

Response to Anonymous Referee #2

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.

Comments to the Author

This manuscript presents the EuFor-reported and EuFor-harmonized datasets, which compile and harmonize forest area, growing stock, harvest, and increment statistics for 38 European countries over 2000-2023, with additional historical observations where available. The authors integrate heterogeneous national forest inventories, forest censuses, State of Europe’s Forests statistics, FAOSTAT, and spatial datasets to produce annual national and subnational time series. The dataset addresses an important need for spatially and temporally consistent forest statistics and has clear potential for forest-use intensity, carbon-budget, and land-use analyses. The manuscript is well structured, and the effort involved in collecting and documenting these fragmented sources is substantial. However, issues remain regarding quality control, the reliability of reconstructed data, validation independence, methodological transparency, and uncertainty characterization.

General comments:

- 1.** The quality-control procedures require strengthening. Much of the source information appears to have been manually transcribed or translated from heterogeneous reports, yet the manuscript mainly describes basic cleaning and visual inspection. A systematic QC workflow should include duplicate detection, range and sign checks, treatment of missing and zero values, verification of unit conversions, detection of temporal discontinuities, national-subnational consistency checks, and stock-harvest-increment plausibility checks. The

manuscript should clearly describe these procedures and summarize their outcomes.

We thank the reviewer for this important suggestion. We have strengthened and clarified the description of our quality-control procedures in the revised manuscript. It now includes dedicated QC sections describing duplicate, missing-value, unit-conversion, aggregation, and national-subnational consistency checks in Section 2.1.1 (lines 204-209) and Section 2.2.2 (lines 334-338).

We did not perform a separate stock-harvest-increment plausibility check on the collected increment data because increment data were not included in the harmonization due to insufficient coverage, as described in Section 2.2 in lines 260-266. However, the stock-harvest-increment plausibility check is actually the main purpose of applying CRAFT to calculate the growing stock time-series. CRAFT calculates the annual increment required to reconcile all available data on biomass stocks and outflows (mortality and harvest), based on growth parameters optimized within empirically derived bounds. The stock-harvest-increment plausibility constraint is therefore the basis of EuFor-harmonized growing stock estimates. The RMSE of the CRAFT model hence provides a measure of how inconsistent or implausible the reported input data streams are in relation to each other (shown in Figure S4 and described in lines 457-461).

- 2.** A substantial proportion of the harmonized dataset is reconstructed rather than directly observed. Approximately one third of the area and harvest series is interpolated or extrapolated, while most growing-stock values are generated using the CRAFT model; additional records involve allocation between forest-use classes. The current calculation flags provide useful provenance but are insufficient to communicate reliability. The authors should provide additional indicators, such as reconstruction method, extrapolation length, temporal distance from the nearest observation, number of observations supporting model calibration, and relevant model-fit diagnostics.

(similar to Reviewer 1, point 4)

We thank the reviewer for this suggestion. Reviewer #1 raised a similar point in point 4. 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.

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.

- 3.** The manuscript currently lacks sufficiently independent validation of the harmonized and reconstructed data. The comparison with State of Europe's Forests (SoEF) cannot be considered fully independent because SoEF data or related information are used at several stages of the harmonization, including area adjustment, growing-stock expansion, and harvest normalization. Moreover, the underlying national sources may overlap. The current agreement is therefore partly expected by construction and should be described as an external consistency assessment rather than independent validation. The authors should conduct out-of-sample validation by withholding a subset of reported observations and evaluating how accurately the interpolation, extrapolation, and modelling procedures reconstruct them. Appropriate metrics, including bias, MAE, and RMSE, should be reported. Where genuinely independent sources are available, additional comparisons would further strengthen the credibility of the dataset.

We thank the reviewer for this important point. Reviewer #3 raised a similar point in comment #2. We agree that the comparison with SoEF should not be considered as an independent validation. We have revised the corresponding sections accordingly and now consistently refer to these analyses as assessments of data source agreement. These sections are intended to compare the EuFor-harmonized database with the only other available pan-European database claiming some degree of internal consistency and comparability across countries. Independent validation with an external dataset would be valuable. However, we are not aware of a consistently available independent dataset at the European scale that could serve this purpose. We deliberately did not use remote-sensing products as a reference, as our objective was to construct a statistics-based database from national and international forest statistics. We have made these points explicit in lines 582-589.

We also agree that out-of-sample validation would be appropriate where sufficient independent observations are available. However, the fragmented nature and limited temporal coverage of the underlying national data do not provide enough observations for meaningful hold-out validation. For countries with only two available NFIs, for example, withholding one observation would leave insufficient data to assess the temporal reconstruction reliably.

We nevertheless consider the reconstructed time series presented in EuFor-harmonized valuable for the intended application, as they provide a consistent and complete representation of changes in forest-use intensity across Europe, where temporal consistency required some trade-off against the accuracy of individual annual estimates.

However, the revised flagging system now provides detailed information on the provenance and processing of each harmonized entry (as described in point 2 above), allowing users who require higher accuracy to filter and select the data accordingly.

- 4.** The CRAFT model is central to the dataset because it generates most harmonized growing-stock values, but its implementation and evaluation require clearer documentation. The authors should report the parameter-estimation procedure, parameter constraints and initialization, treatment of regions with sparse observations, natural-loss assumptions, convergence criteria, and regional model-performance statistics. Temporal or spatial holdout validation should be provided.

We thank the reviewer for this suggestion. We have expanded the documentation of the CRAFT implementation in the SI, including the parameter-estimation procedure, parameter constraints and initialization, natural-loss assumptions, and convergence criteria (lines 67-85).

Regional model performance is now provided as RMSE for each region in the EuFor-harmonized table (Table S5) and is also shown spatially in Figure S4. It is further discussed in the main text (lines 492-501).

As discussed above, we did not perform temporal or spatial holdout validation because the fragmented and limited temporal coverage of the underlying observations does not provide sufficient independent data for meaningful holdout tests, particularly for regions with only a small number of observations.

- 5.** Several harmonization choices require clearer justification and sensitivity assessment. These include the calculation of forest-area harmonization factors when multiple overlapping years exist, treatment of regions without matching observations, assumption that the factors remain constant over time, selection of one source series primarily on the basis of completeness, and unweighted averaging of EuFor-estimate and SoEF-normalized harvest series. Some reported area factors are relatively extreme, suggesting possible differences in definitions or spatial boundaries. The authors should examine these cases and evaluate the sensitivity of the final dataset to alternative harmonization choices.

We thank the reviewer for this comment. As several distinct points are raised in this comment, we address them step by step below.

Regarding the selection between multiple possible data streams per country:

- We have revised the text on the selection process and now include a more explicit explanation and additional information in lines 340-356 in the main text and in the additional Supplementary Dataset DS2.*

Regarding harmonization of forest area:

- We revised the forest-area harmonization and identified several cases where matching was based on extrapolated values. We therefore introduced an additional reconstruction step using original national forest-area data to extend values to the subnational level. This was possible for Finland, Latvia, Slovakia, and Slovenia. For Spain, we replaced the JRC-based matching with SoEF data because the available JRC match relied on extrapolated values. Subsequently, we were able to match JRC data using at least interpolated values, which we consider sufficient given the incremental and relatively small changes in forest area at this temporal scale. The new approach is now explained in Section 2.2.3 lines 368-369.*
- There are no multiple overlapping JRC years for individual regions, as the JRC data provides only one reference year per region. We have made this clear in line 377-378.*
- For Estonia, Great Britain, and Spain, where SoEF was used in lack of sufficient JRC data, multiple country-level reference years were available. Here, we used the average deviation across matched years to avoid imposing the SoEF trend on the primary EuFor data. We have made this transparent in lines 395-398.*
- Harmonization factors were kept constant over time, assuming that differences primarily reflect systematic differences in forest-area definitions. This approach also preserves the temporal dynamics of the underlying primary data rather than imposing additional temporal variation through the harmonization process. We have added this explanation in lines 382-385 & 398-399.*
- Large harmonization factors are primarily associated with differences between the selected forest-area class and the total forest definition used by the JRC. We added Figure S3 showing the selected forest-area classes alongside the harmonization factors to make these differences transparent.*

Regarding harmonization of harvest:

We revised the description of the harvest harmonization procedure in section 2.2.4 to clarify the construction of the EuFor-estimate, SoEF-normalized, and EuFor-harmonized series:

- The EuFor-estimate is based primarily on NFI and national census data, while the SoEF-normalized series uses national SoEF totals complemented by FAOSTAT trends. We refer now to Pilli et al (2025) which showed that both sources have known limitations, including temporal gaps, reporting uncertainties, and potential differences in harvest levels and dynamics.*
- We acknowledge that the equal weighting of the EuFor-estimate and SoEF-normalized series is a methodological choice. We chose a simple unweighted average because no sufficiently independent reference dataset or standardized metric was available to objectively determine source-specific weights across countries. We included text on this in lines 439-442. This provides a transparent and consistent approach while avoiding subjective assumptions about the relative reliability of the two sources.*
- To assess the implications of this choice, we retain both the EuFor-estimate and SoEF-normalized series alongside the final EuFor-harmonized series in the database. The agreement between EuFor-harmonized and SoEF is also shown in Supplementary Figure S7, allowing readers to compare the two harvest estimates.*

- 6.** The uncertainty analysis currently represents cross-source variation rather than the full uncertainty of the harmonized estimates. It does not appear to propagate uncertainty associated with source observations, conversion factors, spatial mapping, interpolation, extrapolation, CRAFT parameters, or manual classification. The term “uncertainty envelope” should therefore be qualified or replaced with “cross-source sensitivity range.”

We thank the reviewer for this suggestion. Following the additional comments from Reviewer 3, we have revised this analysis and renamed it as a cross-source sensitivity test to assess robustness of forest-use intensity. The revised approach is described in Section 2.4.2 and presented in the main text in Section 3.2 (lines 797-825).

Specific comments:

Thank you for the detailed remarks; we have corrected them accordingly.

1. Page 18, line 477: The periods 2000-2005 and 2018-2023 each contain six years when both endpoints are included. They should not be described as “five-year averages.”

We have changed this to “six-year-averages” in line 545.

1. Page 24, line 614: A decline from 311 to 183 corresponds to a reduction of approximately 41%, not 59%. Please correct the percentage.

We have adapted the numbers according to the updated run and checked for correct relative change numbers in line 670.

1. Page 24, line 623: Total European harvest is reported as approximately 2.3-2.8 million $\text{m}^3 \text{yr}^{-1}$. These values appear to represent area-normalized harvest intensity, with units of $\text{m}^3 \text{ha}^{-1} \text{yr}^{-1}$, rather than total harvest, which is several hundred million $\text{m}^3 \text{yr}^{-1}$. Please correct the description, values, and units.

We have corrected the units in lines 686-687.