Dear Chris Folberth,

Thank you for your careful reading and comments on our manuscript. Following your remaining concerns, we substantially and carefully revised the logic, structure, and terminologies used in the manuscript. Virtual trees, the construction of observational benchmarks from these, and the comparison with land surface simulations via the two metrics for each benchmark are, we hope, now much clearer to the reader. This resulted in substantial changes throughout the entire manuscript and its supplementary material. All sections were restructured and revised for logic, clarity. References in text to specific results in the tables, figures, and supplement in support of the discussion have been added. Figures 4-8, clarifying the construction of virtual tree benchmarks and the estimation of skill metrics, were revised; all figure captions were reviewed for clarity and accuracy. The changes in the results table you spotted occurred when we transferred the document from Word to LaTex; we regret this error, and are thankful you noticed the mistake for us to correct in revision. All reported results have been rechecked against the raw data to ensure that the transfer between text editors did not result in other errors.

We hope this revision addresses your remaining concerns with the manuscript. Thank you for your consideration.

Kind regards,
Jina Jeong on behalf of the author team

---

1) L80: Reference to Figure 1: In Fig. 1 you first mention the virtual trees, which have so far not been defined. The caption needs to ensure that the figure can be understood. Add a brief statement on the virtual tree concept and where more information can be found in the manuscript.
We agree. The information was added in short to the caption of Fig. 1.

In addition, the term "maximum virtual tree" is not explained anywhere and only used in Fig. 3 besides here.
We agree. All 'maximum virtual tree' in the text have changed to 'largest virtual tree'.

2) L145-L170: Same issue as above: Fig. 3 is referenced, which requires understanding the concept of virtual trees but virtual trees have not been introduced to the reader and there is no information where their definition can be obtained.

Following your comment we realized that Fig. 3 served too many roles at the same time. This issue was solved by drawing a new figure (Fig. 3) to give an example of ITRDB datasets and referring to Fig. 4 only to explain the challenges and solutions.

3) Caption of Fig. 3: "maximum virtual tree" not clear.

As mentioned earlier, all changed to 'largest virtual tree'.

4) Section 2.2: I recommend to insert a separate (sub-)section on the virtual trees or at least mention the term in the section title, which would make it easier to point readers to this. I.e., rename section 2.2 to e.g. "...and addressing them using a concept of virtual trees" and/or split this section into (a) challenges and (b) the solutions proposed in this manuscript.

We followed this recommendation. The section for "Challenges of using ITRDB data as a long-term benchmark" (section 2.2) was separated from the section "Solutions for the challenges and virtual trees". Furthermore a bullet list was used to enhance the presentation of the virtual trees.

5) L201/L216 etc.: remove "across many sites". An ideal model would have this behavior across all sites.

All occurrences were removed (L207, L222, L234, and L251).

6) Section 2.3 (ii) and Fig. 5: How is the cut-off time period (last ten decades) determined? Section 2.3 (iii) has a thirty year time period and it is mentioned that this may be arbitrary.

It is arbitrary as described in section 2.4 (iii). Text was added to explain that this threshold may vary for different regions in the world (L218-221).

7) L373: I don't understand the reference to Fig. 9 here. Wouldn't Tab. 1 be more appropriate to show the 88 test cases?

The figure was suggested by the referee during the second revision to show the benchmark examples with European biomass network datasets. The figure has now been moved to Supplementary.

8) L380: 72% is not correct anymore as apparently results for site SOB have changed from 0.001 to -0.001 since the last version of the manuscript, which would increase the rate to 77%.

Thank you for noticing this. It was a mistake that happened during changing platforms. The sign was corrected. And after further checks against the raw data, some additional small inconsistencies in the table were found. All were corrected during this revision.

9) L386: Aren't there additional sites that have a positive change?

You are right. Those additional sites (HD2, SOB, and SOR) showed only marginal positive differences therefore there was no further explanation. A sentence was added (L385-386) to clarify the issue.

10) Figure 7 legend: "Simulated biggest output" not clear

We could not find the corresponding phrase in the caption, but 'biggest class' in Fig 5 to 7 was changed to 'largest diameter class' for consistency.

11) Supplement L348: Table S3 does not provide information on "issues of the ITRDB"

You are right. That reference was removed.

12) Supplement L384: "longer" -> "older"?

applied.

13) Supplement L388: "XX mm"?

It was fixed to 1 mm.

14) Supplement L412: "biggest simulation" not clear -> "largest simulated diameter class"?

It was changed following your suggestion.