Modelling species distribution: influences of temporal, spatial, and sampling heterogeneities in data-poor areas. An example from the Kerguelen Plateau

Guillaumot Charlène[1], Alexis Martin[2], Marc Eléaume[2] and Thomas Saucède[3]

[1] Laboratoire de Biologie Marine CP160/15, Université Libre de Bruxelles, F. D. Roosevelt, 50, 1050 Brussels, Belgium
E-mail: charleneguillaumot21@gmail.com
[2] Muséum national d'Histoire naturelle, Département Systématique et Évolution, UMR ISYEB 7205, 57 rue Cuvier, F-75231 Paris Cedex 05, France
[3] UMR 6282 Biogéosciences, Univ. Bourgogne Franche-Comté, CNRS, 6 bd Gabriel F-21000 Dijon, France

The field of species and community distribution modelling has undergone a remarkable increase of interest from ecologists during the last decades. Species Distribution Models (SDM) can constitute seminal conservation tools and usually have three main objectives: (i) estimating species realized ecological niches, (ii) interpolating occurrence point data over continuous maps based on correlative connections with environmental descriptors, and (iii) estimating potential shifts in species distribution associated to past and future environmental changes.

Interpolating species distribution over wide expanses of water like the Southern Ocean can constitute a real input to our understanding of marine life in remote and little-investigated areas. However, there are methodological issues caused by data heterogeneities because our knowledge of species distribution relies on diverse sampling campaigns and strategies. Spatial and temporal heterogeneities can impact the performance of modelling procedures, a critical point when considering the growing interest for the approach.

In the present work, we tested the reliability of different modelling procedures using true occurrence data of four common echinoid species of the Kerguelen Plateau with contrasted ecological niches. We used presence-only data compiled from various sampling campaigns led between 1872 and 2015. We tested the impact of spatial bias, record addition, and temporal variation on model performances, assessed the respective prediction performances and proposed a procedure to correct for the different bias. In addition, potential shifts in species distributions were modelled according to IPCC A2 scenario for 2100; they were compared to the magnitude of present changes (2005-2012).

This work provides a synthetic overview of the potential consequences of data heterogeneities for distribution modelling. It proposes a protocol to test and correct for the bias and gives some clues for improvement when using such procedures in conservation issues.