currently absent as a result of local extinction. However, we considered the number of such species to be very small, such that they would not change our general conclusions.

Another potential bias in our method occurs when the detection date of a species coincides with its colonization of the focal region (i.e., the species was truly absent rather than merely undetected prior to detection). One way we dealt with this potential bias was to exclude non-native and introduced species from the analysis. However, in rare instances, species under consideration may have naturally invaded Singapore and therefore biased our analyses. We are aware of only one species for which this is certainly true (the Common Myna [*Acridotheres tristis*]) (Supporting Information). However, other species for which this is frequently suggested in the literature (Wang & Hails 2007), such as many herons and egrets, would have certainly existed in Singapore at least periodically given their long observation record in neighboring Malaysia and Indonesia. Therefore, although urbanization may have helped their populations increase, we consider their first Singaporean records to be legitimate detection dates and not dates of colonization.

It is tempting to speculate which bird species might constitute the approximately 10 species that we estimate to have gone extinct undetected in Singapore. In the lowland rainforests of the Malaysian state of Johor, just to the north, there are a number of sensitive species, many with cryptic microhabitat specializations, that were never detected in Singapore and are candidates for having gone extinct there undetected. However, we cannot confidently pinpoint which ones are the most likely. Future autecological research, focusing also on historical records from neighboring Indonesian islands to the south, may resolve this.

Our estimates for bird extinctions in Singapore (29.7% observed; 31.0–36.2% observed+undetected) are lower than previous estimates (33.9% observed, 58.5% observed+inferred [Brook et al. 2003]). The reason for this is that we used more conservative criteria to classify species as resident (Supporting Information). This is particularly evident in the discrepancy between our upper bound (36.2%) and Brook et al.'s (2003) upper bound (58.5%), which arises from the assumption by Brook et al. that all lowland species (excluding those from savanna, which does not occur in Singapore) in peninsular Malaysia were originally present in Singapore but became extinct before they could be recorded (this adds 129 putative breeding species to the Singapore list of Brook et al.). With this assumption, Brook et al. (2003) likely overestimated the number of species that have gone extinct before being recorded: Singapore's area is <1% that of lowland peninsular Malaysia and contains only a fraction of the breadth of edaphic conditions and microhabitats present on the mainland. Therefore, Singapore is unlikely to have supported the same number of species.

Extrapolating our lower (observed) and upper (observed+undetected) bounds on bird extinction estimates in Singapore to Southeast Asia with the methods and parameters of Brook et al. (2003), the lower and upper estimates for regional loss of bird species by 2100, assuming deforestation rates of 0.71%/year, are 14% and 16%, respectively, as opposed to 16% and 32% in Brook et al. (2003). We argue that our estimates, for both historical Singaporean and future Southeast Asian extinctions, are likely to be more accurate because they depend on more realistic assumptions.

Our method for estimating undetected extinctions is applicable at any spatial scale, including the scale of a single nature reserve, the national scale (e.g., our case study of Singapore), the regional scale (e.g., Southeast Asia), or the global scale. The data requirements are modest: many museums and herbaria maintain specimen databases, which contain the information necessary to determine dates of detection and extinction for particular groups of species. Our formulas are straightforward to implement in software packages, and we have provided R code that does so. Broad application of our method will contribute to a more accurate picture of how human impacts have impoverished the biological world.

Acknowledgments

R.A.C. and T.F. acknowledge National University of Singapore grant WBS R-154-000-603-112. F.E.R. acknowledges National University of Singapore grants WBS R-154-000-570-133 and WBS R-154-000-583-651.

Supporting Information

A derivation of Eq. 3 showing the estimated proportion of all species that are extinct and a derivation of Eq. 5 showing the error in this estimated proportion arising from imperfect detection (Appendix S1), a derivation of a more general version of Eq. 10 for when extinction rates differ across detected and undetected species (Appendix S2), further details of the simulations we used to verify our formulas (Appendix S3), the full list of Singaporean resident bird species with references (Appendix S4), and R code to implement our methods (Appendix S5) are available online. The authors are solely responsible for the content and functionality of these materials. Queries (other than absence of the material) should be directed to the corresponding author.

Literature Cited

Barnosky AD, et al. 2011. Has the earth's sixth mass extinction already arrived? Nature **471:**51–57.

Blackburn TM, Gaston KJ. 1995. What determines the probability of discovering a species? A study of South American oscine passerine birds. Journal of Biogeography **22:**7–14.

Brook BW, Sodhi NS, Ng PKL. 2003. Catastrophic extinctions follow deforestation in Singapore. Nature **424:**420–423.

Bucknill JAS, Chasen FN. 1927. Birds of Singapore and Southeast Asia. Tynron Press, Scotland.

Ceballos G, Ehrlich PR. 2009. Discoveries of new mammal species and their implications for conservation and ecosystem services. Proceedings of the National Academy of Sciences **106:**3841–3846.

Chasen FN. 1924. On the occurrence of certain alien birds in Singapore. Journal of the Malayan Branch of the Royal Asiatic Society **2:** 68–70.

Corlett RT. 1992. The ecological transformation of Singapore, 1819–1990. Journal of Biogeography **19:**411–420.

Costello MJ, May RM, Stork NE. 2013. Can we name earth's species before they go extinct? Science **339:**413–416.

Gaston KJ, Blackburn TM. 1994. Are newly described bird species small-bodied? Biodiversity Letters **2:**16–20.

Giam X, Scheffers BR, Sodhi NS, Wilcove DS, Ceballos G, Ehrlich PR. 2012. Reservoirs of richness: least disturbed tropical forests are centres of undescribed species diversity. Proceedings of the Royal Society B: Biological Sciences **279:**67–76.

Gibson-Hill CA. 1949. A checklist of the birds of Singapore Island. Bulletin of the Raffles Museum **21:**132–183.

Johnson C. 2006. Australia's mammal extinctions: A 50,000 year history. Cambridge University Press, Cambridge, United Kingdom.

Joppa LN, Roberts DL, Myers N, Pimm SL. 2011. Biodiversity hotspots house most undiscovered plant species. Proceedings of the National Academy of Sciences **108:**13171–13176.

Khan MKbM. 1986. Tigers in Malaysia: prospects for the future. The Journal of Wildlife and Parks **5:**1–23.

Laurance WF, et al. 2014. A global strategy for road building. Nature **513:**229–232.

Pimm SL, Jenkins CN, Abell R, Brooks TM, Gittleman JL, Joppa LN, Raven PH, Roberts CM, Sexton JO. 2014. The biodiversity of species and their rates of extinction, distribution, and protection. Science **344**(6187).

Pimm SL, Moulton MP, Justice LJ, Collar NJ, Bowman DMJS, Bond WJ. 1994. Bird extinctions in the Central Pacific. Philosophical Transactions of the Royal Society B: Biological Sciences **344:**27–33.

Richling I, Bouchet P. 2013. Extinct even before scientific recognition: a remarkable radiation of helicinid snails (*Helicinidae*) on the Gambier Islands, French Polynesia. Biodiversity and Conservation **22:**2433–2468.

Scheffers BR, Joppa LN, Pimm SL, Laurance WF. 2012. What we know and don't know about Earth's missing biodiversity. Trends in Ecology & Evolution **27:**501–510.

Sodhi NS, Brook BW, Bradshaw CJA. 2009. Causes and consequences of species extinctions. Pages 514–520 in Levin SA, editor. The Princeton Guide to Ecology. Princeton University Press, Princeton.

Stork NE. 2010. Re-assessing current extinction rates. Biodiversity and Conservation **19:**357–371.

Tedesco PA, Bigorne R, Bogan AE, Giam X, Jézéquel C, Hugueny B. 2014. Estimating how many undescribed species have gone extinct. Conservation Biology **28:**1360–1370.

Wang LK, Hails CJ. 2007. An annotated checklist of the birds of Singapore. Raffles Bulletin of Zoology Supplement **15:**1–179.

Wilcove DS, Giam X, Edwards DP, Fisher B, Koh LP. 2013. Navjot's nightmare revisited: logging, agriculture, and biodiversity in Southeast Asia. Trends in Ecology & Evolution **28:**531–540.