# Incorporating geospatial data halves chemistry-based timber harvest area predictions



Citation for the original published paper (version of record):
Smith, C., Truszkowski, J., Saunders, J. et al (2026). Incorporating geospatial data halves chemistry-based timber harvest area predictions. Environmental Research Communications, 8(6).
http://dx.doi.org/10.1088/2515-7620/ae7304

# Incorporating geospatial data halves chemistry-based timber harvest area predictions

Environmental Research Communications



OPEN ACCESS

RECEIVED
24 December 2025

REVISED
29 April 2026

ACCEPTED FOR PUBLICATION
26 May 2026

PUBLISHED
22 June 2026



PAPER

# Incorporating geospatial data halves chemistry-based timber harvest area predictions

C C Smith[1,*], J Truszkowski[1,2], J Saunders[1] and V Deklerck[1,3]

1 World Forest ID, Washington, DC, United States of America
2 Chalmers University of Technology, Gothenburg, Sweden
3 Meise Botanic Garden, Meise, Belgium
* Author to whom any correspondence should be addressed.

**E-mail:** charlottesmith0308@outlook.com

**Keywords:** timber origin, sustainable forestry, stable-isotope ratio analysis, harvest location modelling, geospatial analysis

Supplementary material for this article is available online

## Abstract

International demand for forest products is exerting increasing pressure on ecosystems, biodiversity, and carbon storage. In response, international regulatory frameworks are tightening, with many now demanding precise declarations of timber origin. Implementing such regulations will require robust methodologies for scientifically verifying claimed harvest locations of timber. Advances in chemistry-based spatial tools have dramatically improved our capacity to identify timber harvest location. Yet these tools incorporate few environmental constraints that determine whether a location is ecologically viable or operationally feasible for commercial harvesting. We explore how integrating open-access geospatial datasets of key environmental attributes—land cover, topography, and tree cover fraction—can be harnessed to improve existing models. For 323 birch samples in a ~3.9 million km$^2$ study area in Eastern Europe, we found incorporating land-cover restrictions can halve the harvestable range within the predicted harvest area without compromising the accuracy of predictions. Almost 80% of true harvest locations were correctly identified when removing all non-forest land cover from the predicted harvest area. However, imposing high tree cover fraction thresholds substantially reduced correct predictions. These results highlight the benefits of embedding more sophisticated ecological viability and harvest feasibility data into timber harvest location determination models for strengthening the evidentiary basis for legality verification under emerging policy regimes. More broadly, they demonstrate the value of widely available environmental datasets in advancing scientifically rigorous, transparent, and enforceable methods for timber-traceability and regulatory oversight. As global deforestation continues to accelerate biodiversity loss and climate change, ensuring that timber-trade regulations are scientifically enforceable is critical.

## 1. Intro

The international trade in forest products is facilitated by the harvest of 2 billion m$^3$ of timber annually (FAO 2024), with extraction exerting profound pressure on ecosystems, biodiversity, and carbon storage. Market growth has also been accompanied by challenges in ensuring the legality and sustainability of timber supply chains (Brancalion *et al* 2018, Wiedenhoeft *et al* 2019, ITTO 2024). Despite a range of international legislation and trade sanctions seeking to curb deforestation and promote sustainable sourcing (e.g. the United States Lacey Act 2008 Amendment, European Union Timber Regulation 2010, European Union Deforestation Regulation 2023, Australian Illegal Logging Prohibition Act 2012), a shocking quantity of illegal timber still makes its way into the market (INTERPOL Environmental Crime Programme 2012, Kleinschmit *et al* 2016, Nellemann *et al* 2016, Alison *et al* 2022). In an effort to safeguard vulnerable forest ecosystems, laws increasingly require traceability and/or harvest location declarations (European Union 2023). However, such claims are of limited value if they cannot be independently

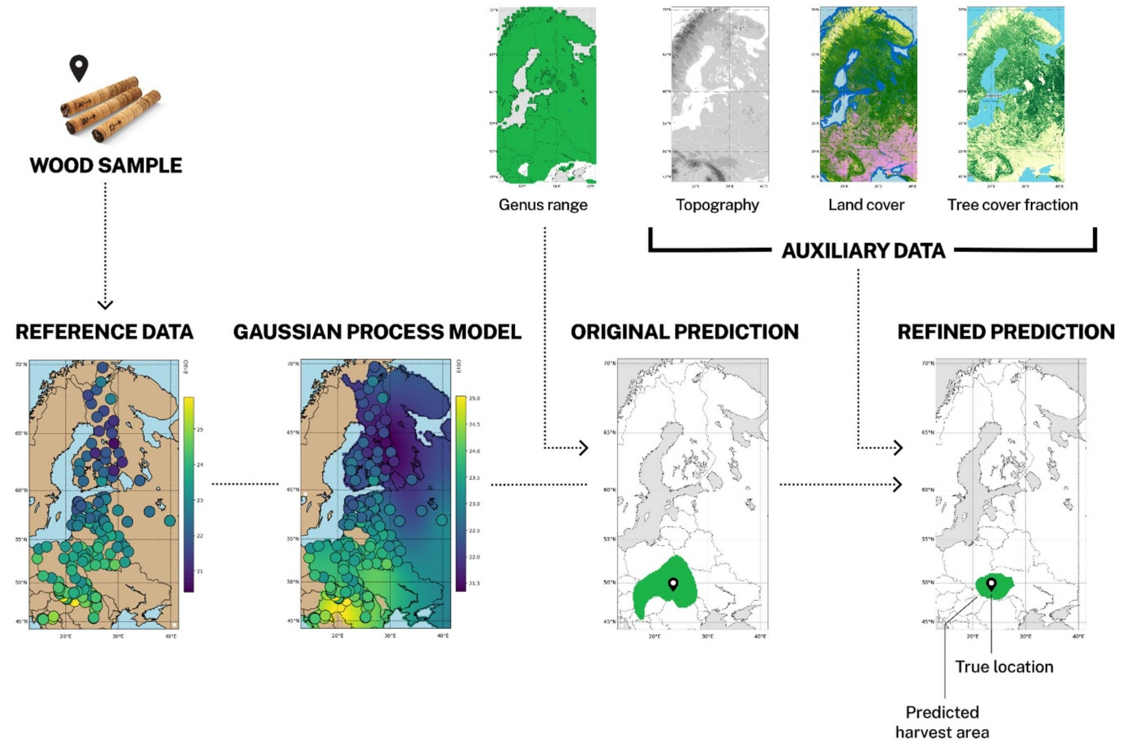


**Figure 1.** Conceptual demonstration of the 'harvestable range' method. (_a) Wood samples with their associated geolocations become (b) chemical reference data, which are used by Gaussian process models to generate (_c) a set of expected chemical values for every location in the study area. Previous studies have used only (d) the theoretical range of the focal genus to limit the spatial extent of (e) the predicted harvest area, which is calculated from the likelihood of observing the chemical values from a sample at a location. This study introduces additional auxiliary datasets—(f) topography, (g) land cover, and (h) tree cover fraction—to (i) identify regions within predicted harvest areas that are suitable for commercial harvest.

verified. Evidence suggests that 20%–50% of traded timber and forest products are sold with fraudulent claims (Wiedenhoeft *et al* 2019, World Forest ID 2025). The effectiveness of these regulations is therefore fundamentally dependent on the availability of robust, science-based methods that can reliably determine the location of timber harvest.

Chemical analysis has emerged as one of the most promising scientific approaches for determining timber harvest location (Watkinson *et al* 2022, Štulc *et al* 2024, Gori *et al* 2025), with recent studies showing a combination of stable isotope ratio analysis (SIRA) and multi-element analysis (MEA) to be highly effective (Mortier *et al* 2024, Aguzzoni *et al* 2025). Chemical classification models have demonstrated the ability of these techniques to discriminate between fixed, distant regions (e.g. Boeschoten *et al* 2023, Aguzzoni *et al* 2025). However, this approach may oversimplify the problem by failing to account for the continuous spatial gradients seen in stable isotope ratios and chemical elements (Mortier *et al* 2024). More recent studies have harnessed these spatial trends using Gaussian Process models, combining them with Bayesian frameworks to accurately predict timber harvest location within large continuous landscapes (figure 1; Mortier *et al* 2024, Truszkowski *et al* 2025).

To date, the effectiveness of the Bayesian approach to harvest location modelling has been inhibited by the use of coarse geographic filters to limit potential harvest areas—notably a reliance on generalised species distribution maps. These maps represent the outer limits of a species' geographical range, often overestimating actual presence by encompassing extensive areas where the focal species does not occur (e.g. urban areas). It is possible to derive more precise species ranges from global tree-occurrence datasets. However this would require a substantial computational investment and data currently contains large geographic gaps and regional biases, even for key commercial taxa (Serra-Diaz *et al* 2017). In conservation biology, the area of habitat (AOH) approach makes use of datasets pertaining to habitat suitability and elevation to refine theoretical species distributions (Rondinini *et al* 2011, Ficetola *et al* 2015, Brooks *et al* 2019). This approach consistently demonstrates that only a fraction of a species' theoretical distribution constitutes suitable habitat (Lumbierres *et al* 2022, Ficetola *et al* 2015), with predicted occurrence accuracy improving for 92% of species (Rondinini *et al* 2011). Given the large distributional ranges

**Table 1.** Samples by country (*Betula* spp.).

| Country | No. of Reference samples |
|---|---|
| Belarus | 35 |
| Croatia | 20 |
| Estonia | 50 |
| Finland | 30 |
| Hungary | 9 |
| Latvia | 56 |
| Poland | 20 |
| Romania | 22 |
| Russia | 34 |
| Slovakia | 21 |
| Ukraine | 60 |
| TOTAL | 357 |

of tree taxa and the particular effectiveness of the AOH approach for wide-ranging species (Lumbierres *et al* 2022), applying AOH-like refinements to timber harvest location models presents an important opportunity to improve the precision and realism of predictions.

In this study, we introduce the concept of 'harvestable range' as a means of generating more realistic predictions of timber harvest location by excluding areas that are either ecologically unsuitable or not viable for commercial harvesting. We use high-resolution auxiliary datasets, such as global land cover and digital elevation, to apply the harvestable range concept to the harvest origin predictions of the Gaussian Process (GP) model developed by Mortier *et al* (2024). Focusing on birch (*Betula* spp.), a genus of high economic importance in Eastern Europe (figure 1), we address two questions: (1) which datasets have the greatest influence on harvestable range? (2) Can these datasets be used to refine timber harvest area predictions by excluding areas that are unsuitable for commercial harvesting?

## 2. Methods

### 2.1. Reference samples

Our geo-referenced sample data consisted of 357 birch samples across 11 countries (table 1, figure 1), including an additional 55 samples compared to Mortier *et al* (2024). All samples were collected in 2022, with the exception of Russian samples which were obtained from a collection conducted in 2005 as the region's ongoing conflict prevented new sample collection. Each sample was measured for 23 chemical values: six stable isotope ratios (see SI) and 14 elements using x-ray fluorescence spectroscopy (XEPOS, SPECTRO analytical instruments; see SI). Full details of sample collection protocols and chemical analysis techniques used are available in Mortier *et al* (2024).

### 2.2. Study area and geospatial model

The study region was defined by a 10 km resolution grid of regularly spaced locations covering the spatial range of our sample data—spanning approximately 10° to 59° degrees longitude and 43° to 71° degrees latitude. Locations >300 km from a reference sample or outside the maximum combined range extent of *Betula pendula* and *Betula pubescens* (Caudullo *et al*) were excluded; leaving a total of 38 789 locations. All 357 reference samples were used to generate the study area and to train the GP regression models outlined below. We ran a sensitivity analysis using a 5 km resolution grid but the marginal improvement in some results did not justify the significant increase in compute cost (see supplementary information).

GP regression models were used to generate an expected value and uncertainty (variance) for each chemical variable at every location in the study area. These models assume that measurements at nearby locations are correlated and control how quickly values change with latitude and longitude. Further details on the regression models can be found in the supplementary information and the full mathematical breakdown is available in Mortier *et al* (2024).

The expected value maps were used to calculate the likelihood of observing the chemical values of a test sample at every location within the study area. Chemical measurements were treated independently, and Bayes' theorem was used to convert the likelihoods into a map of posterior probabilities across the study area. For each sample we calculate three metrics: (1) the 'predicted harvest area' identified as the smallest set of locations that contains 95% of the total probability mass; (2) whether the true harvest location falls within the predicted harvest area; and (3) the 'Area Scored Higher than the true location' (ASH), which calculates the total area that the model gives a higher likelihood than the true harvest location of the sample (Truszkowski *et al* 2025). We performed 5-fold cross-validation and report results as the median and interquartile range (IQR) over all samples from all cross-validation folds; with the exception of whether the true harvest location falls within the predicted harvest area, which is reported as the percentage of samples across all cross-validation folds. See supplementary information for further details.

### 2.3. Harvestable range analysis

We identified areas that are unsuitable for commercial timber harvest using three key environmental constraints (land cover, topography, and tree cover fraction) and evaluated their effect on model outputs. These constraints were chosen based on their ecological relevance for the focal genus (*Betula*) and study area (Eastern Europe), and their influence on the practicality of timber extraction; socioeconomic drivers of harvesting are beyond the scope of this study. Geospatial datasets were selected based on their coverage of the study area, spatial resolution, temporality, and availability in Google Earth engine (GEE). For each dataset, we used GEE to create binary exclusion masks with increasingly stringent thresholds. Harvestable range was defined as the area within each study area location (10 × 10 km grid cell) that remained after applying these masks.

The environmental constraints assessed in this study do not modify the Bayesian posterior probabilities calculated by the GP model as the likelihood of a sample's chemical composition matching a location's predicted chemical value is independent of the proportion of forest cover or other environmental constraint in a given location (10 × 10 km grid cell). The environmental constraints are used to identify the areas within the 'chemically plausible' set of locations where commercial timber harvest is ecologically viable and logistically feasible.

We explored the impact of applying each constraint threshold to model predictions in three ways: (1) by calculating the harvestable range within the predicted harvest area (cells containing 95% of posterior probability); (2) by examining if the true harvest locations are in the harvestable range remaining within the predicted harvest area; and (3) by calculating the harvestable range within the ASH result (locations that have a higher posterior probability than the true harvest location). These accuracy measures were selected because they capture aspects of model performance that are impacted by applying exclusion criteria: ASH quantifies the overall spatial precision of predictions, while the inclusion/exclusion of the true harvest location assesses whether the model result remains operationally useful after constraints are applied.

We excluded the 34 Russian samples collected in 2005 from the harvestable range analysis to avoid introducing a confounding temporal effect when evaluating the performance of exclusion masks. The samples were retained for Gaussian Process modelling, as their chemical measurements remain valid for modelling spatial chemical patterns. Including time series analysis is beyond the scope of this study but is a recommended next step for future research (see Discussion).

#### *2.3.1. Land cover*

For land cover we used the ESA WorldCover 2021 dataset (European Space Agency, 10 m resolution), which offers a high-resolution, globally consistent land cover classification with an overall global accuracy of 76.7%. We derived binary exclusion masks for the ten non-forest land cover classes, and a combined mask to exclude all non-forest land cover simultaneously. We evaluated each class individually to assess which exerted the greatest influence on prediction outcomes. The ESA World Cover 2021 definition of tree cover is 'any geographic area dominated by trees with a cover of 10% or more.' It includes 'areas planted with trees for afforestation purposes and plantations' and 'tree covered areas seasonally or permanently flooded with fresh water' but excludes mangroves.

#### *2.3.2. Topography*

Topographic factors can influence both accessibility for harvest equipment and habitat suitability for forest growth. In this northern boreal–subarctic study area, elevation serves as a useful proxy constraint for commercial feasibility as at higher altitudes both growth rates and mean tree height decline (Prévosto *et al* 1999, Coomes and Allen 2007, Miyajima and Takahashi 2007), and access for commercial harvesting equipment becomes increasingly limited in the absence of road infrastructure at these elevations. However, we acknowledge that elevation is not a universal constraint for harvestable range—for example, road networks may enable access to high-elevation sites, and at lower latitudes, warmer conditions can shift productive growing ranges to higher elevations. Similarly, slope gradient affects the economic viability, safety, and environmental performance of harvest operations, with operational limits ranging from 30% to 60% depending on machine type, with slopes over 50% requiring highly specialised equipment (Visser and Stampfer 2015). We used the COPERNICUS GLO-30 digital elevation model (~30 m resolution; Copernicus) to create elevation masks of ⩾1400 m, ⩾1600 m, ⩾1800 m, and ⩾2000 m. This dataset has an absolute vertical accuracy of <4 m and an absolute horizontal accuracy of <6 m. The GEE function ee.Terrain.slope() was used to derive slope gradient from the same dataset and slope masks were created for ⩾40°, ⩾50°, ⩾60°, and ⩾70. These thresholds reflect the declining probability of harvestable *Betula* spp. stand occurrence and increasing inaccessibility at higher altitudes and on steeper slopes.

#### *2.3.3. Tree cover fraction*

Using the COPERNICUS Land Cover (2019) tree-cover fraction band (100 m resolution, 80.3% thematic accuracy) we exclude areas with <50%, <20%, <10%, and <5% forest cover as a measure of forest occurrence and as a structural forest density proxy (where lower canopy cover correlates with reduced commercial interest). Land-cover type is a strong indicator of whether forest occurs to sufficient extent and density to support commercial harvesting.

### 2.4. Scalability

The approach described here is highly scalable. The GP model analysis took ~2 h, while calculating the posterior probabilities took approximately 6 min. The Earth Engine calculations used to derive constraint coverage per location required a combined compute time of ~520 Earth engine compute unit hours, but with a parallel approach completed in approximately 19 real-world minutes. Importantly, these constraint layers only need to be generated once for a given study area and can then be reused indefinitely for the analysis of individual sample results. Compute times will increase with larger study areas, but the parallelised architecture and one-off nature of the constraint layer generation mean that the method would remain practical for larger regions.

## 3. Results

### 3.1. Unrefined model results

The total study area comprised 38 789 locations for which harvest probability was predicted (figure 4(b)). Each location covered an area of approximately 100 km$^2$, with the total study area equating to 3.9 million km$^2$. Predicted harvest area in the unrefined model (i.e. without applying the exclusion criteria assessed in this study) ranged from 21 700 km$^2$ to 1.2 million km$^2$, with a median predicted area of 441 399 km$^2$ (IQR: 247 599–618 199)—11.3% of the total study area considered by the model. The true harvest location of 83.6% of reference samples fell within their predicted harvest area. Across all samples from all cross validation folds, median ASH was 55 300 km$^2$ (IQR: 10 150–228 699).

### 3.2. Extent of environmental constraints within the study area

Within the study area, 1.7 million km$^2$ (43.1%) of land was classified as non-forest (figure 4(c)), with the three dominant land-cover types being cropland (0.6 million km$^2$), grassland (0.6 million km$^2$), and water (0.2 million km$^2$). The true harvest location of 335 (93.8%) samples were classified as forest. Areas with a tree cover fraction less than 5% accounted for 16.1% (0.6 million km$^2$) of the total land area, increasing to 48.1% (1.9 million km$^2$) when considering tree cover fraction <50%. High-elevation areas were limited, with just 0.7% (0.03 million km$^2$) of the landscape above 1400 m, the most stringent elevation threshold applied in this study. Similarly, steep slopes exceeding 40% represented only 0.2% (0.01 million km$^2$) of the total area.

### 3.3. Impact of environmental constraints on predicted harvest area

Land cover constraints substantially reduced the harvestable range within predicted harvest areas across samples (figure 2(a)). Agricultural lands represented the largest single category, eliminating on average 26.8% (IQR: 10.5%–31.5%; median and IQR over all samples from all cross-validation folds, hereafter) of harvestable range within predicted harvest areas (figure 2(a)). Grasslands were comparable, accounting for a median reduction in harvestable range of 18.2% (IQR: 16.4%–19.4%; figure 2(a)). Water bodies, wetlands, and urban areas had smaller effects, with a median reduction of 2.6% (IQR: 1.0%–6.6%), 1.0% (IQR: 0.5%–2.2%), and 1.6% (IQR: 0.8%–2.2%) of harvestable range, respectively (figure 2(a)).

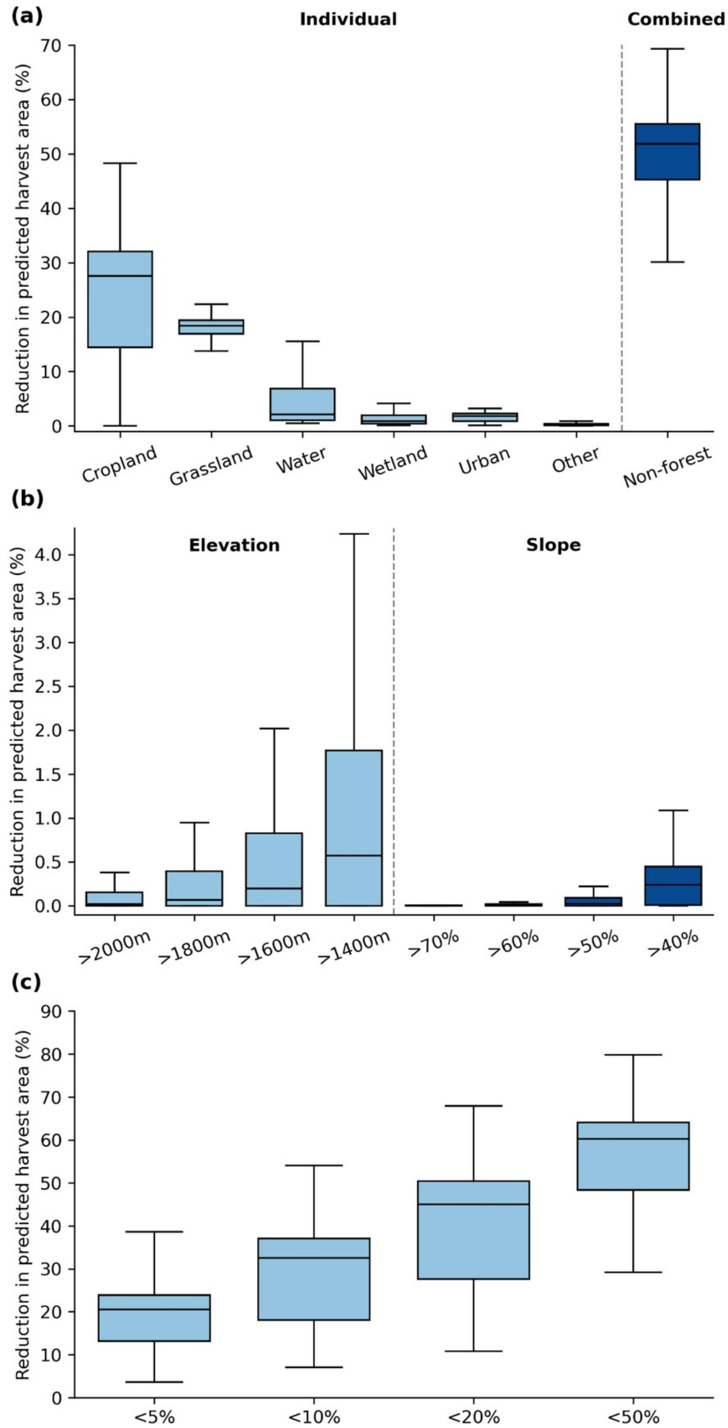


**Figure 2.** Impact of environmental constraints at different thresholds on predicted harvest area. The percentage of land within predicted harvest areas that is excluded when applying environmental constraints; relative to the unrefined model results. (a) The impact of excluding individual land cover classes, and the combined effect of excluding all non-forest classes (ESA WorldCover 2022). 'Other' is a combination of all non-forest classes with <1% coverage in each prediction grid cell. (b) The impact of elevation and slope at different maximum exclusion thresholds, derived from the COPERNICUS GLO-30digital elevation model (Copernicus). (c) Impact of minimum tree cover fraction (COPERNICUS Land Cover, 2019). Boxes extend from the first to third quartile, with a median line. Whiskers extend to the farthest data point lying within 1.5*x* the inter-quartile range from the box.

The influence of other land cover types (shrubland, bare ground, snow, mangrove, lichen) was negligible, with a median combined reduction in harvestable range of 0.2% (IQR: 0.1%–0.4%) (figure 2(a)). Removing all non-forest land cover simultaneously resulted in a median reduction in the harvestable range within predicted harvest areas of 51.0% (IQR: 42.6%–55.2%; figure 2(a)). Excluding regions with a tree cover fraction of <5% reduced harvestable range within predicted harvest area by on average 19.4% (IQR: 10.1%–23.4%), increasing to 31.1% (IQR: 15.9%–36.5%) for <10% tree cover fraction,

**Table 2.** Impact of environmental constraints on predicted harvest area.

| Environmental Constraint | True harvest locations within harvestable range[a] | Harvestable range[a] within predicted harvest area; median (IQR) | Harvestable range[a] within area scored higher; median (IQR) |
|---|---|---|---|
| Unrefined model | 270 (83.6%) | 0.44 million km$^2$ (0.25–0.62) | 55 300 km$^2$ (10 150–228 699) |
| Non-forest cover | 259 (80.2%) | 0.21 million km$^2$ (0.13–0.31) | 24 972 km$^2$ (4793–113 738) |
| Slope >40% | 270 (83.6%) | 0.44 million km$^2$ (0.25–0.62) | 55 300 km$^2$ (10 150–228 699) |
| Elevation >1400 m | 270 (83.6%) | 0.44 million km$^2$ (0.25–0.61) | 55 100 km$^2$ (10 150–228 699) |
| Tree cover <50% | 233 (72.1%) | 0.18 million km$^2$ (0.11–0.25) | 25 435 km$^2$ (4832–98 887) |

[a] For the unrefined model, harvestable range is the entire cell area. For the exclusion masks it is the area of a cell remaining after the mask has been applied.

and 59.2% (IQR: 44.9%–63.6%) for <50% tree cover fraction (figure 2(c)). Even the most stringent topographic exclusions had limited impact on the harvestable range within predicted harvest areas with median reductions of 0.3% (IQR: 0.0%–1.6%) for elevations of >1400 m) and 0.2% (IQR: 0.0%–0.4%) for slopes >40% (figure 2(b)).

When excluding all non-forest land cover, the median reduction in the harvestable range within ASH was to 24 972 km2 (4793 km2–113 738 km$^2$), an average reduction of 50.7% (IQR: 44.4%–59.5%) relative to the unrefined model (figures 3(a) and (b)), and 259 (80.2%; percentage of samples across all cross validation folds) true harvest locations were correctly identified by the harvestable range within the predicted harvest areas (table 2). For the 12 samples located outside the forest mask, the median distance to the nearest forested pixel was just 20.7 m (IQR: 8.3 m–67.9 m). When applying the most stringent tree cover fraction criteria (excluding <50%) the harvestable range within ASH results saw a greater median reduction relative to the baseline model—54.9% (IQR: 43.1%–63.2%; figures 3(e) and (f)). However, just 72.1% of true harvest locations were correctly identified (table 2). At the most lenient tree cover fraction threshold (<5% excluded), 99.1% of samples were correctly identified (table 2). Of the 323 total samples, 3 (0.8%) were situated in areas with a tree cover fraction of <5%, 4 (1.1%) <10%, 10 (2.8%) <20%, and 43 (13.3%) <50% (table 2). Elevation and slope produced negligible changes, with median reduction in the harvestable range within ASH results of <0.01% even at the most stringent thresholds (>1400 m and >40% slope; figures 3(c) and (d)). Topography constraints had no impact on correct predictions as all true harvest locations were less than 1400 m elevation (max: 1137 m) and on less than 40% slope (max: 28%; table 2).

## 4. Discussion

Building upon frameworks developed for Eastern Europe (Mortier *et al* 2024), we present the first assessment of how geospatial datasets can be used to refine the results of spatially explicit timber harvest location models. We show that accounting for land cover classification reduces the total harvestable range within the study area by 43.1% (1.67 million km$^2$; figure 2), while increasing prediction accuracy by an average of 50.7% (reduction of harvestable range within ASH; figure 3). These findings demonstrate the importance of integrating auxiliary datasets to enhance the spatial specificity and reliability of predicted harvest results. In the following discussion, we outline how this approach contributes to recent methodological advances, the logical next steps for future model development, and the broader implications for policies designed to combat deforestation and the illegal timber trade.

### 4.1. Contributions to chemical tracing methodology

Recently developed spatial models incorporating geochemical variation across landscapes (e.g. Mortier *et al* 2024) represent a major advance over classification approaches (e.g. Boeschoten *et al* 2023) which—although effective for discriminating between discrete populations—are of limited use for determining harvest location across extended landscapes. Spatial models offer a more powerful approach for three key reasons. First, they provide a more transparent representation of provenance uncertainty by enabling probabilistic inference across continuous spatial grids, rather than between arbitrary political units where 'national' values have been extrapolated from geographically limited data. Second, they facilitate interpolation into inaccessible areas, such as conflict zones, allowing predictions to be made for unsampled

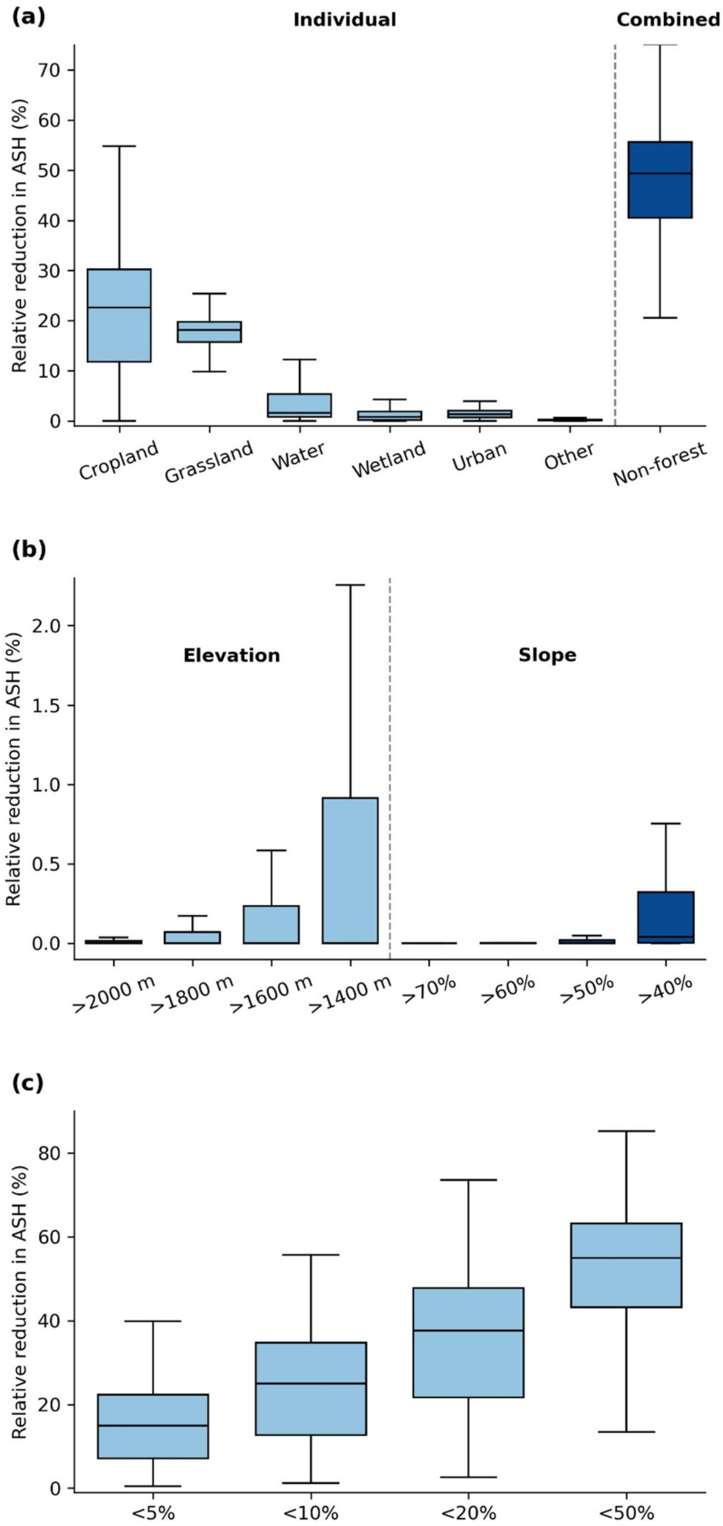


**Figure 3.** Effect of environmental constraints on area scored higher results. The relative reduction in harvestable range within the area scored higher (ASH) results when applying exclusion criteria for land cover, topography, and tree cover fraction. (a) The impact of excluding individual land cover classes, and the combined effect of excluding all non-forest classes (ESA WorldCover 2022).'Other' is a combination of all non-forest classes with <1% coverage in each prediction grid cell. (b) The impact of elevation and slope at different maximum exclusion thresholds, derived from the COPERNICUS GLO-30 digital elevation model (Copernicus). (c) Impact of minimum percentage tree cover thresholds, using the tree-cover fraction band from the COPERNICUS Land Cover dataset (2019). Boxes extend from the first to third quartile, with a median line. Whiskers extend to the farthest data point lying within 1.5x the inter-quartile range from the box.

regions. Limiting inference to areas within a confidence radius ensures that results remain fully supported by empirical data. Finally, the high-resolution probability surfaces of spatial models can be combined with auxiliary datasets to generate more ecologically and operationally meaningful predictions of harvest location—as demonstrated by the51% median reduction in harvestable range within predicted harvest areas observed in this study (figure 2).

Spatial models for determining timber harvest location have previously defined the maximum extent of harvestable range using generalised distributions that represent the geographic extremes of a genus range (Mortier *et al* 2024, Truszkowski *et al* 2025). This is problematic when predicting harvest location as the model may return areas that are chemically plausible but in actuality contain no commercially harvestable forest. This risks undermining the utility of spatial models for enforcing regulations by overstating the possible harvest area. The harvestable range approach developed here combats this by using a variety of environmental filters to delineate regions within predicted harvest areas where commercial timber extraction is both ecologically viable and logistically feasible. We found that this approach can substantially reduce the land considered harvestable within the predicted harvest area without compromising model accuracy (figure 3).

Land cover emerged as the most effective means of constraining harvestable range within predicted harvest area (figure 2). Excluding non-forest land cover halved the harvestable range within predicted harvest area (median 51%; figure 2), while still correctly identifying 80% of true harvest locations. More than 96% of true harvest locations were within the forest cover mask, and the 12 samples that were not, were less than 120 m from a pixel classified as forest (median, IQR: 20.7 m, 8.3 m–67.9 m). This close proximity suggests that the majority of omissions were due to a combination of sample collection at forest edges and habitat transition-zone misclassification in the land cover dataset—a common issue in remote sensing–based land cover classification (e.g. Xu *et al* 2024). In future, this could be mitigated by applying a small buffer to the forest mask, although this would be computationally expensive for large study areas such as this. Tree cover fraction showed similar improvements in ASH and predicted harvest area at the highest threshold (>50% tree cover fraction) but led to a decrease in the number of true harvest locations being retained (72%), illustrating the importance of selecting datasets and restriction thresholds that are appropriate for the focal genus.

### 4.2. Limitations and future research

We have presented a readily implementable advance in spatial modelling of timber harvest location and demonstrate the tangible improvements in reducing the harvestable range within predicted harvest area. Nonetheless, four important limitations remain which offer valuable direction for future research.

First, the land cover datasets used in this study represent a static snapshot in time, which, although temporally appropriate for the samples analysed here, may not reflect forest conditions at the time of harvest in other instances. Ensuring appropriate data selection for the specific scenario being addressed is therefore critical. It should also be noted that single-year data layers may misclassify areas undergoing forestry practices such as rotational clear-cutting (Alonso *et al* 2023). A next step for this approach is to integrate time series data and real-time deforestation alerts (e.g. Global Forest Watch; Hansen *et al* 2013) to facilitate highlighting locations showing evidence of harvesting activity within the typical harvest-to-market timeframe of the focal genus (Pithon *et al* 2012, Popkin 2016). Given regulations often require a declaration of harvest date (e.g. EUDR), this would offer an important advance for supply chain due diligence.

Second, topography was found to have negligible impact on harvestable range within predicted harvest area, as steep and high-altitude terrain represents just 0.2% and 0.7% of the study area, respectively (figure 2). However, in more mountainous regions, such as the Andes or Himalayas, it may have a much greater impact and thus should not be ruled out in future applications of this methodology. Furthermore, high elevations and steep terrain are known to function as ecological barriers and may exert similar influences on geochemical gradients (Chen *et al* 2024), potentially leading to larger differences in chemical composition across shorter straight-line distances than would otherwise be expected. Future research should explore whether incorporating topography when interpolating chemical gradients—rather than applying it post-hoc as a spatial filter—could enhance the representation of natural discontinuities and improve the predictive accuracy of harvest location models, particularly in mountainous regions.

Third, incorporating risk indicators, such as the emergence of new logging roads, accessibility, and patterns of land-use change (Di Lallo *et al* 2017, Berenguer *et al* 2021, Engert *et al* 2025), would allow models to weight areas according to their relative likelihood of recent or ongoing harvest activity. Embedding these indicators within a dynamic modelling architecture—capable of ingesting continually updated datasets on forest disturbance and infrastructure—would enable provenance assessments to evolve alongside the landscapes they represent. While introducing such a component would pose a substantial methodological challenge, it would constitute a major advance. Dynamic, risk-informed modelling frameworks would also provide a foundation for proactive monitoring of high-risk regions, aligning with recent calls to integrate spatial risk assessment into deforestation-free supply chain monitoring (Franca *et al* 2025). Ultimately, this integration would align traceability science more closely with the operational realities of timber governance, trade monitoring, and the practical requirements of due diligence.

Finally, modelling frameworks need to be extended to other key timber-producing regions. This study focuses on Eastern Europe because it is one of the few regions where robust, spatially explicit geochemical prediction models currently exist. While the methodological concept developed here could be applied to any forestry region, the specific constraints analysed in this study may transfer most directly to temperate and boreal forests where species compositions, growing conditions, and commercial harvest constraints are most similar. For example, adapting the method for North America would primarily require the selection of regionally appropriate datasets. However, the tropics, where deforestation pressures and illegal logging risks are most acute (Smith *et al* 2021, Santoro *et al* 2025), are likely to demand substantially different considerations. For example, the vast and continuous tracts of forest in the Amazon present very different accessibility constraints compared to Europe's fragmented forest landscapes (Senf and Seidl 2021, Ma *et al* 2023). Proximity to access routes such as roads and rivers is already being used to predict deforestation risk (Engert *et al* 2025), which suggests they could also serve as powerful predictors of harvestable range. Ecological and operational contexts also differ in subtler ways: elevation constraints, for example, may operate in the opposite direction at lower latitudes, where warmer, drier conditions push growing ranges into higher elevations, potentially requiring minimum elevation thresholds in addition to upper limits. Other landscapes may also benefit from the inclusion of additional datasets not considered here, such as tree height or biomass to further define which trees are actually harvestable. Developing new regional models will require careful selection and calibration of the datasets used to refine predicted harvest area, to ensure they accurately capture the realities of the sourcing landscape.

### 4.3. Implications for policy and regulatory compliance

Our findings demonstrate that spatial harvest location determination model predictions can be immediately and substantially improved through the inclusion of just a single auxiliary dataset: land-cover (figure 4). This simple refinement could greatly increase the plausibility of predicted harvest areas and strengthen the evidentiary reliability of harvest location assessments for timber legality verification. Compliance analysis often focuses on verifying whether declared harvest sites meet regulatory requirements (e.g. were deforested prior to a cut-off date). Yet such verification has limited evidentiary value if the claimed location itself cannot be substantiated. A more robust compliance framework could be built which integrates complementary lines of evidence, first verifying that the claimed harvest location is both chemically plausible and logistically feasible using the geochemical modelling approach refined in this study, then assessing if the location is legally permissible using temporal deforestation monitoring. This sequential logic from chemical plausibility, to logistical feasibility, to legal verification would create a coherent and testable chain of evidence linking a physical sample to its claimed source. Integrating these components within a single analytical framework would offer a more rigorous and transparent basis for due diligence, aligning with recent analyses that emphasise the need to connect harvest location determination directly with deforestation assessment to achieve meaningful traceability outcomes.

### 4.4. Conclusions

Reliable verification of timber harvest location is fundamental in enforcing sustainability and legality requirements in global trade. Where harvest area predictions are inaccurate or uncertain, there is a risk that illegally-harvested timber may enter supply chains undetected, undermining regulatory systems and certification schemes. By integrating 'harvestable range', our approach delivers more precise and defensible harvest area predictions, with direct implications for the implementation of timber regulations. With the necessary land-cover data and computational tools already available, this method could be readily implemented at scale to help address a key barrier to policy enforcement. As global deforestation continues to drive climate change and biodiversity loss, ensuring that these regulations are enforceable is critical. The rollout of deforestation-free trade policies may have been slowed by technical challenges, but the science for validating harvest origin is ready.

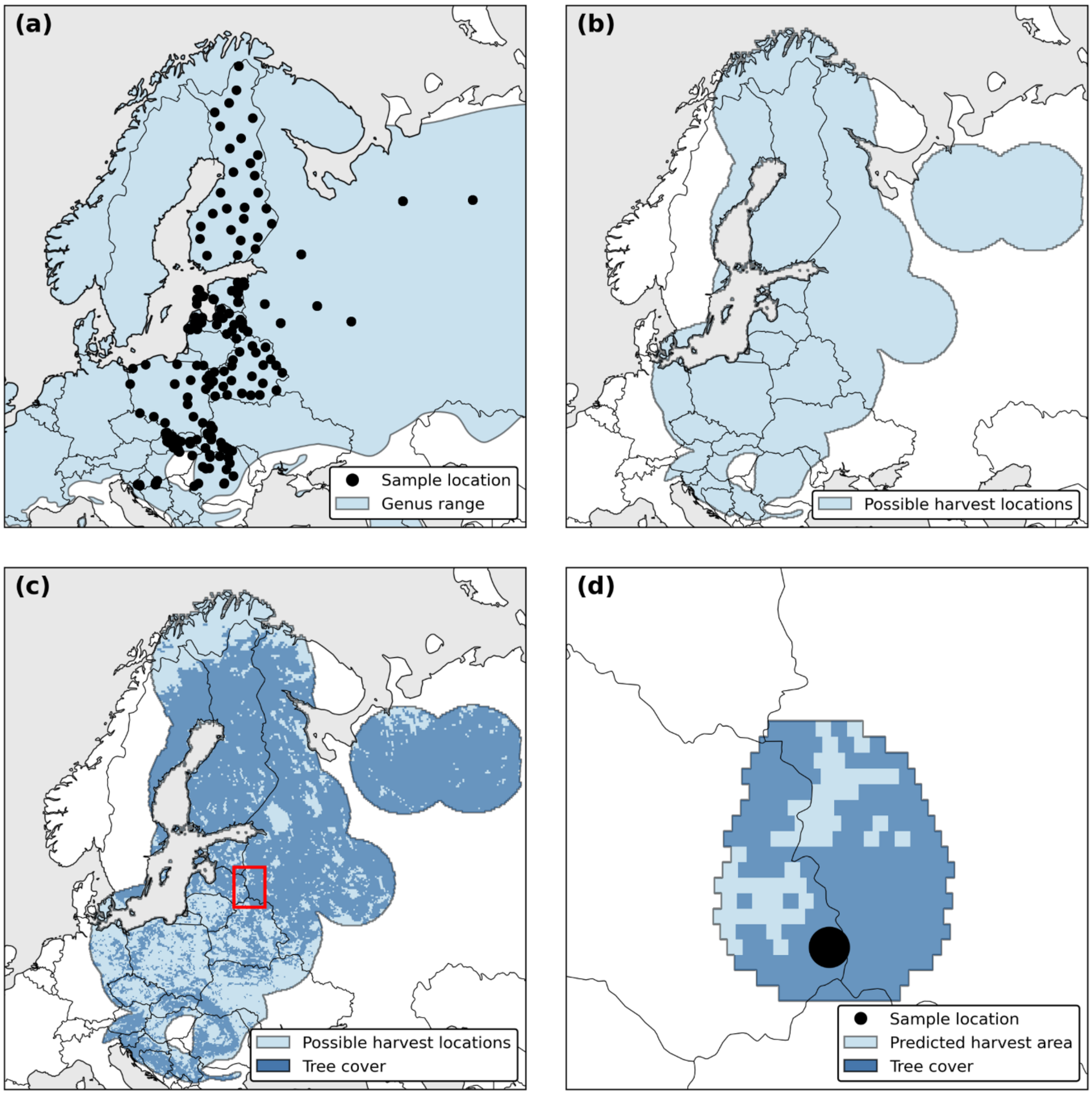


**Figure 4.** Spatial modelling workflow and example of prediction refinement. (a) Ground-truthed sample locations (black points) and the generalised genus distribution map for Betula (see Mortier *et al* 2024). (b) Study area consisting of 38 789, 10 × 10 km grid cells representing potential harvest locations. (c) Area classified as tree cover (dark blue) within the study area (light blue). The red box indicates the region shown in (d). (d) An example of the harvestable range within a predicted harvest area (dark blue), the area within the predicted harvest area that is excluded by the non-tree-cover mask (light blue), and the true harvest location (black marker). Sample locations are obfuscated for collector and landowner safety. Tree cover is classified by the 10 m resolution ESA WorldCover 2021 dataset, displayed here resampled to 1 km resolution; analysis was conducted at full resolution.

## Acknowledgment

This work was funded by World Forest ID. JT is supported by the Swedish Foundation for Strategic Environmental Research (MISTRA).

## Data availability statement

The data cannot be made publicly available upon publication because they contain sensitive personal information. The data that support the findings of this study are available upon reasonable request from the authors.

Model code developed by Mortier *et al* (2024) is available upon request at https://zenodo.org/records/10055203. Code developed for the analysis conducted in this study is available here: https://github.com/WorldForestID/wfid-priors-paper-public.

Supplementary Information available at https://doi.org/10.1088/2515-7620/ae7304/data1.

## Author contributions

C C Smith 0000-0002-3767-6587
Conceptualization (lead), Data curation (lead), Formal analysis (lead), Methodology (lead), Visualization (lead), Writing – original draft (lead), Writing – review & editing (lead)

## References

Aguzzoni A, Giammarchi F, Mundo I A, Voto G, Tonon G, Tirler W and Tomelleri E 2025 Tracing timber origin: geographic provenancing at regional scale with multielement and strontium isotope ratio analyses *For. Ecol. Manage.* **579** 122494
Alison H, Thiago H and Uehara K 2022 *Establishing Fair and Sustainable Forest Economies: Lessons Learned from Tackling Illegal Logging* (Royal Institute of International Affairs)
Alonso L, Rodríguez A, Picos J, Armesto J and Fassnacht F 2023 Challenges in automatic forest change reporting through land cover mapping *Forestry* **96** 155–69
Australian Government 2012 Illegal Logging Prohibition Act 2012 Federal Register of Legislation (available at: www.legislation.gov.au/Details/C2012A00166)
Berenguer E *et al*, 2021 Drivers and ecological impacts of deforestation and forest degradation
Boeschoten L E *et al* 2023 A new method for the timber tracing toolbox: applying multi-element analysis to determine wood origin *Environ. Res. Lett.* **18** 054001
Brancalion P H, de Almeida D R A, Vidal E, Molin P G, Sontag V E, Souza S E X F and Schulze M D 2018 Fake legal logging in the Brazilian Amazon *Sci. Adv.* **4** eaat1192
Brooks T M *et al* 2019 Measuring terrestrial area of habitat (AOH) and its utility for the IUCN red list *Trends Ecol. Evol.* **34** 977–86
Buchhorn M, Smets B, Bertels L, De Roo B, Lesiv M, Tsendbazar N -, E Herold and Fritz S 2019 *Copernicus Global Land Service: Land Cover 100m* **3**
Chen L *et al* 2024 Influence of mountain orientation on precipitation isotopes in the westerly belt of Eurasia *Glob. Planet Change.* **240** 104543
Coomes D A and Allen R B 2007 Effects of size, competition and altitude on tree growth *J. Ecol.* **95** 1084–97
Copernicus Copernicus DEM—Global and European Digital Elevation Model (COP-DEM), GLO-30 [data set] (https://doi.org/10.5270/ESA-c5d3d65)
Di Lallo G, Mundhenk P, Zamora López S E, Marchetti M and Köhl M 2017 REDD+: quick assessment of deforestation risk based on available data *Forests* **8** 29
Engert J E, Souza C M, Kleinschroth F, Bignoli D J, Costa S C P, Botelho J, Ishida F Y, Nursamsi I and Laurance W F 2025 Explosive growth of secondary roads is linked to widespread tropical deforestation *Curr. Biol.* **35** 1641–8
European Union 2010 Regulation (EU) No 995/2010 … timber and timber products on the market *Off. J. Eur. Union L* **295** 23–34
European Union 2023 Regulation (EU) 2023/1115 … commodities and products associated with deforestation and forest degradation *Off. J. Eur. Union L* **150** 206–63
FAO 2024 *Global Forest Products Facts and figure 2023* (FAO) (https://doi.org/10.4060/cd3650en)
Ficetola G F, Rondinini C, Bonardi A, Baisero D, Padoa-Schioppa E and Richardson D 2015 Habitat availability for amphibians and extinction threat: a global analysis *Divers. Distrib.* **21** 302–11
Franca C S S, Persson U M, Cardoso D R R S, Damasceno C, Ward R S and Souza C M 2025 Mapping the limits to timber traceability at origin: the case of Para, Brazil *One Earth* **8** 101447
Gori Y, Deklerck V and Camin F 2025 Stable isotope ratio analysis to determine the geographical origin of timber: a review *TrAC Trends Anal. Chem.* **193** 118448
Hansen M C *et al* 2013 High-resolution global maps of 21st-century forest cover change *Science* **342** 850–3
INTERPOL Environmental Crime Programme 2012 *Green Carbon, Black Trade: Illegal Logging, Tax Fraud and Laundering in the World's Tropical Forests* ed C Nellemann (UNEP/GRID-Arendal)
ITTO 2024 Annual report 2023 (International Tropical Timber Organization)
Kleinschmit D *et al* 2016 Illegal logging and related timber trade: dimensions, drivers, impacts and responses *Int. For. Rev.* **18** 4–27
Lumbierres M, Dahal P R, Soria C D, Di Marco M, Butchart S H M, Donald P F and Rondinini C 2022 Area of habitat maps for the world's terrestrial birds and mammals *Sci. Data* **9** 749
Ma J, Li J, Wu W and Liu J 2023 Global forest fragmentation change from 2000 to 2020 *Nat. Commun.* **14** 3752
Miyajima Y and Takahashi K 2007 Changes with altitude of the stand structure of temperate forests on Mount Norikura, central Japan *J. For. Res.* **12** 187–92
Mortier T *et al* 2024 A framework for tracing timber following the Ukraine invasion *Nat. Plants* **10** 390–401
Nellemann C, *et al* (eds) 2016 *The Rise of Environmental Crime—A Growing Threat to Natural Resources, Peace, Development and Security* (UNEP & RHIPTO)
Pithon S, Jubelin G, Guitet S and Gond V 2012 A statistical method for detecting logging-related canopy gaps using high-resolution optical remote sensing *Int J. Remote Sens.* **34** 700–11
Popkin G 2016 Satellite alerts track deforestation in real time *Nature* **530** 392–3
Prévosto B, Coquillard P and Gueugnot J 1999 Growth models of silver birch (*Betula pendula* Roth.) on two volcanic mountains in the French Massif Central *Ann. For. Sci.* **61** 391–9
Rondinini C *et al* 2011 Global habitat suitability models of terrestrial mammals *Phil. Trans. R. Soc. B* **366** 2633–41
Santoro A, Piras F and Yu Q 2025 Spatial analysis of deforestation in Indonesia in the period 1950–2017 and the role of protected areas *Biodivers. Conserv.* **34** 3119–45
Senf C and Seidl R 2021 Mapping the forest disturbance regimes of Europe *Nat. Sustain.* **4** 63–70
Serra-Diaz J M, Enquist B J, Maitner B S, Merow C and Svenning J-C 2017 Big data of tree species distributions: how big and how good? *For. Ecosyst.* **4** 30
Smith C C, Healey J R, Berenguer E, Young P J, Taylor B, Elias F, Espírito-Santo F and Barlow J 2021 Old-growth forest loss and secondary forest recovery across Amazonian countries *Environ. Res. Lett.* **16** 085009
Štulc A I *et al* 2024 Tracing the origin of wood at the regional scale with dendrochemical markers *Sci. Total Environ.* **957** 177640

Truszkowski J *et al* 2025 A probabilistic approach to estimating timber harvest location *Ecol. Appl.* **35** e3077

United States Congress (1900, as amended 2008) *Lacey Act* vol 16 (U.S.C) pp 3371–8

Visser R and Stampfer K 2015 Expanding ground-based harvesting onto steep terrain: a review *Croatian J. For. Eng.* **36** 321–31

Watkinson C J, Rees G O, Gwenael M C, Gasson P, Hofem S, Michely L and Boner M 2022 Stable isotope ratio analysis for the comparison of timber from two forest concessions in Gabon *Front. For. Glob. Change* **4** 650257

Wiedenhoeft A C, Simeone J, Smith A, Parker-Forney M, Soares R and Fishman A 2019 Fraud and misrepresentation in retail forest products exceeds U.S. forensic wood science capacity *PLoS One* **14** e0219917

World Forest ID 2025 WFID insight briefing (available at: https://learn.worldforestid.org/wp-content/uploads/2025/04/WFID_InsightBriefing_April_2025_v10b-3.pdf) (Accessed 26 November 2025)

Xu X, Li D, Liu H, Zhao G, Cui B, Yi Y, Yang W and Du J 2024 Comparative validation and misclassification diagnosis of 30-meter land cover datasets in China *Remote Sens.* **16** 4330

Zanaga D *et al* 2022 ESA WorldCover 10 m 2021 v200. (https://doi.org/10.5281/zenodo.7254221)