This is a repository copy of *The distribution of species range size: a stochastic process*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/1342/

---

**Article:**


---

**Reuse**
Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher’s website.

**Takedown**
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.
The distribution of species range size: a stochastic process

Kevin J. Gaston¹* and Fangliang He²,³

¹Biodiversity and Macroecology Group, Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK
²Canadian Forest Service, Pacific Forestry Centre, Victoria, BC, Canada V8Z 1M5
³Department of Statistics and Actuarial Science, Simon Fraser University, 8888 University Drive, Burnaby, BC, Canada V5A 1S6

The major role played by environmental factors in determining the geographical range sizes of species raises the possibility of describing their long-term dynamics in relatively simple terms, a goal which has hitherto proved elusive. Here we develop a stochastic differential equation to describe the dynamics of the range size of an individual species based on the relationship between abundance and range size, derive a limiting stationary probability model to quantify the stochastic nature of the range size for that species at steady state, and then generalize this model to the species-range size distribution for an assemblage. The model fits well to several empirical datasets of the geographical range sizes of species in taxonomic assemblages, and provides the simplest explanation of species-range size distributions to date.

Keywords: geographical range size; macroecology; stochasticity; temporal dynamics

1. INTRODUCTION

Species differ enormously in the sizes of their geographical ranges. Some are narrowly distributed, and others occur over areas that may be many orders of magnitude larger (Brown et al. 1996; Gaston 1996). For a given taxonomic assemblage, the frequency distribution of species range sizes (the species-range size distribution) tends to be strongly right-skewed, with the smallest size class being the modal one. That is, most species are rather restricted in their geographical occurrence, and only a few are widespread (Anderson 1977, 1984a,b; Pagel et al. 1991; Gaston 1994, 1996, 1998; Brown et al. 1996; Gaston & Chown 1999).

Much of the discussion of the determinants of the shape of species-range size distributions has focused on the roles of speciation and extinction processes (e.g. Anderson 1985; Flessa & Thomas 1985; Chown 1997; Gaston 1998; Gaston & Chown 1999). Speciation adds new ranges and, depending on its mode, may reduce the sizes of those of the ancestral species (e.g. through vicariance). Extinction removes ranges. Although they clearly determine the number of species that are extant at any one time, it is unlikely, however, that these processes are sufficient themselves to explain the species-range size distributions that are actually observed. Rather, the form of these distributions will be set principally by the temporal dynamics (expansions and contractions) of the range sizes of species between their first appearance and their ultimate demise (Webb & Gaston 2000). These dynamics in turn will be determined by the environmental tolerances and capacities of species (and the effects of selection on these features), how abiotic and biotic conditions change, and how species are able to respond to these changes (e.g. dispersal and colonization abilities). Indeed, the influence of environmental changes on geographical ranges is well established, with local populations being founded and lost, and range limits moving as conditions alter in ecological, let alone evolutionary, time (Burton 1995, 2001; Spicer & Gaston 1999; Dynesius & Jansson 2000; Hewitt 2000).

The importance of environmental factors, and the ability of species to respond, in influencing geographical range sizes raises the possibility of developing a stochastic theory of species-range size distributions. At any one time, the majority of species will have passed beyond the strongly determinate phase of range expansion associated with their initial geographical spread, and the colonization of, and establishment in, new habitats (figure 1). Subsequent changes in range size will reflect both non-random and random temporal variations in the environment. The consequences of both are, however, difficult to predict. Even non-random environmental changes may have complex effects because these changes are themselves spatially heterogeneous and temporally complex (with cycles commonly acting on several time-scales in addition to any broader directional trends), and because for any given species the individuals on which they are acting commonly exhibit both phenotypic and genotypic variation, and thus different responses to a particular alteration of conditions (Spicer & Gaston 1999). Thus, even when local population dynamics and shifts of particular range boundaries may be interpretable in terms of local conditions, changes in the size of whole ranges may often appear essentially stochastic (see § 5).

In this study we first develop a stochastic differential equation (SDE) to describe the dynamics of the range size of an individual species. Based on this SDE, a limiting stationary probabilistic model is derived to quantify the stochastic nature of the range size for that species at steady state. This model is then generalized to the species-range size distribution for an assemblage on the assumption that each individual species follows a similar stochastic process to that shown in figure 1, but equilibrium range sizes vary...
from one species to another. The model is finally tested using several empirical datasets of the geographical range sizes of species in taxonomic assemblages.

2. MODEL

(a) SDE of range size

Let \( x(t) \) be the abundance of a population at time \( t \). The dynamics of \( x(t) \) are traditionally modelled with an ordinary differential equation, such as the logistic growth model,

\[
dx(t) = rx(t)\left(1 - \frac{x(t)}{k}\right)dt,
\]

where \( r \) is the specific growth rate and \( k \) is the environmental carrying capacity supporting the population at \( x(\infty) \), i.e. the population approaches \( k \) as \( t \to \infty \).

In reality, populations seldom grow in a deterministic manner, but rather in a stochastic one. An approach to modelling stochastic growth is to view the specific growth rate as being subject to stochastic environmental fluctuations, which is equivalent to adding a white noise term to equation (2.1) (e.g. May 1974),

\[
dx(t) = rx(t)\left(1 - \frac{x(t)}{k}\right)dt + \sigma x(t)dw(t),
\]

where \( dw(t) \) is Gaussian white noise with \( N(0, dt) \). Here, at steady state (i.e. as \( t \to \infty \)) the population fluctuates along the deterministic equilibrium \( k \). This gives rise to a gamma probability distribution for the population (Dennis & Costantino 1988).

The relationship between the abundance \( x(t) \) and the range size \( y(t) \) of a species can be expressed in terms of a power model (Gaston 1994; Leitner & Rosenzweig 1997; Harte et al. 2001),

\[
y(t) = ax(t)^b,
\]

where \( a \) and \( b \) are parameters. Both abundance \( x(t) \) and range size \( y(t) \) are random variables and dependent on time in equation (2.3). The SDE for \( y(t) \) can therefore be derived according to the following transformation (transforming \( x(t) \) to \( y(t) \)) for the Ito stochastic differential (Karlin & Taylor 1981, p. 347)

\[
dy(t) = \frac{dy(t)}{dx(t)}\left[1 - \frac{x(t)}{k}\right] + \frac{dy(t)}{dt} + \frac{1}{2} \frac{\partial^2 y(t)}{\partial x(t)^2} \sigma^2 x(t)^2 dt + \frac{dy(t)}{\partial x(t)} \alpha x(t) dw(t).
\]  

Substituting the appropriate derivatives of equation (2.3) and further replacing \( dx(t)/dt \) by equation (2.2), we arrive at an SDE for range size,

\[
dy(t) = y(t)\left[\frac{4\beta r + (1 - \beta)\sigma^2}{2\beta^2} - \frac{2\sigma}{\beta k} y(t)\right] dt + \frac{2\sigma}{\beta} y(t) dw(t),
\]

where, in terms of the parameters in equation (2.3), \( a = \left(\frac{1}{b}\right) \) and \( \beta = \frac{1}{b} \).

This SDE (equation (2.5)) describes the stochastic dynamics of the range size of a species, and is of the logistic type, with a trajectory as shown in figure 1.

(b) Steady-state probabilistic distribution of range size

If \( \sigma = 0 \), the SDE (equation (2.5)) becomes a deterministic differential equation. When time is sufficient, the range size \( y(t) \) of a species will reach a stable equilibrium (figure 1), \( y(\infty) = (k/\alpha)^{1/\beta} \). In the stochastic setting, \( \sigma > 0 \), the range size of the species fluctuates above and below this equilibrium. Therefore, at the steady state, the range size approaches an approximately limiting, stationary probabilistic distribution of the form

\[
f(y) = c \exp\left[\frac{2}{\sigma^2} \int_y^\infty \frac{g(z)}{y} dz - 2\log(y)\right], \quad y > 0,
\]

for an SDE \( dy = y\sigma dw(t) + \alpha y dt \) (see Dennis & Costantino 1988). \( c \) is a constant that makes equation (2.6) a probability density function (PDF), i.e. the integration of \( f(y) \) over the support of \( y \) equals 1.

Applying equation (2.6) to the SDE of equation (2.5), results in the stationary PDF for range size \( y \)

\[
f(y) = \frac{\sigma}{\alpha} \frac{\beta}{\Gamma(\delta)} \frac{y^{\delta-1}}{\exp\left(-\frac{\alpha}{\sigma^2} y^\beta\right)} \exp\left(-\frac{c}{k} y^\beta\right), \quad 0 < y < \infty.
\]

Solving for constant \( c \) by setting the integration of equation (2.7) over \( 0 < y < \infty \) to be 1, we obtain a stationary distribution for the range size \( y \) of a species that is a generalized gamma distribution

\[
f(y) = \frac{\beta}{\Gamma(\delta)} \frac{\sigma}{\alpha} y^{\delta-1} \exp\left(-\frac{c}{k} y^\beta\right), \quad y > 0,
\]

where \( \delta = \frac{r}{\sigma} - \frac{3}{4\beta} - \frac{1}{4} \) and \( c = \frac{\alpha r}{\sigma^2} \).

The PDF (equation (2.8)) describes the steady-state distribution of range size for a species that fluctuates above and below the equilibrium range (figure 1). \( \delta \) and \( \beta \) are the two parameters determining the shape of the distribution. The shape of \( f(y) \) shifts from monotonically
Figure 2. The probability densities for the inverse range sizes of species to illustrate the validity of the exponential assumption for equation (2.9). The smooth curves are the fittings of the Weibull distribution to the inverse of range sizes for four assemblages: (a) bumble-bees; (b) New World birds (overall); (c) primates; (d) woodpeckers. See § 4 for descriptions of the datasets.

decreasing to unimodal when \( \delta \) (or \( \beta \)) varies from small to large. On the other hand, \( c/k \) is a position parameter. A small \( k \) (the carrying capacity) (or a large \( c \)) leads the distribution to be concentrated in the small range classes, whereas a large \( k \) makes the distribution more evenly distributed from small to large range classes. If there were time-series of measurements of the range size of a species, the data would be expected to follow this distribution. Unfortunately, such data are not available.

There is variation between species in the growth rate of range size, reflecting differences in their population growth rates. This variation has been accounted for in the SDE (2.2) through the stochastic term, and thus in equation (2.5). If all species had the same equilibrium range size, equation (2.8) would be sufficient to describe the distribution of range sizes for a species assemblage. However, in reality, there is also variation between species in the equilibrium range size, resulting from differences in equilibrium population \( k \). Variation in equilibrium range size can be accounted for by considering the carrying capacity \( k \) in equation (2.8) to be a random variable. To simplify, here we assume the inverse of \( k \) to follow an exponential distribution:

\[
    f\left(z = \frac{1}{k}\right) = \lambda \exp(-\lambda z), \quad z > 0. \tag{2.9}
\]

This assumption is simple yet not unreasonable. This can be approximately corroborated because if we assume that the observed range size of each species had indeed reached the deterministic equilibrium, the inverse of these range sizes (i.e. \( 1/y \)) would be expected to follow the Weibull distribution. The Weibull distribution is derived through variable transformation according to equation (2.3) at \( x(t) = k \). We have found this to be the case for several datasets (including those employed below; see figure 2), probably reflecting the pattern of resource partitioning among the species in an assemblage (see § 5).

A compound distribution of equation (2.8) can be constructed by assuming the exponential distribution of equation (2.9) for \( 1/k \)

\[
    f(y) = \frac{\beta c^\delta}{\Gamma(\delta)} y^{\delta-1} \int_0^1 z^\delta \exp(-cyz) \lambda \exp(-\lambda z) dz. \tag{2.10}
\]

The integration in equation (2.10) leads to the distribution

\[
    f(y) = \beta \delta d \frac{y^{\delta-1}}{(\phi + y)^{\delta+1}}, \quad y > 0 \tag{2.11}
\]

where \( \phi = \lambda/c \). This is the PDF of a Pearson type VI distribution and it accounts for the stochastic effects on the growth rate of range size and for variation in the equilibrium range sizes of an assemblage of species. As for the distribution of typical range size data, equation (2.11) can
show a mode at very small range sizes and be strongly right-skewed, except in the fifth case (dashed line) which has a mode at $y = 5$. (b) The corresponding densities for the log-transformed data.

Table 1. The maximum-likelihood estimates of the parameters of equation (2.11) ($\pm$ s.e.) and the results of KS and $\chi^2$ goodness-of-fit tests to 10 range size datasets.

<table>
<thead>
<tr>
<th></th>
<th>$\beta$ (s.e.)</th>
<th>$\delta$ (s.e.)</th>
<th>$\phi$ (s.e.)</th>
<th>p-value (KS test)</th>
<th>p-value ($\chi^2$-test)</th>
</tr>
</thead>
<tbody>
<tr>
<td>bumble-bees</td>
<td>1.603 (0.0652)</td>
<td>1.137 (0.197)</td>
<td>23.169 (0.687)</td>
<td>0.25</td>
<td>0.38</td>
</tr>
<tr>
<td>New World birds (breed)</td>
<td>1.607 (0.0188)</td>
<td>1.120 (0.0291)</td>
<td>17.738 (0.407)</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>New World birds (overall)</td>
<td>1.464 (0.0185)</td>
<td>1.309 (0.0302)</td>
<td>11.700 (0.399)</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Procellariiforms</td>
<td>1.811 (0.0396)</td>
<td>0.493 (0.0344)</td>
<td>5000.001 (757.079)</td>
<td>0.48</td>
<td>0.12</td>
</tr>
<tr>
<td>mammals</td>
<td>0.903 (0.0369)</td>
<td>0.543 (0.269)</td>
<td>448.531 (2.026)</td>
<td>0.24</td>
<td>0.88</td>
</tr>
<tr>
<td>primates</td>
<td>1.780 (0.0351)</td>
<td>0.303 (0.338)</td>
<td>941 239.652 (86.334)</td>
<td>0.48</td>
<td>0.88</td>
</tr>
<tr>
<td>suckers</td>
<td>0.644 (0.112)</td>
<td>1.471 (1.013)</td>
<td>1137.978 (6.572)</td>
<td>0.43</td>
<td>0.20</td>
</tr>
<tr>
<td>sunfishes</td>
<td>0.740 (0.371)</td>
<td>2.975 (0.500)</td>
<td>5000.027 (928.477)</td>
<td>0.24</td>
<td>0.09</td>
</tr>
<tr>
<td>wildfowl</td>
<td>2.192 (0.0827)</td>
<td>0.441 (0.106)</td>
<td>1248.085 (13.044)</td>
<td>0.086</td>
<td>0.0038</td>
</tr>
<tr>
<td>woodpeckers</td>
<td>1.866 (0.0906)</td>
<td>0.904 (0.0820)</td>
<td>51.478 (6.332)</td>
<td>0.35</td>
<td>0.86</td>
</tr>
</tbody>
</table>

The goodness-of-fit of the model to the observed range data is tested using both the Kolmogorov–Smirnov (KS) test and the $\chi^2$-test. The null hypothesis is: does the sample arise from the hypothesized distribution (equation (2.11))? For continuous data, the KS test is generally more powerful than the $\chi^2$-test and is more likely to reject the null hypothesis.

4. EMPIRICAL EVALUATION OF THE STOCHASTIC MODEL

(a) Data

We have acquired geographical range size data for 10 assemblages of species covering a wide range of taxonomic groups (details of the varied methods employed for the calculation of range sizes are provided in the references cited): (i) bumble-bees—global overall range sizes of 241 species (P. H. Williams, unpublished data; Gaston 1996); (ii) New World birds—breeding range sizes for 3901 species (Blackburn & Gaston 1996); (iii) New World birds—overall range sizes for 3906 species (Blackburn & Gaston 1996); (iv) Procellariiform seabirds—global overall range sizes for 108 species (Chown et al. 1998; Gaston & Chown 1996).
For the majority of the datasets, the fit of equation (2.11) to the range size distributions was quite satisfactory. Although the results of the KS test and of the \( \chi^2 \) goodness-of-fit test were not necessarily in full agreement in every case, for only two datasets (both for New World birds) did both tests fail (table 1). The rejection of these two datasets by the KS test is in a large part due to the large number of species (more than 3900). With such a large sample, even a small difference between the observed cumulative distribution function (CDF) and the hypothetical CDF would result in the rejection of the null hypothesis, although the largest difference between the two CDFs for the datasets is ca. 0.07. Given the difficulty of measuring range sizes for so many species at such a large scale, it is almost inevitable that there are a few 'outlier' species. A third dataset (that for woodfowl) marginally passed the KS test but failed the \( \chi^2 \)-test, which is obviously caused by the unexpectedly high number of species in the third class of the range size distribution (see the PDF for woodfowl of figure 4). A similar explanation applies to the sunfish data, which passed the KS test but failed the \( \chi^2 \)-test.
marginally met the $\chi^2$-test owing to the excessive number of species in the sixth class of the range size distribution (table 1; figure 4). In these cases poor fits are for datasets that do not exhibit a simple frequency distribution, often because the numbers of species per class (and thus overall) are rather low. Nevertheless, regardless of the statistical tests, equation (2.11) captures the overall shape of the distributions for all of the datasets, with a close match in the position of the mode and in the direction of any skew (figure 4).

\section*{c) Derivation of the range size–abundance power model}

The derivation of equation (2.11) is based on the power model (equation (2.3)) for the relationship between range size and population size. The derivation can be reversed, so as to derive the power model from the probabilistic model (equation (2.11)). The parameters $a$ and $b$ of equation (2.3) can be computed from the three parameters ($\beta$, $\delta$ and $c$) of equation (2.11) as

$$a = \left(\frac{4b\delta + \beta + 3}{4\beta c}\right)^{\frac{1}{\gamma}} \quad \text{and} \quad b = \frac{1}{\beta}.$$  

Whilst $\beta$ and $\delta$ can be read directly from equation (2.11), $c$ can only be obtained if we know either $r$ and $\sigma$ of the stochastic logistic equation (2.2) (see notation for equation (2.8)) or $\lambda$ of equation (2.9) (see notation for equation (2.11)). Unfortunately, neither set of parameters is readily determined given a set of range size data. Furthermore, $\alpha$ is a factor that converts abundance ($x$) into a range size for a species. It has to depend on the measurement unit (e.g. m$^2$, km$^2$ or ha) used for the observed range sizes and the activity range of individuals (e.g. m$^2$ may be an appropriate measurement unit for plants, but may not be so for animals). Nevertheless, from equation (2.11) we can at least obtain the exponent of equation (2.3). In other words, we at least are able to predict the observed range sizes up to a proportion: $y \propto x^\alpha$. For example, for the wildfowl data, the maximum-likelihood estimates of the two parameters in equation (2.11) are: $\beta = 2.192$, $\delta = 0.441$ (see ‘wildfowl’ in table 1). From these values, equation (2.3) is obtained as

$$y = 1.015x^{-0.456} \propto x^{-0.456} = \alpha x^{-0.456}.$$  

Although we know the range sizes for the wildfowl in figure 4 are measured as the number of grid cells occupied, with each cell being ca. 611 000 km$^2$, it is still not clear how the conversion factor $\alpha$ should be determined. Nevertheless, if the stochastic model (equation (2.11)) is reasonable, it is expected that the range sizes predicted based on equation (4.1) should be linearly related to the observed range sizes used to parametrize equation (2.11); i.e. equation (4.1) predicts the observed range sizes up to $y \propto x^{0.456}$. This can be tested because independent estimates have been made of the global population size ($x$) of each of the wildfowl species (see Webb et al. (2001) and references therein). Figure 5 shows that the power model (equation (4.1)) predicts the observed range sizes for the wildfowl extraordinarily well, even though in terms of the results of both the KS test and the $\chi^2$-test, equation (2.11) does not provide a particularly good fit (table 1).

\section*{5. DISCUSSION}

The stochastic hypothesis of range size dynamics is generally upheld by the results reported in this study. In particular, equation (2.11) seems to capture reasonably well the pattern of variation in range sizes of species in a taxonomic assemblage.

The formulation of equation (2.11) makes several significant assumptions. First, it assumes a power model for the relationship between population size and range size. The power model has been widely used for describing the range size–abundance relationship (Gaston 1994; Leitner & Rosenzweig 1997; Harte et al. 2001), although it is only one of a number of models that have been employed in this context (He et al. 2002; Holt et al. 2002). The simplicity of the power model allows the derivation of an explicit stationary probabilistic density function for range size that would be difficult to achieve with more complex formulations. However, over the ranges of variation in abundances commonly observed, different range size–abundance models are frequently not strongly differentiated; the power model captures the form of real range size–abundance relationships well (Holt et al. 2002), and we would not anticipate the use of alternatives to change markedly the conclusions drawn herein. Where fits to empirical data are less good, the principal weakness of the power model is likely to lie in overestimating the range sizes of species that have large population sizes (which may often aggregate more strongly than the model indicates), and with possibly some underestimation at small population sizes (Holt et al. 2002). Nonetheless, for the one dataset for which such a test is possible, the wildfowl, the form of the power law range size–abundance relationship assumed from equation (2.11) makes for a reasonable prediction of observed range sizes, particularly given that global estimates of the population sizes of these species are inevitably only approximate (though doubtless better than for any other group of at least moderate species richness), and the range sizes are measured in quite a crude fashion (which is invariably the case at global scales).

Second, equation (2.11) is derived on the assumption that the range sizes of species are at equilibrium with the environment, in as much as whilst range sizes vary, they do so about some equilibrium level. This assumption may break down in the face of strong directional change in environmental conditions (e.g. climate change, habitat destruction (see Parmesan 1996; Burgman & Lindenmayer 1998; Parmesan et al. 1999; Channell & Lomolino 2000)), or in the face of phenotypic or genotypic changes that enable a species to utilize previously unexploited resources and thereby colonize fresh areas (see Lewontin & Birch (1966) for a rare possible empirical example of the latter). The former is the more significant concern, and in the extreme could undermine the model. However, ‘natural’ directional changes in environmental conditions (especially climate) may often result in the geographical ranges of species shifting in location rather than necessarily responding with systematic changes in extent, and there is evidence that species in some assemblages have retained quite stable range sizes over long periods (e.g. Jablonski 1987; Riddle 1996). This may make the
models usually assume a random variable. For simplicity, population dynamics = population dynamics while the latter considers species coloniz-

ation and extinction processes. Following an argument that has more usually been applied to species-abundance distributions (but also non-biological systems, such as the gross national products of different countries), the frequency distribution of carrying capacities can be explained in terms of the action of multiplicative factors and the central limit theorem. If the carrying capacity of each species is a consequence of multiple factors operating essentially independently, and the differences between these capacities are expressed as differences in exponential growth, then one might expect that a right-skewed distribution of species carrying capacities would tend to result, with those for most species tending to be rather small and those of only a small proportion tending to be large (MacArthur 1957; May 1975; Gotelli & Graves 1996; but see Pielou 1975). This makes a great deal of biological sense, in as much as most of the resource bases exploited, and environmental spaces occupied, by species tend to be limited, and only a few tend to be extensive. However, the details of the shape of carrying capacity distributions is extremely difficult to establish, likely to vary from one taxon to another, and any assumptions regarding its form are ultimately somewhat speculative.

The stochastic hypothesis of range size dynamics seems to provide a reasonable prediction of variation in geographical range sizes amongst the species in a taxonomic assemblage. Of course, this does not mean that the processes embodied in the model necessarily give rise to the patterns of variation that are observed. If they do not, the fit of model and data suggest that at worst the actual processes are quite well characterized by the stochastic hypothesis.

We are grateful to T. M. Blackburn, D.-G. Chen, M. P. Pagel, W. Reed, T. J. Webb, P. H. Williams and two anonymous referees for assistance with data, discussion and comments. K.J.G. is a Royal Society University Research Fellow, and F.H.
is supported by the Biodiversity, Climate Change and Ecosystem Processes Networks of the Canadian Forest Service.

REFERENCES


As this paper exceeds the maximum length normally permitted, the authors have agreed to contribute to production costs.