Integrating economic constraints into tree species distributions models

Jean-Sauveur AY   Nicolas MARTIN
Joannès GUILLEMOT   Luc DOYEN   Paul LEADLEY

INRA – UPS – CNRS

Sfécologie 2016
Outline

1 – INTRODUCTION

2 – THEORY

3 – DATA

4 – RESULTS

5 – CONCLUSIONS
Species Distribution Models (SDM)

Very used statistical tool to study natural species distribution
Species Distribution Models (SDM)

Very used statistical tool to study natural species distribution

Probability of presence as a function of bio-climatic variables

\[
\text{Prob}(m_p = 1 \mid X_i) = F(X_i)
\]
Species Distribution Models (SDM)

Very used statistical tool to study natural species distribution

Probability of presence as a function of bio-climatic variables

\[ \text{Prob}(m_p = 1 \mid X_i) = F(X_i) \]

Once \( F(\cdot) \) is estimated, one can predict the probabilities of species presence according to current or projected values of \( X_i \).
Economics of selection bias

SDM are typically estimated on contextual data (inventory).

Major tree species are only observable on forested land uses.
Economics of selection bias

SDM are typically estimated on contextual data (inventory).

Major tree species are only observable on forested land uses.

Not observing a tree species in an agricultural area does not mean that this area has unsuitable bio-climatic conditions.

⇒ Economic choices about land use produce a selection bias.
Contribution of the paper

We develop an econometric **Binary Selection Model** to control the hidden part of tree distributions due to land-use choices.
Contribution of the paper

We develop an econometric Binary Selection Model to control the hidden part of tree distributions due to land-use choices.

We found that classical SDMs can under- or over-estimate the probability of presence, it depends on the tree species.
Contribution of the paper

We develop an econometric Binary Selection Model to control the hidden part of tree distributions due to land-use choices.

We found that classical SDMs can under- or over-estimate the probability of presence, it dependends of the tree species.

We found that modeling land-use selection process is of increasing importance when working at fine spatial resolutions.
Outline

1 – INTRODUCTION

2 – THEORY

3 – DATA

4 – RESULTS

5 – CONCLUSIONS
Source of selection bias

The potential event of interest is unobservable because of the condition of having a Compatible Land Use (forests here):

\[
\text{Prob}(m_p = 1 \mid X_i) \neq \text{Prob}(m_p = 1 \mid X_i, CLU)
\]
Source of selection bias

The potential event of interest is unobservable because of the condition of having a Compatible Land Use (forests here):

\[ \text{Prob}(m_p = 1 \mid X_i) \neq \text{Prob}(m_p = 1 \mid X_i, CLU) \]

Table: What is observed instead of \( m_p \)

<table>
<thead>
<tr>
<th></th>
<th>forest</th>
<th>not forest</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_p = 1 )</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>( m_p = 0 )</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Bias from classical SDMs

The fundamental source of bias comes from the correlation between the errors of the economic and ecological equations:

- positive correlation: **Positive bias** (over-estimation)
- negative correlation: **Negative bias** (under-estimation)
- independent errors: **Without bias**
Presence/absence data

French *Inventaire Forestier National* (2014) at 2, 4 and 8 km resolutions. Regular grid sampling with all forests surveyed:

- For each $1 \times 1$ km site: not surveyed = not forest
Presence/absence data

French *Inventaire Forestier National* (2014) at 2, 4 and 8 km resolutions. Regular grid sampling with all forests surveyed:

- For each $1 \times 1$ km site: not surveyed = not forest

4 tree species: sessile oak \((Q. petrae)\), pubescens oak \((Q. pubescens)\), beech \((F. sylvatica)\) and fir \((A. alba)\)

R package SemiParBIVProbit: Semi-parametric Sample Selection Binary Response Modeling 2013 by Marra and Radice
Outline

1 – INTRODUCTION
2 – THEORY
3 – DATA
4 – RESULTS
5 – CONCLUSIONS
## Significant selection bias

<table>
<thead>
<tr>
<th></th>
<th>Q.petrae</th>
<th>Q.pubescens</th>
<th>F.sylvatica</th>
<th>A.alba</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2 KM</strong></td>
<td>0.536</td>
<td>0.557</td>
<td>-0.486</td>
<td>-0.551</td>
</tr>
<tr>
<td></td>
<td>[0.5, 0.55]</td>
<td>[0.51, 0.57]</td>
<td>[-0.53, -0.43]</td>
<td>[-0.58, -0.51]</td>
</tr>
<tr>
<td><strong>4 KM</strong></td>
<td>0.424</td>
<td>0.494</td>
<td>-0.355</td>
<td>-0.353</td>
</tr>
<tr>
<td></td>
<td>[0.3, 0.48]</td>
<td>[0.41, 0.52]</td>
<td>[-0.41, -0.29]</td>
<td>[-0.42, -0.26]</td>
</tr>
<tr>
<td><strong>8 KM</strong></td>
<td>-0.303</td>
<td>0.536</td>
<td>0.345</td>
<td>0.042</td>
</tr>
<tr>
<td></td>
<td>[-0.49, 0.07]</td>
<td>[-0.54, 0.54]</td>
<td>[0.18, 0.44]</td>
<td>[-0.12, 0.2]</td>
</tr>
</tbody>
</table>

Table: Correlations $\rho$ between errors and 95% CI
Sessile oak at 2 km

(positive correlation)
Sessile oak at 4 km

(positive correlation)
Sessile oak at 8 km

(null correlation)
Beech at 2 km

A. BSM at 2km
B. P−O at 2km
C. P−A at 2km

(negative correlation)
Beech at 4 km

(negative correlation)
Beech at 8 km

(positive correlation)
Synthesis

We known since Ricardo (1821) that best plots of land are first dedicated to crops, hence forests are a residual land use.

Our results are complementary as forests correspond to the best plots of species niche ($\rho > 0$) or the worst plots ($\rho < 0$).

Depending on the correlation, climate change projections from classical SDMs can be over-optimistic or over-pessimistic.