Advances in data modeling extend the reach of chemical timber testing

Chemical timber testing depends on known-origin reference samples, but new sample collections are not always possible where verification is most urgent. Two new papers show how modeling can extend the reach of existing data: by inferring target-genus chemistry from related genera and using environmental data to remove implausible harvest areas.

THE INSIGHT SERIES

World Forest ID's Insight series is designed to communicate the outputs of our longform research, by summarizing data snapshots and interim learning. All research is ultimately published in appropriate peer-reviewed journals.

THE DATA CHALLENGE

Scientific timber testing compares the chemistry of traded products with reference samples collected from trees of known origin, verifying whether they were harvested where claimed. A test is only possible when reference data covers the claimed harvest region, but reference data is not evenly available. In some high-risk timber-producing regions, new collections are limited by security conditions and contested access rights, as well as the cost, time, and practical difficulty of fieldwork.

These constraints often apply in the regions where independent origin verification is most needed, including areas linked to conflict, sanctions, or weak forest governance. This creates a practical challenge for due diligence: the places that most need scientific verification can also be the hardest places to build new reference data.

This Insight summarizes peer-reviewed research that addresses that constraint by making better use of existing data. Multi-genus modeling extends what known-origin samples can tell us across related genera, while growing range refinement makes predicted harvest areas smaller and more actionable. New reference samples remain essential, but this research makes chemical traceability more viable in data-scarce environments and more useful for regulators, certification schemes, and companies responding to traceability legislation such as the EU Deforestation Regulation (EUDR).

1. Truszkowski, J., Boeschoten, L., Mortier, T., et al. (2026). Chemical timber tracing: combining tree-genera information lowers reference data needs and makes harvest location identification more accurate. Annals of Forest Science, 83, 18.

2. Smith, C. C., Truszkowski, J., Saunders, J., & Deklerck, V. (2026). Incorporating geospatial data halves chemistry-based timber harvest area predictions. Environmental Research Communications.

3. Zanaga, D., Van De Kerchove, R., Daems, D., et al. (2022). ESA WorldCover 10 m 2021 v200 (Version v200) [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.7254221

1. Multi-genus modeling expands the range of testable timber products

Timber samples can be traced back to harvest location even where there is limited or no genus-specific reference data available, our new peer-reviewed research published in *Annals of Forest Science* shows[1]. Until now, verifying a timber product's origin claim has required reference samples matching both the declared genus and claimed harvest region.

Our study investigated four commercially traded genera: oak, beech, birch and pine, and showed that while each genus has distinct chemistry, chemical signals follow similar spatial patterns across the landscape. This allows reference samples from one genus to support provenance predictions for another.

This 'multi-genus' modeling approach outperforms single-genus models, improving harvest location accuracy by between 10% and 38%, and detection of misdeclared origins by up to 49%. Critically, it extracts more value from samples already in the collection, expanding the range of testable products without requiring new collection, and freeing resources for wider deployment of the science.

RESULTS - *Multi-genus predictions reduced distance to the true harvest location.*

Oak (Quercus spp.)	*Beech (Fagus spp.)*	*Birch (Betula spp.)*	*Pine (Pinus spp.)*
269 km closer	185 km closer	143 km closer	68 km closer

Based on 1,020 trees, 4 genera, 12 Eastern European countries

HISTORIC SAMPLES KEEP VERIFICATION POSSIBLE AS ACCESS SHIFTS

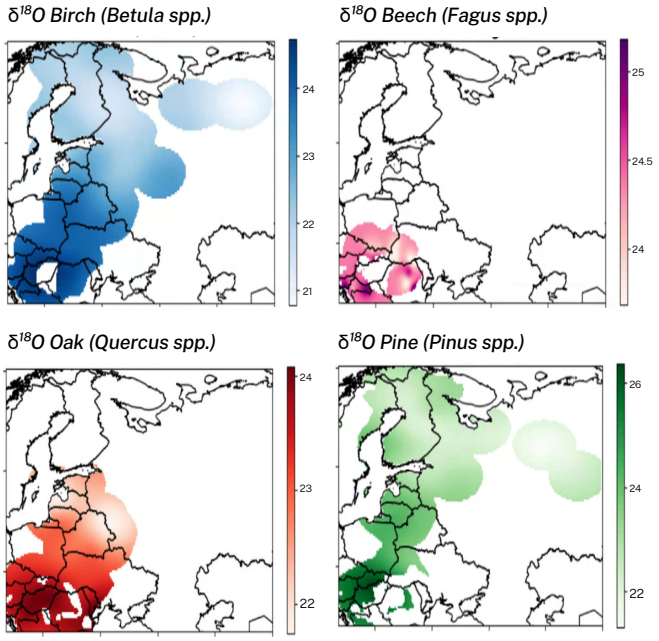
Physical inability to collect reference samples, rather than cost or time, is often the binding constraint. Reference data may exist for one genus but be urgently needed for another, while changing conditions on the ground make new collection impossible.

The regions where field collection is most constrained are often those where misdeclaration risk is highest and the need for verification most acute. For example, much of the Congo Basin, and sanctioned Russia and Belarus, are of particular interest to companies and governments wanting to verify timber origins, but are among the most challenging places to build reference data.

Multi-genus modeling fills that gap: when access to a location is lost, it uses historic reference samples from other genera in the area to support verification for the target genus.

FIGURE 1. — *Different genera, shared landscape inference*

Multi-genus modeling improves timber origin prediction by learning the shared chemical signature of a specific place, while accounting for chemical variation between genera.

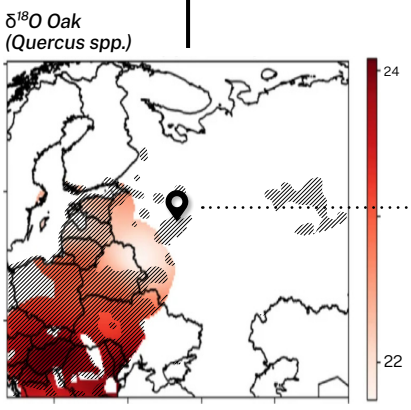


GENUS-SPECIFIC REFERENCE DATA

Each genus has distinct chemistry, but trees growing in the same landscape share chemical patterns.

MULTI-GENUS MODELING

The model learns the shared landscape signal, then uses genus-specific differences to infer the expected chemical pattern for the target genus.



THE RESULT

Where local target-genus reference data are sparse, information from other genera helps infer the expected chemical pattern and improve the estimate of a sample's most likely harvest area.

2. Incorporating land cover data significantly improves harvest area prediction

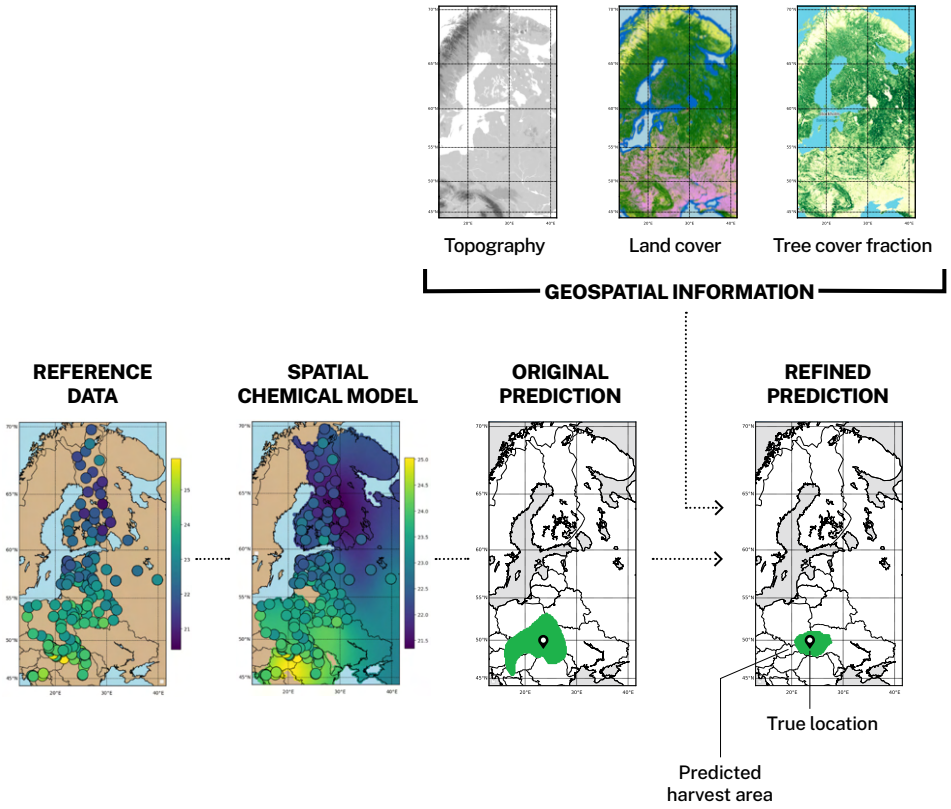
Our second study, published in *Environmental Research Communications*[2], finds that predicted harvest areas can be halved by excluding terrain where commercial timber extraction is not feasible, including cropland, grassland, water bodies, and urban areas.

World Forest ID's unique approach maps chemical signals continuously across landscapes, expressing harvest origin as a predicted area unconstrained by arbitrary political boundaries. The smaller that predicted area, the more precise and evidentially useful the assessment. Excluding non-forest land cover reduced the predicted harvest area by 51%, while correctly identifying 80.2% of true harvest locations.

Like multi-genus modeling, this advancement requires no new data collection and makes models meaningfully more efficient. The underlying land-cover dataset, ESA WorldCover[3], is freely available, making it a cost-efficient route to greater prediction accuracy in due diligence and enforcement contexts.

FIGURE 2. — *Growing range refinement*

A spatial chemical model estimates where a sample's chemistry is plausible. Environmental data layers then exclude non-harvestable or implausible areas, producing a smaller and more actionable predicted harvest area.



What this means for chemical timber testing

FOR REGULATORS AND IMPORT CONTROLS. Multi-genus modeling directly addresses the contexts where risk of sanctions evasion or misdeclaration is highest, and where reference data coverage has historically been scarcest. It extends testing capability to genera in regions where new field collection is impossible due to conflict, political instability, or restricted funds. At the same time, improvements in growing range mapping strengthen the evidentiary value of results, narrowing predicted origin ranges toward the plot-level standard that the EUDR requires, and making findings more actionable under regulations like the United Kingdom Timber Regulation (UKTR) and the United States Lacey Act.

FOR INDUSTRY. For companies conducting due diligence on timber supply chains, the primary barriers to chemical testing have been cost and sample coverage. Multi-genus modeling expands the range of genera that can be tested, including from regions of elevated risk where a verification pathway did not previously exist, reducing exposure for businesses that could not previously demonstrate origin with confidence. Efficiency gains in reference data collection lower the cost of maintaining the underlying platform, supporting wider adoption of scientific traceability as a routine component of supply chain due diligence.

Looking ahead

Multi-genus modeling and growing range refinement leverage existing collections to improve the efficacy of timber traceability. We will now be extending these techniques to tropical forests in Central Africa, South America, and Southeast Asia.

The goal is a reference dataset comprehensive enough to immediately answer any verification inquiry. This achievement still depends on expanding our current reference collection. Companies and certification schemes with access to forests in underrepresented regions can play an important role by donating reference samples. Multi-genus modeling means that each sample donated has an outsized impact: a small number of samples unlocks testing for multiple genera, bringing routine verification of a far wider range of products within reach.

This will give actors across supply chains the ability to identify mislabeling, ensure products meet legal and sustainability standards, and help keep illegal deforestation out of global markets.

EXPERT REVIEW:
Alicia Sullivan, World Forest ID Board Member
Arthur Blundell, World Forest ID Board Member

OUR SUPPORTERS:
Chalmers University of Technology.
Swedish Foundation for Strategic Environmental Research (MISTRA).