Should data be partitioned spatially before building large-scale distribution models?

Patrick E. Osborne *, Susana Suárez-Seoane

Department of Environmental Science, University of Stirling, Stirling FK9 4LA, UK

Abstract

There is growing interest in building predictive models of species distributions over large geographic areas. As larger areas are modelled, however, it is highly likely that heterogeneity in the predictors variable increases and that areas are included where animals respond to habitats in different ways, for example, due to social status. These effects (spatial non-stationary) may weaken model performance. This paper explores whether data partitioning prior to analysis can improve the fit of models and provide ecological insight into distribution patterns. Data on three bird species were modelled for the whole of Spain at 1 km² resolution using logistic regression analysis. Data were partitioned into geographic quarters, concentric rings around the centroid of the distribution, and into random samples for comparison. In all cases, data partitioning produced better models as assessed by Receiver Operating Characteristic curve (AUC) statistics than analysis of the global data set. Inclusion of latitude and longitude improved the global models only when added as smoothed splines but produced different probabilities to the partitioned data. Geographic partitioning is a very crude local modelling approach and we suggest that some form of geographically-weighted regression could offer the best solution to large-scale modelling but is computationally intensive on Geographical Information Systems (GIS) data. It is concluded that simple partitioning by geographic quarters may detect spatial non-stationary and alert the modeller to possible problems; that partitioning into more novel arrangements may be used to test ecological hypotheses; and that data should not be partitioned spatially to build and test models if non-stationary is suspected. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Distribution models; Spatial non-stationary; Logistic regression; Spatial heterogeneity; Birds; Spain

1. Introduction

There is growing interest in building predictive models of species distributions over large geographic areas, often using Geographical Information Systems (GIS) (see reviews by Corsi et al., 2000; Guisan and Zimmermann, 2000). In a pilot project modelling great bustard Otis tarda L. distribution in central Spain, Osborne et al. (2001) warn that two types of difficulties could be encountered in scaling-up to larger regions. First, there are problems related to the consistency of predictor variables over geographic space. This is especially evident in variables derived from Advanced Very High Resolution Radiometer (AVHRR) satellite imagery because of the inherent variability in surface reflectances at different