Comment on gmd-2021-328
Anonymous Referee #1

The authors present the high-resolution regional ocean model IBI-CCS covering a relatively large domain west of Europe, and use it to dynamically downscale the CMIP6 model CNRM-CM6-HR. In the first part of the manuscript, CNRM-CM6-HR, its low-resolution version, and different versions of the downscaling set-up are evaluated mainly against reanalysis data. In the second part of the manuscript, projected changes are compared focusing on sea level, highlighting the effects of dynamical downscaling and a number of corrections that have been applied to the prescribed boundary conditions derived from CNRM-CM6-HR. This has clearly been a major effort, helps to interpret previous downscaling studies for northwestern Europe and offers several new methodological insights. While the model set-up appears to be sound, I have a number of comments on the bias corrections and the evaluation of the model output. Additionally, I think the manuscript is not very well polished. I therefore recommend that the authors check the manuscript to improve the writing (grammar, typo’s, sentence structure) and clarity throughout.

Main comments:

L86 states that the aim of the study is to provide projections of sea-level changes, focusing on methodological aspects. If this is the aim, using only a single (downscaled) GCM is probably insufficient. The specification in L86-L90 suggests, however, that the aim is to evaluate the impact of dynamical downscaling and bias corrections on simulations of sea-level change, while the title suggests that the aim is to ‘evaluate sea-level change’. Other statements in the manuscript imply that the study is meant for presenting and evaluating a regional ocean model that will apparently be used for analyzing extreme sea level in a follow-up study. Altogether, I think that the purpose of the manuscript needs to be more clearly described in the introduction, in the conclusions and possibly in the title as well.

L221-L239: the authors apply a bias correction to the boundary conditions by subtracting the historical mean seasonal cycle of biases, assuming that biases are stationary. However, the seasonal cycle and therefore the associated model biases are likely to change in the future. Could the authors comment on the validity/caveats of their methodology in this light? Showing the size of the bias (corrections) may be insightful as...
Section 3.1: In multiple comparisons in this section, the IBIRYS and IBI-ERAi products are taken as the ground truth for evaluating the added value of downscaling, while those products rely on models and have their own biases as well. A quantitative comparison against observations to support the case that the changes due to downscaling and bias corrections are actually improvements, seems to be missing.

Section 3.1.6: the authors compare the 99th percentile SSH between IBI-CCS_corr and GESLA2 observations and conclude that their model ‘properly’ reproduces the observed extremes, with an ‘error rarely exceeding 20%’. What is missing, however, is their motivation and contextualization for why errors of 10-20% are acceptable. Additionally, the authors should motivate why assessing only this single aspect of the simulation of ESLs is sufficient evidence that IBI-CCS is ‘a suitable tool (L790)’ for analyzing extreme sea levels and projecting their changes, as the authors seem to plan doing in a follow-up study.

Other comments:

L11: “some relevant processes’ - here and elsewhere, the paper would benefit from the authors more precisely formulating what they mean

L15: suggest changing ‘developed for’ to ‘participating in’

L36: here and throughout, it would be good to avoid complicated compounds such as ‘decision-making processes’ (consider ‘decision making’), ‘SL focus’ (‘focus on sea level’), ‘the model spatial resolution’ (‘the spatial resolution of the model’), SL changes spatial variations (‘spatial variations of SLC’), etc.

L45: remove ‘to’ or replace by ‘due to’

L51: could the authors add an average kilometric resolution? (only few CMIP6 models have a quarter degree resolution)

L56: The authors may consider writing out less commonly used abbreviations like DD, as well as the regional features in Figure 1, as these abbreviations may confuse readers unfamiliar with these terms

L60: ‘thanks to DD’ -> ‘using DD’?

L75: is it dynamical downscaling that can overcome this problem or is it the bias corrections?

L81-82: unclear sentence structure

L85: up to this point the authors have made clear what has been done in the literature, but they have not pointed out what has not been done in the literature and therefore the motivation for and the novelty of their model/study is not yet clear.

L89: the adjective ‘high-resolution’ is somewhat misleading here: the resolution may be high compared to other CMIP6 GCMs, but is not that high compared to some regional ocean models. If it would be, there would be less reason to dynamically downscale.

L89: ‘to these aims’

L99: I suggest to change ‘climate change scenario’ to ‘greenhouse gas concentration’
scenarios’ and add a reference. Additionally, the choice to use these two scenarios needs motivation.

L104: ‘regional ocean climate model’ change to ‘regional ocean model’?

L103-L106: could the authors clarify why CNRM-CM6-1-HR specifically is chosen?

L111: does this mean that only the horizontal ocean resolution is increased in IBI-CCS compared to the CNRM-CM6-1-HR model?

L119: ‘to CMIP6 typical resolution’ -> ‘to the typical ocean grid resolution of CMIP6 models’

L130: ‘required to force’ -> ‘for’

L135: ‘approximatively’ -> ‘approximately’

L136-137: needs a reference

L149-155: I suggest to change the order: first introduce the grid (resolution) of the IBI-CCS model and then explain its grid is based on an existing reanalysis

L165: perhaps ‘added value’ here should be replaced by ‘expected areas of improvement’ or alike, since the added value is assessed later on in the manuscript

L188: the reason for the different set-up at the eastern boundary is missing

L189: could the authors specify the type of boundary constraints used for the other variables as well?

L190-191: from where are the initial conditions and the runoff derived? From the GCM? Observations? elsewhere?

Section 2.2.1: could this usefully be merged with the text above table 1, leaving just a separate section describing the version with drift/bias corrections?

L211: could the authors specify how large? Does this have any effect on the SSP runs?

L215: please specify which variables, and could the authors explain how they know a linear fit is ‘indeed appropriate’?

L216: subtracted ‘from’

L269: The structure of the manuscript may benefit from a more descriptive title for this section

L270-272: I think this line of reasoning could be clarified by writing that the frequency shift of ESLs depends on total sea-level rise rather than just the ocean dynamic component, hence additional SLC components need to be incorporated in the model.

L273: the contribution of changes in land-water storage to GMSLR seems to be neglected?

Section 2.3.3: it would help to clearly distinguish between sea level and sea-level change in the manuscript. For instance, is defined as total sea level (not sea-level change) but also as the sum of GMSL”R” and SSH. Sentence structure and grammar also need attention in this section.
Figure 3: this figure shows many different things. To clarify things, the authors could consider splitting it up in multiple figures and discussing them one by one. I am also wondering, if the aim here is to evaluate depth-integrated salinity and heat, why these variables are then not evaluated directly, since their biases may be more intuitive to understand than their meter sea level equivalents. Also, in all maps in the manuscript, it would be helpful to indicate some of the relevant isobaths in the regions (e.g., the shelf break).

Figure 4: apart from panel c, the arrows depicting the currents are too small to recognize.

Figure 6: it would be helpful to mention that (presumably?) all panels here do not include the inverse barometer effect on mean sea level (same for Fig7?). Additionally, can the authors point to any reason for the differences in MDT between (a) and (b), especially in the Celtic Sea?

L462: the authors mention that all models reproduce the observed MDT (6a) 'well'. This assessment needs further justification given the differences between modelled and observed MDT in Figure 6.

L572-574: can the authors explain why it is expected that IBI-CCS_corr matches better with projections from CMIP(57) models than IBI-CCS_raw? One may expect this to be the other way around, assuming that the CMIP models referred to also have deficiencies in resolving the Mediterranean Sea?

Figure 12: can the authors explain why they only computed steric changes 0-2000 m, and what the implications are for comparing results against studies that have integrated down to the full depth of the ocean? Does this mean that manometric + steric change in this case is not equal to total DSLC, since steric changes below 2000 m depth are missing? If so, that should be mentioned. Also, is the change of the inverse barometer effect incorporated in the manometric change here?

L593-611: I do not fully understand the sign change of thermosteric SLC in large parts of the Atlantic between SSP1-2.6 and SSP5-8.5. Additionally, the positive halosteric SLC in the Atlantic implies the whole 0-2000 m is getting fresher everywhere in the deep ocean. Can the authors point to any reasons for these effects? Have columns c and d of Figure 12 been interchanged accidentally, by any chance?

Figures 14 & 15: the white dots do not strongly contrast the shading of the figures, and in many cases cover most of the domain. I suggest choosing another color and only stippling those parts that have insignificant differences.

L634: some caution is needed in attributing all of these differences solely to resolution issues, since resolution is not the only aspect in which the models differ.

L668-669: the authors state that the sea-level projections of IBI-CCS_raw and IBI-CCS_corr (note the typo) are very similar. However, this seems to be contradicted by Figures 15b-d, which show fairly large differences in the deep ocean. Since the differences in DSLC (Figure 15a) are smaller, apparently manometric SLC also differs substantially due to the bias corrections. Could the authors comment on this and explain why the bias corrections appear to affect the steric and manometric changes in the deep ocean in a substantial but compensating manner?

L689: significant differences or significant changes? These plots do not show differences between models do they?

L700-702: could the authors point to any potential physical mechanism behind the
decrease of the magnitude of interannual sea-level variability in the North Sea?

L749-750: it would be good to refer to Hermans et al. (2020) here once more, since they already argued the same thing, based on downscaling two GCMs differing in ocean grid resolution in the region.

L751-752: it would be good to add that to verify this, CNRM-CM6-1 would need to be downscaled as well.

L776-777: Other GCMs may have other biases that may require different corrections. Different types of corrections, for example based on different reanalysis products, have not been tested here. Additionally, the model set-up in this manuscript requires more computational effort than time-slice methods such as used by Jin et al. (2021), which makes it difficult to obtain the large ensembles that are eventually required for comprehensive projections. Some more discussion of the caveats and the wider application of the methods in this study seems warranted.

Finally, the authors may consider adding another test, comparing a simulation with and without tides, for example in terms of seasonal biases and the simulation of mean sea-level change. This may be well outside the scope of the manuscript, but if relatively easily implemented in the IBI-CCS framework, it could give valuable insights into the limitations of models excluding tides in the context of projecting mean sea-level change.