LETTER TO THE EDITOR

A problem with variable selection in a comparison of correlative and process-based species distribution models: Comments on Higgins et al., 2020

Abstract
Comments are presented on an article published in October 2020 in Ecology and Evolution (“Predictive ability of a process-based versus a correlative species distribution model”) by Higgins et al. This analyzed natural distributions of Australian eucalypt and acacia species and assessed the adventive range of selected species outside Australia. Unfortunately, inappropriate variables were used with the MaxEnt species distribution model outside Australia, so that large climatically suitable areas in the Northern Hemisphere were not identified. Examples from a previous analysis and from the use of the freely available spatial portal of the Atlas of Living Australia are provided to illustrate how the problem can be overcome. The comparison of methods described in the Higgins et al. paper is worthwhile, and it is hoped that the authors will be able to repeat their analyses using appropriate variables with the correlative model.

The analysis of suitable environmental conditions for Australian tree species, both in Australia and overseas, has been a recurrent theme since the first species distribution model (SDM) package became available in 1984 (see sections 2.2 and 3.0 in Booth & Muir, 2020). The Higgins et al., (2020) paper analyzed natural distribution data for 664 Australian eucalypt and acacia species. Natural distribution data for 46 selected species were also analyzed to assess whether a process-based or a correlative SDM most effectively predicted the adventive range of these species outside Australia. It was concluded that “the correlative model—MaxEnt—has a superior ability to describe the data in the training data domain (Australia) and that the process-based model—TTR-SDM—has a superior ability to predict the distribution of the study species outside of Australia.” These conclusions may be true, but readers cannot be sure as there is a major problem with the variables selected for use when applying the correlative model outside Australia. This causes the MaxEnt model (Phillips et al., 2006) to fail to identify large environmentally suitable areas in the Northern Hemisphere.

In their methods section, Higgins et al., (2020) described that their process-based modeling used monthly data for minimum, mean and maximum monthly temperature, soil moisture contents, and solar radiation. A soil nitrogen content measure was also used that was held constant across all months. They stated that “This exact same list of environmental variables was used for the MaxEnt model fitting. This ensures that both models use the same information for estimating the species distribution models.”

The problem with using monthly variables in a correlative model for the whole world is that the same months are not comparable across hemispheres. To avoid this problem, climatic classifications, such as the widely used Köppen and Köppen–Geiger systems (Köppen, 1936; Peel et al., 2007), were developed using measures related to periods such as seasons (i.e., summer or winter) and the coldest month. Early SDM studies built on these ideas but provided sets of variables rather than a hierarchical classification system. The first SDM package was called BIOCLIM (Booth et al., 2014; Nix, 1986) and used a set of 12 variables, which was extended to 19 variables in 1996. These variables are listed in the ANUCLIM publication (for current version, see Xu & Hutchinson, 2011), which describes several programs including BIOCLIM. The set of 19 BIOCLIM variables was adopted for use by the WorldClim system (Fick & Hijmans, 2017; Hijmans et al., 2005), which with over 20 000 citations in Google Scholar™ is the most widely used source of climatic data for SDM studies.

Higgins et al., (2020) provided an example of their comparative analyses in their Figure 5. The top map in this figure shows “global projected environmental suitability” based on a MaxEnt analysis of the natural distribution of Acacia saligna in Western Australia. Only very limited areas of climatic suitability are shown in Figure 5 in the Northern Hemisphere. These are mainly in the Middle East, particularly in southwestern Iran and northern Iraq. Their figure shows that there are no occurrence records for A. saligna available from the Global Biodiversity Information Facility (GBIF) in these areas.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2021 CSIRO. Ecology and Evolution published by John Wiley & Sons Ltd.
Some of the main areas in which *A. saligna* is found naturally in Australia are in Csa and Csb regions of the Köppen–Geiger climate classification, that is, hot-warm summer/Mediterranean (winter rainfall) climates. Not surprisingly, there are extensive Csa and Csb regions in the actual Mediterranean region, as well as in California as shown by the maps of Peel et al., (2007). In contrast, the Higgins et al., (2020) MaxEnt analysis indicates almost no climatically suitable areas in the Mediterranean region and none in California, though blue crosses indicate numerous occurrences of *A. saligna* in the Mediterranean (see Figure 1) and some in California as recorded by the GBIF.

The difference between the Higgins et al., (2020) MaxEnt analysis for the Mediterranean region and what might be expected can be seen even more clearly when it is compared with a MaxEnt analysis provided by Thompson et al., (2011) of the Mediterranean region (see Figure 2). This was also based on an analysis of the natural distribution of *A. saligna* in Western Australia. However, Thompson et al., (2011) used variables developed for the BIOCLIM model. These were the following: temperature seasonality (Bio4), mean temperature of the hottest quarter (Bio10), mean temperature of the coldest quarter (Bio11), precipitation seasonality (Bio15), precipitation of the hottest quarter (Bio18), and precipitation of the coldest quarter (Bio19).

Using the spatial portal of the Atlas of Living Australia (ALA, https://www.ala.org.au/, Belbin, 2011), a MaxEnt analysis for *A. saligna* was carried out for the whole world using the six BIOCLIM variables employed by Thompson et al., (2011) (see Figure 3). In contrast to Figure 5 of Higgins et al., it shows extensive areas of climatic suitability around the Mediterranean and limited climatically suitable areas in California.

MaxEnt parameter settings and region selection can have important impacts on analysis outputs (Merow et al., 2013), but there is no indication that these could account for the problem described here. All the analyses described here used default parameter settings. The analyses shown here in Figures 1 and 3 were for the whole world. The Thompson et al., (2011) analysis shown in Figure 2 was clipped to include only the Mediterranean region but produced similar results for the region to those shown in Figure 3.

In conclusion, the use of just monthly variables is not appropriate for global application of correlative models. It is recommended that BIOCLIM variables, such as those provided by the WorldClim database (Fick & Hijmans, 2017; Hijmans et al., 2005), should be used. If researchers wish to use other variables derived from monthly values, it is recommended that they compare their detailed SDM results with the broad regions provided by Köppen–Geiger maps (Peel et al., 2007). If there are gross differences, this suggests more variables may be needed or that the variables being used may not be appropriate.

Higgins et al., (2020) is potentially a very interesting paper, but the Köppen–Geiger classification, the Thompson et al., (2011) paper, and the ALA analysis described here clearly indicate that there is a fundamental problem with the MaxEnt analyses as illustrated by Figure 5 in the Higgins et al., (2020) paper. It is hoped that the authors will be able to repeat their analyses using BIOCLIM variables with the MaxEnt model.

**FIGURE 1** Environmental suitability of *A. saligna* for the Mediterranean region based on a MaxEnt analysis of the natural distribution by Higgins et al., (2020). Dark red indicates an environmental suitability of 0. Blue crosses indicate occurrences of *A. saligna* obtained from the Global Biodiversity Information Facility (GBIF).

**FIGURE 2** Environmental suitability of *A. saligna* for the Mediterranean region based on a MaxEnt analysis of the natural distribution by Thompson et al., (2011). Green-yellow areas are climatically suitable and reddish-brown areas are highly suitable. Note the extensive suitable areas compared with the almost total lack of suitable areas in Figure 1.
**ACKNOWLEDGEMENTS**

I am grateful to Prof. Steven Higgins for some preliminary email exchanges concerning his paper with colleagues. Figures 1 and 2 are reproduced here under the terms of the Wiley Open Access Licence. I thank the Atlas of Living Australia team for the provision of their extremely useful facility. Thanks to Jacqui England and Libby Pinkard for comments on an early draft of this manuscript, as well as the Associate Editor and anonymous reviewers for their helpful comments. Thanks to Chris Foote (Editor-in-Chief, Ecology and Evolution) for advice on submitting a comment piece.

**CONFLICT OF INTEREST**

None declared.

**AUTHOR CONTRIBUTION**

Trevor Booth: Conceptualization (lead); Writing-original draft (lead); Writing-review & editing (lead).

**DATA AVAILABILITY STATEMENT**

The data used to produce Figure 3 were downloaded from the open-access Atlas of Living Australia (https://doi.org/10.26197/ala.2d8c5a9b-39a3-43a6-82ac-a5fd43f36166) and are also available at: https://doi.org/10.5061/dryad.6m905qfztt

Trevor H. Booth

CSIRO Land and Water, Canberra, Australian Capital Territory, Australia

**REFERENCES**

Belbin, L. (2011). The Atlas of Living Australia’s spatial portal. In M. B. Jones, & C. Gries (Eds.), Proceedings of the Environmental Information Management Conference 2011 (EIM 2011, pp. 39–43), Santa Barbara, September 28–29: University of California. Retrieved from spatial.ala.org.au.

Booth, T. H., & Muir, P. R. (2020). Climate change impacts on Australia’s eucalypt and coral species: Comparing and sharing knowledge across disciplines. Wires Climate Change, 11, e657. https://doi.org/10.1002/wcc.657

Booth, T. H., Nix, H. A., Busby, J. R., & Hutchinson, M. F. (2014). BIOCLIM: The first species distribution modelling package, its early applications and relevance to most current Maxent studies. Diversity and Distributions, 20, 1–9. https://doi.org/10.1111/ddi.12144

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology, 37, 4302–4315. https://doi.org/10.1002/joc.5086

Higgins, S. I., Larcombe, M. J., Beeton, N. J., Conradi, T., & Nottebrock, H. (2020). Predictive ability of a process-based versus a correlative species distribution model. Ecology and Evolution, 10(20), 11034–11054. https://doi.org/10.1002/ece3.6712

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. https://doi.org/10.1002/joc.1276

Koppen, W. (1936). Das geographisca System der Klimate. In W. Koppen, & G. Geiger (Eds), Handbuch “der Klimatologie. 1. C. (pp. 1–44) ~ Gebr, Borntraeger.

Merow, C., Smith, M. J., & Silander, J. A. (2013). A practical guide to MaxEnt for modeling species’ distributions: What it does, and why

---

**FIGURE 3** MaxEnt analysis of *A. saligna* produced using the Atlas of Living Australia applying the six BIOCLIM variables used by Thompson et al., (2011). Purple dots indicate natural distribution data from Western Australia used to calibrate the model. The scale at the left of the map indicates climatic suitability with yellow/green areas being relatively suitable and blue unsuitable. Note the extensive climatically suitable areas in the Mediterranean region compared with output from the MaxEnt model shown in Figure 5 of Higgins et al., (2020)

This page contains a diagram showing a MaxEnt analysis of the distribution of *A. saligna* using the Atlas of Living Australia. The diagram includes purple dots indicating natural distribution data from Western Australia, a color scale indicating climatic suitability from yellow/green for relatively suitable to blue for unsuitable, and a comparison with output from the MaxEnt model in Figure 5 of Higgins et al., (2020).
inputs and settings matter. *Ecography*, 36, 1058-1069. https://doi.org/10.1111/j.1600-0587.2013.07872.x

Nix, H. A. (1986). A biogeographic analysis of Australian elapid snakes. In R. Longmore (Ed.), *Atlas of elapid snakes of Australia: Australian flora and fauna series 7* (pp. 4-15). Bureau of Flora and Fauna.

Peel, M. C., Finlayson, B. L., & McMahon, T. A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences*, 11, 1633-1644. https://doi.org/10.5194/hess-11-1633-2007

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190, 231-259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Thompson, G. D., Robertson, M. P., Webber, B. L., Richardson, D. M., Le Roux, J. J., & Wilson, J. R. U. (2011). Predicting the subspecific identity of invasive species using distribution models: *Acacia saligna* as an example. *Diversity and Distributions*, 17, 1001-1101. https://doi.org/10.1111/j.1472-4642.2011.00820.x

How to cite this article: Booth TH. A problem with variable selection in a comparison of correlative and process-based species distribution models: Comments on Higgins et al., 2020. *Ecol Evol*. 2021;11:13609-13612. https://doi.org/10.1002/ece3.7496