Title
opinion: Relationship between local population density and environmental suitability estimated from occurrence data

Permalink
https://escholarship.org/uc/item/9xg7v2mn

Journal
Frontiers of Biogeography, 3(2)

Author
Jiménez-Valverde, Alberto

Publication Date
2011

DOI
10.21425/F5FBG12408

Copyright Information

Relationship between local population density and environmental suitability estimated from occurrence data

Alberto Jiménez-Valverde

Departamento de Biología Animal, Facultad de Ciencias, Universidad de Málaga, 29071 Málaga, Spain.
e-mail: alberto.jimenez@uma.es, alberto.jimenez.valverde@gmail.com; http://www.biogeografia.org/ajimenez/albertojimenezvalverde.htm

The spatial configuration of species' abundance has been a topic of discussion in ecology for a century (see Gaston 2003). On several occasions, the fact that abundance peaks at the centre of the geographic ranges has been reported and has even been accepted as rule of thumb (Sagarin and Gaines 2002). Assuming that local abundance is a by-product of Hutchinsonian niche processes, and that some of the most determinant variables demonstrate spatial autocorrelation and some inter-independence, Brown (1984) argued that the high abundances found at the centre of the ranges of distribution are to be expected. However, the pattern is far from general (Sagarin and Gaines 2002; Gaston 2009); there are several contingent and demographic causes that may disrupt the "abundant centre" distribution hypothesis (Rapoport 1975; Brown 1984). Even when abundance truly reflects environmental suitability, given the potential complexity of the structure and geographic patterns of niches (Soberón and Nakamura 2009), violations of the single and centred environmental optimum may be the rule rather than the exception.

Species distribution models (SDMs) have become popular in recent years, in part because of the increased availability of occurrence data in huge biological databases. Within the usual terms of "ecological niche models" or "habitat suitability models" lies the assumption that, since occurrence data reflect some parts of the niches, these methods estimate something close to the environmental suitability (herein suitability for short) of the species (Soberón and Nakamura 2009). Given that occurrence data are much easier to compile than abundance, an obvious question that arises is whether local suitability can inform us about local abundance. Curiously, few studies have directly explored this topic, and the results of the investigations of Pearce and Ferrier (2001), Nielsen et al. (2005) and Jiménez-Valverde et al. (2009) are not encouraging at all.

The study of VanDerWal et al. (2009) went a step further towards understanding the relationship between suitability and abundance. These authors suggested that the relationship between
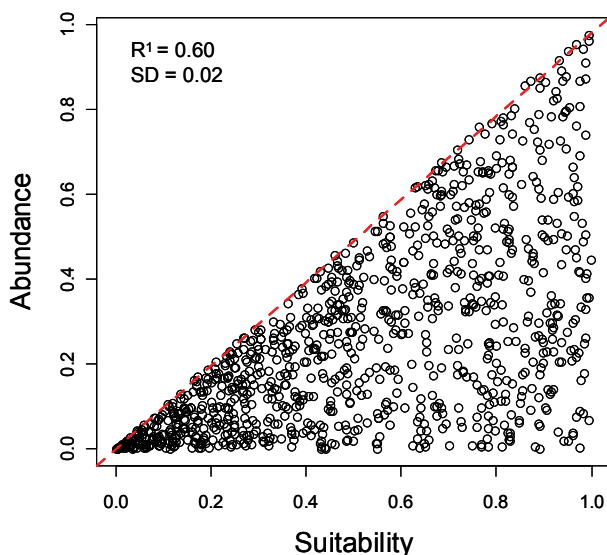

Figure 1. Idealized triangular relationship between suitability and abundance. Two vectors – one representing abundance (y) and another suitability (x) – were generated. In the case of x, 1000 values ranging between 0 and 1 were randomly extracted from a uniform distribution. In the case of y, 1000 values were also randomly extracted from a uniform distribution following the condition $0 \leq y_i \leq x_i$ (i = {1, …, 1000}). Linear quantile regressions were fitted to the 97.5th percentile (red dashed line). From 1000 simulations, the mean goodness-of-fit statistic for quantile regressions (R^1; Koenker and Machado 1999) was computed, with a value of 0.60. Analyses were done in R using the quantreg package (Koenker 2009).

suitability and abundance should be triangular (see Fig. 1). They reasoned that, since many factors may reduce the theoretical maximum density that a species can reach at a certain location, suitability should determine the maximum limit instead of the mean abundance. The authors modelled the distribution of 69 species of vertebrates in the Australian wet tropics region. Presence data for the species were compiled from a variety of sources and several climate and vegetation-related variables were used to fit the SDMs. The authors applied Maxent, a technique that has gained surprising popularity recently, and that uses presence and background (i.e., locations with no information about the occurrence of the focus species) data. Then they studied the relationship between suitability and abundance data derived from standardized sampling.

To test for the expected polygonal relationship, the authors argued that ordinary least square regressions were not suitable. In fact, suitability only accounted for 12% (mean) of the variation in vertebrates' mean abundance. Certainly, suitability was not a good surrogate of abundance. They applied quantile regressions (QR), which are appropriate for dealing with the unequal variation in ecological data associated with limiting variables (Cade and Noon 2003). VanDerWal and collaborators found evidence for a positive relationship between the upper limit of local abundance and suitability. Here, I want to stress four concerns.

First, the calculation of an R-squared measure (explained variance) makes little sense in a QR context (see FAQ() in Koenker 2009). While in a mean regression (ordinary least squares regression) the sum of squared residuals is minimized, in a median regression (QR fitted to the 50th percentile) what is minimized is the sum of absolute residuals. Thus, in QR, the interest is on the weighted sum of absolute residuals (R^1) as a local measure of the goodness-of-fit for a unique quantile (Koenker and Machado 1999). Second, the reported values of R^1 are not generally very large (mean value lower than 0.2), although it has to be taken into account that for an idealized perfect triangular relationship the maximum R^1 that a lin-
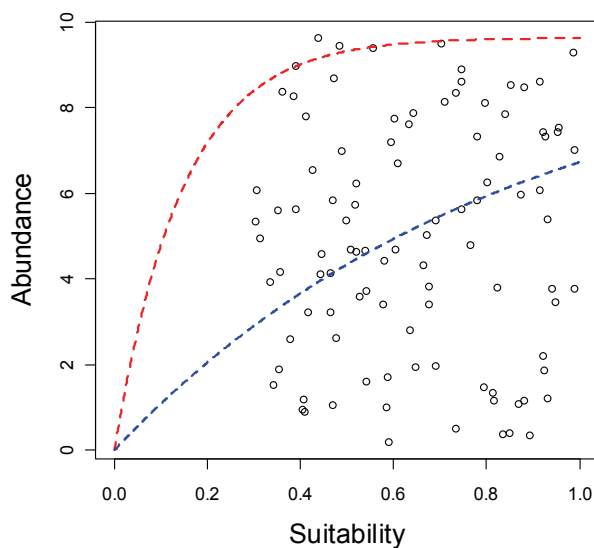

Figure 2. Naive example showing the potentially trivial results of fitting quantile regressions with a fixed (0, 0) intercept. Two uncorrelated vectors – one representing the abundance (*y*) and another the suitability (*x*) – were generated. In the case of *y*, 100 values ranging between 0 and 10 were randomly extracted from a uniform distribution. In the case of *x*, 100 values ranging between 0.3 and 1 were also randomly extracted from a uniform distribution. Nonlinear quantile regressions were fitted to the 95th (red dashed line) and 50th (blue dashed line) percentiles using the function *y* = [max(*y*)-exp(-*bx*)] (in this example, max(*y*) = 10). The parameter *b* was estimated as 6.84 (95th) and 1.20 (50th) and was highly significant in both cases (*p*<0.0001). Analyses were done in R using the `quantreg` package (Koenker 2009).

ear QR fitted to the 97.5th percentile can yield is 0.60 (not 1; see Fig. 1). Third, fitting a function with a fixed intercept at point (0, 0) (as done by the authors with nonlinear QR – one of several functions fitted) may yield statistically significant but at the same time spurious results (see Fig. 2). Finally, VanDerWal and collaborators applied a threshold to convert continuous maps into presence–absence maps, and excluded locations falling below that threshold (i.e. locations with low suitability values) from the abundance–suitability regression analysis. This allowed them to avoid the inclusion of a potentially large number of zero counts in locations with very low suitability values (in which case a positive abundance–suitability relationship may be indicating good discrimination capacity of the SDMs rather than capacity to account for abundance) while still taking into ac-

count absence data within the potential range of the species. However, it would be very informative to know the extent to which the thresholding process excluded locations with very low suitability values but with non-zero species' abundance.

In any case, the fact that linear QR showed higher significant slopes in the upper percentiles, and that R^1 progressively increased, hints at the idea of a polygonal relationship. Clearly, the study of VanDerWal et al. has brought to light a new perspective in the way to approach the analysis of the "abundance–suitability" pattern. Given the difficulty of obtaining abundance data and its potential relevance in conservation practices (Brown et al. 1995), many more studies in this line of research are required to discern under which circumstances, if any, the pattern holds.

Finally, an issue that worries me about the general idea of the "abundance–suitability" pattern is that (assuming well designed sampling, yielding unbiased occurrence data), if the density of points across the grid cells is a reflection of abundance, then a positive relationship between suitability and abundance may be a trivial fact. (Note that, as is usually done in SDM studies, VanDerWal and collaborators aggregated multiple observations at the same location, which certainly helped avoid tautological relationships, but this does not invalidate my argument: I am referring to the density of points *across* the grid cells). My reasoning is that it seems justified to assume that the abundance affects the probability of detecting the species. Therefore, if the probability of detection determines the geographic density of presence points across the grid cells, is the positive relationship between local abundance and suitability that is estimated from well calibrated SDMs a revelatory pattern or just the consequence of circular reasoning?

Acknowledgments

References

Brown, J.H. (1984) On the relationship between abundance and distribution of species. American Naturalist, 124, 255–279.

Brown, J.H., Mehlman, D.W. & Stevens, G.C. (1995) Spatial variation in abundance. Ecology, 76, 2028–2043.

Cade, B.S. & Noon, B.R. (2003) A gentle introduction to quantile regression for ecologists. Frontiers in Ecology and the Environment, 1, 412–420.

Gaston, K.J. (2003) The structure and dynamics of geographic ranges. Oxford: Oxford University Press.

Gaston, K.J. (2009) Geographic range limits: achieving synthesis. Proceedings of the Royal Society of London Series B, 276, 1395–1406.

Jiménez-Valverde, A., Diniz, F., Azevedo, E.B. & Borges, P.A.V. (2009) Species distribution models do not account for abundance: the case of arthropods on Terceira Island. Annales Zoologici Fennici, 46, 451–464.

Koenker, R. (2009) quantreg: Quantile Regression. R package version 4.27.

Koenker, R. & Machado, J.A.F. (1999) Goodness of fit and related inference processes for quantile regression. Journal of the American Statistical Association, 94, 1296–1310.

Nielsen, S.E., Johnson, C.J., Heard, D.C. & Boyce, M.S. (2005) Can models of presence-absence be used to scale abundance? Two case studies considering extremes in life history. Ecography, 28, 197–208.

Pearce, J. & Ferrier, S. (2001) The practical value of modelling relative abundance of species for regional conservation planning: a case study. Biological Conservation, 98, 33–43.

Rapoport, E.H. (1975) Areografía. Estrategias geográficas de las especies. México D. F.: Fondo de Cultura Económica.

Sagarin, R.D. & Gaines, S.D. (2002) The "abundant centre" distribution: to what extent is it a biogeographical rule? Ecology Letters, 5, 137–147.

Soberón, J. & Nakamura, M. (2009) Niches and distributional areas: Concepts, methods, and assumptions. Proceedings of the National Academy of Sciences USA, 106, 19644–19650.

VanDerWal, J., Shoo, L.P., Johnson, C.N. & Williams, S.E. (2009) Abundance and the environmental niche: Environmental suitability estimated from niche models predicts the upper limit of local abundance. American Naturalist, 174, 282–291.

Edited by Núria Roura-Pascual