Received: 4 October 2016 | Revised: 21 April 2017 | Accepted: 18 May 2017
DOI: 10.1111/geb.12611

CORRESPONDENCE

WILEY Global Ecology and Biogeography A Journal of Macroecology

Using *n*-dimensional hypervolumes for species distribution modelling: A response to Qiao et al. (2016)

Benjamin Blonder[1] | Christine Lamanna[2] | Cyrille Violle[3] | Brian J. Enquist[4,5]

[1]Environmental Change Institute, University of Oxford, Oxford, United Kingdom
[2]World Agroforestry Centre, United Nations Avenue, Nairobi, Kenya
[3]CNRS, CEFE UMR 5175, Université de Montpellier – Université Paul Valéry – EPHE, Montpellier Cedex, France
[4]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona
[5]The Santa Fe Institute, Santa Fe, New Mexico

Correspondence
Benjamin Blonder, Environmental Change Institute, University of Oxford, South Parks Road, Oxford OX 1 3QY, United Kingdom.
Email: bblonder@gmail.com

Editor: Pedro Peres-Neto

Funding information
U.K. Natural Environment Research Council, Grant/Award Number: NE/M019160/1; CGIAR research programme on Climate Change, Agriculture, and Food Security (CCAFS); European Research Council (ERC), Grant/Award Number: ERC-StG-2014-639706-CONSTRAINTS

Abstract

Hypervolume approaches are used to quantify functional diversity and quantify environmental niches for species distribution modelling. Recently, Qiao et al. (2016) criticized our geometrical kernel density estimation (KDE) method for measuring hypervolumes. They used a simulation analysis to argue that the method yields high error rates and makes biased estimates of fundamental niches. Here, we show that (a) KDE output depends in useful ways on dataset size and bias, (b) other species distribution modelling methods make equally stringent but different assumptions about dataset bias, (c) simulation results presented by Qiao et al. (2016) were incorrect, with revised analyses showing performance comparable to other methods, and (d) hypervolume methods are more general than KDE and have other benefits for niche modelling. As a result, our KDE method remains a promising tool for species distribution modelling.

1 | INTRODUCTION

We recently proposed a geometrical *n*-dimensional hypervolume method that uses kernel density estimation (KDE) to delineate niche boundaries (Blonder, Lamanna, Violle, & Enquist, 2014). The method has been used to explore functional diversity (Díaz et al., 2016; Lamanna et al., 2014) and community and ecosystem dynamics (Barros, Thuiller, Georges, Boulangeat, & Münkemüller, 2016; Carboni et al., 2016; Loranger et al., 2016) and can be used for species distribution modelling (SDM). Recently, Qiao, Escobar, Saupe, Ji, and Soberón (2016) cautioned against using KDE for SDM applications, because KDE causes high error rates by overfitting sparse or biased datasets. Here we raise four response points.

2 | KDE OUTPUT SHOULD DEPEND ON DATA PROPERTIES, REFLECTING UNCERTAINTY IN THE DATA SAMPLE

The KDE method was criticized because its output depends on the number of observations and the dimensionality of the input data. This is correct, but it is, a useful property. The KDE approach assumes that observed data are a random sample from a true distribution. Given that data are samples from a true distribution, the KDE pads around each sample with a kernel function, whose width is determined by a bandwidth parameter. The shape of the object is determined by thresholding the density function at a certain volume quantile (Blonder, 2016; Blonder et al., 2014). As such, the method predicts the occurrence of a species at niche points close to sampled points, and predicts the absence of a species at niche points further from sampled points. Larger bandwidths or lower thresholds lead to more padding around the data, whereas smaller bandwidths or higher thresholds lead to less padding. Varying the bandwidth and the threshold allows a trade-off between false-positive and false-negative errors. Multiple algorithms for choosing the bandwidth or threshold can guide decisions for satisfying different optimality criteria (Blonder, 2016; Liu, White, & Newell, 2013), or data can simply be resampled to control for variation in sample size.

3 | ALL SDM METHODS ADD ASSUMPTIONS TO CORRECT FOR DATASET SIZE AND BIAS

If an SDM is intended to describe a realized niche of a given taxon, then a method that best fits the observed data is best. If the choice of an SDM is to describe a fundamental niche, then a method that both fits the observed data and predicts other unobserved data is best. If the observed data are an unbiased random sample, then these two problems are equivalent to each other. However observed data can be biased samples, for instance because of climate space availability (Jackson & Overpeck, 2000), species interactions or dispersal limitation (Guisan & Thuiller, 2005), or insufficient sampling effort (Araujo & Guisan, 2006; Merow, Wilson, & Jetz, 2017). Making unbiased predictions from biased calibration data is a general problem for all correlative SDM methods (Araujo & Guisan, 2006). KDE is appropriate when a model of a realized niche is desired or when the data are an unbiased random sample of a fundamental niche.

Complex shapes may arise for both fundamental and realized niches. Although we agree that fundamental niches can have simple shapes describable with simple SDM methods, several studies have delineated approximate fundamental niches for various taxa that show complex non-convex shapes [e.g., for *Daphnia* (Hooper et al., 2008), corals (Hoogenboom & Connolly, 2009) and endotherms (Porter & Kearney, 2009)]. Facilitation also may expand the niche in complex ways by permitting growth in conditions that would otherwise be non-viable (Bulleri, Bruno, Silliman, & Stachowicz, 2016; Guisan & Thuiller, 2005; Stachowicz, 2012). As such, KDE should also be useful for modelling complex shapes for fundamental niches and realized niches.

It is challenging to use biased observed data to make unbiased niche estimates. Given that the nature of the sampling bias may be unknown, the investigator must make additional assumptions about the form of the unbiased distribution. For convex hull SDMs, the assumption is that the observed data provide lower and upper bounds on possible niche values. For generalized linear models, the assumption is that responses along individual niche axes are mostly independent from responses along other axes (i.e., only linear and low-order interaction terms). For KDE, the assumption is that unobserved data are likely to fall close to observed data. It is not surprising that each SDM method works best when its assumptions are valid. As such, Qiao et al., (2016) showed that if a biased sample of data is obtained from a true convex-shaped distribution, then a convex hull method is best for reconstructing it, or that if the true distribution is box shaped, range box methods are best. Testing statistical models on constructed data will have limited generality, because models must ultimately be applied to real data where true statistical properties are inherently unknowable.

4 | KDE HAS REASONABLE STATISTICAL PERFORMANCE

The KDE method was not used correctly by Qiao et al., (2016). For identical data clusters, the KDE method should yield equal padding regardless of the coordinate position of each cluster (Figure 1a,b). This was not seen in their tests because their model prediction was not based on data with the same units and scale as for model building. This error occurred only for their KDE analysis and not for other methods [lines 51/55 of their file functions.r, and lines 47/50 of their file Figure. R (Qiao, Escobar, Saupe, Ji, & Soberón, 2017)]. Their simulations generated data on a unit interval; they then log-transformed data (giving values less than zero) when constructing the hypervolume and then delineated hypervolume boundaries over only the untransformed unit interval. This caused clipping of hypervolume boundaries when prediction occurred only over the untransformed unit interval, ignored data with log-transformed values less than zero and made smaller data clusters appear larger (Figure 1c). This can be replicated with code in Supporting Information Data S1.

Thus, their reported performance metrics were incorrect. Revised results presented by Qiao et al., (2017) show more reasonable performance for KDE. The similarity and volume estimated by KDE are comparable to other methods, and in many cases, KDE sensitivity is higher than for other methods. However, we agree that KDE specificity can be lower than for other methods because of sampling assumptions we described above. KDE specificity would probably also be higher if they had used different bandwidth estimators. The default Silverman estimator they used is appropriate only for normally distributed data and will yield overly broad padding for the multiple clusters or holes in their tests.

However, lower performance on small datasets is expected. We already proposed that KDE approaches be used when $\log_e m > n$, where m is the number of observations and n the number of dimensions (Blonder, 2016). Estimating the shape of any hypervolume with limited data is not recommended because the data required to constrain different shapes grow geometrically with dimensionality. Constraining shapes with fewer data requires additional assumptions, such as convexity.

5 | THE HYPERVOLUME METHOD CAN BE USED WITH OTHER SDM METHODS BESIDES KDE

The hypervolume approach is more generally a geometrical method for describing the shape of any object in an n-dimensional space that can be used for other SDM methods besides KDE. Any correlative SDM

Global Ecology and Biogeography A Journal of Macroecology WILEY

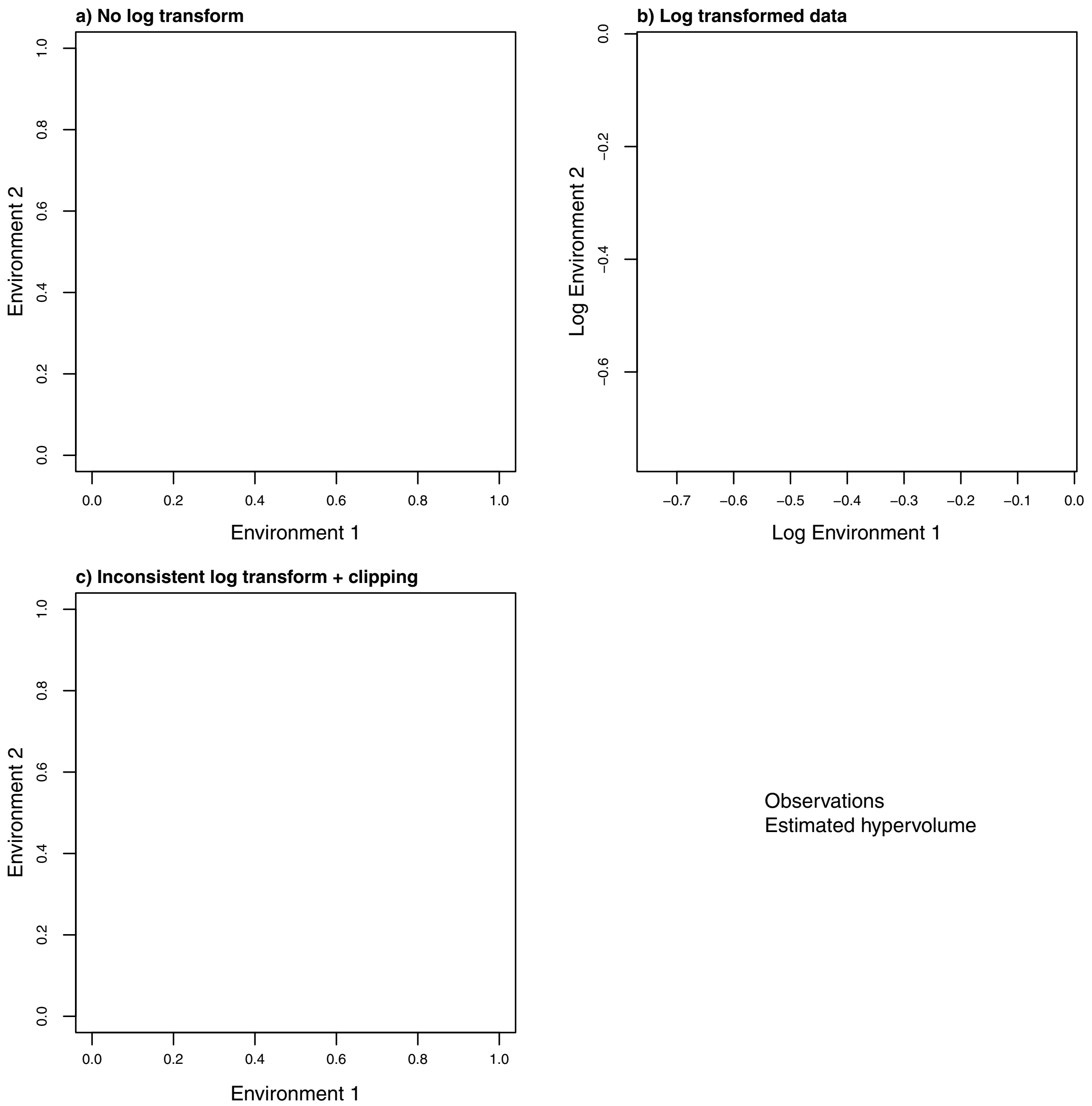

FIGURE 1 Demonstration of problems with inconsistent data transformation in the simulation analysis by Qiao et al., (2016). All panels may be compared with their Figure 1d. A dataset comprising a bimodal realized niche with 1,000 observations (dark purple points) is used to infer a hypervolume (stochastic geometry description comprising light orange points). (a) If the model is built and used for prediction on untransformed data, consistent with their convex hull, minimal volume ellipsoid and range box calculations, the kernel density estimation (KDE) method produces similar padding around each data cluster. (b) If the model is built and used for prediction on transformed data, one data cluster becomes larger than the other, such that the constant padding for all points appears proportionately smaller for the larger cluster (note negative coordinates for axes). (c) If the model is built using transformed data but used for prediction on untransformed data, boundaries appear clipped and skewed, yielding mistakes that incorrectly decrease both sensitivity and specificity

without spatial constraints can be transformed to an *n*-dimensional hypervolume. By generating predictions throughout niche space (Blonder et al., in review), a hypervolume of probability densities can be visualized. Response functions that are often used to describe SDMs (e.g., Guisan & Zimmermann, 2000; Merow, Smith, & Silander, 2013) are one-dimensional slices through these hypervolumes and provide fewer insights into niche geometry.

Visualizing SDMs as hypervolumes provides insights into the behaviour of these methods. We illustrate this by building SDMs for the tree *Quercus alba* using the *sdm* R package (Naimi & Araújo, 2016), for generalized linear models, generalized additive models and boosted regression trees. We use three Worldclim climate axes (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005) coupled to presence/background occurrence data from the BIEN3 database (Enquist, Condit, Peet, Schildhauer, & Thiers, 2016). Code to replicate this analysis is available as Supporting Information Data S2. The methods yield different niche geometries (Figure 2; animated in Supporting Information Movie S1) and have clearly different biological interpretations and implications.

The hypervolume method also can directly calculate the volume quantiles of the hypervolume, locate its position in niche space, determine

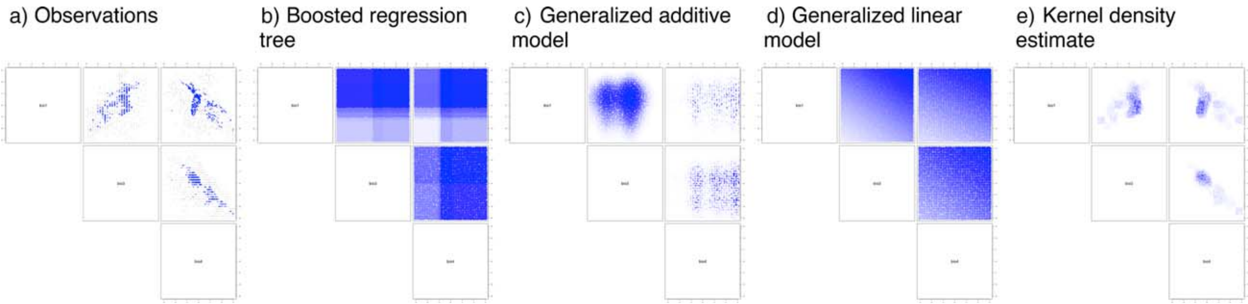

FIGURE 2 Comparison of species distribution modelling methods visualized as *n*-dimensional hypervolumes. All plots show pair plots for scaled and centred three-dimensional niche axes [Bioclim variables 1 (mean annual temperature), 3 (isothermality) and 4 (temperature seasonality)]. Bluer colours indicate higher probability of occurrence. (a) Raw occurrence data for the tree *Quercus alba*, transformed into climate space. Observations are shown as shaded blue dots; pseudo-absences sampled from the North American climate space are shown as grey dots. (b) A boosted regression tree niche model. (c) A generalized additive model. (d) A generalized linear model. (e) A kernel density estimate. Videos showing three-dimensional rotations of these panels are available as Supporting Information Movie S1

overlap with other niches and identify the size and shape of any holes. These tools are useful for assessing niche breadth and extinction risk (Boulangeat, Lavergne, Van Es, Garraud, & Thuiller, 2012), invasion outcomes (Broennimann et al., 2007), response to climate change (Jackson & Overpeck, 2000) or species interactions (Blonder, 2016). We are unaware of other approaches that solve these mathematical problems for arbitrary high-dimensionality objects.

6 | SUMMARY

All correlative SDM approaches are fundamentally limited by the data used to generate them. Biased inputs for observations lead to biased outputs for niche estimation. The KDE method provides an unbiased estimate of a multivariate distribution. The investigator must determine whether this assumption is appropriate. We think that KDE is a viable SDM approach when the observed data are thought to be an unbiased sample of the niche and when a minimal set of parametric assumptions are desired. This approach is appropriate for modelling realized niches and provides a flexible approach for modelling fundamental niches.

ACKNOWLEDGMENTS

B.B. is supported by a U.K. Natural Environment Research Council independent research fellowship (NE/M019160/1). C.L. is supported by the CGIAR research programme on Climate Change, Agriculture, and Food Security (CCAFS). C.V. is supported by the European Research Council (ERC) Starting Grant Project 'Ecophysiological and biophysical constraints on domestication in crop plants' (Grant ERC-StG-2014–639706-CONSTRAINTS).

REFERENCES

Araujo, M. B., & Guisan, A. (2006). Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, *33*, 1677–1688.

Barros, C., Thuiller, W., Georges, D., Boulangeat, I., & Münkemüller, T. (2016). *N*-dimensional hypervolumes to study stability of complex ecosystems. *Ecology Letters*, *19*, 729–742.

Blonder, B. (2016). Do hypervolumes have holes? *The American Naturalist*, *187*, E93–E105.

Blonder, B., Lamanna, C., Violle, C., & Enquist, B. J. (2014). The *n*-dimensional hypervolume. *Global Ecology and Biogeography*, *23*, 595–609.

Blonder, B., Morrow, C. B., Maitner, B., Lamanna, C., Violle, C., Enquist, B., & Kerkhoff, D. (in review) New approaches for delineating boundaries for *n*-dimensional hypervolumes. *Methods in Ecology and Evolution*.

Boulangeat, I., Lavergne, S., Van Es, J., Garraud, L., & Thuiller, W. (2012). Niche breadth, rarity and ecological characteristics within a regional flora spanning large environmental gradients. *Journal of Biogeography*, *39*, 204–214.

Broennimann, O., Treier, U. A., Müller-Schärer, H., Thuiller, W., Peterson, A., & Guisan, A. (2007). Evidence of climatic niche shift during biological invasion. *Ecology Letters*, *10*, 701–709.

Bulleri, F., Bruno, J. F., Silliman, B. R., & Stachowicz, J. J. (2016). Facilitation and the niche: Implications for coexistence, range shifts and ecosystem functioning. *Functional Ecology*, *30*, 70–78.

Carboni, M., Münkemüller, T., Lavergne, S., Choler, P., Borgy, B., Violle, C., . . . Thuiller, W. (2016). What it takes to invade grassland ecosystems: Traits, introduction history and filtering processes. *Ecology Letters*, *19*, 219–229.

Díaz, S., Kattge, J., Cornelissen, J. H. C., Wright, I. J., Lavorel, S., Dray, S., . . . Gorné, L. D. (2016). The global spectrum of plant form and function. *Nature*, *529*, 167–171.

Enquist, B. J., Condit, R., Peet, R. K., Schildhauer, M., & Thiers, B. M. (2016). Cyberinfrastructure for an integrated botanical information network to investigate the ecological impacts of global climate change on plant biodiversity. *PeerJ Preprints*, *4*, e2615v2. doi: 10.7287/peerj.preprints.2615v2

Guisan, A., & Thuiller, W. (2005). Predicting species distribution: Offering more than simple habitat models. *Ecology Letters*, *8*, 993–1009.

Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, *135*, 147–186.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*, 1965–1978.

Hoogenboom, M. O., & Connolly, S. R. (2009). Defining fundamental niche dimensions of corals: Synergistic effects of colony size, light, and flow. *Ecology*, *90*, 767–780.

Hooper, H. L., Connon, R., Callaghan, A., Fryer, G., Yarwood-Buchanan, S., Biggs, J., . . . Sibly, R. M. (2008). The ecological niche of *Daphnia magna* characterized using population growth rate. *Ecology*, *89*, 1015–1022.

Jackson, S. T., & Overpeck, J. T. (2000). Responses of plant populations and communities to environmental changes of the late Quaternary. *Paleobiology*, *26*, 194–220.

Lamanna, C., Blonder, B., Violle, C., Kraft, N. J. B., Sandel, B., Šímová, I., . . . Enquist, B. J. (2014). Functional trait space and the latitudinal diversity gradient. *Proceedings of the National Academy of Sciences USA*, *111*, 13745–13750.

Liu, C., White, M., & Newell, G. (2013). Selecting thresholds for the prediction of species occurrence with presence-only data. *Journal of Biogeography*, *40*, 778–789.

Loranger, J., Blonder, B., Garnier, É., Shipley, B., Vile, D., & Violle, C. (2016). Occupancy and overlap in trait space along a successional gradient in Mediterranean old fields. *American Journal of Botany*, *103*, 1050–1060.

Merow, C., Smith, M. J., & Silander, J. A. (2013). A practical guide to MaxEnt for modeling species' distributions: What it does, and why inputs and settings matter. *Ecography*, *36*, 1058–1069.

Merow, C., Wilson, A. M., & Jetz, W. (2017). Integrating occurrence data and expert maps for improved species range predictions. *Global Ecology and Biogeography*, *26*, 243–258.

Naimi, B., & Araújo, M. B. (2016) sdm: A reproducible and extensible R platform for species distribution modelling. *Ecography*, *39*, 368–375

Porter, W. P., & Kearney, M. (2009). Size, shape, and the thermal niche of endotherms. *Proceedings of the National Academy of Sciences USA*, *106*, 19666–19672.

Qiao, H., Escobar, L. E., Saupe, E. E., Ji, L., & Soberón, J. (2016) A cautionary note on the use of hypervolume kernel density estimators in ecological niche modelling. *Global Ecology and Biogeography*. doi: 10.1111/geb.12492

Qiao, H., Escobar, L. E., Saupe, E. E., Ji, L., & Soberón, J. (2017) Using the KDE method to model ecological niches: A response to Blonder et al., (response). *Global Ecology and Biogeography*. doi:10.1111/geb.12610

Stachowicz, J. (2012). Niche expansion by positive interactions: Realizing the fundamentals. A comment on Rodriguez-Cabal et al. *Ideas in Ecology and Evolution*, *5*, 42–43

BIOSKETCHES

Benjamin Blonder is a plant ecologist interested in science education and community ecology.

Christine Lamanna is a climate change ecologist applying niche modeling and functional diversity methods for adapting smallholder agriculture to climate change in Africa.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Blonder B, Lamanna C, Violle C, Enquist BJ. Using *n*-dimensional hypervolumes for species distribution modelling: A response to Qiao et al. (2016). *Global Ecol Biogeogr*. 2017;00:1–5. https://doi.org/10.1111/geb.12611