

## General comments

This manuscript presents the development of an Atlantis model applied to the Bay of Biscay ecosystem, from its parameterization to a first calibration. It places the model in the context of spatial ecosystem-based management, and concludes that the first calibration met the target of stable biomasses and validated spatial distribution of two species, while emphasizing some data limitation and parameters sensitivity. I recognize that the work done to build this ecosystem model is huge, and has required to integrate a lot of information. However, some of the choices that have been made seem wrong to me, and I do not agree with the validation section. Therefore I suggest to reject the manuscript (but see comments below, hopefully to serve as a basis to improve the model and resubmit a manuscript in the future).

## Specific comments

- Spatial validation

The model is presented as a good tool to better understand ecosystem functioning in 3D. Based in the introduction, I would have expected some results regarding spatial functioning. This was limited to only two species (the authors could have produced spatialized ecosystem, community or trophic indicators for instance), and the validation was not convincing. Indeed, the model is initialized with the average state of 2000-2003 (cf line 85) with repeated oceanographic forcing of the year 2000 (cf line 160), i.e. no interannual environmental variability. Sardine and anchovy biomass distribution is parameterized from eggs distribution from BIOMAN 2018 (cf appendix). The authors then compared the simulated biomass of 2019 and 2020 (i.e. 20 years after initialization with similar environmental forcing) with eggs distribution observed during BIOMAN 2019 and 2020. The comparison performed is more about how similar are BIOMAN 2018 to BIOMAN 2019 or BIOMAN 2020 than how good the model is at fitting observation, specially as there is no interannual variability in the forcings. Therefore you cannot use these results to validate the spatial structure and functioning of this Atlantis model.

The spatial validation was done for the two species that spatial maps were used to define the input distribution. We agree that the spatial validation should have been done for all the groups. Also the maps cannot be used as a direct comparison, but as an indication of whether or not the spatial distribution of these two species fit with the expected overall distribution.

- Model setting

I sometimes wondered what was the rationale behind the grouping of some species into functional groups (e.g. the distinction between « deep sea fishes » (including alfonsino, red bandfish, roughsnout grenadier...) and small demersal fishes (including softhead grenadier) is not always clear to me), but there is no perfect grouping, and I can work with that. The main difficulty I have with the grouping relates to the « representative species » used for parameterization. For instance conger represents large demersal fishes, but even if most abundant in the survey, it is quite different from the other species. Using the most abundant species is a correct way of proceeding, but similarities between species should be considered as well (for instance, if a group counts several gurnard species, maybe the abundance of all the gurnards together should be computed to assess what is the main species). I mostly wonder what are the ecological meaning of using the growth parameter of one species (e.g. grenadier) with maturity and mortalities from another species (e.g. alfonsino) as these species can be quite different. Databases of functional traits exist now, and maybe could be used to ensure the ecological coherence of grouping some parameters of one species with parameters of another. If no other solution exist, this should be at least discussed.

A huge literature review was done, and to be consistent with how all the groups were parameterized, we decided to use the most abundant species of the surveys. Using databases of functional traits to ensure the ecological coherence of grouping some parameters of one species with parameters of another is a good suggestion, and will be considered in the further work on the improvement of the model. We will add a section in the discussion on how the combination of parameters potentially can change the ecological role of the functional group.

Finally, a lot (if not all) recruitment parameters come from the SE-Australian Atlantis model, with no indication about possible similarities between ecosystem functioning and/or taxonomic group. Some details about the generality of this procedure should be provided (it is the same for most Atlantis models – and therefore could be considered as default values – or is it specific to the BoB application, and if so why?) and the impacts of the results should be discussed as well.

Agree. An explanation of why we considered recruitment parameters from SE-Australian Atlantis model will be added in the reviewed manuscript.

- Biomass estimates and targets

Biomass estimates from survey: how did you translate biomass index from survey to absolute biomass estimates? Did you consider that species are homogeneously distributed and use the CPUE from the survey scaled to the entire BoB? Which catchability coefficient did you use? When estimating the biomass of a functional group with several species, did you use the biomass index of this representative species or of all species? How did you couple information from EVHOE and DEMERSALES (for example regarding spatial distribution, eg bivalves and polychaetes, are you sure that both survey have the same catchabilities or that both survey provide the same level of taxonomic details for benthic invertebrates?).

The biomass of the of a functional group was calculated summing up the biomass of all the species. The information from EVHOE survey was used for the area corresponding to the ICES 8c division, whilst the DEMERSALES survey was used for the ICES 8ab division. All this information will be added in the supplementary material of the reviewed manuscript.

Dealing throughout the manuscript both with total biomass (tons) and biomass concentrations (tons/km<sup>2</sup>) is confusing, and I would suggest to use only one if they are equivalent (and if they are not, specify how to go from one to the other). Some initialization values seem odd in regards to the reference (e.g. other large pelagic fishes initialized with biomass of 0.14 t/km<sup>2</sup> supposedly from EwE; but in Corrales et al, this group has a biomass of 0.27 t/km<sup>2</sup> (idem mesopelagic fish: 1.80 t/km<sup>2</sup> in Atlantis – 1.05 in EwE).

The biomass differences are due to some adjustment we did for some of the species as they were too low when taking into account predation pressure.

I think that more biomass estimates could be obtained from stock assessment (e.g. for sole, which is not well caught by GOV trawls), and could even be associated with smaller uncertainty. Why haven't you used that source of information?

If available, biomass estimates were used from surveys instead of stock assessment, as the survey biomass estimates let us define the spatial distribution of the groups. But it's a good suggestion to compare the output from the model to the assessment time series, this will be done in the revised manuscript.

- Seasonality

I couldn't find any information about the seasonality of the parameters (it is mentioned e.g. Line 228 but distribution from once-a-year survey cannot be used for that), so what is variable across seasons and what's is not? How come we can see some seasonality in the simulated biomass of some vertebrates groups but not in the plankton biomass? It seems to be one of the target of the calibration (cf line 351) but doesn't seem to be well represented.

Same distribution was used per season, since, as you comment, the once-a-year survey information cannot be used for this. The revised manuscript will be rewritten to make all explanations clearer.

- Calibration steps and targets

The calibration steps are great, and should be better emphasized I think. I didn't understand if when calibrating the NPZD part of the model the authors used the current structure of the model (with several phytoplankton and zooplankton boxes) and a really simple NPZD model. It would have been interesting to plot the initial simulation, the simulation #14 (end of the first step) and the final simulation to see the evolution of this low trophic levels groups. The targets of each calibration step are not specified, so we wonder what model properties is assessed at each step.

When calibrating NPZD part of the model we used the current structure of the model. The plots of the initial simulation, the simulation #14 (end of the first step) and the final simulation to see the evolution of the low trophic levels groups can be added to the supplementary material in the revised version. Also the targets of each calibration step (i.e., Increasing maximum growth rate because they are not growing enough, increase recruitment because there are too few babies) will be added in the methods section.

Finally I also wonder if one can obtained stable biomass around an observed averaged biomass (with fishing mortality affecting most groups in real life) with no fishing activity/mortality considered in the model and an objective of keeping all mortalities as low as possible (ie even accounting for the non-represented fishing mortality). Is such a state conceptually reachable?

When an Atlantis model is developed without fishing activity (named unfished system), the aim of the calibration process is to produce stable biomass through time under stable-environmental forcing, that is, keep all groups from becoming extinct, and obtain stable biomasses that are not oscillating more than a certain percentage of their initial values (+-50%?).

### Technical corrections

All the technical corrections/comments will be taken into account and corrected in the revised version. Below can be found some clarifications of some comments.

Line 82: why the model starts from Ribadeo and not Cabo de Finisterre as it is the case for Ecopath ?

Ribadeo was decided to be the limit of the modelled area considering information from the BIOMAN and JUVENA surveys area, and previously developed model in the area (Lasalle et al. 2011: "Lower trophic levels and detrital biomass control the Bay of Biscay continental shelf food

## Reply to the comment in blue

web: Implications for ecosystem management”). The definition of the modelled area was done before the development of the EwE for the BoB (Corrales et al. 2022).

Section 2.2: there is a lot of information on the Atlantis framework, but that it is not used in the BoB application. This is misleading, I would suggest to focus on key aspects used in the BoB application and/or emphasize on the modular aspect of Atlantis.

Agree. The information that is not relevant for the study will be removed in the revised version.

Line 124: what is the difference between age-structured biomass pools and age-structured groups ?

Biomass pools can include a small number of life-stage stanzas and fully age-structured, where the average size-at-age of a typical individual is tracked along with the membership of that age group. Typically, biomass pools are used for invertebrate groups and age-structured for vertebrates.

Line 188: what does « nitrogen created as spawn is temporally removed from the model and returned as recruits » mean ?

The model explicitly tracks the flow of nutrients through the trophic levels, typically using nitrogen as the “common currency” (carbon and phosphorous may also be used). We then used the Redfield ratios C:N (carbon to nitrogen) of 5.7 and WW:ASDR (wet weight to ash free dry weight) of 20 to convert from carbon or dry weight to nitrogen.

In fully age-structured groups and age-structured biomass pools nitrogen produced as spawn is temporarily taken out of the model and is returned as recruits after a set larval period. After this time recruits start “arriving” into the model throughout the period set. Recruited individuals are attributed to the first year of the first age group.

Line 197: what are the conversion factors used between nitrogen and wet weight for vertebrates groups ? (See comment above)

Section 2.4: I suggest to present model details as they are used for the BoB application (i.e. if a process is optional and not used – do not present it here).

Agree. Will be corrected in the revised version.

Lines 369-371: reducing cannibalism was already mentioned lines 363-364. The reduction is huge, what does it mean to keep it and not to put it to 0 ?

Was intended to not remove completely the cannibalism, just keep a tiny amount.

- Appendix

Throughout the document: specie should be species

- I couldn't find in Audzijonyte et al. 2017 explanations or guidelines about how to convert the annual mortality values found in the literature into mortality value per day for juveniles and adults to be used in Atlantis – could you explain it?

## Reply to the comment in blue

The annual values found in the literature (per year) was first converted to daily mortality. This daily mortality was then multiplied by  $1e^{-4}$  to obtained natural mortality for adults, and the adult mortality divided by 2 to obtained the mortality for juveniles.

- Table S3 – sardine: what are the implications of having > 99% of 4 age classes mature (age 3 to 6) instead of 100% ; Isn't it too refined compared to the model structure and level of details of other species?

Agree. This will be updated in the revised version.

- Table S8: this table is impressive, both because it reflects the amount of work done but also the rigor followed when calibrating the model. I am sure it can be further exploited in different ways. I was surprised though that when a parameter is calibrated, you don't come back to adjust its value. Does it mean that there is no retroaction effects from the upper trophic levels components on the dynamics of a group?

I'm a bit uncertain what are you referring to. In the supplementary material of the paper Pethybridge et al. (2019) there is a table summarizing common symptoms of improper parameterizations, what do we have to observe and the actions to take (i.e. what parameters we have to tune) to improve the calibration. In each change, the dynamic of the group and the already calibrated groups are analyzed.

- Is there a difference between figure S2 and figure 3 from the main text?

Figure 3 corresponds to the absolute biomass values and figure S2 to the relative biomass values with; the figure S2 was used in the calibration of the model.