

Response to Anonymous Referee #1

black = referee comments

blue = authors answers

General Note from the Authors:

We thank the reviewers for their constructive and helpful comments, which have contributed to improving the database, manuscript and supplementary information.

Note that during the revision, we identified two calculation errors. The SoEF extrapolation was implemented over five rather than three years. The CRAFT calculation used “EuFor-estimate” instead of “EuFor-harmonized” harvest data. Further, we updated the forest area harmonization process, which affected the countries Spain, Latvia, Finland, Slovakia and Slovenia. This slightly changed the results but not their overall interpretation. All analyses and corresponding results have been updated accordingly.

Dear authors,

I have reviewed the study “Consistent and Harmonized Forest Statistics for Forest-use Intensity Analyses Across Europe (2000-2023)”. This manuscript introduces the EuFor database, an open-access compilation of European forest statistics. Using this database, the authors compute four indicators of forest-use intensity (harvest per area, per growing stock, per net annual increment, and per potential NPP) and systematically assess how data uncertainties affect trend interpretation. My comments are as follows.

- 1.** Section 2.2.3 uses deviation factors based on the JRC dataset to adjust EuFor-reported forest areas to the “total forest” definition. However, the JRC data themselves are harmonised products with inherent uncertainties. The manuscript does not discuss how uncertainties in the JRC reference data propagate into EuFor-harmonised area estimates, I recommend the authors add more contents on it.

Thank you for pointing this out. We agree that uncertainties in the JRC data would propagate into the EuFor-harmonised estimates if the harmonised JRC estimates were used directly. However, EuFor uses only the NFI reference forest area data, rather than the JRC harmonized estimates for 2020. As described in the JRC paper, these reference data were provided by the NFI institutions themselves, using a common reference definition. Thus, the differences we find between our independently collected NFI data and the JRC reference data reflect differences between reported NFI datasets themselves rather than uncertainty introduced by the JRC harmonization.

This distinction was not sufficiently clear in the manuscript, and we have now clarified this in Section 2.2.3. lines 374-380.

- 2.** In Section 2.3.2, NPPpot is kept constant at 2018 values for 2019-2020 and at 2020 levels for 2021-2023. Given that NPPpot is climate-sensitive, this constant extrapolation may introduce systematic bias, please add more contents on it.

We acknowledge that keeping NPPpot constant for recent years may introduce uncertainty into the harvest/NPPpot indicator, given its climate sensitivity. Unfortunately, at the time of data compilation, no newer dataset on potential NPP was available to us. We have therefore added a sensitivity analysis using low- and high-NPPpot assumptions, based on the mean of the five lowest and five highest NPPpot values available for each region, respectively, and then compared these with the constant NPPpot assumption used in the main text.

Th sensitivity analysis shows that the magnitude of change between the reference (2000-2005) and latest (2018-2023) periods, along with some temporal pattern classifications, depends on the assumptions used to extrapolate NPPpot. Nevertheless, the resulting spatial patterns, overall direction of intensity change, and agreement in forest-use intensity change across the four indicators remain robust.

We nevertheless acknowledge this as a limitation and have now clarified in the main text that updated NPPpot estimates should be used in future applications where harvest per NPPpot is used as an indicator. Given that the primary purpose of this paper is to describe the database and to demonstrate NPPpot-based harvest intensity as an auxiliary indicator that can be calculated from the database, rather than to provide an in-depth analysis of single forest-use intensity indicators, we consider the assumption appropriate for the present application.

We have added text and figures regarding the sensitivity test of the NPPpot extrapolation assumptions in the text in the following places:

- *Methods, Section 2.3.2, lines 533-536*
- *Results and Discussion, Section 3.2, lines 781-786.*
- *SI in Section S2.2*

- 3.** Section 2.2.2 describes a selection process where one “data stream” per country/indicator is chosen among multiple candidates (different NUTS+ levels, forest classes, sources). Although two prioritization criteria are given, the actual decisions for each country are not documented. I request that a supplementary table be added, listing for each country which candidate streams were considered and the reasons for selecting or rejecting them.

Thank you for pointing this out. We agree that the selection process would benefit from more detailed documentation to improve transparency and reproducibility.

We have therefore added an additional XLSX file (Supplementary Dataset SD2) documenting the data stream selection for each country.

We have also revised Section 2.2.2 in lines 340-355 to describe the selection procedure more explicitly and to indicate where decisions between multiple available data streams were required.

In the process of reviewing the selection, we identified and corrected an error in the summary Table 2, where the forest area data stream for Lithuania was incorrectly listed as stocked forest instead of total forest.

- 4.** The “flag_calc” column is mentioned as a tool for assessing data provenance (lines 820828), but no aggregated reliability metric is provided.

We thank the reviewer for this suggestion. Reviewer #2 raised a similar point in comment #2. We have completely revised the flagging system and replaced the former single “flag_calc” column with dedicated provenance and processing columns for forest area, harvest, and growing stock. These columns document the data source and aggregation, the underlying primary data, reconstruction methods and interpolation/extrapolation source years, harmonization and correction factors, estimated components, and, for growing stock, CRAFT input data and model performance. This provides substantially more detailed information on the provenance and processing of individual observations.

We did not introduce a single aggregated reliability metric, as this would require subjective weighting of the different harmonization and time-series reconstruction steps and of the various correction and expansion factors. Instead, the revised flagging system allows users to select and filter observations according to their specific requirements, for example by excluding reconstructed values, selecting primary-data observations, or filtering CRAFT-derived estimates based on model performance.

The new flagging system is introduced in the main text in lines 250-258. The underlying procedures and processing steps represented by the individual flags are described in Sections 2.2.2-2.2.5 and summarized in the new column-description tables provided in Tables S3-S5. In addition, a statement for best practice is included in the Result and Discussion Section 3.3 in lines 906-913.

- 5.** Although the paper primarily emphasises the construction and utility of this database, I would like the authors to explicitly summarise its innovativeness at the methodological level.

We thank the reviewer for this suggestion. We have added a statement to the conclusion highlighting the methodological contribution of the study in lines 952-954.