Dear Editor, dear reviewers,

thank you for your constructive comments and suggestions for our study. We carefully revised the manuscript and addressed all points raised by the reviewers point by point. Please find our responses below.

Particularly, we included a new random forest regression analysis that links biome change explicitly to temperature and precipitation (current values and change). The analysis demonstrates that temperature change is more important than precipitation change, and that high latitudes are more susceptible to biome change. We also clarified that the main purpose of the ms is a multi-model comparison of potential biome shifts under future conditions, rather than method development, and that our classification approach is necessary for such an analysis. We included more information on temperature responses and bioclimatic limits in different GVMs.

Yours sincerely,

Simon Scheiter and co-authors

Review 1

This manuscript uses global vegetation models (GVMs) to project the likelihood of biome shifts in different climate change scenarios. The study shows that the used gvms are capable of reproducing current biome distribution as biome maps derived from model outputs agree with observation-based biome maps. The gvm biome projections show increasing risks of biomes shifts in higher emission scenarios and that higher latitudes were more likely to have biome shifts than lower latitudes, which the authors attribute primarily to temperature effects.

The manuscript is well written and generally clear, despite the methodology requiring some clarifications and the discussion being excessively long and descriptive in my opinion. I found the study results interesting and useful for scientists interested in climate change. My main criticism is related with how some hypotheses are tested. The authors use a classification based on expert knowledge of temperature limited and not-limited biomes and compare how these groups shift. The problem with this approach is that the actual temperature change in the grid cell is not considered, then it is impossible to claim if the biome really shifted because of temperature effects or some other reason (that is, even a very temperature limited biome still couldn't have shifted because of temperature if there was no temperature changes in that particular grid cell to being with). I believe a more rigorous test should involve assessing the relationship between biome shifts and the actual temperature change in the grid cell.

Thank you for the overall positive feedback and the suggestion for additional analyses to relate biome change and temperature change. We agree that our classification into temperature limited and other biomes is expert-based and created by the authors. As commented, we did not relate biome shifts to actual temperature shifts projected by the climate models. We only referred to previous studies indicating that temperature change is projected to be stronger in higher latitudes and concluded on biome shifts.

However, our hypothesis that predicted biome shifts are more consistent in temperature-limited biomes was based on two observations: 1.) In particular cold-temperature limits are relatively well understood and, therefore, probably well and similarly represented in different GVMs 2.) Climate models simulate very consistent and very strong temperature changes in the high-latitudes (where cold-temperature limits are crucial for biome borders). In other words, the difference in temperature change is a mechanistic component of our hypothesis.

Yet, we included additional random forest regressions to analyze more explicitly if the probability of biome change can be explained by current baseline temperature and precipitation, by the anomalies until the end of the century, or by the climatic zone (cold, temperate, warm). This analysis supports our previous results as it shows that temperature anomalies are more important than precipitation anomalies, and that probabilities of biome change are higher for the cold climatic zone. We conducted this analysis for the entire dataset (i.e., including all GVMs, all RCPs, all 31 biome maps), not only for the specific analysis for the Olson et al. biome presented in the original version. However, we kept the original analysis where biomes were classified into temperature-explained biomes and other biomes by the authors.

We added text to describe those analyses in the Methods (line 217) and in the results (line 299) and we revised the discussion to include those results. We also added Figs. 4 and S8 and Tables S9 and S10 to present those new results.

Specific comments:

L24: Provide references for the claim that biomes are often used to assess vegetation change.

We added references. Huntley et al 2023, Bonannella et al 2023, Tobian et al 2024.

L35: I think this statement needs to be clarified. Model projections of precipitation are more uncertain than temperature projections but that doesn't mean precipitation regimes won't change in the future, only that model capacity to project it is currently more limited/varied between models. Therefore, I think it needs to be clarified that temperature limited biomes could respond more to current climate change projections, which does not necessarily mean they will be the most affected by climate change. In addition, it is important to consider that increased temperatures could also contribute to drought, through temperature effects on atmospheric vapor pressure deficit.

We fully agree about these points and will clarify them. These points are partly related to the arguments for our hypothesis that we describe above. Of course, precipitation changes, in particular in combination with increasing temperatures and evaporative demand, can cause large-scale biome shifts, such as drought- and heat-related forest dieback. There is a large debate how changes in drought and drought-induced tree mortality may influence ecosystems under future climates. Yet, given the challenges and limitations of modeling plant hydraulics and tree mortality, as well as more consistent projections of temperature change in different GCMs, we argue that biome shifts related to temperature should be represented more accurately and consistently in GVMs and have larger effect size.

We revised this statement as suggested, to be more nuanced and concise. We now make clear that current models should be more appropriate to model biome shifts related to temperature than to precipitation change, particularly drought and temperature effects on VPD.

Revised version:

„Both temperature and precipitation changes are expected to cause vegetation changes and shifts in biome patterns. 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. The representation of plant hydraulics and drought-induced tree mortality are also a source of uncertainty in many GVMs, limiting the model's capacity to simulate responses to precipitation, particularly drought and temperature effects on vapor pressure deficit (VPD).“ Line 40

In addition, we explicitly analyzed the effects of temperature and precipitation change on biome change using a random forest regression (see previous point).

L95: I don’t fully understand the justification for using only increased CO_2 scenarios. This was the combination that had the highest number of models available?

Yes, this combination had the highest number of models available. This maximizes comparability across models as all forcings are identical. We added „This combination of scenarios had the highest number of GVMs available, ensuring model comparability.“ Line 123

L98: Please clarify if all these models have a dynamic vegetation component (e.g. PFTs competing for grid-cell space, Prentice et al 2007)? I think it should be better explained how each model handles vegetation dynamics and responses to climate (especially temperature) in supplementary material S1.

The analysis includes both models with (LPJ-GUESS, ORCHIDEE-DGVM, CARAIB) and without (ORCHIDEE, CLM) a dynamic vegetation component. We agree that it is important to state this clearly in the ms. We added this information to the main text. Line 134

We added supplementary tables to describe how vegetation responds to climate and temperature change (e.g., by describing if bioclimatic limits related to temperature are included) and what bioclimatic variables are included in the models (Supplementary Tables S2 and S3). Note that in ORCHIDEE and CLM, simulated changes of LAI of PFTs in response to climate change can also trigger biome changes, even if the PFT distribution is not dynamic.

L115: Is it really sensible to discuss biome shifts in anthropogenic areas (for example Agricultural lands)? You are assuming agricultural lands are currently a natural biome based on potentially minuscule PFT fractions on it? Wouldn't be more sensible to exclude these areas?

We share the concern that human impacts are highly relevant w.r.t. current biome distributions and potential future biome change. Therefore, we do not analyze biome shifts in anthropogenic areas but remove them by (1) removing anthropogenic PFTs from the Fischer et al biome maps, and (2) removing anthropogenic PFTs from GVM results using the scaling approach described in L 115. Thereby, our analyses only include natural biomes (from observation-based products) and natural PFTs (from GVMs). This approach was already described in the original version; we revised the related text and added „Our analyses only consider natural vegetation, both in the GVMs and the observation-based biome data, while all anthropogenic PFTs and land cover types are excluded." Line 146

We think that some readers will find global potential natural biome shifts interesting, even though they are not very relevant in areas currently covered 100% by croplands and cities. Excluding such areas based on some kind of arbitrary threshold would not add much, we believe. Furthermore, it is not impossible that some agricultural areas will be reforested until the end of the century, in fact quite a few scenario studies suggest that the agricultural area might decline in the future in some areas.

L145: Why 2 PFT products? One current and one future? Please clarify

Both PFT products are for present-day vegetation, future maps are not available. The two products differ in the set of PFTs, the method to create the products and the spatial patterns of PFTs. To account for these differences, we included two different products for current conditions. This analysis allows us to evaluate how robust the GVM-based classification is, compared to classification of observation-based PFTs. We clarified this in the text and revised: „Those maps are only available for current conditions, while future changes of PFT maps are not available" Line 173

L187: That doesn't seem like a very rigorous test of the temperature effect.

This classification into temperature driven and not temperature-driven biomes was not based on an analysis but on our knowledge. We argue that this classification is not unique as, to some extent, all biomes are influenced by temperature, precipitation, soil, and other factors. Yet, some biomes such as boreal forests have well-established bioclimatic limits related to temperature (extreme cold and growing season length towards the north and chilling requirements towards the south, which have been well constrained already in the 1990s). Our classification is based on such model assumptions. We added more information on temperature effects and bioclimatic limits in the supplement (Tables S2, S3).

L240: Please show graphically or at least describe these effects better. From the info in the previous paragraph, the difference between scenarios does not appear very large. P-values from traditional statistics are not always very informative when using very large samples that can be obtained from model outputs and will be highly "significant" even when the actual effect is very small.

We added supplementary table S8 showing these results.

Regarding p-values "in the previous paragraph": those were not derived from statistics using the full global spatial information of biome maps and GVMs, but contain only 31 data points for each of the 31 biome maps. Spatial data were aggregated into one point per biome.

L270: Where do this p value comes from and what it refers to? The difference between coefficients of variation?

Yes, in this analysis we compare coefficients of variation as indicators for consistency of biome shift of temperature-limited and other biomes. We reworded this paragraph to make sure that the analysis is well-described in the methods section: „… biome change was slightly more consistent, i.e., the coefficient of variation of areas affected by specific biome transitions was lower, …" Line 335.

We also added that we tested differences: „p-value 0.522 for t-test comparing coefficients of variation". L337

L279: I do not think this claim was rigorously tested in your results. I suggest investigating the temperature changes in grid cells with different likelihood of biome shifts.

Thanks for this suggestion, see our comments above regarding an additional analysis on drivers of biome change. We do not claim to disentangle how strongly the biomes respond to a standardized temperature change.

L308: If fire is included in the GVMs you use and they still have low-agreement with data it shows that either the fire model is not working very well or that other factors not explicitly included in the models (such as herbivores) or even other aspects of the model structure or parametrization do not work well for the biomes here.

We agree that multiple processes may explain data-model disagreement, and that there is still high potential for model development and improvement. Yet, we think that it is beyond the scope of our study to dive deep into each GVM and understand which process explains disagreement. Our study can provide helpful insights for the different modeling groups where mismatches with respect to biomes occur and accordingly what biomes or regions should be prioritized. Those insights are based on a large number of different classification models informed by the 31 biome maps.

We added in the discussion: „Nevertheless, our study identified data-model mismatches for different GVMs and can be used to identify model limitations or to attribute uncertainties to fire, animals, humans or other factors. Doing such detailed analyses for the different models is beyond the scope of our study.“ Line 380

L333: Couldn't these acclimation effects also affect the vulnerability of higher altitude biomes? As far as I know they aren't included in GVMs.

We agree that acclimation effects may also be relevant in higher altitudes. We revised the discussion to account for this point. In the meantime, several studies have addressed this issue with GVMs at the global scale, which we discuss in our revision.

We added: „While thermal acclimation and adaptation of temperature optima could mitigate some warming impacts (Choury et al, 2022), these processes are often not well represented in current GVMs. They often rely on fixed thermal response functions (Wang et al., 2025) to model for example photosynthesis and respiration. Consequently, the lack of dynamic acclimation in models may lead to uncertainty in predicting the resilience of ecosystems under climate change (Wang et al., 2025)." L 417

And:

"In contrast to the tropics, acclimation may not necessarily buffer vulnerability to warming in cold and high-altitude biomes. Recent evidence suggests that alpine species possess a more limited capacity for thermal acclimation compared to other biomes (Harris et al., 2024)." L 450

L385: Do any of the used GVMs use these empirical limits? That would make the high agreement with observation maps less impressive than if were the result of bottom-up vegetation dynamic models as suggested in the beginning of the discussion.

Yes, many GVMs use a set of bioclimatic limits, e.g., frost or minimum temperature thresholds, to constrain climatic niches of PFTs. Those bioclimatic limits are based on empirical evidence and implicitly represent eco-physiological processes that constrain growth once those limits are crossed. PFTs defined by bioclimatic limits and biomes do not fully match, such that the models still produce variables (that are then used for biome classification) not directly used for model parametrization and calibration. So it is not a priori clear, that GVMs using bioclimatic limits, and the underlying ecosystem dynamics will produce LAI and PFT patterns that represent the biomes included in different biome products. Yet, we agree that GVMs where bioclimatic limits emerge from ecological processes bottom-up, that would allow adjustments of bioclimatic limits in response to climatic and atmospheric change would be desirable.

In the revised ms, we added a table with bioclimatic limits in the used GVMs, and we added in the discussion:

„Yet, single PFTs do not match directly with biomes. Hence, it is not a priori clear, that GVMs using bioclimatic limits and ecosystem dynamics represented in GVMs will produce LAI and PFT patterns that represent the biomes included in different biome products. We therefore conclude that bioclimatic limits alone do not explain the good data-model agreement in our study.“ Line 478

L394: It would be important to clarify which processes are affected by temperature in each model in the methods section.

We agree, and we added more information on temperature-related processes and bioclimatic limits in the methods section (see also a previous response). We also included tables S2 and S3 providing information on temperature sensitivity and bioclimatic limits in different GVMs.

Fig. 4. Please define “Man” in the legend.

“Man” is mangrove, we added this (and also SnI - Snow and Ice) in the legend.

Review 2

The authors performed a study to determine potential biome shifts under climate change. They make various projections, using 3 RCP scenarios, 5 GVMs and 31 biome maps. The novelty of this study is that they use random forest classification models to translate GVMs’ LAI results to determine biome distributions.

Overall, the paper is well written. My main concern is regarding the focus of this study in relation to its novelty. Why do you focus so much on the future predictions, while the novelty of this study focuses on how to use GVM output in a more meaningful or coherent way. I would much rather see that the study focuses on that, and describes the differences between biome maps (e.g. why do maps differ so much, and maybe if the future scenarios remain, what causes the larger biome shifts at the borders of biomes – is that really just temperature change?). There are already quite some future biome map studies, and while a more coherent one would be nice, this study doesn't specifically focus on why the maps are different, so this study simply adds to the uncertainty in future biome mapping. I am not saying the study is not relevant, I am merely pointing out that, as I see it, the strong suit of this study is away from RCP scenarios and zooming in on how to make the best of GVM outputs and use them more directly, more consistently. The discussion now reads like a comparison of different papers on what biome changes where, which is not the main focus of this study and therefor perhaps not necessary (and not what you want to present as your main take away for this modelling effort).

We thank the reviewer for this important comment. We agree that the balance between methodological innovation and results on potential biome shifts is critical, and we also agree that understanding why biome maps differ is interesting and relevant. Our starting point was the observation that different GVMs use different sets of PFTs and model-specific biome classification schemes, which makes an objective comparison of modelled biomes for current and future conditions difficult. While comparisons of continuous outputs such as productivity or carbon stocks are straightforward, biomes always involve a post-processing and, somewhat arbitrary, classification step. This classification also leads to a lowest common denominator effect, i.e., the DGVM with the coarsest PFT and biome scheme constrains the overall biome classification. To enable an objective comparison of potential biome changes across models and scenarios, which does not suffer from the artifacts of more classical approaches, we first needed a consistent classification framework, i.e. the random forest approach. We see this method as a necessary step towards robust model comparison rather than as the main focus of the study. Robust biome comparisons require unified classification schemes. Our random forest framework provides such a unified, observation-informed classification of GVM outputs into existing biome schemes, enabling, to our knowledge, the first fully comparable multi-model assessment of future biome dynamics. Our approach therefore does not simply "add to the uncertainty", but demonstrates and quantifies the uncertainty related to biome classification and highlights that the choice of classification strongly affects projected biome shifts under future climate change. Such knowledge is essential for the development of conservation or climate change mitigation strategies because mis-classification may lead to inappropriate management decisions.

Furthermore, we think that it would be a lost opportunity to only focus on these methodological advances. Our results are related to issues like Amazon forest or boreal forest die-back and tipping elements in the Earth climate system, and our model results differ from the current mainstream story that forest systems will tip. Therefore, we think that we need to focus on both aspects: to derive consistent biome changes from several models as a first step, and then the analysis and discussion of potential biome shifts.

In the revised manuscript, we considered the points raised by the reviewer.

(i) We strengthened the introduction to emphasize that the methodological innovation serves the primary goal of delivering robust, comparable projections of biome shifts across multiple GVMs and RCP scenarios. We added:

„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." Line 83

(ii) We kept the focus on future projections and model inter-comparisons as the core results while clarifying how the method enables these analyses.

(iii) We expand the analysis of patterns of biome shifts to better address what drives these transitions. Specifically, we added a regression model that explains the probability of biome change in response to current temperature and precipitation, anomalies until the end of the century, and climatic zone (cold, temperate, warm). This analysis clearly shows that temperature change is more important than precipitation change, and that the probability is higher in the cold zone.

The results are provided in the new Figs. 4 and S8 and in Tables S9 and S10.

Abstract

There are a few questions that come to mind.

We appreciate these clarifications and revised the abstract accordingly.

First, why do we need biome projections if we have 31 observation-based maps already and we have 5 GVMs that project future vegetation. What is the benefit of the biome maps created in this study?

The 31 observation-based maps represent current conditions based on different data and methods, but they do not provide future projections. Future biome distributions can only be inferred from models, but GVM-specific classification schemes differ and are not directly comparable. To do so in a consistent way, we first need to classify GVM results into biomes, evaluate how well they reproduce observation-based biome patterns, and then apply the same classification to future GVM outputs. Our contribution is a multi-model GVM ensemble of future biome projections constructed with a unified classification framework. This enables direct comparison of GVMs within each biome scheme and quantification of how projected biome shifts depend on both model choice and reference biome map.

We reworded the abstract for clarification.

Second, the 4-56% projected change is huge, and as I read it, it is a certainty check of the outcomes: depending on the GVM and observation based biome map the random forest is build on, you get these large variations in projected biome change. The RCP's are simply scenarios you put in, so this is a simulation and not a model check. Thus, this sentence needs to be rephrased, as I probably do not understand what message you are actually trying to convey here.

The model evaluation step in our study is the comparison between the 31 observation-based biome maps and the biome maps derived from classifying current GVM results, where our approach shows high agreement (average κ across all models 0.77, maximum>0.9, see Fig. 1). The 4–56% range refers only to projections and expresses the spread in the fraction of grid cells undergoing biome transitions across GVMs, RCPs and the 31 biome maps used for training of random forest models (403 maps in total). These numbers are not used for model evaluation, as suitable future data for evaluation are not available.

We rephrased and restructured the abstract to make clear that the 4–56% range reflects the full future ensemble and spread across GVMs, RCP scenarios and biome maps (in total 403 maps of future biome projections), and should not be read as a probabilistic uncertainty range for future biome shifts. The fact that this range is large is itself a key result. It illustrates large uncertainties in future biome projections, here arising from different climate scenarios, GVM structures and biome classifications.

Third, what are your novel results? There are other studies that have projected biome changes under climate change, so what is new? I do read that Equatorial rainforest remains stable while other studies find forest dieback (yet I also know a study that projects increasing area for this biome). But what else? What is your method or result different from published literature?

The most important novel results are: (a) the first directly comparable multi-GVM assessment of the susceptibility to biome shift under RCP scenarios using a unified classification framework, (b) the identification of regions with consistently high biome shift susceptibility across models, including under RCP2.6, and (c) the robust pattern across all GVMs that equatorial rainforests remain relatively stable in terms of biome type, whereas the largest and most consistent biome transitions occur in mid- to high latitudes, (d) that also most of the boreal forest remains stable, i.e., no evidence for two highly debated tipping elements in the Earth climate system. We revised the abstract to highlight these points.

Introduction

It doesn’t really become clear why this approach is necessary. Sure, many different models with different outcomes currently exist, but the introduction reads like a summation of what goes wrong instead of a delineation of what these maps are actually used for and how they are created. The observation based biome maps is still an unclear aspect to me, how do you have a global (?right – not mentioned until now) observational map of biomes? Especially since it exists in 31 different versions?

The 31 observation-based biome maps we use were compiled by Fischer et al. (2021); we did not create those observation-based maps ourselves and it is not our aim to compare or assess those observation-based biome maps. Our study builds on this dataset rather than aiming to re-derive or re-evaluate these maps in detail. We use these maps to inform the classification of GVM results into biomes. The fact that at least 31 global biome maps exist (and likely more) with different conceptual bases and biome types illustrates the need for an objective framework to analyze and compare biome projections. This is what we develop and apply here.

In the revised manuscript, we provide more information and examples how global observation-based biome maps are derived from different combinations of floristic information, remote-sensing-based land-cover products, bioclimatic zoning, and expert mapping, and why this has led to multiple parallel global products with differing biome concepts. We also provide examples of how biome maps, both observation-based and modelled, are used (e.g. in assessments of ecosystem services, biodiversity patterns, land-use impacts and biome stability) to better motivate why robust projections are needed.

We added in the introduction:

„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.“ Line 62

„For instance, the Olson et al. (2001) and Dinerstein et al. (2017) maps are based on biogeographic zonation and species distributions, and the Whittaker and Likens (1975) is based on temperature, precipitation and plant community distributions. More recent biome maps are often informed by remote sensing products. For instance, Buchhorn et al. (2020) used multi-spectral reflectance data, and Higgins et al. (2016) used a range of parameters including productivity and height. Biome maps are widely used in climate change and biodiversity research. For instance, the Olson et al., (2001) map is often used as a reference biome map for biome-specific analyses or by international initiatives such as IPBES or WWF.“ Line 70

Additionally, the uncertainties mentioned in the introduction around GVMs seep through in the future predictions. Not only the aftermath (going from model results on biomes) is what causes uncertainty, but also various aspects within the GVMs. How do you propose that is handled by your random forest? I don’t understand how a more direct link between GVM output and biomes helps reduce most of the uncertainty. Relatively speaking, how much of the uncertainty in GVM results comes from the final aggregation of data into biomes?

We agree that uncertainties from GVM structure and forcing propagate into future projections, and we see this as additional motivation for an objective framework that allows us to compare biome shifts across a range of GVMs and biome schemes. Our approach does not reduce the underlying GVM or climate-forcing uncertainty; rather, it standardizes the classification step so that these uncertainties become more transparent and comparable. The results can help guide model development, for example by highlighting biomes and regions where agreement with observation-based maps is particularly low.

In the revised version, we explicitly distinguish three main uncertainty sources: (i) GVM process representation and parameterization, (ii) climate forcing and scenario uncertainty, and (iii) the biome-classification step. Our random forest approach primarily addresses (iii) by eliminating subjective, model-specific biome classification rules and applying a consistent, observation-informed scheme across models. While we cannot robustly quantify the exact fraction of total uncertainty attributable to each component, our ensemble results show that both GVM structure and the choice of the biome map used to inform the random forest models substantially influence projected biome shifts.

We added:

„While GVM uncertainties related to different climate forcings (Ahlström et al., 2012), climate change scenarios (Martens et al., 2021; Tobian et al., 2024), and the representation and parametrization of vegetation (Pappas et al., 2013) have been analyzed in previous studies, we address uncertainties related to the post-processing step of biome classification, by eliminating subjective, model-specific biome classification rules and applying a consistent, observation-informed scheme across models.“ Line 100

Methods

Is there a way to indicate how LAI and PFT are linked?

Yes. In the revised methods section, we added a brief subsection explaining that each GVM simulates LAI separately for each PFT and that LAI of a given PFT emerges from its simulated performance under the prevailing environmental conditions.

We added:

„The GVMs simulate state variables such as carbon pools or LAI separately for each PFT in response to environmental conditions, competition and disturbances such as fire (Prentice et al., 2007). The cover of different PFTs within grid cells can be static and informed by land cover maps (ORCHIDEE, CLM4.5) or change dynamically in response to biotic and abiotic factors (LPJ-GUESS, ORCHIDEE-DGVM, CARAIB).“ Line 132

We describe in the methods that we first weight PFT-specific LAI by cover fraction and then rescale to represent only natural vegetation, and that the resulting vector of PFT-specific LAI values is used as predictor input for the random forest classification.

The biome classification was done for each combination of GVM result and biome map. Why not make one model, to enhance comparison? You mention that the aftermath of GVM results into biomes creates much uncertainty, but now you do the same but with a random forest instead of the method GVM people decided to use. What is the difference?

Separate random forest models per biome map are necessary because the 31 maps define incompatible biome types (different numbers and definitions of classes). Constructing a single merged biome scheme would itself require a series of subjective decisions and would likely introduce additional uncertainty. Our goal is instead to enable systematic GVM comparisons within each established biome scheme and, in a second step, to assess the sensitivity of projected shifts across schemes. The key advance over model-specific schemes is that we use identical predictors and algorithms across all GVMs and train them objectively against the observation-based maps, instead of relying on model-specific thresholds. For more detailed analyses, we then select one widely used map (Olson et al.) as a common reference. We clarified this logic in the revised manuscript.

We added:

„In our analysis, we created random forest models for each combination of the 31 observation based biome maps and the GVMs and did not aggregate them into consensus biome or GVM maps. This approach is necessary because the biome maps and GVMs use incompatible biome and PFT types, respectively. Aggregation into consensus maps would itself require a series of subjective decisions and would introduce additional uncertainty. Further, our approach allows us to explicitly quantify uncertainties related to biome classification." Line 159

I wonder how the random forest models are dependent on the prevalence of the different biomes. Are biomes that are more frequent on a global map not also predicted more often, simply because the change that a 'large' biome occurs is bigger? Random forests can only predict 'small' biomes in the case that the predictors there are extremely specific.

Thank you for this comment, this is an important methodological point. For the Olson map we already report class-specific performance metrics (biome-wise κ values and confusion matrices) and discuss how global biome prevalence affects classification performance. In the revised version, we state this in the methods/results and we systematically assess whether the number of grid cells covered by a biome type is related to the κ value across biomes.

We added in the methods:

„We further calculated κ for each biome type in each F31 biome map and related it to the number of grid cells covered by the respective biome type, to analyze if the performance of the random forest model is related to biome coverage." Line 196

We added a supplementary figure (Fig S4).

We added in the results:

„The κ values per biome were related to size of the respective biome, i.e., biomes with large cover had higher κ values, while biomes with smaller cover showed large variation with values between zero and one (Fig. S2)." Line 264

Results

The results presented in Figure S2 are the basis of the rest of the results, right? I am not sure it is best placed in the SI in that case. Also, the fact that some GVMs have such large disagreements, I find it tricky to understand what this means and why all GVMs are all still used for the rest of the analyses.

We agree that Figure S2 is important, as it shows the spatial agreement between observation-based biomes and biomes derived from the random forest models for all 31 maps. We prefer to keep it in the Supplement for two reasons: (i) the associated κ values summarizing this agreement are already presented in Fig. 1 in the main text, which highlights differences among the 31 biome maps, and (ii) the subsequent results focus on projected biome shifts (Fig. S3, aggregated in Fig. 2), which are central to the objectives of our study. Given our primary aim of quantifying uncertainties in projected biome shifts, we see Fig. 2 as the key figure.

We retained all GVMs, including those with lower agreement, because we explicitly aim to document and analyze ensemble spread and uncertainty rather than focusing only on the best-performing model.

Additionally, you mention that biome shifts are mostly present at borders of biomes. That is quite logical, but what I would like to see (perhaps in the discussion you mention it already) is that the tropical regions don't change.

We agree that biome shifts at boundaries are expected, and we will emphasize more clearly in the text and figure captions that the grey areas in Figs. 2 and 5 illustrate the relative stability of tropical and boreal regions in terms of biome type. As the reviewer indicates above, large-scale biome tipping would also imply biome changes in the core areas of a biome.

In the results section and the figure captions for Figs. 2 and 5 we added that the grey areas represent areas with no or low susceptibility to biome shifts (I.e. biome stability).

Great impact of climate change in e.g. the Amazon has been reported, but this is not reflected. How do GVMs deal with novel climate combinations? Is there an upper limit in climate suitability for the warmest biomes or lower limit for the driest biomes? Or are all warmer or drier conditions simply outputted as the biome with the warmest or driest climate? This links to the statement in lines 322-325 in the discussion.

The GVMs indeed include implicit or explicit upper limits for climate suitability. For example, in many models NPP decreases at high temperatures due to the temperature response of photosynthesis and respiration, which constrains the thermal niche of vegetation. Similarly, limited water availability reduces growth and can trigger drought-induced mortality. However, representation of drought impacts and mortality remains a challenge in GVM development, and potential heat- or drought-related decreases in productivity are often partially compensated by CO_2 fertilization, which can contribute to tropical forest stability in the models. How novel climates influence biomes in our analysis is determined by the implementation of ecophysiological processes in the GVMs; biome types themselves are not a model output but are derived in a post-processing step via our classification. It is generally assumed that, within limits, process-based GVMs can be meaningfully extrapolated to novel climates because they are built on mechanistic formulations rather than on empirical and static climate envelopes. Note, that these models are also often applied to climates different from today, such as the Last Glacial Maximum or the Pliocene to Miocene warm period. They are designed to be applicable to novel climates, albeit, of course, with larger uncertainties than when applying them under today's climate, for which we also have most data to test the models.

We added these points in the revised Discussion:

"Ecopysiological processes and bioclimatic limits drive the performance of different PFT at different temperatures. For instance, at low temperatures bioclimatic limits may constrain establishment of a specific PFT and unusual extreme cold can also kill trees, while at high temperatures respiration may reduce productivity. Climate change therefore has PFT-specific effects on the productivity, LAI, spatial coverage, and may thereby trigger biome transitions. In GVMs, the climatic niche of PFTs is generally preserved, and novel climates do not imply acclimation and adaptation. Whether novel and non-analogue biome states, characterized by novel combinations of PFT composition emerge in GVMs, remains to be tested. Uncertainties persist regarding the extent to which CO_2 fertilization offsets physiological stressors (Walker et al., 2021) in emerging environmental niches and the models' ability to capture extreme heat and drought remains a key challenge for model development." Line 420

In line 219, you mention susceptibility category and link this to consensus between models. I don't understand that. Biome change and model agreement is something else, and the link you refer to is not made specific.

Here, "susceptibility" refers to the number (or fraction) of model–biome-map combinations (5 GVMs x 31 biome maps) that project a biome change in each grid cell, i.e., an ensemble change frequency. Higher susceptibility thus means that more combinations project a change, and lower susceptibility means that more combinations project stability. This is distinct from data–model agreement for current conditions shown e.g. in Fig 1.

We reworded this sentence in the revised version to clarify that this line refers to biome change and model consensus regarding projected biome change.

I am wondering about the comparison with the Olsen map you make in section 3.3. I know that somewhere in the methods you describe that nothing changes in biome structure between early 2000 and 2020, but it is strange that you look into the future until 75 years from now (2099) and ignore the first 20 years. Coming back to section 3.3, you describe that the olson map is reproduced with high agreement. But that map was made before 2001, so should the GVM result from the current climate really match 'such an old map' that well?

We did not intend to imply that nothing changed in biome structure between the early 2000s and 2020. In Sect. 2.1 we state: "We ignored potential biome shifts after the publication of biome maps and compared all maps to the model results for current conditions (2006–2020)." As noted there, the biome maps in Fischer et al. (2021) were published between 1964 and

2020, and a map published in, for example, 2001 is based on data from earlier years. Thus, none of the maps is perfectly matched to 2006–2020 conditions. To maintain consistency, we chose a single reference period (2006–2020) for all GVMs and used this for training and evaluation against all 31 maps, rather than using different temporal subsets for different maps. We agree that this introduces a temporal mismatch. We clarified this issue in the Methods section and state that such temporal mismatches may introduce uncertainties.

Our interpretation of the high agreement with the Olson map is therefore that GVMs reproduce the broad biogeographic patterns encoded in that map, not that they replicate the exact present-day vegetation. We added this point to the discussion. Line 530

I do not understand the difference between figure 2 and 5.

The analysis proceeds in two steps. First, we build random forest models for every combination of the five GVMs and the 31 observation-based biome maps and apply them to the three RCP scenarios. Figure 2 summarizes these full ensemble results by showing, for each grid cell and RCP, how many of these combinations project a biome change. In a second step, we focus on one specific biome map, the widely used Olson et al. map, for a more detailed analysis of transitions between biome types. Figure 5 shows only these Olson-based results. The difference between the figures is therefore that Fig. 2 integrates over all 31 biome maps, whereas Fig. 5 shows detailed results for a single scheme. We reworded the Fig. 5 legend for clarification.