nature sustainability

Analysis

https://doi.org/10.1038/s41893-024-01491-8

Assessing timber trade networks and supply chains in Brazil

Received: 14 June 2024

Accepted: 11 November 2024

Published online: 6 January 2025

Check for updates

Luis Gustavo Nonato [1], Victor Russo[1], Bernardo Costa[1], Felipe Moreno-Vera [2], Guilherme Toledo[1], Osni Brito de Jesus[1], Robson Vieira[3], Marco Lentini[4] & Jorge Poco [2]

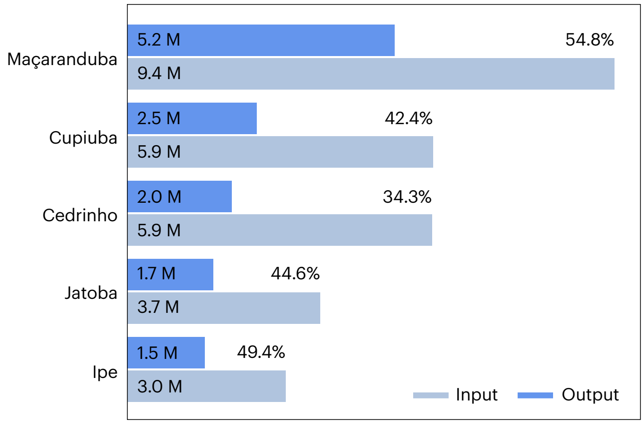
Forest degradation in the Brazilian Amazon is driven by factors such as fire, mining and illegal logging. The Brazilian government has implemented control mechanisms to combat illegal timber extraction that have positively impacted deforestation rates. Under these regulations, all wood products, from raw logs to processed lumber, must be registered in control systems before transportation. This allows analysis of wood products transported between companies over time. However, the existence of three partially integrated control systems complicates a full analysis of the timber market. This study integrates data from these systems to create timber trade networks, which help identify companies or groups operating outside expected standards. We also propose a method to trace likely supply chains of timber companies, addressing long-standing government concerns about timber traceability. Among the results, we show that certain timber trade networks have components that operate without connections with licensed forests, suggesting that unregistered timber is input into those components, which is illegal. Additionally, we illustrate how supply chain analysis can considerably enhance customer confidence in the legality of purchased timber products.

Brazilian Amazon forest degradation has long been a source of great concern for environmental agencies in Brazil and abroad because of its impact on biodiversity[1], environment[2,3] and even on the control of infectious diseases[4]. Forest degradation is influenced by various factors[5], but highly impacted by activities such as livestock farming, crop production (mainly soy) and mining, whose occupied areas grew more than 100%, 170% and 200% from 2001 to 2020 (https://brasil.mapbiomas.org/), respectively. Those activities are known to foster illegal logging[6], a key driver of rising CO_2 emissions in the Amazon[7]. In 2006, the Brazilian government introduced a computational system to control legal logging activities in an effort to reduce illegal timber extraction and forest degradation. This system positively impacted deforestation rates but did not fully eliminate illegal logging. The system is divided into three subsystems: SINAFLOR, which covers most Brazilian states; SISFLORA-MT, for Mato Grosso; and SISFLORA-PA, for Pará. These systems require any wood product, from raw logs to processed lumber, to be registered with a document of forest origin (DOF) or a forest guide (GF) for transportation, even within the same company.

Each DOF and GF records information about the volume, species and products being transported, along with the origin, destination and transportation date. Using this information, a timber trade network (TTN) can be created, with nodes representing entities in the timber market and edges showing trade relationships. This network allows for the calculation of timber volumes traded between entities and helps estimate a company's 'mass balance', the ratio of outflow to inflow volumes.

Timber is expected to enter a TTN through nodes representing licensed forests, which are authorized logging areas explicitly

[1]University of São Paulo, Institute of Mathematical and Computational Sciences, São Carlos, Brazil. [2]Fundação Getúlio Vargas, School of Applied Mathematics, Rio de Janeiro, Brazil. [3]RFV Consultancy and Systems, São Paulo, Brazil. [4]Imaflora, Programa de Cadeias Florestais, Piracicaba, Brazil. e-mail: gnonato@icmc.usp.br



	Nodes	Edges	Concessions
TTN—578 components	543,860	779,077	7,450
Main component	542,587	778,029	6,559
TTN—463 components	467,760	640,010	6,985
Main component	466,810	639,222	6,216
TTN—308 components	685,668	860,934	5,387
Main component	684,808	860,232	4,652
TTN—729 components	155,216	270,640	8,282
Main component	153,637	269,379	7,083
TTN—809 components	54,970	95,878	6,720
Main component	53,001	94,377	5,427

Fig. 1 | Input/output timber volume and associated TTNs. The bar chart on the left shows the input (light blue) and output (dark blue) timber volume (in a million m^3) for five high-value hardwoods traded in Brazil. The table on the right provides information about the TTN associated with each type of wood. The white rows show the total number of nodes, edges and licensed forests present in the TTN, while the grey rows provide the same information but is restricted to the largest connected component.

identified in the GFs and DOFs as such. Consequently, the volume of timber entering into a TTN can also be computed from the GFs and DOFs. All timber or lumber flowing through the network must originate from licensed forests. This implies that every node in a TTN should have a path that begins at a licensed forest and extends to the node.

A crucial aspect of TTNs is their ability to address a critical issue in combating illegal logging: tracing the likely routes that timber follows from its origin to its point of sale, the so-called likely supply chains. By identifying these routes, TTNs can assess the operational scale of businesses, determine the volume of timber transported in the chains, and identify the likely forest origins of the wood. This analysis supports more confident decision-making regarding the legality of the traded timber.

We integrated data from SINAFLOR, SISFLORA-MT and SISFLORA-PA spanning 2010 to 2020 to construct TTNs for selected high-value hardwood species. Using these networks, we developed a method to calculate the k most likely supply chains for timber companies. This approach allows for identifying the businesses involved in each chain, including the likely licensed forests where the timber originated. The analysis reveals groups of companies operating without links to licensed forests and detects mass balance discrepancies in specific companies, indicating potential illegal activities.

SINAFLOR transportation data were obtained from the Brazilian Institute of Environment and Renewable Natural Resources (IBAMA) public data portal (https://dadosabertos.ibama.gov.br/dataset/dof-transportes-de-produtos-florestais), while SISFLORA-MT and SISFLORA-PA data were acquired through collaborative agreements with the environmental regulatory bodies of the respective states. The processing and integration of timber transportation data from these three systems is a main contribution of this work, opening pathways to confronting illegal timber trade in the Amazon. Previous studies on identifying illegal logging have either taken a global approach[8,9], overlooking individual enterprises, or focused on a narrow set of entities[10], thus lacking a comprehensive view of the national context. Moreover, previous studies have not provided a general methodology for identifying the likely supply chains of enterprises within the timber industry. Our approach is a substantial step towards addressing these gaps.

In summary, our methodology reveals that if the timber transportation data were fully reliable and meticulously controlled, the curbing of illegal logging would become achievable by examining TTNs and likely timber supply chains. This approach offers a viable alternative to costly and potentially ineffective traceability resources, an avenue currently under consideration by Brazilian inspection agencies.

Results

Analysing the TTNs

The GFs and DOFs explicitly identify transportation originating from licensed forests. Therefore, we can compute the volume of each wood species that enters the TTN. We assume that timber (or wood products) only leaves the TTN through end-consumer nodes. The bar chart in Fig. 1 shows the input and output volumes of TTNs for five high-value hardwoods traded in Brazil: Ipe, Jatoba, Cedrinho, Cupiuba and Maçaranduba. The values inside the bars are the input and output volumes in million m^3 from 2010 to 2020. The values on the top of the input bars correspond to the balance ratio (output/input). Notice that a substantial amount of wood vanishes, mainly due to losses inherent in the sawing process. The depicted balances follow what is reported in the literature[10] (35–54% waste, depending on the wood species).

The table in Fig. 1 provides information about the TTN associated with each wood species. The white rows display the overall count of nodes, edges and licensed forests in the TTN. The grey rows provide the same information but is restricted to the largest connected component of the TTN. All TTNs have a main connected component containing ~99% of the nodes, except Ipe wood, whose main component contains 96% of the nodes.

Figure 2 shows the five largest connected components of the TTN of Ipe. Timber companies making up the larger connected component are spread nationwide, while the smaller components, which we call fragmented components, are concentrated in particular Brazilian states. This behaviour is also observed in the other wood species. The total volume of wood entering the fragmented components corresponds to less than 2% of the total volume entering the TTN.

Figure 2, right, depicts the fragmented component with 21 nodes. This fragmented component lacks a licensed forest node within it. The two source nodes (red circles in Fig. 2, right) are regular timber trade companies solely delivering timber, which raises suspicions. To enable this behaviour, all timber exchanged within this fragmented component should be stored in the company's yards, particularly in the yards of the highlighted red companies. Trading timber without a connection to a licensed forest could be sustained briefly with timber stored in the yards. However, the highlighted fragmented component operated for 9 years, suggesting that timber is input into it without proper registration, which is illegal. Similar behaviour is observed in fragmented components of other TTN species.

 https://doi.org/10.1038/s41893-024-01491-8

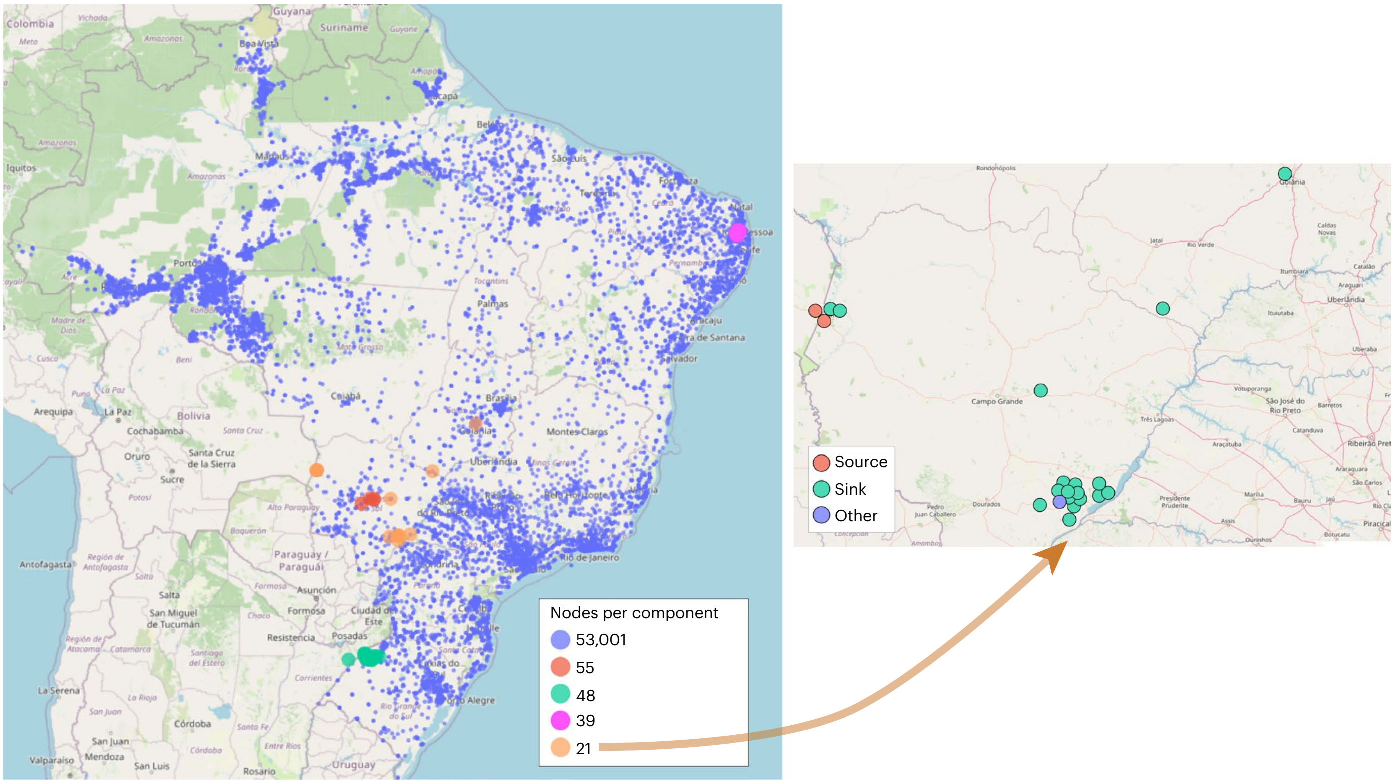


Fig. 2 | The five largest components of the TTN of Ipe. The largest connected component (in blue) is spread nationwide, while fragmented components are concentrated in particular Brazilian states (each highlighted with a different color). On the right is a zoomed-in view of the fragmented component with 21 nodes, not connected to a licensed forest. Maps generated using Plotly (https://plotly.com/python/tile-map-layers/).

Odd mass balances and supply chains

As mentioned before, losses due to sawing processes are expected. Therefore, the mass balance of timber companies (outflow/inflow) is supposed to be less than 1 or, in the best case, close to 1. This is the case for most companies in the TTN of Ipe, where about 94% of the companies, excluding licensed forests and end consumers, have mass balance within the expected range (Supplementary Fig. 2). A similar behaviour is observed for other species.

Nonetheless, there are enterprises whose behaviour is not so simple to explain. As a case study, we examine a company with a mass balance of ~13, meaning that 13 times more units of timber left the company compared to what entered it. We will denote this company as A and it is located in Pará state, as depicted by the red circle in Fig. 3. Company A was fully operational between 2010 and 2020, receiving 3,665 m^3 of timber while forwarding ~46,530 m^3. Timber only input company A in 2011 and 2012, but wood products outflow from it over the 11 years, mostly for export. Moreover, ~45% of its output corresponds to processed timber products, 30% is waste for energy purposes (sawing loss) and 25% is raw logs. So, company A production does not justify a mass balance of ~13, as the waste is considerable.

To further analyse company A, we computed its five most likely supply chains. The snapshots in Fig. 3 right show the geolocation of the enterprises involved in the chains, which are numbered from 1 to 7 in the snapshots. Notice that the five most likely chains originate from five distinct licensed forests (licensed forests correspond to the end nodes of the chains not highlighted in red), which collectively input 18,762 m^3 of raw logs into the chains, considerably less than the 46,530 m^3 exiting company A. Moreover, three out of the five licensed forests are located close to the municipality of Altamira, which recorded the highest deforestation rate in Amazon in 2020 (https://brasil.mapbiomas.org/en/2021/07/09/regiao-norte-lidera-desmatamento-no-brasil/).

Upon searching infraction notices available on the IBAMA website (https://dadosabertos.ibama.gov.br/dataset/fiscalizacao-auto-de-infracao), we found that five out of the eight entities involved in the supply chains, including company A, were fined for misconduct between 2010 and 2020. In particular, company A faced nine fines, most of which related to submitting fake information into the control systems and exporting forest products without compliance with the mandatory 'Regime especial de transporte' required for exporting timber products. Companies 2 and 5, which operate as intermediary companies in four supply chains, had 2 and 21 infraction notices from 2010 to 2020, respectively. Licensed forests 3 and 6 had one fine each in the period. Therefore, not only does company A bear questionable compliance but also most of its market partners are involved in misconduct.

Discussion

The importance of timber traceability has leveraged some initiatives from governmental and non-governmental bodies in Brazil. IBAMA released in 2020 the 'DOF + traceability'[10,11], a system that assigns a unique code to each log harvested in the forest, and this code must accompany all products derived from the log until their final destination. Using stable isotope signature[12] is another alternative explored by non-governmental agencies to match the unique composition of the timber with those found in specific forest regions. Both DOF + traceability and stable isotopes bear drawbacks that hamper their practical use, such as the possibility of fraud in the unique code generation and felling trees with the same isotope signature near licensed forests, as illegal logging tends to occur around authorized areas[13].

Unlike those intricate control mechanisms, the methodology proposed in this work leverages existing data to pinpoint timber companies or groups of companies whose practices diverge from expected standards, thereby raising concerns about their legal compliance. By analysing TTNs, consumers and government agencies can identify

Fig. 3 | Top-five most likely supply chains for company A. The chains start in five different licensed forests (1, 3, 4, 6 and 7). Maps generated using Plotly (https://plotly.com/python/tile-map-layers/).

groups of companies operating without connections to licensed forests, an indication that control systems are being circumvented. Furthermore, analysing the most likely supply chains helps identify the likely sources of the wood and the assessment of the entities involved. As shown in the case study, companies involved in misconduct often collaborate with other non-compliant enterprises, exposing trade networks that facilitate illegal logging activities.

The analytical capability enabled by the proposed methodology enhances confidence in the legality of wood products. This analytical capacity is crucial for combating illegal trade, as reputable consumers can verify not only the credibility of the companies from which they are purchasing wood-based products but also the supply chains in which these companies are involved.

The methodology could achieve even greater analytical power with additional information on timber-processing sawmill efficiency and the quantities of wood species stored in company yards. With this additional data, it would be possible to make precise estimations of the expected volumes of timber flowing through each supply chain, serving as a deterrent against the infiltration of illegal timber products. It is important to note that illegal logging is also fuelled by weak institutional mechanisms that enable corruption and sustain organized crime networks. Effectively addressing this issue requires confronting these deep-rooted systemic challenges as well.

In conclusion, alongside unique code generation and isotope signature analysis, the proposed TTN approach emerges as another powerful tool in the fight against illegal timber trade. This is not only relevant for meeting emerging global market demands but also underscores the critical role of public authorities in leveraging existing tools to address the issue effectively.

The proposed methodology has been incorporated into a freely accessible web-based application available in https://timberflow.org.br/.

Methods

Data compiling and integration

Two distinct categories of forestry areas are legally designated for logging activities. The first category encompasses areas where landowners are granted specific authorizations to conduct logging. The second category comprises concessions for exploring logging on public lands. To simplify our discussion, we collectively refer to both authorization types as licensed forests.

The collected data from SINAFLOR, SISFLORA-MT and SISFLORA-PA resulted in 31,913,611 DOFs and 18,601,146 GFs. To ensure seamless data compatibility across the three systems, we created a unified list of timber companies by standardizing information from the DOFs and GFs. This unification process primarily relied on key parameters such as company name, tax number (CPF/CNPJ), municipality and location coordinates (latitude and longitude). We additionally used location coordinates to establish a unique identifier for licensed forests, as different areas might be linked to the same exploration authorization. End consumers also play a crucial role in the timber trade market, as they are the final destination of the transportation flow tracked in the system. Entities identified in the GFs and DOFs solely as destinations are considered end consumers, including those marked as 'exportation'. End consumers make up ~85% of the entities in the TTN. After integrating data from the three control systems, licensed forests, timber companies and end consumers are uniquely identified as key 'actors' within the timber trade ecosystem.

The transport data have also been processed to accommodate differences between systems, ending up with 13 attributes for each transportation record. The first two attributes are index keys for the source and destination entities related to the transportation. Index keys for the transported species and forest products are also present in the dataset. Efforts were made to standardize and aggregate timber products and species names, including both scientific and common names. This standardization ensured consistency in representing wood types and products, making interpretation easier. The integrated dataset also brings the barcode identification number of the transportation document and the system of origin (SINAFLOR, SISFLORA-PA and SISFLORA-MT). The transported volume and its corresponding unit of measure (mostly m^3), DOF/GF issue date, which we interpret as transportation date, and the status of the document (DOFs and GFs may have different statuses, for instance 'cancelled' and 'active') are also recorded. The remaining attributes account for the authorization number, authorization type and registration type.

In summary, the data compilation process gave rise to a dataset consisting of four interconnected subsets: (1) a subset describing the licensed forests, timber trade companies and end consumers; (2) the transportation data subset; (3) a list of wood species; and (4) a list of forest products. Further details about the integration process are provided in Supplementary Information.

TTNs and supply chains

The origin–destination information in the DOFs and GFs enables the representation of a specific wood species' TTNs as a directed graph with three types of nodes: sources (licensed forests), intermediates (timber trade companies) and sinks (end consumers). Licensed forests are source nodes with only outflow edges, timber companies have both inflow and outflow edges and end consumers are sink nodes with only inflow edges.

In mathematical terms, the TTN of a wood species S is a directed graph $G_S = \{V, E, W\}$, where the vertex set $V = V_f \cup V_c \cup V_e$ is the union of the node sets V_f, V_c and V_e corresponding to licensed forests, timber companies and end consumers, respectively. The collection of directed edges, denoted as E, represents timber transactions. A directed edge from a node v_i to v_j indicates transportation of wood products of species S from v_i to v_j. The entries w_{ij} in the matrix W encapsulate the total volume of species S transported from v_i to v_j.

The mass balance of each node $v_j \in V_c$ is defined as the ratio between outflow and inflow volumes, that is, $\mathrm{mb}_{v_j} = \sum_i w_{ji} / \sum_i w_{ij}$. In our context, a supply chain of a node $v \in V_c \cup V_e$ is a sorted sequence of nodes $\mathrm{sc}_v = \{v_{c_0}, v_{c_1}, \dots, v_{c_s}\}$, where $v_{c_s} = v$, $v_{c_0} \in V_f$ (a licensed forest), $v_{c_i} \in V_c$, $i = 1, \dots, s-1$, $v_{c_i} \neq v_{c_j}$ for all $i \neq j$, with a directed edge from $v_{c_{i-1}}$ to v_{c_i} for $i = 1, \dots, s$. In other words, a supply chain of v is a directed path without loops, starting in a licensed forest and ending in v. The total timber volume transported throughout a supply chain is given by $\sigma(\mathrm{sc}_v) = \sum_i w_{c_{i-1} c_i}$, $i = 1, \dots, s$.

Supposedly, timber may only enter a TTN from licensed forest nodes. Timber leaves a TTN in two ways: due to losses inherent to the sawing process or sales to end consumers. Therefore, the mass balance mb_{v_j} of each node $v_j \in V_c$ should always be ≤1. Ratios >1 might occur if the companies stored wood in their yards for a certain period. Such a situation should only occur occasionally, not lasting for long periods (except when the company possesses an extensive stockpile of wood, which is uncommon).

Most likely supply chains

Given the impossibility of precisely tracing the pathways of wood products, we alternatively compute the most likely supply chains bolstering timber trade companies (or end consumers). A plausible hypothesis is to assume that, among all possible supply chains, the shortest ones with the greatest total transported volume are the most likely. In mathematical terms, given a node $v \in V_c \cup V_e$ we search for chains satisfying:

$$\min_s \max\{\sigma(c) \mid c = \{v_{c_0}, v_{c_1}, \dots, v_{c_s} = v\}\} \tag{1}$$

If we flip the directions of TTN edges and multiply the volumes by −1 (volumes associated with edges become negative), the problem defined in equation (1) can be solved by a variant of Dijkstra's algorithm called single-source shortest-path problem[14]. Such an algorithm has a subpolynomial computational cost, and it finds the minimum weight (the reason why we use negative volumes) paths from the source v to all other nodes in G_S (in the same connected component as v), including licensed forests, thus computing timber supply chains for v.

However, multiple supply chains of a node v may originate from a particular licensed forest v_f. The volumes transported through these chains might exceed those of the supply chains connecting v to other licensed forests, therefore, they must be considered. We resort to Eppstein's algorithm[15] to compute the additional chains connecting v to v_f. Eppstein's algorithm computes the k minimum weight paths between two nodes. In summary, the k most likely supply chains of a node v can be computed as follows: (1) apply Dijkstra's algorithm to find the shortest (minimum weight) chain between v and all licensed forest nodes; (2) get the resulting chain with minimum total volume and its corresponding licensed forest node v_f; (3) compute the k minimum weight chains between v and v_f using Eppstein's algorithm and compare the total volumes of the k chains with those resulting from Dijkstra's algorithm; (4) if the volumes of the k chains from Eppstein's algorithm are smaller than the second smaller Dijkstra path, then finish the process. Otherwise, hold the chains from Eppstein's algorithm that have smaller volumes than Dijkstra's second smaller and repeat steps (2)–(4) considering now the second smaller Dijkstra path and so on until k chains are obtained. The proposed algorithm is detailed in Supplementary Information.

The process described above provides the k 'shortest' supply chains with the smaller (negative) volumes between v and licensed forests, which we assume to be the most likely supply chains of v. Turning the total volume of each computed chain back to positive, we have the shortest chains with the greatest volumes.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The SINAFLOR database is available from the IBAMA public data portal: https://dadosabertos.ibama.gov.br/dataset/dof-transportes-de-produtos-florestais. Data from SISFLORA-PA and SISFLORA-MT were obtained through collaborative agreements with environmental regulatory bodies of Pará and Mato Grosso states and are not publicly available. The integrated transportation data for Ipe obtained as a result of this work are available via GitHub at https://github.com/lgnonato/Timber-Chain. Owing to the terms of the agreement made with the regulatory body of Mato Grosso, the released data do not include timber companies and transportation in the Mato Grosso state. For possible access to the complete dataset, please contact www.imaflora.org/.

Code availability

The code to build the TTN of Ipe and compute supply chains is available via GitHub at https://github.com/lgnonato/Timber-Chain.

References

1. Barlow, J. et al. Anthropogenic disturbance in tropical forests can double biodiversity loss from deforestation. *Nature* **535**, 144–147 (2016).
2. Nobre, C. A. et al. Land-use and climate change risks in the Amazon and the need of a novel sustainable development paradigm. *Proc. Natl Acad. Sci. USA* **113**, 10759–10768 (2016).
3. Pearson, T. R., Brown, S., Murray, L. & Sidman, G. Greenhouse gas emissions from tropical forest degradation: an underestimated source. *Carbon Balance Manag.* **12**, 3 (2017).
4. Ellwanger, J. H. et al. Beyond diversity loss and climate change: Impacts of amazon deforestation on infectious diseases and public health. *An. Acad. Bras. Cienc.* **92**, e20191375 (2020).
5. Lapola, D. M. et al. The drivers and impacts of Amazon forest degradation. *Science* **379**, eabp8622 (2023).
6. da Silveira Bueno, R., Marchetti, L., Cocozza, C., Marchetti, M. & Salbitano, F. Could cattle ranching and soybean cultivation be sustainable? A systematic review and a meta-analysis for the Amazon. *iFOREST* **14**, 285–298 (2021).
7. Gatti, L.V. et al. Increased Amazon carbon emissions mainly from decline in law enforcement. *Nature* **621**, 318–323 (2023).
8. Brancalion, P. H. et al. Fake legal logging in the Brazilian Amazon. *Sci. Adv.* **4**, eaat1192 (2018).
9. Chen, B. et al. Modeling illegal logging in Brazil. *Res. Math. Sci.* **8**, 29 (2021).
10. Franca, C. S., Persson, U. M., Carvalho, T. & Lentini, M. Quantifying timber illegality risk in the Brazilian forest frontier. *Nat. Sustain.* **6**, 1485–1495 (2023).

11. Conama Resolution No. 497, 19 August 2020 (Conama, 2020); https://conama.mma.gov.br/index.php?option=com_sisconama&task=arquivo.download&id=796
12. Deklerck, V. Timber origin verification using mass spectrometry: challenges, opportunities, and way forward. *Forensic Sci. Int.* **3**, 100057 (2023).
13. Finer, M., Jenkins, C. N., Sky, M. A. B. & Pine, J. Logging concessions enable illegal logging crisis in the Peruvian Amazon. *Sci. Rep.* **4**, 4719 (2014).
14. Cormen, T. H., Leiserson, C. E., Rivest, R. L. & Stein, C. *Introduction to Algorithms* (MIT, 2022).
15. Eppstein, D. Finding the *k* shortest paths. *SIAM J. Comput.* **28**, 652–673 (1998).

Acknowledgements
We thank R. Contreras, D. de Almeida, J. Mattos, J. M. Cardoso da Silva and F. Nunes da Silva for their contribution in the early stages of data preparation and database construction. We thank IMAFLORA for its financial and technical support. This work was also partially supported by CNPq (grant nos. 307184/2021-8 (L.G.N.) and 311144/2022-5 (J.P.)), FAPESP (grant nos. 2022/09091-8 (L.G.N.) and 2021/07012-0 (J.P.)), FAPERJ (grant no. E-26/201.424/2021 (J.P.)) and FGV/EMAp (J.P.). The opinions, hypotheses, conclusions and recommendations expressed in this material are the responsibility of the authors and do not necessarily reflect CNPq, FAPESP, FAPERJ and FGV views.

Author contributions
All authors contributed extensively to the work presented in this paper. V.R., B.C., F.M.-V., G.T. and O.B.d.J. implemented the database integration, wrote the supply chain code and ran most of the experiments. R.V. designed the data integration and collected the SINAFLOR data. M.L., J.P. and L.G.N. designed the experiments and analysed the results. R.V., M.L. and L.G.N. wrote the paper and prepared it for submission.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41893-024-01491-8.

Correspondence and requests for materials should be addressed to Luis Gustavo Nonato.

Peer review information *Nature Sustainability* thanks Mohammad Reza Ghaffariyan, Franco Perazzoni and Niels Soendergaard for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2025

nature sustainability

Supplementary information https://doi.org/10.1038/s41893-024-01491-8

Assessing timber trade networks and supply chains in Brazil

In the format provided by the authors and unedited

Contents

1 Data Integration Process

A major challenge in this work was integrating and standardizing data from the three timber control systems operating in Brazil, namely SISFLORA-PA, SISFLORA-MT, and SINAFLOR, into a cohesive dataset. As illustrated in Figure 1, the integrated dataset comprises four tables connected by indices. Specifically, The *transportation* table includes columns with the indices of the origin and destination enterprises recorded in the *Timber Enterprise* table. It also contains indices for the type of wood and product transported, which are described in the *Wood Type* and *Wood Products* tables, respectively. The transportation table also includes information regarding the volume, date, and status of the transport.

Table 1 shows, in each column, the list of attributes provided by each timber control system. The attributes highlighted in bold are the ones used to uniquely identifying each enterprise within the timber trade ecosystem. The unique identification of timber companies proved particularly challenging due to variations in the recording of names, tax numbers, and even municipality names. These records are often inconsistently formatted and spelled, complicating efforts to achieve accurate identification. Notice that the geolocation information (latitude and longitude) of the companies are only available in the DOFs, thus, to further improve the identification of each enterprise, we extracted geolocation information from PDF files publicly available in the SISFLORA-PA system. In cases where geolocation was not possible, we assigned the geographic coordinates of the enterprise as being the same as the municipality where it resides. In general terms, we first built a list with all enterprises containing name, tax number, municipality, state, and geolocation (when available) of the enterprises contained in the tree systems. The same enterprise can appear multiple times in the list due to variations in the spelling and format of the record. To eliminate duplicates we search, for each element in the list, the best matches in terms of name, tax number, municipality, state, and geolocation. The match is computed using the `regex` Python package.

The Wood Type and Wood Products tables were manually compiled with assistance from the forest engineer involved in the project. Generating the transportation table was more straightforward; transports between different states are identified by the DOF number listed in the GFs, while transports within the same state are uniquely recorded in either the GFs or DOFs.

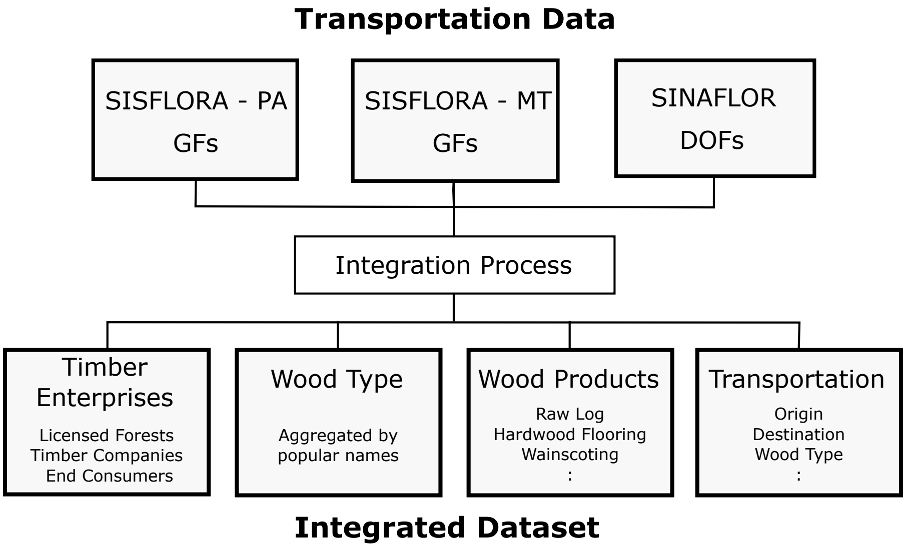


Figure 1: Data from the three control systems operating in Brazil, SISFLORA-PA, SISFLORA-MT, and SINAFLOR, are integrated into a single standardized dataset made up of four components: a table of unique enterprises (licensed forests, timber trade companies, and end consumers), a table with standardized wood types (species aggregated according to genus), a table with standardized wood products, and the transportation data itself.

SISFLORA-PA	SISFLORA-MT	SINAFLOR
tipo_gf, numero_gf, status_gf, ceprof_remetente, **empreendimento_remetente**, **cpf_cnpj_remetente**, ie_remetente, **uf_remetente**, **municipio_remetente**, tipo_destino, ceprof_destinatario, **empreendimento_destinatario**, **cpf_cnpj_destinatario**, ie_destinatario, **uf_destinatario**, **municipio_destinatario**, zona_alfandegada, tipo_zona_alfandegada, uf_zona_alfandegada, municipio_zona_alfandegada, n_dvpf_origem, n_autorizacao, processo, codigo_barras, dt_emissao, dt_recebimento, gf_id, codigo_tora, nome_cientifico, nome_popular, codigo, produto, volume, unidade, preco_total, ferroviario, fluvial, maritimo, rodoviario, Vazio	CC-SEMA Remetente, **Empreendimento Remetente**, **CNPJ/CPF Remetente**, Inscrição Estadual Remetente, **UF Remetente**, **Município Remetente**, CC-SEMA Destinatário, **Empreendimento Destinatário**, **CNPJ/CPF Destinatário**, Insc. Estadual Destinatário, **UF Destinatário**, **Município Destinatário**, Complemento Destinatário, Exportação, Tipo, Número, Status, N DVPF, Origem, N Autorização, Processo, Código de Barras, Emissão, Ultima Modificação, Nota Fiscal, Nome Científico, Nome Popular, Class., Produto, Volume, Unid, Preço Total, Preço Unit., Rodoviário, Ferroviário, Hidroviário	id, **nome_remetente**, **cpf_remetente**, **uf_origem**, **municipio_origem**, ctf_remetente, tipo_origem, nome_patio_origem, numero_serie_autex, numero_autorizacao_original, tipo_autex, orgao_emissor_autex, dt_validade_autex, numero_di, orgao_emissor_di, dt_validade_di, nome_porto_entrada, pais_origem, numero_autesp, orgao_emissor_autesp, dt_validade_autesp, **latitude_origem**, **longitude_origem**, **uf_destino**, **municipio_destino**, **nome_destinatario**, **cpf_destinatario**, ctf_destinatario, nome_patio_destino, **latitude_destino**, **longitude_destino**, nome_porto_saida_pais, municipio_porto_destino, uf_porto_destino, pais_destino, dt_emissao, ano, dt_validade_inicial, dt_validade_final, ultima_transacao, dt_ultima_transacao, numero_oferta, numero_serie_dof, codigo_controle_dof, rota_transporte, produto, nome_cientifico, nome_popular, unidade, volume, valor_reais

Table 1: List of attributes provided by the three timber control system, SISFLORA-PA, SISFLORA-MT, and SINAFLOR in the GFs and DOFs, respectively. The attributes highlighted in bold are the ones used to uniquely identifying the enterprises within the timber trade ecosystem.

2 k-Most Likely Supply Chains Computation

Algorithm 1 describes the computation of the k most likely supply chains of a given timber company $v \in V_c \cup V_e$. As detailed in the main manuscript, V_f, V_c, and V_e account for the set of nodes representing forest concessions, timber companies, and end consumers, respectively. E and W are the directed edge set and corresponding weights of a Timber Trade Network - TTN.

Algorithm 1 k-most likely supply chains

Require: $G_{\mathcal{S}} = \{V_f \cup V_c \cup V_e, E, W\}$; $v \in V_c \cup V_e$; k

Flip edges, directions and the sign of their corresponding volumes

$\mathcal{C} \leftarrow \emptyset$ ▷ $\mathcal{C}$ holds the most likely chains

$C \leftarrow$ Dijkstra($G_{\mathcal{S}}$,v) ▷ Shortest chains with greatest volume from v to each forest concession

$C \leftarrow sort(C)$ ▷ Sort chains in ascending order of negative volumes; $C[i][0]$ corresponds to v and $C[i][-1]$ to a forest concession $v_f \in V_f$ in the i-th chain $C[i]$, $i = 0, \ldots, |C| - 1$; $|C|$ is the number of chains in C.

$count \leftarrow 0$
if $k > |C|$ **then**
  $k \leftarrow |C|$
end if
$i \leftarrow 0$
while $count < k$ **do**
  $\mathcal{C} \leftarrow \mathcal{C} \cup C[i]$ ▷ Add the likely chain $C[i]$ to $\mathcal{C}$
  $count \leftarrow count + 1$
  $i \leftarrow i + 1$
  $P \leftarrow$ Eppstein($G_{\mathcal{S}}$,C[i-1][0],C[i-1][-1],k) ▷ P holds the k-shortest chains with greatest volume from $C[i-1][0]$ to $C[i-1][-1]$
  $P \leftarrow sort(P)$ ▷ Sort chains from P in ascending order of negative volumes
  $j \leftarrow 1$
  while $\sigma(P[j]) < \sigma(C[i])$ and $count < k$ **do** ▷ $\sigma(\cdot)$ is the total volume of the chain
    $\mathcal{C} \leftarrow \mathcal{C} \cup P[j]$ ▷ Add the likely "secondary" chain $P[j]$ to $\mathcal{C}$
    $count \leftarrow count + 1$
    $j \leftarrow j + 1$
  end while
end while
return($\mathcal{C}$)

3 Odd Mass Balances

Losses due to sawing processes are expected, so the mass balance of timber companies (outflow/inflow) is supposed to be less than or close to one in the best case. The histogram in Figure 2a shows this is the case for most companies in the Ipes's TTN, where about 94% of the companies, excluding licensed forests and end consumers, have mass balance within the expected range (similar behavior is observed for other species). Note the peaks close to 0.5, which aligns with typical sawing losses, and another peak near 1.0, likely corresponding to companies that trade unprocessed logs. Mass balances smaller than 0.5 may be due to a high degree of wood processing, low efficiency in the sawing process, or wood storage practices, whereas values slightly greater than 1.0 may be related to the trading of stocked timber in the period.

Figure 2b shows that some companies exhibit "spurious" mass balances, exceeding 1,000, indicating they sent out much more timber units than received. The substantial discrepancies warrant attention, but they may not necessarily indicate illegal conduct. For instance, the two companies with mass balances exceeding 1,000 (rightmost in Figure 2b) were nearly inactive between 2010 to 2020, receiving

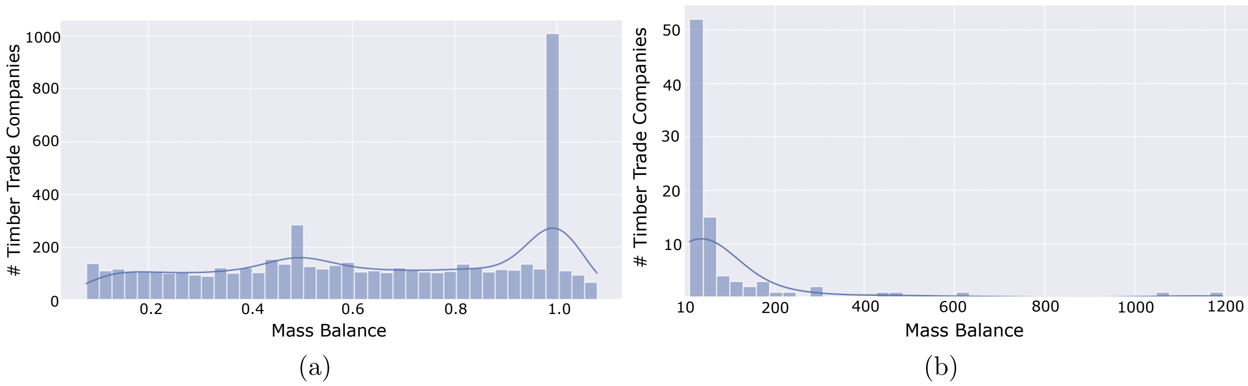


Figure 2: Mass balance of companies in the Ipe's TTN.

approximately $2m^3$ of timber and dispatching around $2,000m^3$ during this period, a volume that could be stored in their yards.