



## **Accounting for selection bias in species distribution models: An econometric approach on forested trees based on structural modeling**

Jean-Sauveur Ay (1), Joannès Guillemot (2), Nicolas K. Martin-StPaul (3), Luc Doyen (4), and Paul Leadley (2)  
(2) ESE Laboratory (UMR 8079). Université Paris Sud, CNRS, AgroParisTech. Orsay, France., (1) CESAER. INRA. Dijon, France., (3) URFM. INRA. Avignon, France., (4) GREThA. CNRS. Bordeaux, France.

Species distribution models (SDMs) are widely used to study and predict the outcome of global change on species. In human dominated ecosystems the presence of a given species is the result of both its ecological suitability and human footprint on nature such as land use choices. Land use choices may thus be responsible for a selection bias in the presence/absence data used in SDM calibration. We present a structural modelling approach (i.e. based on structural equation modelling) that accounts for this selection bias. The new structural species distribution model (SSDM) estimates simultaneously land use choices and species responses to bioclimatic variables. A land use equation based on an econometric model of landowner choices was joined to an equation of species response to bioclimatic variables. SSDM allows the residuals of both equations to be dependent, taking into account the possibility of shared omitted variables and measurement errors. We provide a general description of the statistical theory and a set of application on forested trees over France using databases of climate and forest inventory at different spatial resolution (from 2km to 8 km). We also compared the output of the SSDM with outputs of a classical SDM in term of bioclimatic response curves and potential distribution under current climate. According to the species and the spatial resolution of the calibration dataset, shapes of bioclimatic response curves the modelled species distribution maps differed markedly between the SSDM and classical SDMs. The magnitude and directions of these differences were dependent on the correlations between the errors from both equations and were highest for higher spatial resolutions. A first conclusion is that the use of classical SDMs can potentially lead to strong miss-estimation of the actual and future probability of presence modelled. Beyond this selection bias, the SSDM we propose represents a crucial step to account for economic constraints on tree species distribution that will help to assess the trade-offs and opportunities arising from global change and refine adaptive management strategies.