Bayesian additive regression trees (BART) applied to global scale species distribution models (SDMs): comparing present and future projections.

Alba Fuster-Alonso[1], *M. Grazia Pennino*[2], *Xavier Barber*[3], *J. Maria Bellido*[2], *David Conesa*[4], *Antonio López-Quilez*[4], *Jeroen Steenbeek*[5], *J. Carlos Baez-Barrionuevo*[2], *Villy Christensen*[5,6] *and Marta Coll*[1,5]

[1]afuster@icm.csic.es, Instituto de Ciencias del Mar (ICM-CSIC),

[2]Instituto Español de Oceanografía (IEO-CSIC),

[3]Centro de Investigación Operativa, Universidad Miguel Hernández (UMH),

[4]Departamento de Estadística e Investigación Operativa (VaBar), Universidad de Valencia

[5]Ecopath International Initiative (EII),

[6]Institute of the Oceans and Fisheries, University of British Columbia

Marine Ecosystem Models (MEMs) have been developed to analyse the past and future dynamics of life in the oceans. One of such efforts is EcoOcean, a complex, mechanistic and spatio-temporal explicit MEM of the global oceans based on a trophodynamic core. EcoOcean requires as inputs the species native ranges and suitable habitats, and for key environmental conditions, species' functional responses and time-varying maps delivered by Earth System Models (ESMs). The different sources of uncertainty in these inputs may influence the validity and accuracy of EcoOcean results. For this reason, our study explores the use of global Species Distribution Models (SDMs) to reduce the uncertainty associated with these inputs.

A promising new alternative to traditional SDMs classification tree methods is the Bayesian Additive Regression Trees (BART). BART is a non-parametric Bayesian regression approach based on a sum-of-trees model. Then, in order to model the presences/pseudo-absences data to obtain EcoOcean inputs, the model applied in this work was as follows:

$$Y_i \sim Ber(\pi_i), \qquad i = 1, \dots, n\,,$$

$$\phi^{-1}(\pi_i) = \sum_{j}^{m} g_j\left(\boldsymbol{X}, T_j, M_j\right),$$

where, Y_i is our response variable (presence/pseudo-absence of species) in each observation *i* associated with a Bernoulli probability distribution; π_i is the probability of presence linked to the predictor by a link function ϕ^{-1}; then, g_j is the *j-th* tree of the form $g_j(\boldsymbol{X}; T_j, M_j)$, where m is the total number of trees, $\boldsymbol{X}$ is a vector of multiple covariates, T_j represents a binary tree structure consisting of a set of interior decision rules and a set of terminal nodes, and $M_j = \{\mu_{1j}, \dots, \mu_{jb}\}$ denote a set of parameter values.

Finally, to test BART's capability as an SDM on a global scale. We performed a suitability study for two globally distributed functional groups: 1) marine turtles and 2) tunas. Our results show that BART is a powerful approach to predict the potential distribution of target species, as well as their relationship with key environmental variables, on a global scale, and the outputs obtained are potentially useful for informing EcoOcean.

Keywords: BART, global scale, climate change.