Modelling spatial distribution of small pelagic fishes' biomass using boosted regression trees hurdle models

Laura Julià[1], *Maria Grazia Pennino*[2], *Jose M. Bellido*[3], *Marta Coll*[4,5] *and Francisco Ramírez*[5]

[1]lauraj@icm.csic.es, Department of Renewable Marine Resources, Institute of Marine Sciences (ICM-CSIC), Barcelona, Spain

[2]Instituto Español de Oceanografía (IEO-CSIC), Centro Oceanográfico de Vigo, Spain

[3]Instituo Español de Oceanografía (IEO-CSIC), Centro Oceanográfico de Murcia, Spain

[4]Ecopath International Initiative Research Association, Barcelona, Spain

[5]Department of Renewable Marine Resources, Institute of Marine Sciences (ICM-CSIC), Barcelona, Spain

Small pelagic fishes (SPFs) in the northwestern Mediterranean are key elements of marine ecosystems: they represent a significant fraction of fish catches and are the main source of food for a large number of species. Evaluating the spatial distribution of the biomass available for these species allows to assess the food availability and enables to make conservation decisions. In this study, we implemented species distribution models (SDMs) for the most abundant SPFs, the European anchovy (*Engraulis encrasicholus*), the European sardine (*Sardina pilchardus*) and the sardinella (*Sardinella aurita*), in order to predict spatially-explicit biomasses. The study area includes the waters of the Iberian continental shelf within the geographical sub-area (GSA) 06. Species data comprises surveys carried out between June and July during the period 1998 to 2017 (both included) by the MEDITS (MEDIterranean Trawl Survey) oceanographic surveys. The SDMs were adjusted using Boosted Regression Trees (BRT) technique, defining sea surface temperature (ºC), sea surface salinity (PSU), net primary productivity (mg/m3), sea depth (meters) and year as predictor variables of the models. To deal with spatial autocorrelation, an autocovariable derived from the residuals of the model, was included to the environmental variables in the model (BRT-RAC method). Since the biomass had a long-tailed and zero-inflated distribution, we also used hurdle models. This approach deals with high numbers of zeros by using a two-step modelling process. In the first one, presence/absence data were modelled using a Binomial distribution in order to obtain a prediction of presence probability of the studied species; in the second, biomass data were modelled using a Gaussian distribution only in areas where species were expected to be present. Finally, we predicted the annual (mean of June and July) probability of occurrences as well as the back-transformed biomass for the years between 1998 and 2019 and then we combined both of them. As a result, our study determined the environmental influence on the biomass distribution of the three species, identified important marine areas for each species, and evaluated their species distribution trends over last decades.

Keywords: Species distribution modelling, BRT-RAC hurdle models, small pelagic fishes