## Title

Using the KDE method to model ecological niches: a response to Blonder et al., (*response*)


## Authors

Huijie Qiao[1*], Luis E. Escobar[2*], Erin E. Saupe[3,4], Liqiang Ji[1] and Jorge Soberón[5]

## Affiliations

[1]*Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing, 100101 China.*

[2]*Department of Fisheries, Wildlife and Conservation Biology, University of Minnesota, St. Paul, Minnesota, 55108 USA;*

[3]*Department of Geology & Geophysics, Yale University, New Haven, Connecticut, 06511 USA.*

[4]*Department of Earth Sciences, S Parks Road, Oxford, OX1 3AN UK*

[5]*Biodiversity Institute and Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, Kansas, 66045, USA.*

*Corresponding author: Luis E. Escobar, Phone: (+1) 612-626-1831. Email: lescobar@umn.edu. Address: 135 Skok Hall, 2003 Upper Buford Circle St. Paul, MN 55108-6074.


**Running title:** ‘Holes’ in fundamental niches

**Dear Editor:**

Recently, we noted (Qiao et al. *in press*) that multivariate kernel density estimation (KDE) may not outperform other methods for estimating hypervolume geometries, and that under certain circumstances the algorithm would be unable to detect ‘holes’ in environmental space, as Blonder et al. (2014) had proposed. In our original note (Qiao et al. *in press*), we explained that KDE (i) is sensitive to both sample size and environmental dimensionality; (ii) may overestimate niche volumes in low dimensions and constrict niche volume estimates in high dimensions; and (iii) is useful only to the extent that the realized niche is sought and not the fundamental niche. Here, we additionally note that bandwidth is a crucial parameter for KDE, and its selection needs to be evaluated rigorously and all assumptions stated clearly.

In their response to our comments, Blonder et al. (*response*) indicated that (i) we made a data transformation mistake that may invalidate our conclusions regarding the KDE method; (ii) the stochastic geometry approach has additional conceptual benefits for niche modeling; (iii) concerns over error rates are consistent with established mathematical understanding of KDE; and (iv) the debate over whether niche modeling approaches should model realized or fundamental niches is general to all correlative methods, not just KDE. We address these points below, which we divide into two main categories: methodological and conceptual/theoretical concerns.

***Methodological concerns***

Blonder et al. (*response*) criticized our transformation of units during the modeling process. We had log-transformed the data when constructing KDE hypervolumes following the log-transformed data framework in the ’hypervolume’ R package demo code (see https://cran.r-project.org/web/packages/hypervolume/index.html), developed by Blonder et al. (2014). We agree, however, that this transformation is not necessary if hypervolumes are subsequently delineated and plotted in untransformed space. Consequently, we re-ran our analyses using untransformed units and two bandwidth configurations: the bandwidth estimated using the 'estimate_bandwidth()' function in the 'hypervolume' package from Blonder et al. (2014), and the value obtained from this function when divided by two, as used in Blonder et al. (*response*).

Our results remained consistent using the default bandwidth from estimate_bandwidth(). That is to say, our previous conclusions (Qiao et al. *in press*) hold regardless of whether we use transformed or untransformed data: KDE overestimated niche volumes in low environmental dimensions, underestimated niche volumes in high dimensions, was unable to detect holes in environmental space with low sample size (Figs. S2c and S3c), and was plagued by decreased sensitivity (Fig. S4). The other evaluation metrics, including specificity, hyper-volumes, and the Jaccard similarity index (Figs. S5-S7), show patterns similar to those reported in our previous analyses (Qiao et al. *in press*).

In their new comparison of hypervolume methods, Blonder et al. (*response*) tuned the KDE bandwidth by halving the ‘default’ bandwidth to achieve good model fit and to identify ‘holes’ in environmental space (see lines 16-17 in Data S1 of Blonder et al. *response*). Our reanalysis using this smaller bandwidth also detected holes (Fig. S2.3c and S3.3c), but at the cost of increased type II error (Fig. S1.3a, S2.3a and S3.3a). This is an important intervention in the model parameterization that should not be overlooked: a pragmatic *a posteriori* parameterization was necessary for KDE to effectively reconstruct the niche of the virtual species. Of course, when dealing with data from real species with unknown niche shapes, bandwidth selection would be more complex. Blonder et al. (*response*) did not justify why they modified the bandwidth selection during their experiments, which may reduce the merit of their response. The bandwidth is a critical parameter for KDE, which we noted in our original manuscript, and bandwidth selection deserves further research. We recommend that researchers using KDE explain their assumptions during bandwidth selection, explore a series of bandwidths configurations, and present the results of these models for more informed conclusions.

***Theoretical concerns***

The remaining points made by Blonder and colleagues (*response*) are primarily conceptual in nature. Blonder et al. (2014 and *response*) argue that fundamental niches can have holes and complex shapes in higher dimensions. While we argue this is still far from certain, what is most relevant for discussions herein is that niches in high dimensions may be highly clustered in the central regions of environmental space (Drake, 2015), such that any holes in niche 'hyperspace' are, once again, difficult to identify and determine. Blonder et al. (*response*) quote several references in the literature to suggest fundamental niches may have complex forms. However, our own perusal of these references indicate that these estimates (obtained either from first-principle models or experimental data) have either simple convex shapes (Hoogenboom & Connolly, 2009; Fig. 4; Sterck et al., 2011, Fig. 1B; Porter & Kearney, 2009, Fig. 5), or the data presented include just a few dozen points, making the estimation of 'complex shapes' a doubtful exercise at best (Vetaas, 2002, Fig. 2). This point, however, may be moot. KDE can be a useful method to fit shapes in multivariate spaces: interpretation about the meaning of the shapes may be best left to the researchers.

We note that an overfitted KDE may be no more informative than using the original species occurrence records to identify occupied environments, and a complex model with high fit (e.g., narrow KDE bandwidth) would be redundant. Comparison of KDE with other, more physiologically-realistic methods (e.g., range bagging; Drake, 2015), is warranted. We agree that KDE is a promising method and should be included in the toolbox of ecological niche modelers. KDE has both pros and cons that may not be present in other algorithms, making it complementary and not opposed to other methods. Indeed, our original cautionary note (Qiao et al. *in press*) was inspired by our interest in preventing the adoption of single algorithms as "silver bullets" for characterizing fundamental and/or realized niches for any given species. The algorithm of choice depends on the nature of the research question, as Blonder et al. (*response*) also note, and on the nature of the research data.

## References

Blonder B, Lamanna C, Violle C, et al (2016) Using n-dimensional hypervolumes for species distribution modeling: A response to Qiao et al. (2016). Glob Ecol Biogeogr. In press.

Drake JM (2015) Range bagging: A new method for ecological niche modelling from presence-only data. J R Soc Interface 12:20150086.

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.

Porter, W.P. & Kearney, M. (2009) Size, shape, and the thermal niche of endotherms. Proc Natl Acad Sci USA, 106, 19666-19672.

Qiao H, Escobar LE, Saupe EE, et al (*in press*) A cautionary note on the use of hypervolume kernel density estimators in ecological niche modelling. Glob Ecol Biogeogr. doi: 10.1111/geb.12492

Sterck, F., Markesteijn, L., Schieving, F. & Poorter, L. (2011) Functional traits determine trade offs and niches in a tropical forest community. Proc Natl Acad Sci USA, 108, 20627-20632.

Vetaas, O.R. (2002) Realized and potential climate niches: a comparison of four Rhododendrontree species. J Biogeogr, 29, 545-554.