Accounting for land-use selection bias in tree species distribution models

Jean-Sauveur Ay, Joannès Guillemot, Nicolas Martin-StPaul, Luc Doyen, Paul Leadley

Poster: GCC53C0547

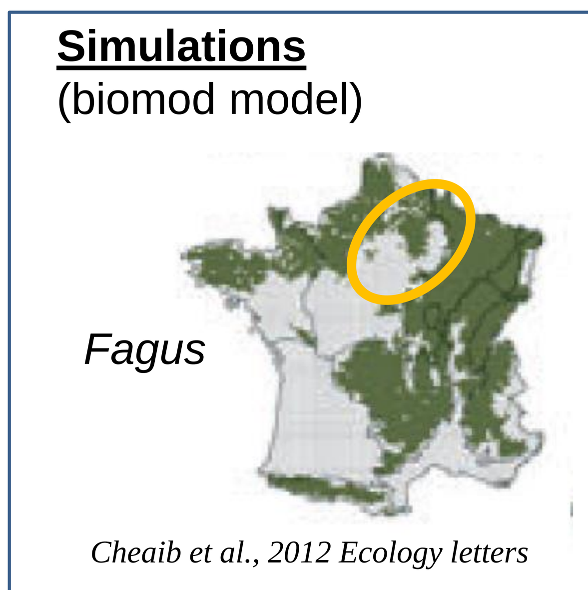
Accounting for land-use selection bias in tree species distribution models

Ay Jean-Sauveur; Guillemot Joannes; Martin-StPaul Nicolas K.; Doyen Luc & Leadley Paul

Mail of the corresponding authors:
nicolas.martin@paca.inra.fr
jean_sauveur@hotmail.fr
joannes.guillemot@u-psud.fr

Introduction

- Species distribution models (SDMs) are common tools to assess climate changes impacts on tree species **[1]**. They compute the probability of presence of a species given abiotic environmental conditions (T, ETP, …).
- Most classical SDMs do not account for land-use choices although it can affect calibration and projection : Absence data is recorded on suitable site if the land-use is incompatible with forest (e.g. crop, urban…).
- **We developed a bivariate distribution model (BDM) accounting for land use through economic rules and tested the hypothesis that classical SDMs are subjected to a selection bias as illustrated in Figure 1. Because the effect of land use is supposed to depend on the spatial scale [1], we performed empirical tests at 2 different spatial resolutions.**



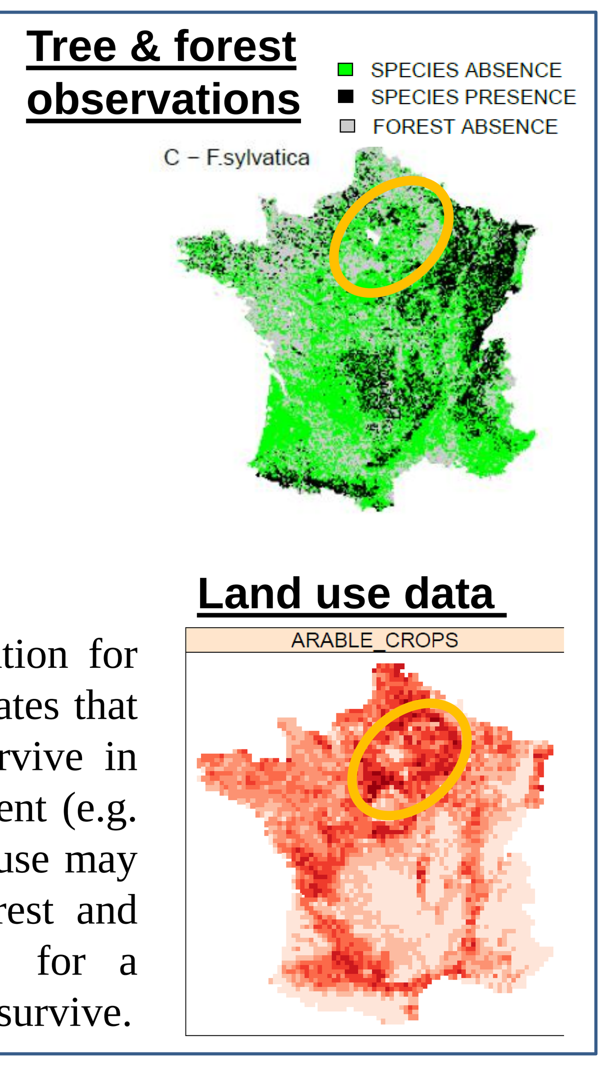


Figure 1: Classical SDM simulation for beech under current climate indicates that this tree species is unable to survive in many places where forest are absent (e.g. orange circle). In such areas landuse may have trigered the absence of forest and may therefore be responsible for a missestimation of beech ability to survive.

Theory, materials & methods

THEORY: Let's $\mu_i = f_p(X_i) - \varepsilon_i$ represents the abiotic requirements of a given species. The event $m_p = 1$ of « potentially observed the species with the environment X_i » occurs *if* $\mu_i > 0$. So, SDMs aim to model :

$$Prob(m_p = 1|Xi) = Prob(\mu_i > 0) = Prob(\varepsilon_i < f_p(X_i)) \quad (1)$$

In anthropized areas the event $m_p = 1$ is observable only for compatible land use ($m_l = 1$). We assume a random utility framework: landowners choose a landuse that provides the highest utility (U), so they choose forest if: $Uf > Ua \Leftrightarrow z \equiv Uf - Ua > 0$. In terms of probability:

$$Prob(m_l = 1|Xi, Wi) = Prob(z_i > 0) = Prob(\xi_i < f_l(X_i, W_i)) \quad (2)$$

Equations (1) and (2) can be combined to obtain a bivariate distribution model (**BDM**) that account for land use:

$$Prob(m_p = 1|Xi) = Prob(m_p = 1|Xi, m_l = 1) + Prob(f_p(X_i) > \varepsilon_i \cap f_l(X_i, W_i) \geq \xi_i) \quad (3)$$

From equation 3, the selection bias of classic SDMs appears to be a function of the joint distribution of errors (ε, ξ), as shown in Figure 2.

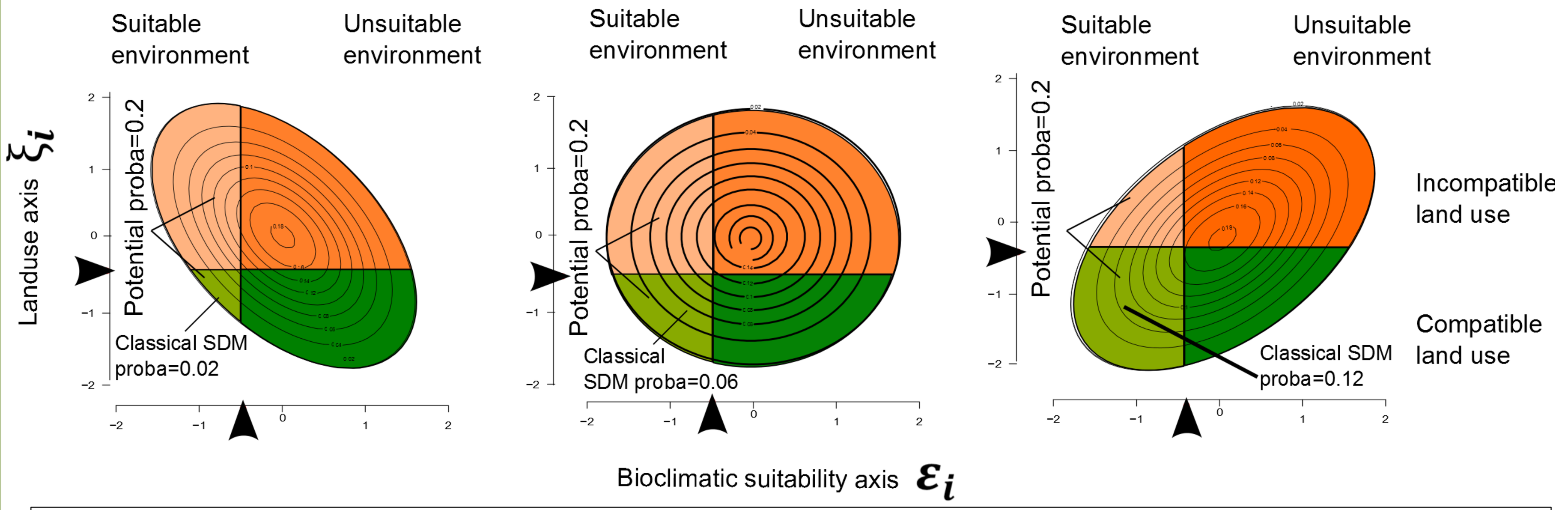


Figure 2: Illustration of the biais (i.e difference between the potential probability of presence and what is computed by classical SDMs) and its dependance to the joint distribution of errors (contour plot) for a given site. The site taken for this example is characterised by bioclimatic suitability ($f(X_i)$, see equation 1) and landuse compatibility ($f(X_i$, Wi), see equation 2) that are indicated by the spikes on both axis.

MATERIALS & METHODS: BDM and classical SDMs were implemented and compared at 2 different resolutions for 2 tree species over France. GLM and GAM were used (package semiParBivarProbit in *R*,**[3]**). **The french national forest inventories** (2005–2012) was used to get data of tree distribution (presence/absence) and landuse (forest *vs*. non forest). **The analysis SAFRAN** (daily surface climate for the 1960-2012 period) was used to compute bioclimatic indices (MAT, annual rainfall, ETP etc…). Downscaling at 2 and 4km was performed following **[4]**. Approximation of **monetary returns** for forest, crops and urban area were used to calibrate the econometric equation of landuse according to **[5]**.

Results & Discussion

We found significant correlations between errors (ρ<0 for *Fagus* and ρ >0 for *Quercus*) is strongly related to the intensity and the sign of the selection bias. Accordingly, classical SDM overestimated (for *Quercus*) or underestimated (for *Fagus*) the response to climate (Figure 3, Figure 4). This bias is highly visible at 2km spatial resolution and decreases at coarser resolution (Figure 3, Figure 4).

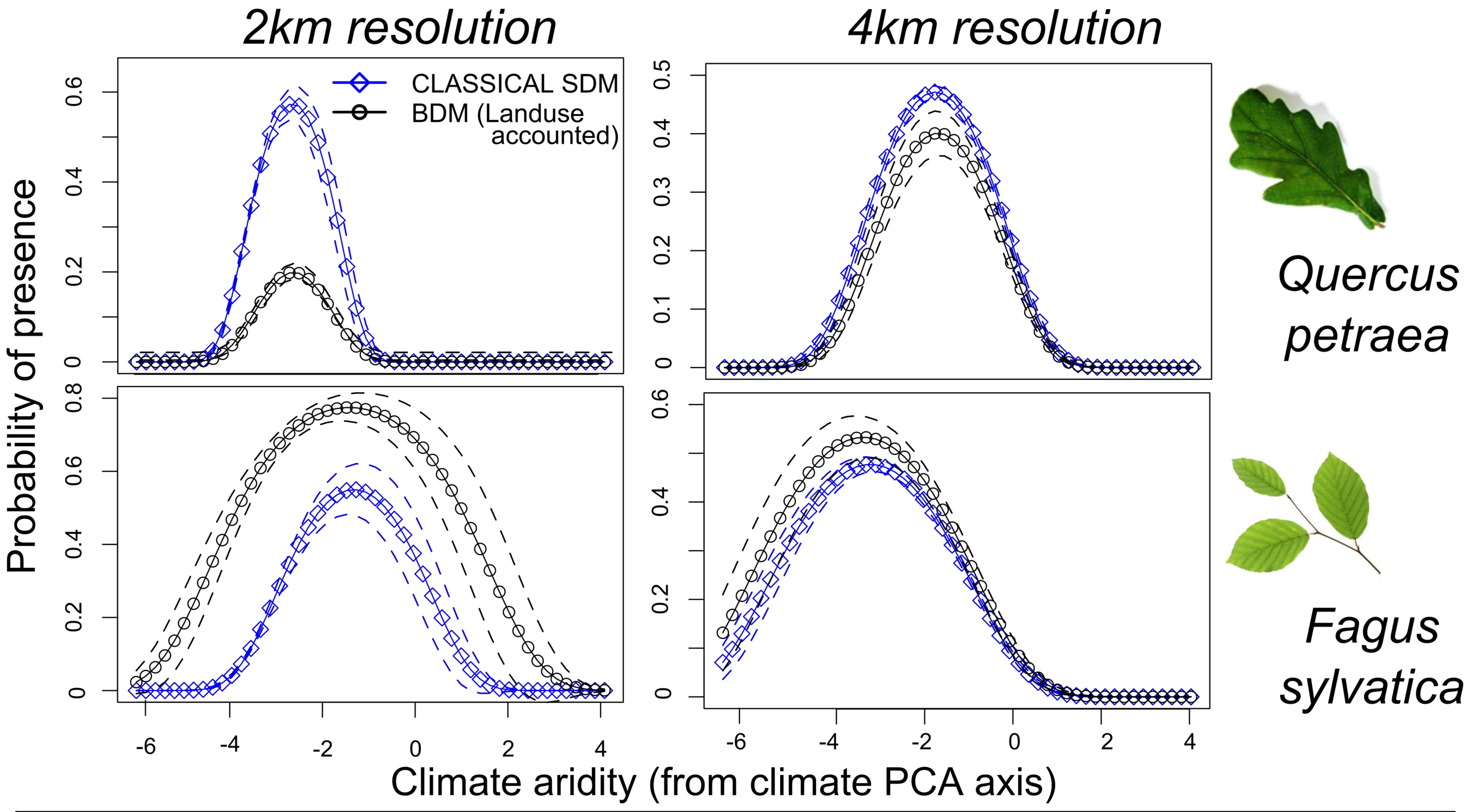


Figure 3: probability of presence as a function of temperature-aridity gradient for the SDM and the BDM, oak and beech at 2 spatial resolution (2 km and 4km).

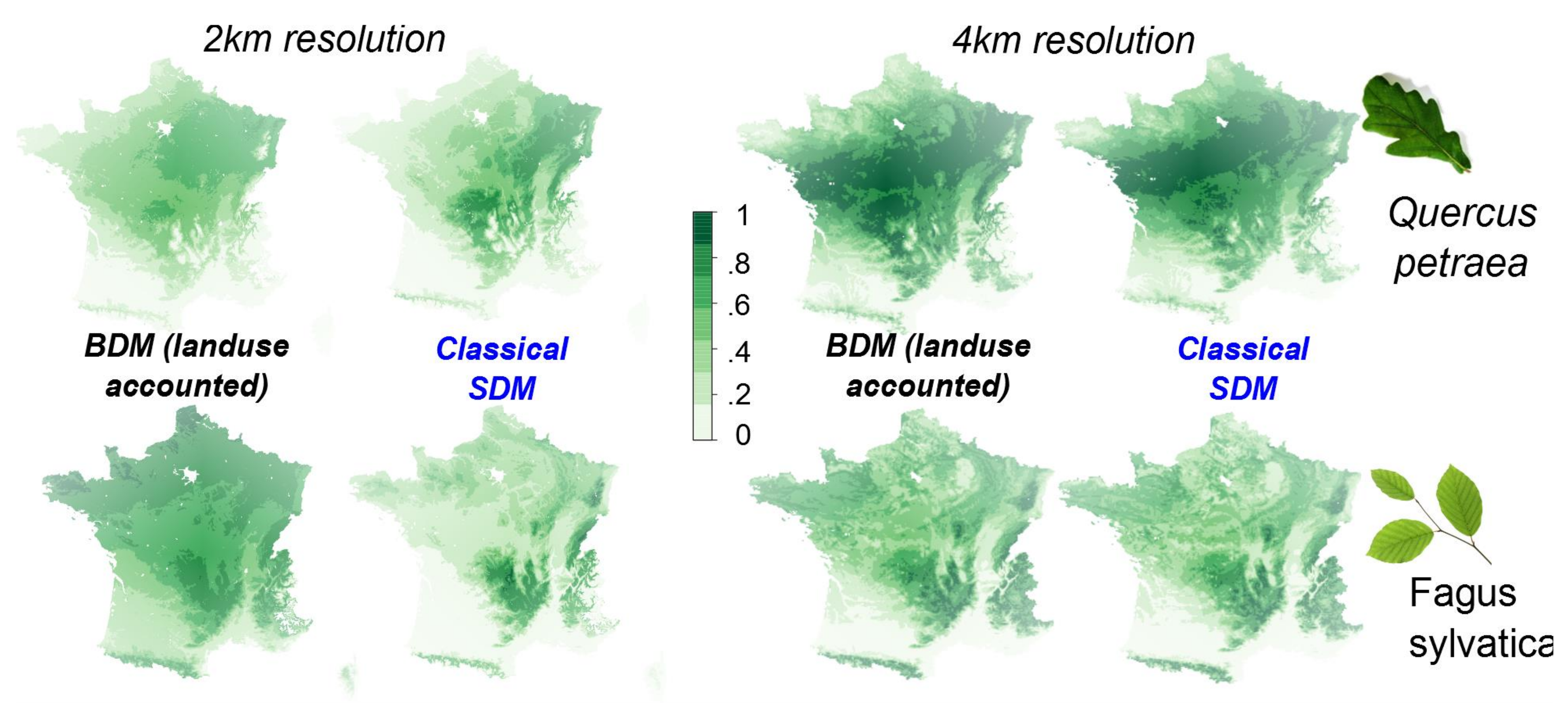


Figure 4: Maps of probability of presence simulated under current conditions for *Quercus* and *Fagus* at 2 spatial resolutions.

Conclusion

- Classical SDMs, that do not account for land use, are subjected to a selection bias that affect predictions of tree probability of presence.
- A bivariate model (BDM) helps accounting for this bias.
- The BDM offers opportunity to integrate economic decisions into projections of climate change impacts.

REFERENCES: **[1]** Cheaib A, et al. *Ecol Lett*. 2012;15(6):533-44; **[2]** Thuiller *et al*. *J Biogeogr*. 2004 31:353-361. **[3]** Marra & Radice SemiParBIVProbit: R package version 3.2-8, 2013. **[4]** *Ruffault et al*., *Theor Appl Climatol*. 2013. **[5]** Ay *et al*., *Clim Change*. 2014;126(1-2):13-30.

Acknowledgements: This work has been funded by the FRB (*Fondation de Recherche sur la Biodiversité*), GDF–SUEZ and the GIS (Groupement d'Intérêt Scientifique) "Climat, Environnement et Société" through the MOBILIS and the Humboldt projects.