

# Reply to Anonymous Referee #2

B. Abis and V. Brovkin

Dear Referee,

Thank you for your time and for your valuable comments and suggestions. In this document, we will provide an answer to your comments and queries, highlighting how we will modify the manuscript in view of your suggestions.

Best regards,  
B. Abis and V. Brovkin

- C1. “I found the paper difficult to read, mainly due to the many abbreviations. I think that the authors could delete loads of them and just write down the whole names. Further my major concern is the structure of the results. It is unclear what is expected, some parts are discussion already, while some crucial results are not introduced. The authors should take more time to present their results.”
- A1. Thank you for your valuable opinion. We thought we would simplify the paper by introducing some abbreviations. However, from your comments it seems like we actually made it harder to read. We will try to reduce the number of abbreviations and write plain sentences when possible. Regarding the structure of the results, we found it difficult to separate some results from their interpretation and discussion. However, following your suggestion, we will restructure them. We will expand the Discussion section and simplify the results one, especially the GAMs Results section, taking more time to introduce the results and moving all the discussion and interpretations in the Discussion section.
- C2. “Please try to minimize the abbreviations. Is it needed to mention them already in the abstract?”
- A2. We understood it is necessary to introduce abbreviations in the abstract from BG manuscript regulations. However, following your suggestion, we will reformulate the abstract so that it will not make use of them.

- C3. “The aim it to study the impact on the tree cover fraction by eight environmental factors. I think you do not prove that it is the impact; you only link them following a statistical approach? So I would be in favour to change the aim.”
- A3. Dear referee, the meaning of that particular sentence is that we want to quantify the impact on tree cover, since the primary role that environmental variables exert on the vegetation has already been studied by many before us. To avoid ambiguity, we will follow your suggestion and rephrase our aim so that it is clear that we study the link between the various distributions.
- C4. “It is unclear to me why the authors didnt use a seasonal variable in here. I think that seasonality in the temperature and rainfall will probably tell more than the averages and minimum values.”
- A4. Dear Referee, you are right in saying that seasonal variables play an important role in the boreal forest dynamics. For this reason, we actually included several indicators that account for seasonality. In particular, the spring soil moisture measures water availability during the thawing period, when plants have access to a deeper active layer and can start to use unfrozen water, whereas the growing degree days above 0°C are a proxy for the extent and intensity of the plant growing season. In fact, growing degree days are a measure of heat accumulation, and many developmental events of plants depend on it. Hence, by using degree days above 0°C it is possible to estimate the influence of the growing season regardless of differences in temperatures from year to year. On the other hand, we agree that monthly data would provide a finer representation of the different seasonal aspects, however, due to the already high number of variables, such analysis would increase too much the degrees of complexity of the problem, going beyond the scope of the paper. Nonetheless, we recognise that the lack of details about the datasets used and the role of the environmental variables in the boreal forest biome makes it harder for the reader to understand our motivations. For these reasons, we decided to include in the manuscript information about the definition, role, and importance of the variables used for the analysis.
- C5. “What is permafrost distribution? I am not a specialist on this, but it would be helpful to add more information on the selected environmental variables and also add units in [Table 1](#).”
- A5. We agree that units are necessary and we will add them to Table 1. Furthermore, as stated in the answer to comment 4, we will include detailed information on all the variables used. Regarding your question on permafrost, the zonation index shows to what degree permafrost exists only in the most favourable conditions or nearly everywhere.
- C6. [\[Page 3, line 30\]](#): “How many RS-cells is 0.05degree?”
- A6. Dear Referee, throughout the entire paper, we make reference to rectangular LONLAT grids. In particular, on a global level, 0.05° correspond to a grid with 7200 × 3600 griddcells with side length of ~5.5 km. This translates into 1400000 (2800 × 500) griddcells for North America, and 1760000 (4400 × 400) griddcells for Eurasia. The numbers for the 0.5° grid are the same divided by 100.

- C7. “Also add all abbreviations in [Table 1](#)”.
- A7. We did not understand this comment, as abbreviations are already present in Table 1. We will, however, improve the caption for the table.
- C8. [\[Page 4, line 3\]](#): “Of course both data sets are highly correlated, but more interesting is to see the anomalies”.
- A8. You are right. To deal with this aspect, we made a full analysis of the differences in results due to the use of the two datasets. The findings are already reported in the supplementary material. However, due to the restricted amount of anomalies, the core results regarding transitions zones are essentially the same.
- C9. [\[Page 5, line 13\]](#): “just call EV environmental variables. These changes will highly improve the reading”.
- A9. We will implement your suggestion to improve readability and reduce abbreviations.
- C10. [\[Page 6, line 1\]](#): “we associate every grid cell? Which grid cell is this the 0.05 degree or the RS-grid cell?”
- A10. Dear referee, in practical terms, there is only one geographical grid used throughout the analysis to which all variables (including tree cover) refer to. It is a rectangular LONLAT grid with  $0.5^\circ$  resolution. At every location (what we call gridcell, indicated by its longitude and latitude) we associate the values of all the environmental variables for that specific location and, at this particular step, the value given by the classification.
- C11. [\[Page 7, line 3\]](#): “Here you start referring to the table by not interpreting what we see but only saying that we have uncertainty bands. I think that you first need to introduce what we see highest explained variance is found for NA\_ E. (In the text this is mentioned as NAE. . . please just use the whole name, you also do this later on with Eastern North Eurasia for instance). And that this differs per region etc. Then you should also make a column with average values for all data. A question I do have is if the differences in explaining variance per region are dependent on the range of the environmental variables. With a larger range you would expect a large explaining variance.”
- A11. Dear Referee, as the answer to this particular comment is somewhat lengthy, we will structure it in points.
- A11.1 We agree that starting this section with a description of the findings would be an improvement, and we will implement this change in the paper.
- A11.2 Furthermore, as already agreed, we will minimise the use of abbreviations in the entire paper, to improve clarity and ease of reading.

- A11.3 We think that adding the average result for every variable would not be relevant, as there are clearly differences within the four regions. These differences would not be apparent from the average that will be consistently be a low number. However, we will add the average per variable per region, so that the distinction between regions will still be clear.
- A11.4 Regarding your question on a larger range of variables, we are not sure whether the question relates to the number of variables, or the spanned range of values for the single variables, so we will provide an answer to both interpretations. I) There are some factors that we could not include in the analysis, as stated in the discussion, such as the role of grazers (or other disturbances), and the role of nutrients. However such data are either not available, or their role is still under discussion. Furthermore, to improve our results, new variables must have a strong regional effect, and this effect must not be connected with the one of the variables already considered. For the same reason, the GAM results using all 8 variables, or only the 6 used in latter part of the paper, are very similar, and introducing new variables does not improve the results. Henceforth, we assume that the improvement of additional variables could only be minor. II) The case of larger ranges for the variables would only make sense when considering a larger geographical range. This would at the same time increase the extent of the biome analysed, including areas at mid-latitudes or at more than 70N. Doing so would introduce different plant species and vegetation controls, resulting into a different problem entirely that would, most likely, require a revised set of variables. Thus, it is difficult to make predictions on the outcome of such analysis, but we hypothesise that the increased complexity would not benefit the explained variance.
- C12. “Results-vs Discussion: I realize that the above section should not be too much discussion. From [page 7, lines 13–15 to page 8, lines 1–20](#) onwards you have a mix of results and discussion. These should be separated and should have a new section in the discussion chapter.”
- A12. Dear referee, we understand your point and we will try to implement your suggestion, separating findings and interpretations.
- C13. “Phase-space results: Again take more time to introduce [Fig2a](#). Is it not only the phase-space but also the KDE? Also mention in the text what these intersections are. Mention what the colours are and how we need to interpret [Fig2a](#). After that introduce [Fig2b](#).”
- A13. Thank you for your valuable suggestion, we will introduce and explain the figures more carefully, improving the captions as well.
- C14. “I also like to see a correlation matrix how the different EVs are correlated with each other. It is now unclear why you show Eastern North Eurasia with the combination of the two and why not a combination of other variables. I think that MAR and mean\_TMIN are highly correlated as they are placed on one line, meaning overlap of information.”
- A14. Following your comment, we will include a correlation matrix in the supplementary material, as we think it would distract the reader from the flow of information. Regarding the figure, it is only

meant as an example of the fact that within the regions, for some variables, e.g., precipitation and minimum temperature in Eastern North Eurasia, it is possible to find a clear separation between the three vegetation states (regardless of correlations), whereas for some other pairs, this separation is not clear and we find intersections. Hence, the choice of Northern Eurasia with those specific variables was aimed at exemplifying this point with a figure. We take your point that this information is not clearly conveyed. Hence, we will specify it within the manuscript.

- C15. “3.3: I do not understand the part at [page 9, lines 5–10](#). I can see that you are interested in grid cells having similar EVs but not similar tree cover. However I am confused how to read [Table 3](#), why is that you have four columns? If you mention a number of classes ([page 9, line 8](#)), what do you mean?”
- A15. We included Table 3 as a summary of the possible vegetation states found during the analysis. However, from your comment we realise that it causes confusion due to its structure. For this reason, we will reshape it, including only three columns and making it clear that they correspond only to the possible monostable, bistable, and fire-disturbed vegetation states. The classes we refer to at [page 9, line 8](#) are the 19 classes reported in Table 4, i.e. the classes that allow for bistable states. To make sure this sentence does not cause confusion, we will rephrase it to explicitly mention it.
- C16. “It is very interesting to see how these data are clustered. I have problems with reading the different colours in the legend. Also some symbols have a black line and others not, but unclear if this relates to the fire or non-fire disturbed states or does it relate to single stable vs bistable data points? Can you also see some spatial patterns of data which have the same bistability, but now currently in a different mode?”
- A16. We found hard to retrieve a colourblind-safe colour-scheme with eight colours. However, all the symbols should have the same structure and only different colours. We will try to improve clarity and increase the size of the legend markers. Regarding your second question, with our setup it is only possible to detect states with bistability when they are in a different mode.
- C17. “[Figure 4](#): It is Silvermans test (two words).”
- A17. You are right. The typo is due to the name of the package used for the implementation, which is the one the plot refers to. We will correct this for consistency and correctness. Thank you for your comment.
- C18. “Treeless state: I agree with your statement that tree cover below 20% is difficult to measure with RS. Therefore I have my doubts about the results of [Fig. 4](#). Why is it that you use in that detail the tree cover fractions below 20%? What do you want to show with these figures. There is not much text about [figure 4](#), so is it needed or can you directly introduce [figure 5](#). Although for this figure, the same holds for the data  $<20\%$ .”

- A18. Dear Referee, we understand your confusion about this. We will clarify better in the text and in the captions the meaning of these figures. To answer your questions, the plot in [figure 4](#) and the Silverman's test are meant only to show that there is a clear separation between the modes. And this can be clearly seen when looking at the decrease in frequency happening around 20% tree cover (on both sides). The unsuitability of the MODIS tree cover fraction product below 20% comes into play only when trying to make fine assessments at high resolution. However, in this particular case, only the generic distribution is important, i.e., the fact that there is a peak below 20% tree cover. This information is reliable, as if the tree cover would have been higher, it would have been measured with higher precision by the RS instruments, hence we can conclude that the peak is present. On the other hand, [figure 5](#) is related to the internal variability of the tree cover fraction dataset. The modal peaks are in fact more spread apart than what internal variability alone could cause, making them significant. Again, the precise distribution below 20% tree cover is not extremely important, only the fact that it is below such threshold.
- C19. "Do not understand your statement on [page 14, line 6](#); what kind of feedbacks? You didnt study this, so why is it that they might or might not play a role?"
- A19. We are referring to the main feedbacks happening between vegetation, environmental variables, and climate. However, we understand your point that we did not discuss them. For this reason, we will include details of the main feedbacks within the introduction and within the expanded description of the environmental variables. So that we can make a clearer reference to them.
- C20. "I found the discussion on N-cycling, decomposition, fertilisation a bit too much detail in comparison to the work you have presented. You have now linked it to soil type, and if you would be more interested in Nitrogen then you could have used modelled maps from DGVMs (as LPJ-Guess) or use maps or soil organic matter. I think that there are more important things to discuss, for instance why individual set of EVs are different between the regions, why fire is that important, which regions are more sensitive to a change in temperature then others, or a change in permafrost depth etc. etc. So keep the discussion more related to your own findings."
- A20. Dear Referee, this part of the discussion was intended to show that there are other factors playing a role in the boreal forest biome. However, we will implement your suggestion by reducing its extent and by expanding the discussion related to our findings. Additionally, we will move here part of the discussion presented in the GAMs results section.