

# A Stepwise Assessment of Parsimony and Entropy in Species Distribution Modelling <sup>†</sup>

Raimundo Real <sup>1</sup> and Alba Estrada <sup>2</sup>

<sup>1</sup> Biogeography, Diversity and Conservation Research Team, Department of Animal Biology, Universidad de Málaga, Málaga, Spain

<sup>2</sup> IREC (CSIC-UCLM-JCCM), Ronda de Toledo s/n, Ciudad Real, Spain

<sup>†</sup> Presented at the Entropy 2021: The Scientific Tool of the 21st Century, 5–7 May 2021; Available online: <https://sciforum.net/conference/Entropy2021/>.

Published: 5 May 2021

Entropy is an intrinsic characteristic of the geographical distribution of a biological species. A species distribution with higher entropy involves more uncertainty, i.e., is more gradually constrained by the environment. Species distribution modelling tries to yield models with low uncertainty, but normally has to produce them by increasing their complexity, which is detrimental for another desirable property of the models, parsimony. By modelling the distribution of 18 vertebrate species in mainland Spain, we show that entropy may be computed along the forward-backward stepwise selection of variables in Generalized Linear Models to check whether uncertainty is reduced at each step. This allows selecting the model that best combines the complementary characteristics of certainty and parsimony. This also allows to disentangle the entropy due to the intrinsic uncertainty of the species distribution from that due to failure in the model specification. A reduction of entropy was produced asymptotically in each step of the model, with some exceptions. This asymptote could be used to distinguish the entropy attributable to the species distribution from that attributable to model misspecification. We discussed the differential suitability of Shannon and fuzzy entropy for this end. The use of Shannon entropy in distribution modelling has not biogeographical sense, because it computes probability of presence as if the species were only present in one cell of the study area. Fuzzy entropy has not such restriction and always has values between zero and one, which produces results that are commensurable between species and study areas. Fuzzy entropy is also more correlated with AUC values. Using a stepwise approach and fuzzy entropy may be helpful to counterbalance the uncertainty and the complexity of the models. The model yielded at the step with the lowest entropy combines reduction of uncertainty with parsimony, which results in high efficiency.



© 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).