

Point process model solutions to presence-only data problems

Ian W. Renner^a and David I. Warton^b

^aSchool of Mathematical and Physical Sciences
University of Newcastle
Callaghan, NSW, Australia
Ian.Renner@unsw.edu.au

^bSchool of Mathematics and Statistics
University of New South Wales
Sydney, Australia
David.Warton@unsw.edu.au

Keywords: species distribution models

Abstract: Often, the best available data for species distribution modelling is so-called “presence-only” data, which consists of a list of observed species locations with no corresponding information about species absences. A number of methods have been developed to create species distribution models using presence-only data, in particular MAXENT and pseudo-absence logistic regression. Most of these methods face challenges in implementation, interpretation and checking model assumptions.

These challenges can be addressed by instead using Poisson point process models, which have been shown to be equivalent to pseudo-absence logistic regression (Warton and Shepherd 2010) and MAXENT (Renner and Warton 2013). In particular, I will show how the point process model framework offers to solutions to the questions of how to choose pseudo-absences, how to account for sampling bias (Warton, Renner and Ramp 2013), how to interpret model output, how to assess model assumptions, and how to choose the LASSO penalty to regularise model coefficients.

References

- Renner, I.W. and Warton, D.I. (2013) Equivalence of MAXENT and Poisson point process models for species distribution modeling in ecology. *Biometrics*, 69:274–281.
- Warton, D.I. and Shepherd, L.C. (2010) Poisson point process models solve the “pseudo-absence problem” for presence-only data in ecology. *Annals of Applied Statistics*, 4:1383–1402.
- Warton, D.I., Renner, I.W., and Ramp, D. (2013) Model-based control of observer bias for the analysis of presence-only data in ecology. *PloS one*, 8:e79168