Missing and presumed lost: extinction in the ocean and its inference

Woollcott K. Smith[1] and Andrew R. Solow[2]*

[1]*Statistics Department, Temple University, Philadelphia, PA 19122, USA*
[2]*Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA*

Corresponding Author: tel: +1 508 289 2746; fax: +1 508 457 2184; e-mail: asolow@whoi.edu.

The number of modern extinctions in the ocean is unknown. The actual demise of the last individual of a species is essentially unobservable, so extinction can only be inferred. Statistical methods are described for inferring extinction from sighting records, species–area considerations, and taxonomic samples collected at two different times. The methods are illustrated using a variety of real datasets, including a sighting record of the Caribbean monk seal and results from three surveys of benthic invertebrates.

Keywords: Atlantic Shelf, benthic communities, species extinction, statistical sampling.

Introduction

No-one knows how many modern extinctions there have been in the ocean. Carlton *et al.* (1999) conservatively identified just 12 marine species—five birds, three mammals, and four invertebrates—declared extinct based on the time elapsed since their last sighting. Then, based on a survey of documentary material, Dulvy *et al.* (2003) increased this number to 133 extinctions, of which 21 were global. However, after a close review of the same material, Monte-Luna *et al.* (2007) argued that this estimate may be too high by a factor of 2.

Carlton *et al.* (1999) discussed three issues that arose in assessing marine extinctions: taxonomy, geography, and sampling. The first concerns the true taxonomic status of the species under consideration, the second the spatial scale of potential species loss, and the third the incompleteness of the observational record. Here, we focus exclusively on the last of these issues. Because actual extinction events, i.e. the demise of the last individual of a species, are essentially unobservable, extinction needs to be inferred. The purpose here is to review some statistical methods for this inference. Specifically, we consider inference about extinction based on sighting records, species–area considerations, and taxonomic sampling. In the first two cases, the methods reviewed are not new to the literature, but in the third case, the methods appear to be novel.

Inferring extinction from a sighting record

The most common approach to inferring the extinction of a particular species is based on the record of its sightings. This was the approach used by Carlton *et al.* (1999), Dulvy *et al.* (2003), and Monte-Luna *et al.* (2007). The basic question in inferring extinction from a sighting record is how long a species has to go unsighted before it is safe to conclude that it is extinct. Until recently, this question was answered using *ad hoc* rules, e.g. declaring a species extinct if it had gone unsighted for

50 years (Groombridge, 1993). Although conclusions based on such rules are by no means necessarily wrong, the case of the coelacanth that went unrecorded for all of human history until a living specimen was found in 1938 (Smith, 1939) is salutary.

The confidence with which extinction can be inferred from a given period during which a species is not sighted depends on the expected rate at which it would be sighted were it not extinct. In many situations, it is reasonable to assume that this rate is roughly proportional to the product of sighting effort and species abundance. The term sighting effort may be misleading because it does not necessarily imply directed effort. As with the coelacanth, sightings of many rare or cryptic species are simply by chance. Returning to the main point, unless sighting effort falls to zero, the only way that the sighting rate can fall to zero is for abundance to fall to zero, i.e. for the species to become extinct. However, even with constant sighting effort, abundance can fall to the point that the expected sighting rate is very low and confidence in extinction remains low even if the interval since the last sighting is very long. Therefore, it is necessary to base formal inference about extinction on an explicit statistical model of the sighting record.

Solow (2005) reviewed formal statistical approaches to inference about extinction based on a sighting record. These approaches can be used to test the null hypothesis that the species is extinct against the alternative hypothesis that it is not. One of the approaches reviewed in Solow (2005) assumes a constant pre-extinction sighting rate; another assumes that the pre-extinction sighting rate declines exponentially, e.g. as a result of declining abundance. As an illustration, we apply these two approaches to the sighting record for the Caribbean monk seal (*Monachus tropicalis*). This species was among the 12 identified by Carlton *et al.* (1999) as extinct. The modern sighting record for the Caribbean monk seal consists of just five sightings: in 1915, 1922, 1932, 1948, and 1952. Both methods require that the start of the observation period is known. Here, we take the start of the observation

period as 1915, and omit the initial sighting. This leaves a very sparse record and no test can be expected to have higher power. The observed significance level (or p-value) under the assumption of a constant pre-extinction sighting rate is simply

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

where t_n is the time (since the start of the observation period) of the most recent sighting, T the length of the observation period, and n the number of sightings in the record. For the Caribbean monk seal, $p = ((1952 - 1915)/(2011 - 1915))^4 = 0.022$. By conventional standards, this result is significant, so it is safe to conclude that the Caribbean monk seal is extinct.

Under the assumption of an exponentially declining pre-extinction sighting rate, the observed significance level is given by

$$p = \left(\frac{F(t_n)}{F(T)}\right), \tag{2}$$

where

$$F(x) = 1 - \sum_{i=1}^{[s/x]} (-1)^{i-1} \binom{n}{i} \left(1 - \frac{i\,x}{s}\right)^{n-1}, \tag{3}$$

s is the sum of the n sighting times, and $[\cdot]$ denotes the integer part. For the Caribbean monk seal sighting record, $p = 0.17$, which by conventional standards is not significant. In essence, allowing for an exponential decline in pre-extinction sighting reduces the extent to which even a 57-year gap in the sighting record supports the hypothesis of extinction. We note that there is no evidence in this sighting record of a declining sighting rate before 1952, so the assumption of there being one is certainly questionable.

The basic point here is that the extent of support that a gap in a sighting record lends to the hypothesis of extinction depends on what can be assumed about the behaviour of the pre-extinction sighting rate. This behaviour, which *inter alia* depends on sighting effort and pre-extinction abundance, can be complicated and, without additional information, difficult to assess from the sighting record itself.

Solow (2005) also described a test for extinction using the most recent sightings in a record containing many sightings. This method is based on a remarkable asymptotic distributional result for extreme order statistics from any distribution with a finite endpoint, so avoiding the need to specify a model for the pre-extinction sighting rate. Let $t_n > t_{n-1} > \cdots > t_{n-k+1}$ be the k most recent of a large set of n sighting times in the observation period $(0, T)$. The approximate significance level for this test is

$$p = \exp\left(-k\left(\frac{T - t_n}{T - t_{n-k+1}}\right)^{1/\hat{v}}\right), \tag{4}$$

where

$$\hat{v} = \frac{1}{k-1} \sum_{i=1}^{k-2} \log \frac{t_n - t_{n-k+1}}{t_n - t_{n-i+1}}. \tag{5}$$

Note that neither the beginning of the observation period nor n needs to be known to apply this test. As an illustration, suppose

the sighting record of the Caribbean monk seal actually represents the five most recent of a larger number of sightings. The value of $\hat{v}$ is 0.76 and the significance level is

$$\exp\left(-5\left(\frac{2011 - 1952}{2011 - 1915}\right)^{1/0.76}\right) = 0.072,$$

which, by conventional standards, is ambiguous as to extinction. Of course, the treatment of this sighting record as the five most recent of a much larger set is not justified, and the asymptotic result is open to question even as an approximation.

Inferring extinctions from species–area considerations

As on land, one of the chief threats to species in the marine environment is habitat alteration, e.g. the degradation and loss of coral-reef habitat. A common approach to estimating species loss attributable to habitat alteration is through the species–area curve (Pimm and Raven, 2000). Let $s(a)$ be the number of species contained in a region of area a. It is commonly assumed that $s(a) \propto a^z$ for some constant $z < 1$. It follows that reducing the area by a factor c will reduce the species number by a factor c^z. The exponent z is commonly taken as ~ 0.3 so, for example, a 90% reduction in area corresponds to a 50% reduction in species number. Reaka-Kudla based her prediction that up to 1.2 million reef species would be lost by the middle of the 21st century on this approach (Malakoff, 1997). A general discussion of the species–area curve and marine conservation was provided by Neigel (2003).

The species–area curve describes the rate at which the number of species contained in a region increased with the area of the region. Interest in this relationship dates back at least to Watson (1835). A significant landmark in establishing the centrality of this relationship with ecology was the publication of the monograph by MacArthur and Wilson (1967) on island biogeography. In this early work, the species present on each island were known or assumed to be known. More commonly, the species–area relationship is estimated from taxonomic counts within small spatial samples or quadrats (Colwell and Coddington, 1994). In this case, the behaviour of the sample species–area curve reflects two factors. First, as the spatial extent of the quadrats increases, the coverage of the heterogeneity of the region with respect to the species present also increases. Second, as the number of quadrats increases, the chance of observing spatially rare species also increases. Only the first of these is relevant to assessing extinction, whereas the second is a sampling artefact.

Suppose that n quadrats of unit area are sampled and the species within each are recorded. Suppose further that a total of s species is found in the union of these quadrats, and let n_j be the number of quadrats containing species j. The sample species–area relationship is given by

$$s_{\text{sample}}(m) = s - \sum_{j=1}^{s} \frac{\binom{n - n_j}{m}}{\binom{n}{m}}, \tag{6}$$

for $m = 1, 2, \ldots, n$, and where $\binom{a}{b} = 0$ if $a < b$ (Solow and Smith, 1991). The quantity $s_{\text{sample}}(m)$ is precisely the expected

number of species contained in m sample quadrats sampled at random without replacement. Of course, quadrats are not islands. If instead each quadrat is properly seen as a sample from a different island (or another subregion) within a larger region of interest, then it is almost certain that the species overlap between quadrats will underestimate the species overlap between subregions. Depending on how species are distributed across subregions, the effect of this on the way in which species actually accumulate with increasing area can be very large.

Suppose that each quadrat represents a sample from a subregion. Let x_{ij} be the number of individuals of species j contained in quadrat i, $x_j = (x_{1j} x_{2j} \cdots x_{nj})$ be the vector of quadrat counts for species j, and K_j be the unknown number of subregions that contain species j. By analogy to the sample species–area relationship in Equation (4), Chu *et al.* (1998) defined the hidden species–area relationship as

$$s_{\text{hidden}}(m) = s - \sum_{j=1}^{s} \frac{E\left(\binom{n - K_j}{m} | x \right)}{\binom{n}{m}}, \quad (7)$$

where $x = (x_1 x_2 \cdots x_s)$ is the complete set of observed species counts. The hidden species–area curve is the conditional expected value of the number of species contained in a random sample of m subregions, given the observed species counts. If each quadrat is treated as a subregion (and not as a sample from a subregion), then $K_j = n_j$ and the hidden species–area relationship in Equation (7) is the same as the sample species–area relationship in Equation (6). Otherwise, $s_{\text{hidden}}(m) \geq s_{\text{sample}}(m)$ and $s_{\text{sample}}(m)$ will overestimate the rate at which species accumulate with area.

To take this argument further, it is necessary to specify a statistical model for the way in which individuals are distributed within subregions. Chu *et al.* (1998) presented one such model. The model operates in the following way. First, $K_1, K_2, \ldots, K_s$ are assumed to be independent and identically distributed with a zero-truncated beta-binomial distribution with parameters α and β. The beta-binomial distribution extends the familiar binomial distribution to allow for variability between species in the occurrence probability within subregions. The zero truncation ensures that the species are present in at least one subregion. Second, given K_j, species j is distributed at random to K_j of the n subregions. Third, the total observed number $\sum_{i=1}^{n} x_{ij}$ of individuals of species j is distributed among these K_j sample quadrats according to a Dirichlet multinomial distribution with common dispersion parameter γ. The Dirichlet multinomial distribution is the extension of the beta-binomial distribution to the situation in which >2 outcomes are possible. Note that in this model, the process by which species are distributed in subregions is decoupled from the process by which individuals are distributed in samples within subregions.

The parameters α, β, and γ of this model can be estimated by the method of maximum likelihood—technical details are provided in Chu *et al.* (1998)—and these estimates can be used to estimate the hidden species–area relationship in Equation (5). Computer code for performing these calculations is available from the first author. A key assumption of this model is that the dispersion parameter γ is the same for all species. Roughly speaking, this allows the variability of positive quadrat counts of

abundant species to provide information about the zero counts of rare species.

As an illustration, we applied this approach to data from the classic benthic survey described in Grassle and Maciolek (1992). This example involves ten stations located on a 180-km transect along the 2100-m isobath off the coast of New Jersey. The data at each station, which are summarized in Table 1, represent pooled results over six sampling times and three replicate cores per sampling time. Figure 1 shows the sample species–area relationship in Equation (6). This increases from 297 species in a single sample to 647 species in all ten samples and conforms reasonably well to a power model with exponent 0.3. The estimated hidden species–area relationship, which is also shown in Figure 1, is strikingly different, increasing only slightly from around 590 species in a single subregion to 647 in all ten subregions. In essence, under the model outlined above, the data are consistent with a situation in which the species tend to be cosmopolitan across subregions, but within subregions, individuals are distributed patchily.

Returning to the question of assessing species loss, it is clear that failing to account for unsampled species in estimating the regional species–area relationship can have a dramatic effect on conclusions about the effect of habitat loss. Accounting for unsampled species is certainly a challenge and requires specification of a statistical model. The model of Chu *et al.* (1998) is

Table 1. Summary statistics for ten deep-sea sampling stations.

Station	Individuals	Species	Singleton species	Unique species
1	7 093	301	92	28
2	8 155	329	97	28
3	6 632	325	89	26
4	7 683	281	80	17
5	7 178	308	82	16
6	5 224	267	69	17
7	6 070	304	86	23
8	2 559	225	67	12
9	5 956	278	74	13
10	7 739	351	114	42
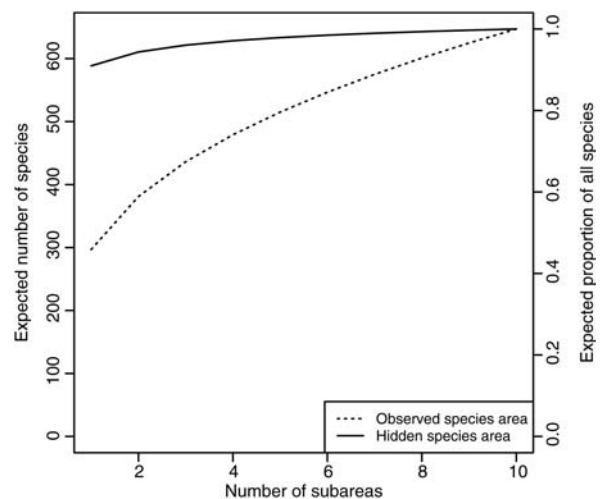
Figure 1. Observed species–area curve (dotted line) and estimated hidden species–area curve (solid line) for deep-sea benthic data.

only one possibility, and there is need for more work in this area and, in particular, in the development of models that reflect what is known about biogeographical processes.

Inferring extinction from direct species sampling

The most direct way to infer extinction in a community or other group of species is by observing the community over time. Here, we focus on the situation in which taxonomic samples are taken at two times, and the question of interest is whether any of the species observed in the earlier sample have become extinct by the time of the later sample. Within this situation, we consider two cases. In the first, species counts are available for both times. In the second case, only a species list is available for the earlier time. In both cases, the main difficulty is that the absence of a species in the second sample can be attributed either to extinction or to sampling variability.

Suppose that an earlier sample contains a total of s species, and let $E_1, E_2, \ldots, E_s$ be the species counts for those species. A second sample is taken later. Let $L_1, L_2, \ldots, L_s$ be the species counts for those same s species in the second sample. Note that E_j must be >0 for all j (otherwise, the species is not observed in the earlier sample), but that L_j can be zero. Let $C_j = E_j + L_j$ be the combined count for species j in both samples. We assume that, conditional on the observed value c_j of C_j, L_j has the probability mass function

$$\text{prob}(L_j = l_j | C_j = c_j) = \binom{c_j}{l_j} \prod_j^{l_j} (1 - \Pi_j)^{c_j - l_j}, \quad (8)$$

where the random probabilities $\Pi_1, \Pi_2, \ldots, \Pi_s$ follow a mixture of a point mass q at zero and a beta distribution with parameters a and b over the unit interval. This model is a modification of one proposed by Smith et al. (1996), who called it a delta–beta binomial model, for estimating species overlap. The basic idea is that the point mass at zero accommodates zero species counts associated with extinction, whereas the beta component accommodates variability in the sampling probabilities of species that are not extinct.

Let $l_1, l_2, \ldots, l_s$ be the observed species counts in the second sample. The log-likelihood of the complete sample is given by

$$\log L(q, a, b) = \sum_{j=1}^{s} \log p(l_j | c_j), \quad (9)$$

where

$$p(l_j | c_j) = q\,\delta_j + (1 - q) \binom{c_j}{l_j} \frac{B(l_j + a, c_j - l_j + b)}{B(a, b)} \quad (10)$$

is the conditional probability mass function of L_j given $C_j = c_j$, where $\delta_j = 1$ if $l_j = 0$ and 1 otherwise, and B is the beta function. This log-likelihood can be used to test the null hypothesis $H_0 : q = 0$ that all zero counts at the later time are attributable to sampling variability against the one-sided alternative hypothesis $H_1 : q > 0$ that some may be attributable to extinction. The likelihood ratio (LR) statistic for testing H_0 against H_1 is

$$\Lambda = 2(\log L(\hat{q}, \hat{a}, \hat{b}) - \log L(0, \hat{a}(0), \hat{b}(0))), \quad (11)$$

where $\hat{q}$, $\hat{a}$, and $\hat{b}$ are the unrestricted maximum likelihood (ML)

estimates of q, a, and b, and $\hat{a}(0)$ and $\hat{b}(0)$ are the ML estimates of a and b with q fixed at zero. The null hypothesis can be rejected at approximate significance level α if the observed value of Λ exceeds the upper 2α-quantile of the χ^2 distribution with 1 degree of freedom. Computer code for performing this test is available from the first author.

As an illustration, we applied this method to some data from a benthic survey conducted by the Massachusetts Water Resources Authority in Massachusetts Bay. These data, which are described in more detail in Maciolek et al. (2008), consist of two samples collected in 1996 and 2007 near the end of the Boston Harbour sewage outfall. This outfall began operating in 2000, and the samples were taken as part of a large monitoring programme designed to identify and measure any effects of the outfall. The data are presented in Figure 2, where the abundances (on a log-scale) in 2007 of the 203 species found in 1996 are plotted against their abundances in the 1996 sample. Of the species found in 1996, 49 were not found in 2007. Most of these were rare in the 1996 sample, but some were not.

For these data, the unrestricted ML estimate of q is 0.06, i.e. an estimated 6% of the species observed in 1996 were no longer present in 2007. This contrasts with the nearly 25% of species observed in 1996 that were not observed in 2007. The value of the LR statistic is 3.77 for an approximate p-value of $\sim$0.026. Therefore, by conventional standards of significance, we reject the null hypothesis of no extinctions. Two points are worth emphasizing here. First, the spatial scale of the sampling is very small and provides only a highly local picture of species loss. Second, there is no reason to believe that any local species loss was caused by the operation of the outfall. In particular, the 2007 sample contains 36 species that were not present in 1996. Considerably more information is needed about the natural rate of species turnover at this site before any issues of attribution can be addressed.

Finally, we consider the case in which only a species list is available for the earlier time and species counts are available only for
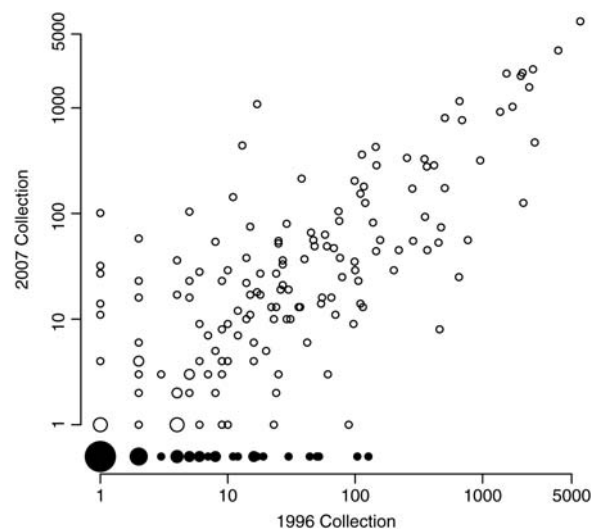
Figure 2. Sample abundances (at a log-scale) in 2007 of 203 benthic species found in 1996 at the Boston Harbour outfall site vs. their abundance in 1996. Larger points indicate species numbers >1. Species with zero abundance in 2007 are indicated as filled circles just above the horizontal axis.

the later time. Because species counts are not available for both times, it is not possible as above to focus on how they are distributed between the two times conditional on their sum. Instead, we model the species counts in the later period using a stochastic abundance distribution (Chao and Bunge, 2002) extended to allow for extinction. A convenient stochastic abundance model is the Poisson lognormal (Bulmer, 1974). Under this model, the species count L_j for species j has a Poisson distribution with stochastic mean Θ_j that follows a lognormal distribution with mean μ and variance σ^2. As above, we extend this model to allow for extinction by including an additional probability q of a zero count.

The log-likelihood of the later sample is

$$\log L(q,\ \mu,\ \sigma^2) = \sum_{j=1}^{s} \log p(l_j). \tag{12}$$

Here,

$$p(l_j) = q\,\delta_j + (1-q)f\,(l_j;\ \mu,\ \sigma^2), \tag{13}$$

where $\delta_j = 1$ if $l_j = 0$ and 0 otherwise, and $f\,(l_j;\ \mu,\ \sigma^2)$ is the Poisson-lognormal probability mass function evaluated at l_j. The LR statistic for testing the null hypothesis $H_0: q = 0$ of no extinction against the one-sided alternative hypothesis $H_1: q > 0$ is

$$\Lambda = 2\,(\log L(\hat{q}, \hat{\mu},\ \hat{\sigma}^2) - \log L(0,\ \hat{\mu}(0),\ \hat{\sigma}^2(0))), \tag{14}$$

where $\hat{q}$, $\hat{\mu}$, and $\hat{\sigma}^2$ are the unrestricted ML estimates of q, μ, and σ^2, and $\hat{\mu}(0)$ and $\hat{\sigma}^2(0)$ are the ML estimates of μ and σ^2 under the restriction that $q = 0$. As above, H_0 can be rejected at approximate significance level α if the observed value of Λ exceeds the upper 2α-quantile of the χ^2 distribution with 1 degree of freedom. Computer code for performing this test is also available from the first author.

As an illustration, we applied this method to some data on benthic invertebrates on Georges Bank off the coast of Massachusetts. To begin with, we used the Ocean Biogeographic Information System to compile a list of 154 benthic invertebrate species from specimens collected before 1960. We then extracted the counts of these species from the 1981/1982 surveys conducted by Frederick Grassle and Nancy Maciolek (Maciolek and Smith, 2009). The 1981/1982 counts for these 154 species, of which 42 (or just over 27%) are zero, total 169 772 individuals. The unrestricted ML estimates of q, μ, and σ are 0.17, 3.14, and 2.97, respectively, and the ML estimates of μ and σ under the restriction that $q = 0$ are 2.10 and 3.64. The value of the LR statistic is 4.47, with an approximate significance level of 0.017. Therefore, we can reject the null hypothesis that there have been no extinctions. As a check of the model fit, Table 2 reports the relative frequencies of different values of species counts along with the frequencies fitted under H_1. Note that the fitted value of the frequency of zero counts is guaranteed to equal the observed value. In overall terms, the fit appears to be good, so we conclude that of the 42 species that were unobserved in 1981/1982, an estimated 26 are no longer present. A possible explanation for this species loss is the intensive bottom-trawling on Georges Bank during the 1960s and the 1970s (Collie *et al.*, 1997).

Table 2. Observed and fitted frequencies to species counts from the 1981/1982 survey of Georges Bank.

Count	Observed frequency	Fitted frequency
0	0.273	0.273
1	0.045	0.054
2	0.039	0.037
3	0.013	0.028
4	0.032	0.023
5	0.013	0.019
6–10	0.091	0.066
11–50	0.162	0.172
51–100	0.084	0.071
101–500	0.123	0.133
501–1 000	0.039	0.040
1 000–10 000	0.071	0.068
>10 000	0.013	0.017

Discussion

We began by stating that no-one knows how many modern extinctions there have been in the ocean. Of course, for practical purposes, the precise number of modern marine extinctions cannot be known and the real question is whether there is any credible estimate. The answer to this question is also negative, and our purpose has been therefore to take a step towards explaining why this might be. The basic reason for the lack of an estimate is that it is difficult to generate one. The main difficulty is that extinction is a singular event and the observational basis for inferring it is very thin. To make progress, it is necessary to supplement the observational record with fairly substantial statistical assumptions. As illustrated here, this allows one to conduct formal inference about extinction. This level of quantitative rigour is relatively new to questions on extinction. The results of these methods may be sensitive to the underlying statistical model, and the thinness of the observational record poses a challenge to the type of model-validation exercise that would otherwise be possible. For this and other reasons, there is a strong need to base the statistical models on biological knowledge rather than mathematical convenience.

Further, an argument can be made for moving away from a focus on numbers and instead developing a deeper understanding of the processes underlying the extinction of different types of organism in different habitats facing differing types of perturbation. Beyond its scientific value, such an understanding could support both a qualitative assessment of the likely extinction impacts of more easily observed perturbations, as well as measures that can be taken to avoid or mitigate them.

Acknowledgements

References

Bulmer, M. G. 1974. On fitting the Poisson lognormal distribution to species-abundance data. Biometrics, 30: 101–110.

Carlton, J. T., Geller, J. B., Reaka-Kudla, M. L., and Norse, E. A. 1999. Historical extinctions in the sea. Annual Review of Ecology and Systematics, 30: 515–538.

Chao, A., and Bunge, J. 2002. Estimating the number of species in a stochastic abundance model. Biometrics, 58: 551–559.

Chu, C. S., Smith, W. K., and Solow, A. R. 1998. Estimating hidden species–area relationships. *In* American Statistical Association Proceedings in Statistics and the Environment, pp. 125–129.

Collie, J. S., Escanero, G. A., and Valentine, P. C. 1997. Effects of bottom fishing on the benthic megafauna of Georges Bank. Marine Ecology Progress Series, 155: 159–172.

Colwell, R. K., and Coddington, J. A. 1994. Estimating terrestrial biodiversity through extrapolation. Philosophical Transactions of the Royal Society of London, Series B, 345: 101–118.

Dulvy, N. K., Sadovy, Y., and Reynolds, J. D. 2003. Extinction vulnerability in marine populations. Fish and Fisheries, 4: 25–64.

Grassle, J. F., and Maciolek, N. J. 1992. Deep-sea species richness: regional and local diversity estimates from quantitative bottom samples. The American Naturalist, 139: 313–341.

Groombridge, B. 1993. The IUCN Red List of Threatened Animals 1994. IUCN, Gland, Switzerland.

MacArthur, R. H., and Wilson, E. O. 1967. The Theory of Island Biogeography. Princeton University Press, Princeton, NJ.

Maciolek, N. J., Doner, S. A., Dahlen, D. T., Diaz, R. J., Hecker, B., Hunt, C., and Smith, W. K. 2008. Outfall benthic monitoring interpretive report: 1992–2007. Massachusetts Water Resources Authority Report, 2008-20, Boston, MA.

Maciolek, N. J., and Smith, W. K. 2009. Benthic species diversity along a depth gradient: Boston Harbor to Lydonia Canyon. Deep Sea Research II, 56: 1763–1764.

Malakoff, D. 1997. Extinction on the high seas. Science, 277: 486–488.

Monte-Luna, P., Lluch-Belda, D., Serviere-Zaragoza, E., Carmona, R., Reyes-Bonilla, H., Aurioles-Gamboa, D., Castro-Aguirre, J. L., *et al.* 2007. Marine extinctions revisited. Fish and Fisheries, 8: 107–122.

Neigel, J. E. 2003. Species-area relationships and marine conservation. Ecological Applications, 13: 138–145.

Pimm, S. L., and Raven, P. 2000. Biodiversity: extinction by numbers. Nature, 403: 843–845.

Smith, J. L. B. 1939. A surviving fish of the order Actinistia. Nature, 143: 455–456.

Smith, W. K., Solow, A. R., and Preston, P. E. 1996. An estimator of species overlap using a modified beta-binomial model. Biometrics, 52: 1472–1477.

Solow, A. R. 2005. Inferring extinction from a sighting record. Mathematical Biosciences, 195: 47–55.

Solow, A. R., and Smith, W. K. 1991. Detecting cluster in a heterogeneous community sampled by quadrats. Biometrics, 47: 311–317.

Watson, H. C. 1835. Remarks on the Geographical Distribution of British Plants. Longmans, London.