

Dear Dr. Buscardo,

Thank you for the reviews and the opportunity to submit a revised version. Following the suggestions and recommendations, we revised our manuscript. Particularly, we added a specific statement of the main goal and we revised the introduction to improve the relation to the hypotheses. Please find our point-by-point responses below.

Yours sincerely,
Simon Scheiter and co-authors

Dear Dr. Simon Scheiter,

We have received the comments by one of the original reviewers and a new one. While the original reviewer is satisfied with the way you have addressed his comments, the new reviewer has additional suggestions as to how to improve the quality of your manuscript.

Among them you are asked to clarify what the main goal is and what the specific objectives of your study are and relate them directly to the hypotheses tested. Please do so, taking into account the new reviewer's comments.

Thank you.

Sincerely,

Erika Buscardo

Reviewer 1

(no comments)

Reviewer 2

I am reviewing this manuscript for the first time. After reviewing it as it were a new submission, I found that it is a very interesting piece of work that is useful to scientists working with vegetation-climate relationships and climate change. In general, I found it is well-written and the methods are well executed. However, I found that the main goal of the manuscript was neither clear nor well-justified and is not fully aligned with the hypotheses. I was also disturbed by the word "objective" in the title, as if objectivity would be an asset, but then realizing that the methods are plenty of arbitrary decisions that are far from being "objective". While I would tend to agree that objectivity should be pursued, subjectivity is always present and this work is not an exception. Some of these points are in line with comments made by previous reviewers and have been already addressed in the revised version I read, though perhaps not to the extent it would be required to meet acceptance for

publication. I would like to develop these points further in the hope that are of use to the authors for improving the manuscript.

Thank you for your positive feedback and constructive suggestions to improve the manuscript. We addressed all the points, and provide our responses after the more specific comments below.

While the main objective, purpose, aim or goal of the manuscript is not explicitly stated, a justification for this work is found on lines 61-64, 67-68, 77-81 and 83-85:

"This makes objective and direct comparisons of modeled biome patterns and biome change difficult. We argue that an objective biome classification applicable to any GVM is required to facilitate multi-model comparisons of future biome shifts and a quantification of uncertainties related to biome classification."

"A further caveat for analyzing modeled biome distributions and to evaluate models based on observation-based biome maps is that multiple observation-based biome or land cover maps were developed"

"The biome classification scheme applied to GVM results and the observation-based biome map used for data-model comparisons can therefore influence data-model agreement and projected rates of biome shifts under future climate conditions (Scheiter et al., 2024a). Misclassification of biomes may result in inappropriate conclusions for conservation and management policies for ecosystems (Kumar et al., 2020)."

"To achieve robust and comparable projections of biome shifts and quantify uncertainties in biome projections across GVMs, climate change scenarios and existing observation-based biome classification schemes, we first developed a novel and objective biome classification framework."

This is followed by what was actually done, but the connection between these arguments and the reason for doing what was done is missing. I believe it requires stating the goals of the manuscript.

We added a main aim both in the abstract and the introduction: "Here, we aimed to robustly quantify and map future biome shifts under climate change using a reproducible, multi-model ensemble approach."

Furthermore, the four hypotheses are poorly aligned with these justifications, except perhaps hypothesis 1, which is hard to falsify as "large variations" can potentially be anything.

In hypothesis 1, we agree that "large variation" is hard to falsify and also not a surprising result. We reworded the hypothesis to make clear that we expect differences in data-model agreement between GVMs and biome map combinations.

Hypothesis 2 predicts that biome stability and susceptibility patterns are consistent across models, but the meaning of "model" in this context is not clear to me and is perhaps

something that is not expected to be related to a biome type but rather to the geographical location of the climate change as predicted by the RCP scenarios, which appears to be confirmed by the results, though not properly evaluated.

In hypothesis 2, we replaced “models” by “GVMs” for clarification. We also think that Fig. 2 clearly illustrates agreement and disagreement of GVMs regarding biome transitions. In large areas, none of the GVM-biome map combination indicates biome transitions, while at the other extreme, many combinations indicate biome transitions (the different categories in the map). To analyze what drives the susceptibility to biome change, we conducted the random forest regression with different drivers related to precipitation, temperature, temperature zone, location (latitude, longitude) and Fig. 4 shows that current temperature, latitude and current precipitation are the most important factors. We think that these formal statistical analyses are a proper evaluation and support our hypotheses. Yet, we would appreciate additional and specific suggestions how to evaluate this “properly”.

Hypotheses 3 and 4 are connected with "well-understood biome temperature limits" which I found hard to define, difficult to assess and poorly developed in the methodology.

We agree that hypotheses 3 and 4 are not well connected to the ms text provided in the reviewer’s comment above. Yet, in the introduction, we also state “Yet, under the current climate model projections with more consistency in temperature change than in precipitation change, we expect that boreal forests and other biomes with clear temperature-related bioclimatic limits should respond more consistently to climate change across different GVMs and climate change scenarios than biomes primarily limited by moisture or disturbances, or those without clear bioclimatic limits.” This is a justification for hypotheses 3 and 4. We revised the introduction to be more specific:

“We used regression models to relate the susceptibility to environmental drivers. As temperature changes are more consistent in climate change projections (Shi et al., 2021), and PFTs in GVMs are often defined by temperature-related bioclimatic limits, we expected more consistent responses to temperature than to precipitation, and in biomes associated with temperature-related bioclimatic limits. The representation of plant hydraulics and drought-induced tree mortality are a source of uncertainty in many GVMs, limiting the model's capacity to simulate responses to precipitation change, particularly drought, and temperature effects on vapor pressure deficit (VPD).”

We agree that the definition of temperature-related biomes in the case study for the Olson biome map is subjective, and we fully transparent about this in our ms; we state that this is expert-based and that other authors may use a different approach. To include a less subjective approach to analyze main drivers of biome change, we also conducted random forest regressions including all GVMs and biome classification schemes and a set of different predictors such as climate drivers, latitude and longitude. This analysis does not include expert-based classification but is purely based on modeled biome changes and environmental driver, including all modeling results in our study.

My suggestion regarding these points would be to make it clear which is the major goal and the specific objectives of this work and how do they directly connect with (testable) hypotheses, and, unless the methods to assess hypotheses 3 and 4 are made more robust

(for example, fully clarifying how temperature driven biomes were chosen beyond saying "based on literature and expert knowledge" [line 240], and testing the changes against a random classification scenario), remove the temperature evaluation part, as it looks weak (only one biome map was considered [Olson et al. 2001], the distinction between mainly limited by temperature or other factors is too black and white, the results are not very insightful) and it does not appear to be fully related to the (unstated) objective of the work. Regarding hypothesis 2, I would like to suggest evaluating the extent to which stability and susceptibility patterns are related to, or are even the reflection of, the magnitude of climatic change in RCPs scenarios (which is not the same as testing for RCP scenario) and so they are or are not biome- or model-dependent or are just a matter of geographical location. I believe this could be done per climatic variable and/or aggregating them in a PCA or something similar.

As described above, we added a clear statement on the main aim of our study. We also revised the introduction to better relate to hypotheses 3 and 4 (see previous response). We used two different approaches to address those hypotheses: a formal regression model, and a more subjective approach where we classified biomes into temperature-related or not.

We reworded the description of our classification into temperature-related biomes for clarification (changes in italics): "This classification is based on literature and expert-knowledge of the authors, and other authors may use another classification. *We did, however, not test the sensitivity of our results to this classification as we used random forest regressions to analyze biome changes for the full model ensemble in response to multiple predictors (see previous section).*"

We prefer to keep both approaches. The black-white approach for the Olson map should be seen as a case study for a specific and well-known biome map, in contrast to the random forest regressions applied to all combinations of 31 biome maps and 5 GVM. Repeating the temperature-related classification for all 31 biomes would be possible for further evaluation and sensitivity analyses, but this would be challenging and beyond the scope of our study.

The methods are plenty of subjective or arbitrary decisions. For example, land use, nitrogen deposition and fertilizer input were included as fixed, though they may change in time and potentially affect results; choice of LAI as biome classification variable; ignoring bare ground and anthropogenic PFTs (which is not methodologically developed in supplement S1 as one may have expected); not aggregating biome maps into a consensus map. This would be perfectly fine to me if the claim of an "objective" classification were not there, so my suggestion is to remove that claim from the manuscript, including the title. For the title, an option would be "Machine learning biome classification...", which would better reflect what was actually done.

We agree that there are subjective decisions in our approach. Following the recommendation, we avoid the word "objective" in the ms by deleting it or replacing it by "reproducible". Yet, we do not agree that our decisions are arbitrary; our decisions are justified in the methods and the discussion. For example, we describe in the methods why we select LAI for classification (as it is a common variable for biome classification) or why we use specific GVM model scenarios from the ISIMIP repository (because data availability is

maximized compared to other scenarios, and decisions regarding scenarios were already made by the ISIMIP consortium while developing well-motivated settings and scenarios).

Given the comment regarding objectivity, also in the first reviewer comment, we would have been interested in how a real objective approach would look like from the reviewer's perspective. We would think it's just impossible to do any scientific experiment or analysis without making decisions. As the reviewer states "subjectivity is always present". From that perspective, objectivity is just impossible.

Minor comments:

188: Random forest analysis needs more explanation regarding the setup, such as the division between training and test data.

We clarified the random forest setup: "We used the 'randomForest' default settings. For each random forest, 2500 trees were created. Model performance was evaluated using out-of-bag (OOB) error estimation for each tree, therefore we did not split data into training and testing subsets."

236, 253 and elsewhere: Please provide versions of all employed packages and software.

We added version numbers for R packages.