Manuscript Information:

- **Classification:** Biological Sciences
- **Number of text pages for figures and tables:** 1
- **Word and character counts:**
  - **Words in abstract:** 108
  - **Characters in manuscript**
    * Characters: 11,696
    * Figure 1, one column, 8 cm high: 180*8 = 1440
    * Figure 2, one column, 8 cm high: 180*8 = 1440
    * Equations, 8x60 = 480
    * Space allowance, equations and figs = 1200
    * **Total Characters in paper:** = 15,824
On the pattern of discovery of introduced species

Christopher J. Costello* and Andrew R. Solow†

January 5, 2003

*Corresponding author. Donald Bren School of Environmental Science & Management, 4410 Donald Bren Hall, UC Santa Barbara, Santa Barbara CA 93106, Tel: (805) 893-5802, Fax: (805) 893-6113, costello@bren.ucsb.edu
†Marine Policy Center, Mail Stop 41, Woods Hole Oceanographic Institution, Woods Hole MA 02543
Introductions of non-indigenous species can have significant effects. It is commonly claimed that the rate of species introductions to the US has increased over time. This claim is based in part on the increasing rate of discoveries of introduced species. This discovery rate is influenced by factors other than the introduction rate. These include the sampling rate and population growth in the introduced species. In this paper, we show that the discovery rate can increase even when there is no increase in either the introduction rate or the sampling rate. This suggests that the basis for some claims regarding an increasing rate of introductions may be invalid.

The introduction of non-indigenous species to both the terrestrial and marine environments can have significant ecological and economic effects. These introductions pose a major threat to biological diversity (Williamson, 1999). Fifty thousand non-indigenous species are believed to exist in the US alone, where their predation on and competition with native species has been implicated in the decline of at least 400 of 958 species on the Endangered Species List (Wilcove et al., 1998; Pimentel et al., 2000). The annual damage caused by introduced species in the US has been estimated to be more than $100 billion (Pimentel et al., 2000).

It is often claimed that the rate of species introductions to the US has increased over time. This claim is based, in part, on the observation that the rate of discovery of introduced species has increased over time. For example, Cohen and Carlton (1998) showed that, even after correcting for extraordinary collecting effort, the time series of the cumulative number of introduced species discovered in the San Francisco Estuary is convex - that is, increasing at an increasing rate - over the period 1850-1995; Figure 1 of this paper is based on Cohen and Carlton’s adjusted time series. On the basis of this finding, they concluded that the rate of increase of actual introductions had also increased over time.

The discovery curve for introduced species reflects, in part, the process by which specimens are collected. In interpreting this curve, it is necessary to account for this collection process. For example, the rate at which introduced species are discovered will depend, to some extent, on collection effort. Cohen and Carlton (1998, p. 556) recognized this and excluded from their analysis species that were discovered through “an extraordinary collection effort.” In this paper, we go a step further and use a simple model of the discovery process to show that an increasing rate of discoveries need not imply an increasing rate of introductions even when collection effort is constant. This somewhat surprising result suggests that, by itself, a convex discovery curve cannot be taken as evidence of an increasing introduction rate.

A model of discovery

Let the random variable \( Y(t) \) be the cumulative number of introduced species that have been discovered by time \( t \). This random variable is given by:

\[
Y(t) = \sum_{j=1}^{n(t)} I_j(t)
\]

where \( I_j(t) \) is the binary random variable taking value 1 if the \( j^{th} \) introduced species is discovered by time \( t \) and 0 otherwise and \( n(t) \) is the cumulative number of species introduced by time \( t \). It follows that the expected value of \( Y(t) \) is:

\[
E[Y(t)] = \sum_{j=1}^{n(t)} \text{prob}(I_j(t) = 1)
\]

\[
= n(t) - \sum_{j=1}^{n(t)} q_j(t)
\]

where \( q_j(t) = 1 - \text{prob}(I_j(t) = 1) \) is the probability that the \( j^{th} \) species has not been discovered through time \( t \). We will refer to \( E[Y(t)] \) as the expected discovery curve.

To proceed, it is necessary to specify a model for \( q_j(t) \). We will adopt the following simple model. The basic unit of sampling is the individual organism. Individuals are sampled at
random from an effectively infinite collection of individuals representing both indigenous and introduced species. Sampling is random in the sense that, if the $k^{th}$ individual is sampled at time $t$, then the probability that it belongs to the $j^{th}$ introduced species is the relative abundance $\pi_j(t_k)$ of this species at this time. It follows that:

$$q_j(t) = \frac{r(t)}{\prod_{k=1}^{n(t)} (1 - \pi_j(t_k))}$$  \hspace{1cm} (3)

where $r(t)$ is the cumulative number of individuals sampled by time $t$ and, upon combining (2) and (3) that:

$$E[Y(t)] = n(t) - \sum_{j=1}^{n(t)} \frac{r(t)}{\prod_{k=1}^{n(t)} (1 - \pi_j(t_k))}$$  \hspace{1cm} (4)

The expression in (4) shows that, in qualitative terms, the expected discovery curve depends on (1) the temporal pattern of introductions, (2) the temporal pattern of sampling, and (3) the temporal pattern of relative abundances of the introduced species. As noted, interpretations of empirical discovery curves have focused almost exclusively on the pattern of introductions. Specifically, convexity of empirical discovery curves is taken as evidence that $n(t)$ is also convex - that is, that the rate of introductions has increased over time. This conclusion is occasionally tempered by noting the potential importance of variations over time in discovery effort. However, in the next section, we show that the expected discovery curve is typically convex even when both the introduction rate and the sampling rate are not increasing.

**Some results**

In this section, we present some results for the case where the rate at which individuals are sampled is constant. Specifically, we will assume that $r(t) = rt$ and, for convenience, we will rescale time so that $r = 1$. This amounts to measuring time in units of the constant sampling interval. To begin with, suppose that $n$ species are introduced at $t = 1$, and that no subsequent introductions take place, so that $n(t) = n$ for all values of $t$. Also, suppose that these species are identical, in the sense that $\pi_j(t) = \pi(t)$ for all values of $j$. In this case, the expected discovery curve is:

$$E[Y(t)] = n \left( 1 - \prod_{k=1}^{t} (1 - \pi(k)) \right)$$  \hspace{1cm} (5)

A necessary and sufficient condition for convexity of (5) is:

$$\pi(t) > \frac{\pi(t - 1)}{1 - \pi(t - 1)}$$  \hspace{1cm} (6)

for all $t$. Provided that the relative abundance of each introduced species increases, but remains small (so that the denominator of (6) is close to 1), over the period of observation, the expected discovery curve will be convex even with no new species introductions.

This result implies that the expected discovery curve will also be convex if species with identical growth patterns are introduced at a constant rate, so that $n(t) = nt$, provided that:

$$\pi_j(t) > \frac{\pi_j(t - 1)}{1 - \pi_j(t - 1)}$$  \hspace{1cm} (7)

for all $j$ and $t > j$, where $\pi_j(t)$ is the relative abundance at time $t$ of a species introduced at time $j$ for all $j$ and $t > j$. This follows from the previous case because the overall expected discovery curve is the sum of expected discovery curves for the species introduced at each time. Provided that (7) holds, each of these curves will be convex and, as the sum of convex curves is also convex, so will be the overall expected discovery curve. Note that, even if all species have the same pattern of post-introduction population growth, their relative abundances will, in general, depend on the time of introduction, if only because of the growth in the number of introductions. The condition in (7) will be met if, over the period of observation, the post-introduction abundance of each introduced species remains a small fraction of, but grows faster than, the total abundance of both indigenous and introduced species. This
condition is sufficient, but not necessary for convexity under a constant introduction rate. For example, suppose that there are \( N \) indigenous species with average abundance scaled to \( 1 \). At each time, \( n \) species are introduced at average abundance \( 1 \) and there is no population growth thereafter. Under these assumptions, the expected discovery curve is:

\[
E[Y(t)] = n \left( t - \sum_{j=1}^{t} \prod_{k=j}^{t} \left( 1 - \frac{1}{N + nk} \right) \right)
\]  \hspace{1cm} (8)

It can be shown by induction that (8) is convex provided \( N > 1 \).

We have shown that, even if the rate of species introductions is constant, the expected discovery curve can be convex both when post-introduction population growth is rapid and when there is no post-introduction population growth. It is not surprising that the same result can hold for the intermediate case in which the population of each introduced species first grows to, and then remains at, a constant level. For example, suppose that there are \( N = 15,000 \) indigenous species with average abundance 2. Each time period (measured in months in this example), \( n = 2 \) species are introduced at abundance 0.002 and thereafter each population grows according to the logistic model with carrying capacity \( K = 2 \) and intrinsic growth rate \( r \). In Figure 2, expected discovery curves are shown for \( r = 0.1, 1, \) and 10. The corresponding times until the population reaches 0.99\( K \) are 115, 11.5, and 1.15, respectively. Parameter values have been chosen to demonstrate that our model can yield discovery rates similar to those depicted in Cohen and Carlton. Other combinations of parameters could yield similar results. In all cases, the expected discovery curve is convex over the period of observation. As the total relative abundance of introduced species grows, there is a tendency for the expected discovery curve to become linear. For the obvious reason, this occurs faster the greater \( r \).

**Discussion**

The goal of this paper has been to show that the pattern of discoveries of introduced species may not reflect the true pattern of introductions. In broad terms, the basic result is that the expected discovery curve for introduced species will be convex even if the introduction rate is constant provided the relative abundance of each introduced species remains small in absolute terms. This result is quite robust and, in particular, does not depend on the details of the distribution of relative abundances. Thus, after correcting for sampling effort, an increasing rate of discoveries of introduced species does not necessarily imply that the rate of introductions has also increased. This is a theoretical result and does not imply that the actual rate of introduction of species to the US (or elsewhere) has not increased over time. Instead, it points out that this conclusion cannot be based on the convexity of the empirical discovery curve alone. To interpret the pattern of discoveries and obtain rigorous empirical estimates of the true rates of introduction would require careful analysis of both the sampling process and the population processes of introduced species beyond the scope of this paper. We have shown, though, that this model can yield discovery rates that closely correspond to those empirically observed. As an alternative to detailed modeling, conclusions about the rate of introductions can be based on other types of information.

The sampling model on which our results are based is widely used - either explicitly or implicitly - in the analysis of sampled species data. While this model is clearly extremely simple, there is nothing in this simplicity that unfairly favors these results.

**Acknowledgments**

We thank Tin Klanjsek for useful suggestions about the convexity results and Carol McAusland and an anonymous referee for helpful comments.
References


Figure 1: Cumulative counts of non-indigenous species in San Francisco Bay estuary, 1850-1995 correcting for extraordinary collection effort. Based on Figure 1B from Cohen and Carlton, 1998.
Figure 2: Expected cumulative discovery curves through time under the logistic growth model. Each curve represents a different value of the intrinsic growth rate, \( r \).