We thank the reviewer for their comments. Our responses below are in bold.

Specific comments

I think the most critical point the authors should address is the lack of a scientific "discussion". There is no single reference to other works in Section 3. For instance, how do these results compare with Kim and Bae (2021)? Are these results far away from similar works with CMIP5 like Rahimi et al. (2020)? Here I only refer to those citations in the Introduction, but perhaps there exists more literature that can be discussed. Keeping together "Results and Discussion" is possible, but in this case I think Section 3 only includes a description of the outcome of the author's own analyses.

As requested, a longer Discussion section is included which refers to a these additional papers and others.

Also, besides using references for the discussion section, I think the authors should revise the use of references throughout the manuscript. For instance in the Introduction there are no references about CMIP6 or the Paris climate targets. Also when the authors say in line 60 that the scheme has had many alterations, they could provide a list of some publications in parentheses with as e.g..

We have included these suggested references and expanded our reference list for the paper.

Is Table 1 and the modifications described in lines 62-65 the same as in Peel et al. (2007)? Or are these the author's own modifications to what Peel et al. (2007) do? Please also consider including a citation in the caption of Table 1 if this is indeed taken from the reference.

These classifications follow Peel et al. (2007) exactly. Although this paper is cited directly prior to the table, it will also now be referenced in the table caption.

Authors should motivate better their model selection process. This is especially important since the 6 chosen models include repeated model components. How different are, for instance, CanESM5 and CanESM5-CanOE in terms of simulated (atmospheric variables)
monthly precipitation and temperature values? What are the "data management" reasons that lead to these 6 models being selected over other models?

These models were chosen based on the requirements for our analysis. Models needed to provide a "land" (sftlf) mask, and to output the required fields (monthly-mean temperature and precipitation) over each year of the 250 year runs, in a single file.

Along with the previous comment, please include a table with the specifications of the data used. Since the model output data is a critical component of this study, it is important that some characteristics (e.g., spatial resolution, time step, citation to technical paper) can be readily seen and compared in a table.

A table with these key characteristics of each model will be included in the revised manuscript.

Please discuss possible shortcomings of downscaling model output to a finer grid (0.5°), in case this was done when the model output is at a coarser scale. This is why a table with some specifications of the data could be useful.

We now include the following text: “The model output data is typically at a coarser resolution than the underlying 0.5° climatology. The anomaly corrected fields therefore contain spatial variability that is solely due to the underlying climatology at scales which are not resolved by a model. This also implies that the diagnosed changes in bioclimatic types (which are dependent on the model anomalies) tend to be somewhat smoother at these finer spatial scales”.

Please consider providing maps like those in Appendix A but using the "streamlined" version.

We include maps for the streamlined classification in the Appendix, as suggested.

In multiple occasions the word "significant" is used. I think that with the data the authors have it should be possible to perform some statistical significance tests. I think such tests could increase the strength of the results. Without the tests, please consider alternating with synonyms like "substantial", "marked", "large" or alike.

As requested, “significant" is only used to refer to statistical significance in the revised paper. The suggested synonyms are used elsewhere.

Technical details

Please consider re-organising some paragraphs in the Introduction. Paragraph starting on line 35 may fit better after the description of the classification systems. As a side note, this paragraph does not mention any references about CMIP6.

We have revised this paragraph to include references to CMIP6, as suggested.

I would recommend to move the sentence about previous applications of KG (line 43) to the end of Section 2.1.

We agree, this change has been implemented.

Abstract contains references. Are they urgently required?
Good point. These references have been removed from the abstract.

Consider prefixing "ensemble mean" with "multimodel", because some times "ensemble" could be referring to a group of runs or data.

We agree that this change improves clarity and we have therefore implemented it as suggested.

Please include spaces between numbers and their units, and use units in exponential notation (e.g., cm month⁻¹)

Done.

I think authors should mention vegetation in some places. At least when explaining the KG classification system on line 32.

As suggested, we have expanded on the description of the KG classification system to give examples of the vegetation types that it suggests for particular climates.

Perhaps Section 2.4.1 should actually be a subsection of or follow the Section on the traditional KG scheme (Section 2.1).

Done. We agree that following the traditional scheme is a more logical position for the streamlined scheme, so this change has been made.

Please expand on what the anomaly corrections are.

The following text has been added to clarify this: “Model outputs are anomaly corrected to agree with the observational over the period 1901-1931. This is done by calculating anomalies relative to that period for each model and then adding these anomalies to the observational climatology.

Consider using "averaging" instead of "meaning".

We prefer to continue to use “meaning” as it more accurately describes the averaging process used.

L3: "hinder" -> maybe "limit" is an alternative. Many detailed assessments can be made in spite of inter-model spread.

As suggested, “hinder” has been replaced with “limit”.

L4: why capitalisation in "Earth System Models".

Earth System Models is now only capitalised in its first use, subsequently it has been replaced with ESM.

L4: remove "very".

Done.

L5: "will" -> "would".

Done.
L21: Is it correct "regional areas"?

**Changed to “areas”**.

L31: missing umlaut in Köppen name.

**Corrected**.

L43: maybe this "most popular" should be referenced. Otherwise "popular" should suffice.

**Changed to “highly popular”**.

review "more intuitive".

**Changed to “intuitive”**.

L52: review this sentence.

**Changed from**:

"Furthermore, we utilise the KG system as an exploratory technique for understanding CMIP6 model output, the KG classification scheme has been previously applied to CMIP5 data to evaluate simulations (Phillips and Bonfils, 2015)"

**to**:

"We utilise the KG system as an exploratory technique to visualize CMIP6 model output. The KG classification scheme has been previously applied to CMIP5 data to evaluate simulations (Phillips and Bonfils, 2015)"

L60: check citation is \textcite, not \parencite.

**Fixed**.

L72: please use active voice: "we do not expect".

**Done**.

L76: check if "correctly" can be replaced by "following observations".

**Changed to “of observational data”**.

L80: is it above pre-industrial levels or above the reference period 1901--1931?

**Changed to reference period**.

- L81: remove "model" after "CMIP6".

**Done**.

- L90: review this sentence.

**Changed from**
“A key goal of bioclimatic classifications is to illustrate climate change in a way that is more intuitive for many people.”

To

“A key goal of bioclimatic classifications is to illustrate climate change in a way that is intuitive.”

- L110: remove period after "Figure 1".
  
  Done.

- L116: (and Fig. 1 and 2 captions) consider "averaging process" instead of "meaning".

  See above. We prefer “meaning” as it is less ambiguous.

- L117: 'lagging' as in having lower values?

  “Lagging” as in following the same trend as the individual models but behind them in time and with reduced variability, this has been clarified in the revised manuscript.

- L119, L120: check use of "correctly", that can always be related to "according to observation-based KG".

  Now refers to “observational “.

- L129: check "real world climate".

  Changed to “observed climate data”.

- L134: use cross-referencing with the appendices (e.g., Appendix A).

  Done.

- L135: "shows" -> "suggests". "will be" -> "could be".

  As suggested, these changes have been made.

- Fig. 3 caption: maybe place "reference period" before CMIP6 ssp585 runs.

  A semicolon has been added for clarity:

  “Anomaly plot of ensemble mean CMIP6 ssp585 runs for; the reference period, 1.5K, 2K, and 4K of global warming with the traditional Köppen-Geiger classification system applied.”

- L162: "will" -> "could".

  Done.