What do we model when we model species distributions?: An ecological and statistical perspective

Geert Aarts[a], John Fieberg[b] and Jason Matthiopoulos[c]

[a] IMARES and Aquatic Ecology and Water Quality Management
Wageningen University and Research Center
't Horntje - Texel, The Netherlands
geert.aarts@wur.nl

[b] Department of Fisheries, Wildlife and Conservation Biology
University of Minnesota
St. Paul, MN, United States
jfieberg@umn.edu

[c] Institute of Biodiversity Animal Health and Comparative Medicine
University of Glasgow
Glasgow, United Kingdom
Jason.Matthiopoulos@glasgow.ac.uk

The need to understand processes shaping species distributions has resulted in a vast increase in the diversity of spatial wildlife data and statistical models. One may aggregate location data into spatial units (e.g. grid cells) and model the resulting counts or presence–absences as a function of environmental covariates. Alternatively, point data may be modelled directly, by combining the individual observations with a set of random or regular points reflecting habitat availability, a method known as a use-availability, presence – pseudo-absence or case–control design. Although species distribution models fitted to such diverse data are widely used, the ecological literature is not explicit about the statistical and ecological interpretation of their parameter estimates and predictions.

This study's objective is to illustrate that under certain assumptions, models fitted to spatial point, count and presence–absence methods can all be motivated by the same underlying spatial inhomogeneous Poisson point process (IPP) model, in which the intensity function is modelled as a log-linear function of covariates. This result also applies to models fitted using Maxent.

In geographical space, IPP model predictionsare proportional to the expected density of observations, or usage. In environmental space, IPP models are parameterized in terms of the ratio of habitat use over availability. Thus, positive coefficients are assumed to indicate preferential selection for particular environmental conditions. However,when an essential habitat type is already sufficiently abundant to meet an individual's needs, increasing

theavailability of this habitat type can lead to negative coefficients, suggesting an apparent avoidance. Such changes in the model coefficients as a function of habitat availabilitycan be captured bygeneralized functional response (GFR) models. Hence, GFRs explicitly estimate the influence of habitat availability on usage, andthus can improve spatial predictions in novel habitats, and may signpost habitats that are critical for the organism's fitness.