**Supplementary Material 2**

Supplementary material 2 (SM2) presents a very brief description of the models found during the systematic review of ecological models applied to seamounts ecosystems. We present a short description of Model categories (i.e., broader aim of the model), followed by Types of Models (i.e., type of framework/software).

**Model category**

|  |  |
| --- | --- |
| Acronym | Description |
| BHM | Benthic Habitat Modelling  It is a specific type of Habitat Suitability Model, targeting bottom areas, generally including sediment surface and some sub-surface layers. |
| EBM | Ecosystem-Based Model  Aims to understand and represent the interactions and dynamics within an ecosystem; it captures complex relationships between various components of an ecosystem, including living organisms and their environment. |
| HSM | Habitat Suitability Model  Predict the suitability of a location for a species, or group of species, based on their observed relationship with environmental conditions; overall habitat quality. |
| IBM | Individual-Based Model  Are based on the explicit representation of individual organisms. They are developed for problematics where individual variability, local interactions, and adaptive behaviour are essential. |
| PM | Population Model  Are related with individual-level responses (vital rates in demographic terminology or life history traits in eco-evolutionary terms) to changes in population density and structure. |

|  |  |
| --- | --- |
| Acronym | Description |
| SDM | Species Distribution Model  Are spatial distributions of different taxa, considering the ecological requirements, ecological responses, and distribution areas; predicted likelihood of the occurrence of a species. |
| SDM/HSM | Species Distribution Model/Habitat Suitability Model  Combinative approach where SDMs provide useful information in terms of habitat suitability and help to identify areas where the predicted distribution overlaps with the most suitable habitats; very useful for prioritise conservation efforts in regions where the species is both likely to occur based on its known distribution and where the habitat quality is highest. |

**Type of Model**

|  |  |
| --- | --- |
| Model Type | Description |
| BCM | Bayesian cluster modelling  It is a statistical approach that combines Bayesian statistics and cluster analysis to model and analyse data with the goal of identifying groups or clusters within the dataset; it deals with complex datasets that exhibit patterns of similarity or grouping among observations, accounting for uncertainty, and incorporating prior knowledge. |
| Bio-optical | Are based on absorption spectra and/or pigment data, to differentiate phytoplankton size groups |
| BRT | boosted regression tree  It is a machine learning algorithm that combines two techniques: decision/regression trees (models that relate a response to their predictors by recursive binary splits) and boosting (an adaptive method for combining many simple models to give improved predictive performance); it deals with complex relationships and non-linear interactions in the data. |
| BTM | benthic terrain modeler  Aim to understand and classify the benthic environment, by using bathymetric data. |

|  |  |
| --- | --- |
| Model Type | Description |
| dbRDA | distance-based redundancy analysis  It is a multiple regression model (i.e., models the effect of an explanatory matrix X (n x p) on a response matrix Y (n x m)), that carries out constrained ordinations on data using non-Euclidean distance measures; allows to model the effect of an explanatory matrix on a response matrix, rather than a single response variable. |
| DISTLM | distance-based linear models (or distance-based regression models or distance-weighted regression models)  linear model where explanatory information is coded as distances between individuals, i.e., accounts for the distances between data points when estimating regression coefficients. |
| DM | distribution model  aim to predict SDM but can also be used to predict the distribution of habitats (e.g., fish habitats, habitat mapping based on habitat-forming species distribution, among others); it is a correlative approach that use discrete distribution data and full spatial coverage of environmental data to explain and predict patterns of distribution. |
| DSM | density surface models  It is a two-stage approach for estimating spatially varying density from distance sampling data; provide detailed estimates of species distribution and abundance across a landscape. |
| ENFA | environmental niche factor analysis  compare the distributions of the eco-geographical variables (EGV) between the presence data set (species distribution) and the whole area (global distribution); it summarises many EGV into a few uncorrelated factors retaining most of the information. |
| ENM | ecological/environmental niche model  aim to reconstruct the relationships between species and the environments where they occur and allow us to identify unexplored areas in geography where these species might be present. |

|  |  |
| --- | --- |
| Model Type | Description |
| EwE | Ecopath with Ecosim  It is a modelling approach that combines software for ecosystem trophic mass balance analysis (Ecopath), with a dynamic modelling capability (Ecosim); Ecosim models can be replicated over a spatial map grid (Ecospace), accounting for spatial dispersal/advection effects. |
| GAM | generalised additive model  It is a specific category of linear model that allows to model non-linear data (continuous or binary), interactions, and complex patterns, while maintaining coherence; uses smoothing functions, such as splines or kernel functions, that allow for flexible curves and surfaces. |
| GDM | generalised dissimilarity modelling  Explore and understand patterns of species turnover or dissimilarity across geographic or environmental gradients; models how ecological communities change as environmental conditions change; accepts any species list and a suite of environmental layers and then bends the environmental layers to best align with the species distributions. |
| GLM | generalised linear model  Are used to model relationships between predictor variables and response variables while accounting for the specific characteristics of the data and their underlying distribution; it allows the linear model to be related to the response variable via a link function and allowing the magnitude of the variance of each measurement to be a function of its predicted value. It is more flexible than linear regression as it works when the output variables are not continuous or unbounded; it allows changes in unconstrained inputs to affect the output variable on an appropriately constrained scale. |
| GLMM | generalised linear mixed model  Combines generalized linear models (GLMs) and linear mixed models (LMMs); are used to analyse data with non-normal response distributions, account for random effects, and model the relationships between multiple predictor variables and the response variable. |
| HMM | hidden Markov model  probabilistic frameworks where the observed data are modelled as a series of outputs (or emissions) generated by one of several (hidden) internal states; describe and analyse sequences of observations that are assumed to be generated by an underlying hidden process with unobservable states. |

|  |  |
| --- | --- |
| Model Type | Description |
| LTL | low trophic levels model  Usually it is a 3-component nutrient-phytoplankton-zooplankton (NPZ) model or 4-component NPZD (NPZ-Detritus) biological model; they can be more complex (e.g., European Regional Seas Ecosystem Model (ERSEM). |
| Maxent | maximum entropy  It is a machine learning algorithm used for SDM; aims to find the distribution that maximizes entropy (information uncertainty) while satisfying the constraints imposed by the available data; it is able to handle presence-only data; it uses niche modelling to predict the distribution of a species from the probability of finding it within raster squares, based on environmental variables and recorded locations. |
| PDE | Partial-differential equation models  Sets of equations describing the evolution of a physical quantity, not only with time, but also according to a structure variable such as space. |
| RF | random forest  It is an ensemble machine learning algorithm used for both classification and regression problems; consists in building multiple decision trees and combining their predictions to improve accuracy and reduce overfitting; can provide accurate results with minimal parameter tuning. |
| SEAPODYM | spatial ecosystem and population dynamics model  It is an Advection-Diffusion-Reaction (ADR) equation-based model that couples a physical-biological interaction model at basin scales, combining a forage (prey) production model with an age-structured population model of targeted (predator) species. |
| Schaefer model | It is a bioeconomic model mainly used within the fishing industry and designed for calculating the maximum sustainable yield. This model considers factors such as biological growth rates, carrying capacity, as well as total and marginal costs and revenues. |

|  |  |
| --- | --- |
| Model Type | Description |
| SSRW | Bayesian state-space random walk  Aims to infer the hidden states that generate the observed measurements; it estimates and predicts hidden states in a time series based on observed data; generally applied to sequential data where the underlying states evolve over time; incorporates uncertainty and prior knowledge into the modelling process. |
| Stepping stones | Aims to model dispersal and gene flow under the assumption of isolation by distance, where migration is non-random and constrained by distance. The primary occurrence of gene flow is between adjacent subpopulations, resulting in neighboring populations displaying a higher genetic similarity to each other. |
| Tweedie | Tweedie  It is a type of exponential dispersion models and are often used as distributions for GLMs or GLMMs to model and analyse data with the appropriate distributional characteristics;  model parameters typically involve statistical techniques like maximum likelihood estimation or Bayesian inference. |
| VGPM | vertically generalised productivity model  It is a model commonly used to estimate primary production within the ocean, as it can be applied to chlorophyll-a data from satellites and has a relatively simple design. |